Kapitel 3

Likelihoodschätzer im klassischen Fall

Es sei $(\Omega, \mathscr{A}, \mathscr{P}, X)$ ein statistisches Experiment mit dem Stichprobenraum $E = \mathbb{R}^n$, $\mathscr{E} = \mathscr{B}^n$ und $X = (X_1, X_2, \ldots, X_n)$ bestehe aus reellwertigen, bezüglich jedem $\mathbb{P} \in \mathscr{P}$ unabhängigen und identisch verteilten Zufallsgrößen $X_k,\ k = 1, \ldots, n$. Die Familie $\mathscr{P}$ sei parametrisiert:
$\mathscr{P} = (\mathbb{P}_\vartheta,\ \vartheta \in \Theta),\ \Theta \subseteq \mathbb{R}^k$, und
$F_\vartheta(x) := \mathbb{P}_\vartheta(X_1 \leq x),\ x \in \mathbb{R},\ \vartheta \in \Theta$ bezeichne die Verteilungsfunktion von X_1 bezüglich $\mathbb{P}_\vartheta$.
Dann hat X die Verteilungsfunktion

$$F_\vartheta^X(x_1, \ldots, x_n) = \prod_{m=1}^{n} F_\vartheta(x_m) = \mathbb{P}_\vartheta\big(X \in (-\infty, x_1] \times \ldots \times (-\infty, x_n]\big).$$

Besitzt F_ϑ eine Dichte f_ϑ, so hat X die Dichte

$$f_\vartheta^X(x_1, \ldots, x_n) = \prod_{m=1}^{n} f_\vartheta(x_m).$$

Das bedeutet $\mathbb{P}_\vartheta(X \in B) = \int_B f_\vartheta^X\, dx$ für alle Borelmengen $B \in \mathscr{B}$.
Ist X_1 diskret verteilt mit $\mathbb{P}_\vartheta(X_1 = a_m) = p_m(\vartheta)$, so ist auch X diskret verteilt und es gilt

$$\mathbb{P}_\vartheta\Big(X = (a_{m_1}, \ldots, a_{m_n})\Big) = \prod_{r=1}^{n} p_{m_r}(\vartheta)$$

In beiden Fällen nennt man (bei festgehaltener Stichprobe $x = (x_1, \ldots, x_n)$ bzw. $x = (a_{m_1}, \ldots, a_{m_n})$)

$$L_n(\vartheta; x_1, \ldots, x_n) = \prod_{m=1}^{n} f_\vartheta(x_m) \ \text{ bzw. } = \prod_{r=1}^{n} p_{m_r}(\vartheta), \quad \vartheta \in \Theta,$$

die *Likelihood-Funktion* des statistischen Experiments.

Maximum-Likelihood-Schätzer (ML-Schätzer)

Es sei x eine (konkrete) Stichprobe aus $E = \mathbb{R}^n,\ x \in \operatorname{supp}(X)$.

Definition 3.1. Jeder Wert $\hat{\vartheta}(x)$ aus Θ, der die Likelihoodfunktion $L_n(\cdot\,; x)$ bei gegebenem x maximiert, heißt ein *Maximum-Likelihood-Schätzwert* für ϑ auf der Basis der Stichprobe x:

$$\hat{\vartheta}_n(x) := \arg\max_{\vartheta\in\Theta} L_n(\vartheta\,; x) = \arg\max_{\vartheta\in\Theta} \log L_n(\vartheta\,; x)$$

Offenbar ist $\hat{\vartheta}_n(\,\cdot)$ eine Stichprobenfunktion. Setzt man die mathematische Stichprobe X ein, so erhält man einen sogenannten *Maximum-Likelihood-Schätzer* (kurz: ML-Schätzer) $\hat{\vartheta}_n(X)$.

> *R.A. Fisher: Maximum-Likelihood-Prinzip*
> „Finde diejenigen Voraussetzungen, die das Beobachtete mit großer Wahrscheinlichkeit nach sich ziehen und fasse Zutrauen, daß diese Voraussetzungen die wirksamen sind.“

Bemerkung: Maximum–Likelihood-Schätzer sind häufig einfach auszurechnen, haben vielfach gute Eigenschaften, existieren aber nicht immer oder sind nicht eindeutig. ML-Prinzip ist ein sehr allgemeines Prinzip, kann auch bei stochastischen Prozessen angewendet werden.

Maximum-Likelihood-Gleichungen

Unter der Voraussetzung, daß $L_n(\vartheta\,;\ x_1,\ldots,x_n)$ für jedes $(x_1,\ldots,x_n) \in \operatorname{supp}(X)$ bezüglich ϑ differenzierbar ist, sind

$$\frac{\partial}{\partial\vartheta_r} L_n(\hat{\vartheta}_n\,;\ x_1,\ldots,x_n) = 0 \qquad r = 1,2,\ldots,k \qquad \text{(„ML-Gleichungen“)} \quad (3.1)$$

notwendige Bedingungen für $\hat{\vartheta}_n$, ein ML-Schätzer zu sein, sofern das Maximum von L_n im Inneren von Θ angenommen wird.
Anstelle L_n führt man

$$l_n(\vartheta\,;\ x_1,\ldots,x_n) = \log L_n(\vartheta\,;\ x_1,\ldots,x_n) \quad \text{ein.} \qquad \text{(„Loglikelihoodfunktion“)}$$

Äquivalent zu (3.1) ist

$$\frac{\partial}{\partial\vartheta_r} l_n(\hat{\vartheta}_n\,;\ x_1,\ldots,x_n) = 0 \qquad r = 1,2,\ldots,k. \qquad (3.2)$$

Die Funktion

$$\dot{l}_n(\vartheta\,;\ x_1,\dots,x_n) := \Big(\frac{\partial}{\partial\vartheta_r} l_n(\vartheta\,;\ x_1,\dots,x_n), \quad r = 1,2,\dots,k\Big), \quad \vartheta \in \Theta$$

nennt man *Scorefunktion* des statistischen Experiments $(\Omega, \mathscr{A}, \mathscr{P}, X)$.
Es gilt

$$\begin{aligned} l_n(\vartheta\,;\ x_1,\dots,x_n) &= \operatorname{grad} l_n(\vartheta\,;\ x_1,\dots,x_n) \quad \text{und} \\ \mathbb{E}_\vartheta\Big(\dot{l}_n(\vartheta\,;\ X_1,\dots,X_n)\Big) &= \mathbb{E}_\vartheta\Big(\frac{\dot{L}_n}{L_n}\Big) = \underset{\mathbb{R}^n}{\int\cdots\int} \dot{L}_n(\vartheta\,;\ x_1,\dots,x_n)\,dx_1\dots dx_n \\ &= \operatorname{grad} \underset{\mathbb{R}^n}{\int\cdots\int} L_n\,dx_1\dots dx_n \\ &= 0 \end{aligned}$$

(Hier haben wir vorausgesetzt, daß Differentiation nach ϑ und Integration bezüglich x vertauschbar sind.)

Bemerkung:

$$l_n(\vartheta\,;\ X_1,\dots,X_n) = \sum_{m=1}^{n} \log f_\vartheta(X_m)$$

ist eine Summe unabhängiger und identisch verteilter Zufallsgrößen.

$$\dot{l}_n(\vartheta\,;\ X_1,\dots,X_n) = \operatorname{grad}_\vartheta l_n(\vartheta\,;\ X_1,\dots,X_n)$$

ist eine zentrierte Summe unabhängiger identisch verteilter Zufallsvektoren.
Für jedes feste ϑ sind $L_n(\vartheta\,;\ X_1,\dots,X_n)$, $l_n(\vartheta\,;\ X_1,\dots,X_n)$ und $\dot{l}_n(\vartheta\,;\ X_1,\dots,X_n)$ Stichprobenfunktionen.

Definition 3.2. Ist $E = \{x \in \mathbb{R} \mid f_\vartheta(x) > 0\}$ (bzw. $E = \{a \in \mathbb{R} \mid \mathbb{P}_\vartheta(X = a) > 0\}$) unabhängig von ϑ, so nennt man für ϑ, $\alpha \in \Theta$ und $x \in E$ den Quotienten

$$\frac{L_n(\alpha\,;\,x)}{L_n(\vartheta\,;\,x)} \qquad \text{den } \textit{Likelihoodquotienten.}$$

Ist $\log \frac{f_\alpha}{f_\vartheta}$ bezüglich F_ϑ integrierbar, so gilt

$$\frac{1}{n}\log\frac{L_n(\alpha\,;\,X_1,\dots,X_n)}{L_n(\vartheta\,;\,X_1,\dots,X_n)} \xrightarrow{\mathbb{P}_\vartheta -\text{f.s.}} \int \Big[\log\frac{f_\alpha}{f_\vartheta}\Big] f_\vartheta\,dx =: -K(F_\vartheta\,,\,F_\alpha)$$

$K(F_\vartheta\,,\,F_\alpha)$ heißt *Kullback-Information von F_ϑ bezüglich F_α.*

Lemma 3.3. *Es gilt:*

$$K(F_\vartheta\,,\,F_\alpha) \geq 0$$
$$K(F_\vartheta\,,\,F_\alpha) = 0 \Longleftrightarrow F_\vartheta = F_\alpha$$

Im Fall, daß X_1 eine diskrete Verteilung besitzt, gilt

$$K(F_\vartheta\,,\,F_\alpha) = -\sum_m \Big[\log \frac{p_m(\alpha)}{p_m(\vartheta)}\Big] p_m(\vartheta).$$

Beweis: (nur für den Dichtefall)
Die Funktion $h(x) = x\log x + 1 - x$ ist für $x > 0$ und $x \neq 1$ positiv und nur für $x = 1$ gleich Null. Folglich gilt:

$$\int \Big(\frac{f_\alpha}{f_\vartheta}\log\frac{f_\alpha}{f_\vartheta} + 1 - \frac{f_\alpha}{f_\vartheta}\Big) f_\vartheta\,dx \geq 0$$

und $K(F_\vartheta\,,\,F_\alpha) = 0$ impliziert $f_\alpha = f_\vartheta$.
Also gilt für $\alpha \neq \vartheta$ die Beziehung $-K(F_\vartheta\,,\,F_\alpha) < 0$ und somit

$$\log\frac{L_n(\alpha\,;\,X_1,\ldots,X_n)}{L_n(\vartheta\,;\,X_1,\ldots,X_n)} \xrightarrow{\mathbb{P}_\vartheta\,-\text{f.s.}} -\infty \quad \text{für } n \longrightarrow \infty,$$

mit anderen Worten, für $\alpha \neq \vartheta$ gilt

$$\frac{L_n(\alpha\,;\,X_1,\ldots,X_n)}{L_n(\vartheta\,;\,X_1,\ldots,X_n)} \xrightarrow{\mathbb{P}_\vartheta\,-\text{f.s.}} 0 \quad \text{für } n \longrightarrow \infty.$$

Andererseits ist offensichtlich der Quotient für $\alpha = \vartheta$ gleich Eins. □

Lemma 3.3 ist noch einmal ein Argument für die Vernünftigkeit des Maximum-Likelihood-Schätzers: $L_n(\alpha; X_1,\ldots,X_n)$ wird für α fernab von ϑ vergleichsweise zu $L_n(\vartheta; X_1,\ldots,X_n)$ klein sein (mit wachsendem n konvergiert der Quotient ja gegen Null), für α in der Nähe von ϑ auf Grund der Stetigkeit von $\alpha \longrightarrow L_n(\alpha; X_1,\ldots,X_n)$ nahe Eins. Verwendet man $\hat{\vartheta}_n(X_1,X_2,\ldots,X_n)$ als Schätzer für ϑ, so wird man also erwarten können, daß dieser Schätzer in der Nähe von ϑ liegt.

Eigenschaften der Maximum-Likelihood-Schätzer:

Wir geben hier zwei wichtige Eigenschaften von Maximum-Likelihood-Schätzern für den Fall von Stichproben an, die aus unabhängigen, identisch verteilten Zufallsgrößen bestehen. Für die keineswegs einfachen Beweise sei auf die Literatur verwiesen.

a) KONSISTENZ:

Im allgemeinen ist der Maximum-Likelihood-Schätzer nicht erwartungstreu, das heißt, es gilt i.a. nicht $\mathbb{E}_\vartheta\big(\hat{\vartheta}_n(X_1, X_2, \ldots, X_n)\big) = \vartheta$. Die folgende Eigenschaft der Konsistenz besagt aber, daß man für große Stichprobenumfänge den Schätzer $\hat{\vartheta}_n$ mit großer Wahrscheinlichkeit in der Nähe von ϑ finden wird. (Wir beschränken uns mit der Formulierung auf den Fall, daß $\mathbb{P}_\vartheta^{X_1}$ eine Dichte $f_\vartheta(x)$ hat.)

- sei Θ eine kompakte Teilmenge von $\mathbb{R}^n$
- es gelte $\{x \in \mathbb{R} \,:\, f_\alpha(x) > 0\}$ unabhängig von $\alpha \in \Theta$
- wenn $\alpha \neq \vartheta$, so $F_\alpha \neq F_\vartheta$ (Identifizierbarkeit des Modells bei ϑ)
- für alle $x \in \mathbb{R}$ sei $\alpha \longrightarrow f_\alpha(x)$ stetig
- es existiere eine $\mathbb{P}_\vartheta$-integrierbare Zufallsgröße H mit

$$\sup_\alpha |\log f(\alpha, X_1)| \leq H(\omega) \qquad \mathbb{P}_\vartheta\text{-f.s.}$$

Aussage 3.1. *Unter den genannten Bedingungen ist jeder Maximum-Likelihood Schätzer $\hat{\vartheta}_n$ konsistent im Sinne von*

$$\mathbb{P}_\vartheta\big(\| \hat{\vartheta}_n - \vartheta \| > \epsilon\big) \longrightarrow 0 \qquad \forall \epsilon > 0 \quad \forall \vartheta \in \Theta$$

Beweis:

s. Dacunha-Castelle, Duflo II, S. 126 f.

b) ASYMPTOTISCHE NORMALITÄT:

Wir stellen weiter einige Voraussetzungen an unser statistisches Modell.

Definition 3.4. Es sei $(\Omega, \mathscr{A}, \mathscr{P}, X)$ ein statistisches Modell mit $\mathscr{P} = (\mathbb{P}_\vartheta,\ \vartheta \in \Theta)$, $\Theta \subseteq \mathbb{R}^k$ und es sei $\vartheta \in \Theta$.
Dann heißt $(\Omega, \mathscr{A}, \mathscr{P}, X)$ *regulär bei* ϑ, falls Θ eine Umgebung von ϑ ist, und falls $L_n\big(\cdot\,;\, X(\omega)\big)$ wie folgt gewählt werden kann:

H1) In einer Umgebung V von ϑ mit $V \subseteq \Theta$ ist die Funktion $\alpha \longrightarrow L_n\big(\alpha\,;\, x\big)$ für jedes x zweimal stetig differenzierbar.

H2) grad $\log L_n\big(\vartheta\,;\, X(\cdot)\big)$ ist ein zentrierter Zufallsvektor mit endlichen zweiten Momenten bezüglich $\mathbb{P}_\vartheta$.

Außerdem gilt

$$\begin{aligned}
&\mathbb{E}_\vartheta\Big(\frac{\partial}{\partial\vartheta_i}\log L_n(\vartheta\,;\,X)\cdot\frac{\partial}{\partial\vartheta_j}\log L_n(\vartheta\,;\,X)\Big)\\
&=\mathbb{E}_\vartheta\Big(\frac{\partial^2}{\partial\vartheta_i\,\partial\vartheta_j}\log L_n(\vartheta\,;\,X)\Big)\\
&=: I_n^{(i,j)}(\vartheta)
\end{aligned}$$

Die $k\times k$ -Matrix $I_n(\vartheta) := \big(I_n^{(i,j)}(\vartheta)\big)_{i,j=1,\dots,k}$ heißt *Fisher-Informationsmatrix* für ϑ auf der Basis von $X=(X_1,X_2,\dots,X_n)$.

H3) $I_n(\vartheta)$ ist invertierbar.

Wir kehren zum ML-Schätzer zurück, betrachten aber nur den Fall, daß X_1 unter jedem $\mathbb{P}_\vartheta$ eine Dichte f_ϑ besitzt.

$$l_n(\vartheta\,;\,X)=\log L_n(\vartheta\,;\,X)=\sum_{m=1}^n \log f_\vartheta(X_m)$$

Wir setzen

$$\begin{aligned}
&Y_n^i := \frac{\partial}{\partial\vartheta_i} l_n(\vartheta)=\sum_{m=1}^n \frac{\frac{\partial}{\partial\vartheta_i}f_\vartheta(X_m)}{f_\vartheta(X_m)} \qquad \text{und}\\
&Y_n := (Y_n^i)_{i=1,\dots,k} = \operatorname{grad} l_n(\vartheta)
\end{aligned}$$

Die Vektoren

$$\Big(\frac{\frac{\partial}{\partial\vartheta_i}f_\vartheta(X_m)}{f_\vartheta(X_m)}\Big)_{i=1,\dots,k}$$

bilden für $m\geq 1$ bezüglich $\mathbb{P}_\vartheta$ unabhängige, identisch verteilte zentrierte Zufallsvektoren mit der Kovarianzmatrix $I_1(\vartheta)$ (Beachte H3).
Nach dem zentralen Grenzwertsatz für zufällige Vektoren gilt:

$$\frac{1}{\sqrt{n}}Y_n(\vartheta)\xrightarrow{d(P_\vartheta)}\mathcal{N}_k(0\,,\,I_1(\vartheta))$$

(Dacunha-Castelle, Duflo I, Seite 225).

Diese Eigenschaft führt nach einer Reihe weiterer Rechnungen auf die folgende

Aussage 3.2. *Es sei $(\Omega,\mathscr{A},\mathscr{P},X)$ ein an der Stelle $\vartheta\in\Theta$ reguläres statistisches Modell. Die Verteilung $\mathbb{P}_\vartheta^X$ habe bezüglich eines dominierenden Maßes μ die Dichte $f_\vartheta(x)$, $x\in E$, $\vartheta\in\Theta$. Es sei weiterhin $(X_1,X_2,\dots,X_n)$ eine klassische mathematische Stichprobe aus einer nach $\mathbb{P}_\vartheta^{X_1}$ verteilten Grundgesamtheit (d.h., $X_1,X_2,\dots,X_n$ seien unabhängige, identisch nach $\mathbb{P}_\vartheta^{X_1}$ verteilte Zufallsvariablen). Weiterhin gelte*

H4) *Es existiert eine Umgebung V von ϑ, $V \subseteq \Theta$, und eine $\mathbb{P}_\vartheta^{X_1}$-integrierbare Funktion H auf $\mathbb{R}^k$ mit*

$$\Big|\frac{\partial}{\partial\vartheta_i}\frac{\partial}{\partial\vartheta_j}\log f_\vartheta(x)\Big| \leq H(x) \quad \vartheta \in V, \quad i,j = 1,\ldots,n$$

Bezeichnet $\hat{\vartheta}_n(X_1, X_2, \ldots X_n)$ einen Maximum-Likelihood-Schätzer für ϑ, so gelte $\hat{\vartheta}_n \xrightarrow{\mathbb{P}_\vartheta} \vartheta$ (Konsistenz).

Dann haben wir:

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta) \xrightarrow{d(P_\vartheta)} \mathcal{N}\big(0, I^{-1}(\vartheta)\big) \text{ und}$$
$$I(\vartheta)\sqrt{n}(\hat{\vartheta}_n - \vartheta) - \frac{1}{\sqrt{n}}\operatorname{grad} l_n(\vartheta) \xrightarrow{P_\vartheta} 0.$$

Zum Beweis dieser Aussage sei ebenfalls auf Dacunha-Castelle, Duflo II, S. 127, verwiesen.

Beispiele 3.1. a) Normalverteilung

Es seien $X_1, \ldots, X_n$ unabhängige, identisch $\mathcal{N}(\mu, \sigma^2)$-verteilte Zufallsvariablen. Es sei $\vartheta = (\mu, \sigma^2)^T \in \mathbb{R} \times (0, \infty) =: \Theta$
Folglich erhalten wir

$$\log f_\vartheta(x) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2}\frac{(x-\mu)^2}{\sigma^2},$$
$$\operatorname{grad}\log f_\vartheta(x) = \begin{pmatrix} \dfrac{x-\mu}{\sigma^2} \\ -\dfrac{1}{2\sigma^2} + \dfrac{(x-\mu)^2}{2\sigma^4} \end{pmatrix},$$
$$\dot{l}_n(X_1, X_2, \ldots, X_n) = \begin{pmatrix} \displaystyle\sum_{m=1}^n \frac{(X_m - \mu)}{\sigma^2} \\ \displaystyle -\frac{n}{2\sigma^2} + \frac{1}{2}\sum_{m=1}^n \frac{(X_m-\mu)^2}{\sigma^4} \end{pmatrix}$$

$\dot{l}_n = 0$ liefert also die Lösung $\hat{\vartheta}_n(X_1, \ldots, X_n) = (\hat{\mu}_n, \hat{\sigma}_n^2)^T$, wobei

$$\hat{\mu}_n(X_1, X_2, \ldots, X_n) = \frac{1}{n}\sum_{m=1}^n X_m =: \bar{X}_n$$
$$\hat{\sigma}_n^2(X_1, X_2, \ldots, X_n) = \frac{1}{n}\sum_{m=1}^n (X_m - \bar{X}_n)^2$$

Dieses Modell ist regulär im oben genannten Sinne.

b) Verschobene Exponentialverteilung

Es seien $X_1, \dots, X_n$ unabhängige, identisch verteilte Zufallsvariablen mit Dichte

$$f_\vartheta(x) = \mathbf{1}_{[\xi,\infty)}(x)\lambda \exp\{-\lambda(x-\xi)\}, \quad x \in \mathbb{R}\,.$$

Es sei $\vartheta = (\xi, \lambda)^T \in \mathbb{R} \times (0, \infty) =: \Theta$
(Skizzieren Sie die Dichte!)
Die Dichte $f_\vartheta(x)$ ist bei festem x nicht bezüglich ϑ differenzierbar. Bei festem x ist $f_\vartheta(x)$ für $\xi = x$ und $\lambda = \frac{1}{x-\xi}$ maximal. Folglich erhalten wir

$$L_n(\vartheta\,;\, X_1, \dots, X_n) = \mathbf{1}_{[\xi,\,\infty)}\big(\min\{X_1, \dots, X_n\}\big)\lambda^n \exp\big\{-\lambda \sum_{m=1}^{n} X_m + \lambda\, \xi\, n\big\}$$

und somit $\hat\vartheta_n(X_1, \dots, X_n) = \arg\max\limits_{\vartheta\in\Theta} = L_n(\vartheta\,;\, X_1, \dots, X_n)$

$$= \begin{pmatrix} \min\{X_1, \dots, X_n\} \\ \Big(\bar X_n - \min\{X_1, \dots, X_n\}\Big)^{-1} \end{pmatrix},$$

also

$$\hat\xi_n = \min\{X_1, \dots, X_n\} \quad \text{und} \quad \hat\lambda_n = \Big(\bar X_n - \min\{X_1, \dots, X_n\}\Big)^{-1}.$$

Der Schätzer $\hat\vartheta_n$ ist in diesem Fall konsistent aber nicht asymptotisch normalverteilt. Die Regularitätsvoraussetzung H2) ist verletzt.

Ein einfacher Fall stochastischer Prozesse

1 Autoregressives Schema erster Ordnung:
Es sei $(\epsilon_n,\ n \geq 1)$ eine Folge reellwertiger, unabhängiger, identisch verteilter Zufallsgrößen, $X_0 = x_0$ und x_0, α seien reelle Zahlen. Wir definieren

$$X_n = \alpha X_{n-1} + \epsilon_n, \qquad n \geq 1.$$

Die Folge $(X_n\ n \geq 1)$ heißt *autoregressive Folge erster Ordnung* oder *AR(1)-Folge.*
ϵ_1 habe die Dichte f, die überall auf $\mathbb{R}$ positiv sei.
Dann besitzt auch die Stichprobe $X := (X_1, X_2 \dots, X_n)$ eine Dichte

$$f^X(x_1, x_2, \dots, x_n) = \prod_{m=1}^{n} f(x_m - \alpha x_{m-1}) = L_n(\alpha\,;\, x_1, x_2, \dots, x_n)$$

und es gilt

$$l_n(\alpha\,;\,x_1,x_2,\dots,x_n)=\sum_{m=1}^{n}\log f(x_m-\alpha x_{m-1}),\quad \dot{l}_n=-\sum_{m=1}^{n}x_{m-1}\frac{\dot{f}(x_m-\alpha x_{m-1})}{f(x_m-\alpha x_{m-1})}.$$

Die ML-Gleichung lautet:

$$\sum_{m=1}^{n}\frac{X_{m-1}\dot{f}(X_m-\hat{\alpha}_n X_{m-1})}{f(X_m-\hat{\alpha}_n X_{m-1})}=0.$$

Im Spezialfall $\epsilon_1\sim\mathcal{N}(0,1)$ gilt

$$l_n(\alpha\,;\,X_1,X_2,\dots,X_n)=\alpha\sum_{m=1}^{n}X_mX_{m-1}-\frac{\alpha^2}{2}\sum_{m=1}^{n}X_{m-1}^2$$

$$\text{und indem man}\quad \dot{l}_n=\sum_{m=1}^{n}X_mX_{m-1}-\alpha\sum_{m=1}^{n}X_{m-1}^2=0\quad\text{setzt,}$$

bekommt man einen Maximum-Likelihood-Schätzer für α:

$$\hat{\alpha}_n=\frac{\sum_{m=1}^{n}X_mX_{m-1}}{\sum_{m=1}^{n}X_{m-1}^2}.$$

Es gilt

$$\hat{\alpha}_n-\alpha=\frac{\sum_{m=1}^{n}X_{m-1}(X_m-\alpha X_{m-1})}{\sum_{m=1}^{n}X_{m-1}^2}=\frac{\sum_{m=1}^{n}X_{m-1}\epsilon_m}{\sum_{m=1}^{n}X_{m-1}^2}.$$

Man beachte, daß

$$\Big(\sum_{m=1}^{n}X_{m-1}\epsilon_m\Big)_{n\geq 1}\ \text{ein Martingal ist und}\ \Big(\sum_{m=1}^{n}X_{m-1}^2\Big)_{n\geq 1}\ \text{seine bedingte Varianz darstellt:}$$

$$Var_{\wp_n}\big(\sum_{m=1}^{n}X_{m-1}\epsilon_m\big)=\mathbb{E}_\vartheta\Big(\big(\sum_{m=1}^{n}X_{m-1}\epsilon_m\big)^2\Big|\wp_{n-1}\Big)=\sum_{m=1}^{n}X_{m-1}^2.$$

$$\big(\wp_n:=\sigma(X_1,X_2,\dots,X_n)=\sigma(\epsilon_1,\epsilon_2,\dots,\epsilon_n),\quad n\geq 1\big)$$

Für die Untersuchung der asymptotischen Eigenschaften von $(\hat{\alpha}_n,\ n\geq 1)$ für $n\longrightarrow\infty$ bietet sich also die Martingaltheorie an.