Kapitel IV

L^2-differenzierbare statistische Modelle

Reinhard Höpfner

Vorlesung Mathematische Statistik

Wintersemester 2004/2005 und Sommersemester 2008

Institut für Mathematik, Johannes Gutenberg Universität Mainz

13.01.05 und 09.05.08

Übersicht zu Kapitel IV :

A. L^r-differenzierbare statistische Modelle

B. Le Cam’s Zweites Lemma für iid Beobachtungen

In diesem Kapitel diskutieren wir 'Glattheit' allgemeiner statistischer Modelle in einem L^2–Sinn. Dies wird eine Verallgemeinerung der in 1.2 definierten Begriffe Score und Information erlauben. Zugleich wird man sehen, daß in iid Modellen unter schwachen Voraussetzungen eine Entwicklung der log-Likelihoodratios in Termen von Score und Information lokal in kleinen Umgebungen eines beliebigen Punktes ϑ möglich wird.

A. L^r-differenzierbare statistische Modelle

4.1 Vorbemerkung: Betrachte zuerst ein dominiertes Experiment

$$(\Omega, \mathcal{A}, \mathcal{P} := \{P_\vartheta : \vartheta \in \Theta\}) \ , \quad \Theta \subset \mathbb{R}^d \text{ offen} \ , \quad \mathcal{P} \ll \nu$$

mit Dichten

$$\chi_\vartheta \ := \ \frac{dP_\vartheta}{d\nu} \ , \quad \vartheta \in \Theta \ .$$

Sei $r \geq 1$ fest. Trivialerweise gilt $\chi_\vartheta^{1/r} \in L^r(\Omega, \mathcal{A}, \nu)$ für alle $\vartheta \in \Theta$.

1) Im üblichen (Fréchét-) Sinn im normierten Raum $L^r(\Omega, \mathcal{A}, \nu)$ heißt die Abbildung

$$(\diamond) \qquad \Theta \ \ni \quad \xi \ \longrightarrow \ \chi_\xi^{1/r} \quad \in \ L^r(\Omega, \mathcal{A}, \nu)$$

differenzierbar an der Stelle $\xi = \vartheta$ *mit Ableitung* $\widetilde{V}_\vartheta$ wenn (+) und (++) gelten:

$$(+) \qquad \widetilde{V}_\vartheta \ \text{ hat Komponenten } \ \widetilde{V}_{\vartheta,1}, \ldots, \widetilde{V}_{\vartheta,d} \ \in L^r(\Omega, \mathcal{A}, \nu) \ ,$$

$$(++) \qquad \frac{1}{|\xi - \vartheta|} \ \left\| \chi_\xi^{1/r} - \chi_\vartheta^{1/r} - (\xi - \vartheta)^\top \widetilde{V}_\vartheta \right\|_{L^r(\nu)} \ \longrightarrow \ 0 \quad \text{für} \quad |\xi - \vartheta| \longrightarrow 0 \ .$$

2) Wir zeigen: In statistischen Modellen folgt aus (+) und (++) notwendig

$$(*) \qquad \widetilde{V}_{\vartheta,i} \ = \ 0 \quad \nu\text{-fast sicher auf } \{\chi_\vartheta = 0\} \ , \ 1 \leq i \leq d \ .$$

Zum Nachweis dieser Behauptung betrachtet man Folgen $(\xi_n)_n$ in Θ von Gestalt $\xi_n = \vartheta \pm \delta_n e_i$, wobei e_i den i-ten Einheitsvektor in $\mathbb{R}^d$ bezeichnet, und $(\delta_n)_n$ eine beliebige strikt positive Nullfolge. Dann zeigt (++) für $n \to \infty$

$$\left\| 1_{\{\chi_\vartheta = 0\}} \left[\frac{1}{\delta_n} \chi_{\vartheta + \delta_n e_i}^{1/r} \ - \ \widetilde{V}_{\vartheta,i} \right] \right\|_{L^r(\nu)} \ \longrightarrow \ 0 \quad , \quad \left\| 1_{\{\chi_\vartheta = 0\}} \left[\frac{1}{\delta_n} \chi_{\vartheta - \delta_n e_i}^{1/r} \ + \ \widetilde{V}_{\vartheta,i} \right] \right\|_{L^r(\nu)} \ \longrightarrow \ 0 \ .$$

Aus der $L^r(\nu)$-Konvergenz erhält man nach Auswahl geeigneter Teilfolgen eine ν-fast sichere Konvergenz entlang dieser Teilfolgen: also gibt es $(n_k)_k$ mit

$$\text{auf } \{\chi_\vartheta = 0\} \quad : \quad \frac{1}{\delta_{n_k}} \chi_{\vartheta + \delta_{n_k} e_i}^{1/r} \ \longrightarrow \ \widetilde{V}_{\vartheta,i} \ , \quad \frac{1}{\delta_{n_k}} \chi_{\vartheta - \delta_{n_k} e_i}^{1/r} \ \longrightarrow \ -\widetilde{V}_{\vartheta,i} \ , \quad k \to \infty \ .$$

Die links stehenden Dichten sind stets nichtnegativ, also sind auf dem Ereignis $\{\chi_\vartheta = 0\}$ sowohl $\widetilde{V}_{\vartheta,i}$ als auch $-\widetilde{V}_{\vartheta,i}$ nichtnegativ, also muß die Ableitung $\widetilde{V}_\vartheta$ die Eigenschaft $(*)$ besitzen.

3) Definiere nun ein V_ϑ mit Komponenten $V_{\vartheta,1}, \ldots, V_{\vartheta,d}$ durch

$$V_{\vartheta,i} \;:=\; 1_{\{\chi_\vartheta > 0\}}\, r\, \chi_\vartheta^{-1/r}\, \widetilde{V}_{\vartheta,i}\;, \quad 1 \le i \le d\,.$$

Dann gilt $V_{\vartheta,i} \in L^r(\Omega, \mathcal{A}, P_\vartheta)$ für $1 \le i \le d$, und wegen $(*)$ kann man nach Abänderung von $\widetilde{V}_\vartheta$ auf einer ν-Nullmenge in $\mathcal{A}$ auch

$$\widetilde{V}_{\vartheta,i} \;=\; \frac{1}{r}\, \chi_\vartheta^{1/r}\, V_{\vartheta,i}\;, \quad 1 \le i \le d$$

schreiben. Man ist also von $\widetilde{V}_{\vartheta,i} \in L^r(\Omega, \mathcal{A}, \nu)$ zu neuen Zufallsvariablen $V_{\vartheta,i} \in L^r(\Omega, \mathcal{A}, P_\vartheta)$ übergegangen, mit denen Fréchét-Differenzierbarkeit der Abbildung $(\diamond)$ aus Schritt 1) die folgende Gestalt annimmt.

4.1' Definition: Sei $\Theta \in I\!R^d$ offen. Sei $\mathcal{P} = \{P_\xi : \xi \in \Theta\}$ eine dominierte Familie von Wahrscheinlichkeitsmaßen auf $(\Omega, \mathcal{A})$ mit ν-Dichten χ_ξ, $\xi \in \Theta$. Sei $r \ge 1$. Dann heißt $\mathcal{P}$ *L^r-differenzierbar in ϑ mit Ableitung V_ϑ* falls gilt:

i) $$V_\vartheta \text{ hat Komponenten } \quad V_{\vartheta,1}, \ldots V_{\vartheta,d} \;\in\; L^r(\Omega, \mathcal{A}, P_\vartheta)\;,$$

ii) $$\frac{1}{|\xi - \vartheta|^r} \int_\Omega \left| \chi_\xi^{1/r} - \chi_\vartheta^{1/r} - \frac{1}{r}\, \chi_\vartheta^{1/r}\, (\xi - \vartheta)^\top V_\vartheta \right|^r d\nu \;\longrightarrow\; 0 \quad , \quad |\xi - \vartheta| \to 0\,.$$

4.1" Bemerkungen: a) Hat man für ein $r \ge 1$ die L^r-Differenzierbarkeit eines statistischen Modells an der Stelle ϑ im Sinn von 4.1', und ist zusätzlich für jedes feste $\omega \in \Omega$ die Abbildung

$(\circ)$ $$\Theta \;\ni\; \xi \;\longrightarrow\; \chi_\xi(\omega) \quad \in\; [0, \infty)$$

stetig und partiell differenzierbar, so erhält man wie in der klassischen Definition 1.2

$$V_{\vartheta,i}(\omega) \;=\; 1_{\{\chi_\vartheta > 0\}}(\omega)\, \left(\frac{\partial}{\partial \vartheta_i} \log \chi \right) (\vartheta, \omega)\;, \quad 1 \le i \le d$$

ν–fast sicher (und damit P_ϑ–fast sicher), für jedes $\vartheta \in \Theta$.

Dies sieht man, indem man in 4.1' Folgen $\xi_n = \vartheta + \delta_n e_i$ betrachtet (e_i der i-te Einheitsvektor in $I\!R^d$, und $\delta_n \downarrow 0$) und Teilfolgen $(n_k)_k$ auswählt, entlang der sogar ν-fast sichere Konvergenz

$$\left| \frac{\chi_{\vartheta + \delta_{n_k} e_i}^{1/r} - \chi_\vartheta^{1/r}}{\delta_{n_k}} \;-\; \frac{1}{r} \chi_\vartheta^{1/r} V_{\vartheta,i} \right|^r \;\longrightarrow\; 0\,, \quad k \to \infty$$

gilt. Unter ($\circ$) hat man dann $\nabla(\chi^{1/r})(\vartheta,\cdot) = \frac{1}{r}\chi_\vartheta^{1/r} V_\vartheta$ ν–fast sicher, wobei auf dem Ereignis $\{\chi_\vartheta = 0\}$ notwendig $\nabla(\chi^{1/r})(\vartheta,\cdot) = \widetilde{V}_\vartheta$ nach ($*$) aus 4.1 verschwindet, und damit $r\,1_{\{\chi_\vartheta=0\}}\,\nabla(\log(\chi^{1/r}))(\vartheta,\cdot) = V_\vartheta$. Der letzte Ausdruck aber ist unabhängig von $r \ge 1$.
b) Als Beispiel einer für jedes $r \ge 1$ an jeder Stelle ϑ L^r-differenzierbaren Verteilungsfamilie, welche nicht punktweise im Sinne von a) glatt ist, betrachte man das von der symmetrisierten Exponentialverteilung $F(dx) = \frac{1}{2}e^{-|x|}dx$ erzeugte Lokationsmodell $\{F_\vartheta(\cdot) := F(\cdot - \vartheta) : \vartheta \in I\!R\}$. Die Ableitung an der Stelle ϑ ist $V_\vartheta = -\frac{1}{2}\,\mathrm{sgn}(\cdot - \vartheta)$.
c) Das Integral in 4.1' ii) – und damit auch die Gestalt der Ableitung V_ϑ in 4.1' i) – ist unabhängig von der Wahl des dominierenden Maßes ν. In der Tat, betrachtet man zwei dominierende Maße $\mathcal{P} \ll \nu \ll \widetilde{\nu}$, und Dichten χ_ϑ bezüglich ν, $\widetilde{\chi}_\vartheta$ bezüglich $\widetilde{\nu}$, so gilt $\widetilde{\chi}_\vartheta = \chi_\vartheta \frac{d\nu}{d\widetilde{\nu}}$, und der Faktor $\frac{d\nu}{d\widetilde{\nu}}$ kürzt sich aus der 4.1' ii) entsprechenden Darstellung mit $\widetilde{\chi}_\xi, \widetilde{\chi}_\vartheta, \widetilde{\nu}$ sofort heraus.

Indem man nun noch die in 4.1' gemachte Voraussetzung der Dominiertheit vermeidet, kommt man zu der allgemeinen Definition von L^r-Differenzierbarkeit in einem statistischen Modell.

4.2 Definition: Sei $\Theta \in I\!R^d$ offen. Sei $\mathcal{P} = \{P_\xi : \xi \in \Theta\}$ eine (nicht notwendig dominierte) Familie von Wahrscheinlichkeitsmaßen auf $(\Omega, \mathcal{A})$. Für $\xi', \xi \in \Theta$ bezeichne $L^{\xi'/\xi}$ eine Festlegung der Likelihoodratio von $P_{\xi'}$ bezüglich P_ξ wie in 3.1 a) definiert. Sei $r \ge 1$.
Dann heißt $\mathcal{P}$ *L^r-differenzierbar in ϑ mit Ableitung V_ϑ* falls gilt:

i) $$V_\vartheta \text{ hat Komponenten } \quad V_{\vartheta,1}, \ldots V_{\vartheta,d} \;\in\; L^r(\Omega, \mathcal{A}, P_\vartheta)\,,$$

ii)a) $$\frac{1}{|\xi - \vartheta|^r} P_\xi\left(L^{\xi/\vartheta} = \infty\right) \longrightarrow 0 \quad , \quad |\xi - \vartheta| \to 0\,,$$

ii)b) $$\frac{1}{|\xi - \vartheta|^r} \int_\Omega \left| \left(L^{\xi/\vartheta}\right)^{1/r} - 1 - \frac{1}{r}(\xi - \vartheta)^\top V_\vartheta \right|^r dP_\vartheta \longrightarrow 0 \quad , \quad |\xi - \vartheta| \to 0\,.$$

4.2' Bemerkungen: a) Man überzeugt sich, daß es genügt, die Bedingungen ii)a) und ii)b) entlang von *Folgen* $(\xi_n)_n \subset \Theta$ nachzuprüfen, welche gegen ϑ konvergieren.
b) Sei $(\xi_n)_n$ eine gegen ϑ strebende Folge. Dann ist die Teilfamilie $\{P_{\xi_n}, n \ge 1,\, P_\vartheta\}$ abzählbar und damit dominiert (etwa durch $\nu := P_\vartheta + \sum_{n\ge1} 2^{-n} P_{\xi_n}$), und man wählt Dichten $\chi_{\xi_n}, n \ge 1,\, \chi_\vartheta$ von $P_{\xi_n}, n \ge 1,\, P_\vartheta$ bezüglich ν. Innerhalb dieser Teilfamilie schreibt man

$$\int_\Omega \left| \chi_{\xi_n}^{1/r} - \chi_\vartheta^{1/r} - \frac{1}{r}\chi_\vartheta^{1/r}(\xi_n - \vartheta)^\top V_\vartheta \right|^r d\nu \quad = \quad \int_{\{\chi_\vartheta = 0\}} \ldots + \int_{\{\chi_\vartheta > 0\}} \ldots$$

$$= \int_{\{\chi_\vartheta=0\}} dP_{\xi_n} + \int_{\{\chi_\vartheta>0\}} \left| \left(\frac{\chi_{\xi_n}}{\chi_\vartheta}\right)^{1/r} - 1 - \frac{1}{r}(\xi_n-\vartheta)^\top V_\vartheta \right|^r \chi_\vartheta \, d\nu$$

$$= P_{\xi_n}\left(L^{\xi_n/\vartheta}=\infty\right) + \int_\Omega \left| \left(L^{\xi_n/\vartheta}\right)^{1/r} - 1 - \frac{1}{r}(\xi_n-\vartheta)^\top V_\vartheta \right|^r dP_\vartheta$$

da nach 3.1 gilt

$$L^{\xi_n/\vartheta} = \frac{\chi_{\xi_n}}{\chi_\vartheta} 1_{\{\chi_\vartheta>0\}} + \infty \, 1_{\{\chi_\xi=0\}} \quad (P_{\xi_n}+P_\vartheta)\text{-fast sicher.}$$

Also sind *in Restriktion auf die Teilfamilie* $\{P_{\xi_n}, n \geq 1, P_\vartheta\}$ die Eigenschaften 4.2 ii)a)+b) äquivalent zur Eigenschaft 4.1' ii). Dies gilt für beliebige Folgen $\xi_n \to \vartheta$. Also sind die Definitionen 4.1' und 4.2 konsistent.

4.3 Beispiel: Sei $(\Omega, \mathcal{A})$ beliebig, schreibe $\mathcal{M}^1(\Omega, \mathcal{A})$ für die Menge aller Wahrscheinlichkeitsmaße auf $(\Omega, \mathcal{A})$. Sei $r \geq 1$ fest. In Analogie zu Beispiel 1.3 betrachten wir im nichtparametrischen Modell $\mathcal{E} = (\Omega, \mathcal{A}, \mathcal{P})$ *einparametrige Pfade* durch einen beliebig festgelegten Punkt $P \in \mathcal{P}$, in Richtungen

$$g \in L^r(P) \quad \text{mit} \quad \int_\Omega g \, dP = 0 \, .$$

a) Sei zuerst g beschränkt auf Ω, setze $M := \sup |g|$. Dann ist die durch

$$(*) \qquad \mathcal{S}_P^g := \left(\Omega, \mathcal{A}, \left\{Q_h : |h| < \frac{1}{M}\right\}\right), \quad dQ_h := (1+h\,g)\,dP$$

definierte einparametrige Familie L^r-differenzierbar an jeder Stelle $|h| < \frac{1}{M}$ mit Ableitung

$$V_h = \frac{g}{1+h\,g} \quad \text{beschränkt und damit in } L^r(\Omega, \mathcal{A}, P_\vartheta);$$

dies erhält man mit Zwischenwertsatz und dominierter Konvergenz aus

$$\lim_{h'\to h} \frac{(1+h'\,g)^{1/r} - (1+h\,g)^{1/r}}{h'-h} = \frac{1}{r}(1+h\,g)^{\frac{1}{r}-1} g \, .$$

Statistisch am wichtigsten ist die Ableitung $V_0 = g$ an der Stelle $h = 0$.

b) Genau wie im zweiten Teil des Beispiels 1.3 kann man durch Trunkationsargumente die Beschränktheitsvoraussetzung an die Funktion g vermeiden, und erhält so für jede Richtung $g \in L^r(P)$ mit $E_P(g) = 0$ einen einparametrigen Pfad $\mathcal{S}_P^g$ durch P in Richtung g, welcher an der Stelle $h = 0$ L^r-differenzierbar ist mit Ableitung $V_0 = g$. □

Der folgende Satz ist – gegeben die Diskussion in 4.1 und 4.2' – eher klassisch. Einen Beweis findet man in Witting (1985, p. 175–176).

4.4 Satz: Ist für ein $r \geq 1$ die Familie $\mathcal{P} = \{P_\xi : \xi \in \Theta\}$ L^r-differenzierbar in $\xi = \vartheta$ mit Ableitung V_ϑ, so ist $\mathcal{P}$ auch für jedes $1 \leq s \leq r$ L^s-differenzierbar in $\xi = \vartheta$, mit derselben Ableitung V_ϑ.

Notwendig sind die in 4.1' oder 4.2 auftauchenden Ableitungen zentriert:

4.5 Satz: Ist $\mathcal{P} = \{P_\xi : \xi \in \Theta\}$ L^1-differenzierbar in $\xi = \vartheta$ mit Ableitung V_ϑ, so gilt $E_\vartheta(V_\vartheta) = 0$.

Beweis: Betrachte einen Einheitsvektor e_i in $\mathbb{R}^d$, eine Nullfolge $\delta_n \downarrow 0$, und wähle ein dominierendes Maß ν für $\{P_{\vartheta+\delta_n e_i}, n \geq 1, P_\vartheta\}$ sowie zugehörige Dichten. Aus

$$\frac{\chi_{\vartheta+\delta_n e_i} - \chi_\vartheta}{\delta_n} - \chi_\vartheta V_{\vartheta,i} \quad \longrightarrow \quad 0 \quad \text{in } L^1(\nu) \text{ für } n \to \infty$$

erhält man $\int \chi_\vartheta V_{\vartheta,i}\, d\nu = 0$ und also $E_{P_\vartheta}(V_{\vartheta,i}) = 0$. □

Unter L^2-Differenzierbarkeit an der Stelle ϑ ist die Ableitung V_ϑ – zentriert nach 4.4+4.5 – per Voraussetzung in $L^2(P_\vartheta)$. Unter der klassischen Zusatzvoraussetzung ($\circ$) aus 4.1 ist damit die L^2–Ableitung V_ϑ im Sinne der Definition 1.2 der Score in ϑ, und $E_\vartheta(V_\vartheta V_\vartheta^\top)$ die Fisher-Information in ϑ. Die folgende Verallgemeinerung von 1.2 verzichtet auf ($\circ$).

4.6 Definition: Sei $\mathcal{P} = \{P_\xi : \xi \in \Theta\}$ L^2-differenzierbar in $\xi = \vartheta$ mit Ableitung V_ϑ. Dann nennt man V_ϑ den *Score in* ϑ und

$$J_\vartheta \;:=\; E_\vartheta\left(V_\vartheta\, V_\vartheta^\top\right)$$

die *Fisher-Information in* ϑ.

Beschreibt man die Geometrie eines L^2-differenzierbaren statistischen Experimentes durch den Hellinger-Abstand (der bereits in Teilkapitel 1C benutzt wurde), so kommt der Fisher-Information ein herausregende Bedeutung zu: sie wird ausdrücken (vgl. Satz 4.8 unten), wie gut Parameterwerte ξ von ϑ unterschieden werden können, falls sie sich aus einer festen Richtung u kommend dem Referenzpunkt ϑ nähern.

4.6' Definition: Sind Q und Q' zwei Wahrscheinlichkeitsmaße auf $(\Omega, \mathcal{A})$, so heißt

$$H^2(Q,Q') \;:=\; \int \frac{1}{2}\left(\sqrt{\chi'} - \sqrt{\chi}\right)^2 d\nu \;=\; \int \frac{1}{2}\left(\sqrt{dQ'} - \sqrt{dQ}\right)^2$$

(für ein beliebiges Q, Q' dominierendes Maß ν auf $(\Omega, \mathcal{A})$, und beliebige Festlegungen χ, χ' der Dichten $\frac{dQ}{d\nu}, \frac{dQ'}{d\nu}$) *quadrierter Hellinger-Abstand zwischen* Q *und* Q'.

Bei dem letzten Ausdruck auf der rechten Seite handelt es sich um eine symbolische Schreibweise, die die Unabhängigkeit der Größe $H^2(Q,Q')$ von der Wahl des Q, Q' diminierenden Maßes ausdrückt. Der Hellinger Abstand ist eine Metrik auf dem Raum $\mathcal{M}^1(\Omega, \mathcal{A})$ aller Wahrscheinlichkeitsmaße auf $(\Omega, \mathcal{A})$, siehe auch Strasser (1985, Ch. I.2) und Tsybakov (2004, Ch. 2.4) für Abstandsbegriffe für Wahrscheinlichkeitsmaße. Man beachte, daß durch Wahl des Faktors $\frac{1}{2}$ in Definition 4.6' der Hellinger-Abstand durch 1 nach oben beschränkt ist.

4.7 Hilfssatz: In jedem statistischen Modell $\mathcal{P} = \{P_\xi : \xi \in \Theta\}$ gilt ohne weitere Voraussetzungen

$$\sqrt{L^{\xi'/\xi}} \in L^2(P_\xi) \text{ , und } \quad E_\xi\left(\sqrt{L^{\xi'/\xi}} - 1\right) \;=\; -H^2(P_{\xi'}, P_\xi) \;.$$

Beweis: 1) Stets gilt $\sqrt{L^{\xi'/\xi}} \geq 0$. Als Dichtequotient stimmt $L^{\xi'/\xi}$ nach 3.1 für jedes $P_{\xi'}$ und P_ξ dominierende Maß ν auf $(\Omega, \mathcal{A})$ und für beliebige Festlegungen $\chi_{\xi'}, \chi_\xi$ von ν–Dichten $(P_\xi + P_{\xi'})$–fast sicher überein mit $\frac{\chi_{\xi'}}{\chi_\xi} 1_{\{\chi_\xi > 0\}} + \infty\, 1_{\{\chi_\xi = 0\}}$. Also ist die Aussage $\sqrt{L^{\xi'/\xi}} \in L^2(P_\xi)$ trivial. Insbesondere sind alle Erwartungswerte in der untenstehenden Zeile (+) wohldefiniert.
2) Die elementare Gleichung $\sqrt{a} - 1 = \frac{1}{2}(a-1) - \frac{1}{2}(\sqrt{a}-1)^2$ für $a \geq 0$ zeigt nun

$$(+) \qquad E_\xi\left(\sqrt{L^{\xi'/\xi}} - 1\right) \;=\; \frac{1}{2} E_\xi\left(L^{\xi'/\xi} - 1\right) \;-\; \frac{1}{2} E_\xi\left(\left(\sqrt{L^{\xi'/\xi}} - 1\right)^2\right) \;.$$

Wie im Beweis von 3.10 liefert aber

$$E_\xi\left(1 - L^{\xi'/\xi}\right) \;=\; P_{\xi'}\left(L^{\xi'/\xi} = \infty\right)$$

das Gewicht des P_ξ-singulären Anteils von $P_{\xi'}$, und dessen Träger stimmt $(P_{\xi'} + P_\xi)$-fast sicher überein mit $\{\chi_\xi = 0\}$. Also steht auf der rechten Seite von (+)

$$-\frac{1}{2} P_{\xi'}\left(\chi_\xi = 0\right) \;-\; \frac{1}{2} \int_{\{\chi_\xi > 0\}} \left(\sqrt{\chi_{\xi'}} - \sqrt{\chi_\xi}\right)^2 d\nu \;=\; -\frac{1}{2} \int_\Omega \left(\sqrt{\chi_{\xi'}} - \sqrt{\chi_\xi}\right)^2 d\nu$$

womit die Aussage bewiesen ist. □

4.8 Satz: Sei $\mathcal{P} = \{P_\xi : \xi \in \Theta\}$ L^2-differenzierbar in $\xi = \vartheta$ mit Ableitung V_ϑ. Für jede Folge $(\xi_m)_m$ in Θ mit den Eigenschaften

$$\delta_m := |\xi_m - \vartheta| \to 0 \quad , \quad u_m := \frac{\xi_m - \vartheta}{|\xi_m - \vartheta|} \longrightarrow u \in S_{d-1}$$

(mit Schreibweise S_{d-1} für die Einheitssphäre in $\mathbb{R}^d$) gilt für $m \to \infty$:

i) $$E_\vartheta\left(\left(\sqrt{L^{\xi_m/\vartheta}} - 1\right)^2\right) \sim \delta_m^2 \, \frac{1}{4} \, u^\top J_\vartheta \, u$$

ii) $$H^2\left(P_{\xi_m}, P_\vartheta\right) \sim \delta_m^2 \, \frac{1}{8} \, u^\top J_\vartheta \, u \, .$$

Beweis: Seien ν, χ_{ξ_m}, χ_ϑ wie in 4.2' b) gewählt. Für $m \to \infty$ approximiert man wegen L^2-Differenzierbarkeit in ϑ zunächst

$$\frac{\sqrt{\chi_{\xi_m}} - \sqrt{\chi_\vartheta}}{|\xi_m - \vartheta|} \quad \text{durch} \quad \frac{1}{2}\sqrt{\chi_\vartheta}\, u_m^\top V_\vartheta \quad \text{in } L^2(\nu) \, .$$

Dabei gilt $V_\vartheta \in L^2(P_\vartheta)$. Da die Folge $(\xi_m)_m$ den Punkt ϑ aus der Richtung u anstrebt, kann weiter in $L^2(\nu)$

$$\frac{1}{2}\sqrt{\chi_\vartheta}\, u_m^\top V_\vartheta \quad \text{durch} \quad \frac{1}{2}\sqrt{\chi_\vartheta}\, u^\top V_\vartheta$$

ersetzt werden. Hieraus folgt insbesondere für $m \to \infty$

$$\frac{2}{\delta_m^2} H^2\left(P_{\xi_m}, P_\vartheta\right) = \int_\Omega \left(\frac{\sqrt{\chi_{\xi_m}} - \sqrt{\chi_\vartheta}}{|\xi_m - \vartheta|}\right)^2 d\nu \longrightarrow \frac{1}{4} E_\vartheta\left(\left(u^\top V_\vartheta\right)^2\right) = \frac{1}{4} u^\top J_\vartheta \, u$$

nach Definition der Fisher-Information in 4.6. Damit ist die Behauptung ii) des Satzes bewiesen. Weiter gilt nach Voraussetzung in 4.2 ii)a)

$$P_{\xi_m}\left(L^{\xi_m/\vartheta} = \infty\right) = o(\delta_m^2)$$

und daher für $m \to \infty$ auch

$$\begin{aligned} 2\, H^2\left(P_{\xi_m}, P_\vartheta\right) &= \int_{\{\chi_\vartheta > 0\}} \left(\sqrt{\chi_{\xi_m}} - \sqrt{\chi_\vartheta}\right)^2 d\nu + P_{\xi_m}\left(L^{\xi_m/\vartheta} = \infty\right) \\ &= E_\vartheta\left(\left(\sqrt{L^{\xi_m/\vartheta}} - 1\right)^2\right) + o(\delta_m^2) \, . \end{aligned}$$

Daraus folgt wegen ii) auch die Behauptung i) des Satzes. □

Man beachte, daß wegen Kompaktheit der Sphäre S_{d-1} zu jeder gegen ϑ konvergenten Folge $(\xi_n)_n \subset \Theta$ Teilfolgen $(\xi_m)_m$ und Richtungen $u \in S_{d-1}$ existieren, so daß die Voraussetzungen des Satzes nach Übergang zu Teilfolgen erfüllt sind.

4.9 Bemerkungen: a) Der Kern der Aussage von Satz 4.8 ist: lokal an jeder Stelle ϑ, an der das Modell L^2-differenzierbar ist, wird die Geometrie des statistischen Experimentes (im Sinne des Hellinger-Abstandes) durch die Fisher-Information festgelegt.

b) Als Kovarianzmatrix ist die Fisher-Information J_ϑ stets symmetrisch und nichtnegativ definit. Besitzt J_ϑ einen Eigenwert 0, so kann der Hellinger-Abstand Folgen von Parameterwerten $(\xi_m)_m$, die den Punkt ϑ aus Eigenrichtungen u zum Eigenwert 0 anstreben, nicht mit Rate $|\xi_m - \vartheta|$, $m \to \infty$, von ϑ unterscheiden. An dieser Stelle wird klar, welche Bedeutung der *Invertierbarkeit* der Fisher-Information zukommt.

c) Ist die Fisher-Information an der Stelle ϑ strikt positiv definit, so stimmt die Geometrie des statistischen Experiments (im Sinne des Hellinger-Abstandes) lokal in ϑ bis auf Multiplikation mit geeigneten Konstanten mit dem gewöhnlichen euklidischen Abstand auf $\Theta \subset I\!R^d$ überein.

B. Le Cam's Zweites Lemma für iid Beobachtungen

Le Cam (1969, 1990) hat gezeigt: L^2-Differenzierbarkeit eines statistischen Modells an einer Stelle ϑ impliziert, daß bei unabhängiger Versuchswiederholung lokal in kleinen Umgebungen von ϑ log-Likelihoodratios asymptotisch eine einfache Struktur aufweisen (wir werden in Kapitel V sehen, daß diese die Struktur eines 'schönen' Normalverteilungsexperiments ist). Ergebnisse dieses Typs nennt man oft ein 'zweites Le Cam'sches Lemma'. Auch in allgemeinerem statistischen Kontext als den hier betrachteten iid Modellen kann man 'zweite Le Cam'sches Lemmata' beweisen (ergodische Markovprozesse unter zeitkontinuierlicher oder zeitdiskreter Beobachtung, Punktprozeßmodelle, ...).

4.10 Voraussetzungen für Le Cam's Zweites Lemma: a) Im Einzelexperiment

$$\mathcal{E} \;=\; (\Omega\,,\ \mathcal{A}\,,\ \mathcal{P} = \{P_\xi : \xi \in \Theta\})\ , \quad \Theta \subset I\!R^d \quad \text{offen}$$

bezeichnet $L^{\xi'/\xi}$ die Likelihoodratio von $P_{\xi'}$ bezüglich P_ξ.

b) Wird ein Punkt des Parameterraumes Θ mit ϑ bezeichnet, so wird für ϑ vorausgesetzt:

$$\mathcal{P} = \{P_\xi : \xi \in \Theta\} \text{ ist } L^2\text{-differenzierbar in } \xi = \vartheta \text{ mit Ableitung } V_\vartheta .$$

Die Fisher-Information an der Stelle ϑ im Einzelversuch ist $J_\vartheta = E_\vartheta\left(V_\vartheta\, V_\vartheta^\top\right)$.

c) Im Produktexperiment

$$\mathcal{E}_n \;=\; \left(\Omega_n := \mathop{\times}_{i=1}^{n} \Omega\,,\ \mathcal{A}_n := \bigotimes_{i=1}^{n} \mathcal{A}\,,\ \left\{P_\xi^n := \bigotimes_{i=1}^{n} P_\xi : \xi \in \Theta\right\}\right)\ ;$$

werden Likelihoodratio und log-Likelihoodratio von $P_{\xi'}^n$ bezüglich P_ξ^n mit $L_n^{\xi'/\xi}$ und $\Lambda_n^{\xi'/\xi}$ bezeichnet. Punkte in Ω_n werden als $\omega = (\omega_1, ..., \omega_n)$ geschrieben.

4.11 Le Cam's Zweites Lemma: Mit Voraussetzungen und Bezeichnungen aus 4.10 gilt

$$\Lambda_n^{(\vartheta + h_n/\sqrt{n})/\vartheta} \;=\; h_n^\top S_n(\vartheta) \;-\; \frac{1}{2}\, h_n^\top J_\vartheta\, h_n \;+\; o_{(P_\vartheta^n)}(1)\ , \quad n \to \infty$$

für jede beschränkte Folge $(h_n)_n$ in $I\!R^d$, wobei

$$S_n(\vartheta)(\omega_1, ..., \omega_n) \;:=\; \frac{1}{\sqrt{n}} \sum_{j=1}^{n} V_\vartheta(\omega_j)\ , \quad \omega = (\omega_1, ..., \omega_n) \in \Omega_n$$

der mit $\frac{1}{\sqrt{n}}$ skalierte Score im Produktexperiment an der Stelle ϑ ist, und folglich gilt

$$\mathcal{L}\left(S_n(\vartheta) \mid P_\vartheta^n\right) \;\longrightarrow\; \mathcal{N}\left(0, J_\vartheta\right)$$

(schwache Konvergenz in $I\!R^d$, für $n \to \infty$), nach Definition der Fisher-Information in 4.6.

Der Beweis (in 4.15 unten) wird durch eine Serie von Hilfssätzen vorbereitet, stets unter den Voraussetzungen 4.10.

4.12 Hilfssatz: Für jede beschränkte Folge $(h_n)_n$ in $I\!R^d$ und $\xi_n := \vartheta + h_n/\sqrt{n}$ gilt in $\mathcal{E}_n$

$$\sum_{j=1}^{n} \left(\sqrt{L^{\xi_n/\vartheta}}(\omega_j) - 1\right) \;=\; \frac{1}{2\sqrt{n}} \sum_{j=1}^{n} h_n^\top V_\vartheta(\omega_j) \;-\; \frac{1}{8}\, h_n^\top J_\vartheta\, h_n \;+\; \varrho_n(\omega)$$

wobei $\varrho_n = o_{(P_\vartheta^n)}(1)$ für $n \to \infty$.

Beweis: 1) Vorbemerkung: nach 4.7 und 4.8 ist dies eine Zerlegung der Summe auf der linken Seite in ihren Erwartungswert (zweiter Term auf der rechten Seite) plus eine Summe aus zentrierten L^2-Variablen, auf die man den zentralen Grenzwertsatz anwenden will.

2) L^2-Differenzierbarkeit im Experiment $\mathcal{E}$ an der Stelle ϑ liefert für Restterme

$$r_\vartheta(n,h) \;:=\; \frac{1}{|h|/\sqrt{n}} \left[\sqrt{L^{(\vartheta+h/\sqrt{n})/\vartheta}} - 1 - \frac{1}{2} \left(\frac{1}{\sqrt{n}}\, h \right)^{\top} V_\vartheta \right]$$

direkt aus 4.2 ii)b) die im folgenden an mehreren Stellen benutzte Aussage

$$(\diamond) \qquad \sup_{|h|\le C} E_\vartheta \left(r_\vartheta(n,h)^2 \right) \;\longrightarrow\; 0 \quad \text{für } n\to\infty, \text{ für beliebiges } C<\infty\,.$$

3) Wir betrachten mit den eben definierten Resttermen die Differenz aus linker Seite und erstem Term auf der rechten Seite in der Aussage des Hilfssatzes:

$$(*) \qquad \sum_{j=1}^{n} \left[\sqrt{L^{(\vartheta+h/\sqrt{n})/\vartheta}} - 1 - \frac{1}{2} \left(\frac{1}{\sqrt{n}}\, h \right)^{\top} V_\vartheta \right] (\omega_j) \;=\; \sum_{j=1}^{n} \left[\frac{|h|}{\sqrt{n}}\, r_\vartheta(n,h) \right] (\omega_j)$$

für $h \in I\!R^d$. Für beliebige Wahl einer Konstanten $C<\infty$ zeigt dann $(\diamond)$

$$\sup_{|h|\le C} n\, Var_\vartheta \left(\frac{|h|}{\sqrt{n}}\, r_\vartheta(n,h) \right) \;\le\; C \sup_{|h|\le C} E_\vartheta \left(r_\vartheta(n,h)^2 \right) \;\longrightarrow\; 0$$

für $n\to\infty$. Also ist für jede beschränkte Folge $(h_n)_n$ unter $(P^n_\vartheta)_n$ die Folge von Zufallsvariablen auf der linken Seite von $(*)$

$$\sum_{j=1}^{n} \left[\sqrt{L^{(\vartheta+h_n/\sqrt{n})/\vartheta}} - 1 - \frac{1}{2} \left(\frac{1}{\sqrt{n}}\, h_n \right)^{\top} V_\vartheta \right] (\omega_j)\,, \quad n\to\infty$$

asymptotisch äquivalent zur Folge ihrer Erwartungswerte; wegen 4.7 und 4.4+4.5 weiß man

$$E_\vartheta \left[\sqrt{L^{(\vartheta+h_n/\sqrt{n})/\vartheta}} - 1 - \frac{1}{2} \left(\frac{1}{\sqrt{n}}\, h_n \right)^{\top} V_\vartheta \right] \;=\; -H^2 \left(P_{\vartheta+h_n/\sqrt{n}}, P_\vartheta \right)$$

und erhält so eine Darstellung der Folge auf der linken Seite von $(*)$ als

$$(**) \qquad -n\, H^2 \left(P_{\vartheta+h_n/\sqrt{n}}, P_\vartheta \right) \;+\; o_{(P^n_\vartheta)}(1)\,, \quad n\to\infty\,.$$

Mit 4.8 ii) schreibt man $(**)$ äquivalent um in die Form

$$(***) \qquad -\frac{1}{8}\, h_n^\top J_\vartheta\, h_n \;+\; o_{(P^n_\vartheta)}(1)\,, \quad n\to\infty$$

(zum Nachweis der Äquivalenz wählt man aus Teilfolgen von $(h_n/\sqrt{n})_n$ weitere Teilfolgen aus, die den Voraussetzungen von Satz 4.8 genügen, d.h. die sich dem Punkt ϑ mit Rate $\frac{1}{\sqrt{n}}$ aus einer festen Richtung $u \in S_{d-1}$ annähern). Mit $(*)$ und $(***)$ ist der Hilfssatz bewiesen. $\square$

4.13 Hilfssatz: Für jede beschränkte Folge $(h_n)_n$ in $\mathbb{R}^d$ und $\xi_n := \vartheta + h_n/\sqrt{n}$ gilt in $\mathcal{E}_n$

$$\sum_{j=1}^{n} \left(\sqrt{L^{\xi_n/\vartheta}}(\omega_j) - 1 \right)^2 \;=\; \frac{1}{4}\, h_n^\top J_\vartheta\, h_n \;+\; \varrho_n(\omega)$$

wobei $\varrho_n = o_{(P_\vartheta^n)}(1)$ für $n \to \infty$.

Beweis: Dies ist im Kern ein starkes Gesetz der großen Zahlen, siehe 4.8 i).

1) Mit auf Ω_n definierten Resttermen

$$\varrho_n \,,\; \widehat{\varrho}_n \,,\; \widetilde{\varrho}_n \,,\; \ldots \;=\; o_{(P_\vartheta^n)}(1) \quad \text{für } n \to \infty$$

gilt nach Definition der Fisher-Information

$$\frac{1}{n} \sum_{j=1}^{n} \left(V_\vartheta\, V_\vartheta^\top \right)(\omega_j) \;=\; E_\vartheta \left(V_\vartheta\, V_\vartheta^\top \right) \;+\; \widehat{\varrho}_n(\omega) \;=\; J_\vartheta \;+\; \widehat{\varrho}_n(\omega)$$

und damit wegen Beschränktheit der Folge $(h_n)_n$ auch

$$(+) \qquad \sum_{j=1}^{n} \left(\frac{1}{\sqrt{n}}\, h_n^\top V_\vartheta(\omega_j) \right)^2 \;=\; h_n^\top J_\vartheta\, h_n \;+\; \widetilde{\varrho}_n(\omega) \,, \quad n \to \infty \,.$$

(Beachte, daß man aus jeder Teilfolge von $(h_n)_n$ eine konvergente Teilfolge auswählen kann.)

2) Schreibt man im Einzelexperiment $\mathcal{E}$ mit den Notationen aus dem Beweis von 4.12

$$(++) \qquad \sqrt{L^{\xi_n/\vartheta}}(\omega_j) - 1 \;=\; \frac{1}{2\sqrt{n}}\, h_n^\top V_\vartheta(\omega_j) \;+\; \frac{|h_n|}{\sqrt{n}}\, r_\vartheta(n, h_n)(\omega_j)$$

und berücksichtigt man $(\diamond)$ aus dem Beweis von 4.12

$$\sup_{|h| \le C} E_\vartheta \left(r_\vartheta(n,h)^2 \right) \;\longrightarrow\; 0 \quad \text{für } n \to \infty \,,$$

so führt Quadrieren jeder Seite von $(++)$ und Summation über $1 \le j \le n$ mit Cauchy-Schwartz, $(+)$ und $(\diamond)$ zu

$$\sum_{j=1}^{n} \left(\sqrt{L^{\xi_n/\vartheta}}(\omega_j) - 1 \right)^2 \;=\; \sum_{j=1}^{n} \left(\frac{1}{2\sqrt{n}}\, h_n^\top V_\vartheta(\omega_j) \right)^2 \;+\; \varrho_n(\omega) \,, \quad n \to \infty \,.$$

Zusammen mit Schritt 1) ist der Hilfssatz bewiesen. □

4.14 Hilfssatz: Für jede beschränkte Folge $(h_n)_n$ in $\mathbb{R}^d$ und $\xi_n := \vartheta + h_n/\sqrt{n}$ gilt in $\mathcal{E}_n$

$$\Lambda_n^{\xi_n/\vartheta}(\omega) \;=\; 2 \sum_{j=1}^{n} \left(\sqrt{L^{\xi_n/\vartheta}}(\omega_j) - 1 \right) \;-\; \sum_{j=1}^{n} \left(\sqrt{L^{\xi_n/\vartheta}}(\omega_j) - 1 \right)^2 \;+\; \varrho_n(\omega)$$

wobei $\varrho_n = o_{(P_\vartheta^n)}(1)$ für $n \to \infty$.

Beweis: 1) Mit einer elementaren logarithmischen Entwicklung

$$\log(1+z) = z - \frac{1}{2}z^2 + o(z^2)\ , \quad z \to 0$$

sollte man für $n \to \infty$ bis auf vernachlässigbare Restterme haben

$$\begin{aligned} \Lambda_n^{\xi_n/\vartheta}(\omega) &\approx \sum_{j=1}^{n} \log L^{\xi_n/\vartheta}(\omega_j) \\ &= 2\sum_{j=1}^{n} \log\left(1 + \left(\sqrt{L^{\xi_n/\vartheta}} - 1\right)(\omega_j)\right) \\ &\approx 2\sum_{j=1}^{n}\left(\sqrt{L^{\xi_n/\vartheta}} - 1\right)(\omega_j) \ - \ \sum_{j=1}^{n}\left(\sqrt{L^{\xi_n/\vartheta}} - 1\right)^2(\omega_j) \end{aligned}$$

wobei die Asymptotik des ersten und des zweiten Terms auf der rechten Seite nach 4.12 und 4.13 bereits bekannt ist. Man muß aber die beiden Approximationen in der obigen Kette rechtfertigen.

2) Wir präzisieren das erste '$\approx$' in Schritt 1) zu

$$\Lambda_n^{\xi_n/\vartheta}(\omega) \ = \ \sum_{j=1}^{n} \log L^{\xi_n/\vartheta}(\omega_j) \quad \text{für } P_\vartheta^n\text{-fast alle } \omega = (\omega_1, ..., \omega_n) \in \Omega_n \ :$$

wähle zu $P_{\xi_n}, n \geq 1$, P_ϑ ein dominierendes Maß ν und Dichten $\chi_{\xi_n}, n \geq 1$, χ_ϑ. Im Produktexperiment ist $A_n := \{\omega \in \Omega_n : \chi_\vartheta(\omega_j) > 0 \text{ für alle } 1 \leq j \leq n\} \in \mathcal{A}_n$ eine Menge von vollem Maß unter P_ϑ^n. In $\mathcal{E}_n$ stimmt $L_n^{\xi_n/\vartheta}$ als Dichtequotient aber $(P_{\xi_n}^n + P_\vartheta^n)$–fast sicher überein mit

$$\omega \quad \longrightarrow \quad 1_{A_n}(\omega) \prod_{j=1}^{n} \frac{\chi_{\xi_n}(\omega_j)}{\chi_\vartheta(\omega_j)} \ + \ \infty \, 1_{A_n^c}(\omega)\ .$$

Weiter stimmen $\Lambda_n^{\xi_n/\vartheta}(\omega) = \log(L_n^{\xi_n/\vartheta}(\omega))$ und $\sum_{j=1}^n \log L^{\xi_n/\vartheta}(\omega_j)$ auf A_n als $[-\infty, +\infty)$-wertige Zufallsvariable überein. Damit ist die erste Approximation in 1) gesichert.

3) Wir betrachten die zweite Approximation in 1) und setzen

$$\varrho(z) \ := \ \left|\log(1+z) - z + \frac{1}{2}z^2\right| \ , \quad z \in (-1, \infty)\ .$$

Mit Kurzschreibweise

$$\begin{aligned} Z_{n,j}(\omega) &:= \left(\sqrt{L^{\xi_n/\vartheta}} - 1\right)(\omega_j)\ , \quad \omega = (\omega_1, ..., \omega_n) \in \Omega_n \\ Z_n(\omega_1) &:= \left(\sqrt{L^{\xi_n/\vartheta}} - 1\right)(\omega_1)\ , \quad \omega_1 \in \Omega \end{aligned}$$

werden wir zeigen, daß in der Folge der Produktexperimente $(\mathcal{E}_n)_n$ gilt:

$$(+) \qquad \sum_{j=1}^{n} \varrho(Z_{n,j}) \;=\; o_{(P_\vartheta^n)}(1)\,, \quad n \to \infty\,.$$

Zum Beweis von (+) unterscheidet man zwischen 'kleinen' und 'großen' Werten der $Z_{n,j}$. Wegen $\varrho(z) = o(z^2)$ für $z \to 0$ gibt es zu beliebig kleinem $\varepsilon > 0$ ein $\delta = \delta(\varepsilon) > 0$ so daß

$$|\varrho(z)| \;<\; \varepsilon\, z^2 \quad \text{auf } \{|z| \le \delta\}\,.$$

Mit Hilfssatz 4.13 sieht man, daß Beiträge zu (+) der Bauart

$$\sum_{j=1}^{n} \varrho(Z_{n,j}) 1_{\{|Z_{n,j}| \le \delta\}}$$

im Vergleich zu

$$\sum_{j=1}^{n} Z_{n,j}^2(\omega) \;=\; \sum_{j=1}^{n} \left(\sqrt{L^{\xi_n/\vartheta}}(\omega_j) - 1\right) \;=\; \frac{1}{4}\, h_n^\top J_\vartheta\, h_n \;+\; \varrho_n(\omega)$$

durch geeignete Wahl von ε und δ im Limes für $n \to \infty$ unter jede noch so kleine Schranke gedrückt werden können. Es bleiben also nur Beiträge zu (+) der Art

$$\sum_{j=1}^{n} \varrho(Z_{n,j}) 1_{\{Z_{n,j} > \delta\}}\,, \quad \sum_{j=1}^{n} \varrho(Z_{n,j}) 1_{\{Z_{n,j} < -\delta\}}$$

zu betrachten, für beliebig kleine Werte von $\delta > 0$. Wir betrachten zuerst den Fall positiver betragsmäßig 'großer' Werte der $Z_{n,j}$. Assoziiert man zu jedem $\delta > 0$ ein $\gamma = \gamma(\delta) > 0$ durch

$$\gamma \;:=\; \inf\{\varrho(z) : z > \delta\} \;>\; 0\,,$$

so gilt nach Definition

$$\sum_{j=1}^{n} \varrho(Z_{n,j})\, 1_{\{Z_{n,j} > \delta\}} \;>\; \gamma \quad \Longleftrightarrow \quad Z_{n,j} > \delta \text{ für mindestens ein } j = 1, \ldots, n.$$

Wir betrachten die Wahrscheinlichkeit dieses Ereignisses unter P_ϑ^n:

$$\begin{aligned}
&P_\vartheta^n\,(\; Z_{n,j} > \delta \text{ für mindestens ein } j = 1, \ldots, n \;) \quad \le \quad n\; P_\vartheta\,(Z_n > \delta) \\
&\le\; n\; \left[P_\vartheta \left(Z_n > \delta\,,\; \frac{1}{2\sqrt{n}}\, h_n^\top V_\vartheta \le \frac{\delta}{2} \right) \;+\; P_\vartheta \left(\frac{1}{2\sqrt{n}}\, h_n^\top V_\vartheta > \frac{\delta}{2} \right) \right] \\
&\le\; n\; \left[P_\vartheta \left(\left| \sqrt{L^{\xi_n/\vartheta}} - 1 - \frac{1}{2}(\xi_n - \vartheta)^\top V_\vartheta \right| > \frac{\delta}{2} \right) \;+\; P_\vartheta \left(\left| h_n^\top V_\vartheta \right| > \delta\sqrt{n} \right) \right]\,.
\end{aligned}$$

Nun hat man wegen Beschränktheit der Folge $(h_n)_n$ ganz elementar

$$(h_n^\top V_\vartheta)^2 \;\le\; \sum_{i,j} |h_{n,i}||V_{\vartheta,i} V_{\vartheta,j}||h_{n,j}| \;\le\; \sum_{i,j} |h_{n,i}|\, \frac{1}{2}(V_{\vartheta,i}^2 + V_{\vartheta,j}^2)\, |h_{n,j}| \;\le\; K^2 \sum_{i=1}^{d} V_{\vartheta,i}^2$$

mit einer geeigneten positiven Konstante K, und erhält wegen $V_\vartheta \in L^2(P_\vartheta)$ für $n \to \infty$

$$n\, P_\vartheta \left(\left| h_n^\top V_\vartheta \right| > \delta \sqrt{n} \right) \;\le\; (K/\delta)^2\, E_\vartheta \left(\left[\sum_i V_{\vartheta,i}^2 \right] 1_{\{[\sum_i V_{\vartheta,i}^2] > n\,(\delta/K)^2\}} \right) \;\longrightarrow\; 0\,.$$

Nun bleibt noch zu zeigen, daß für beschränkte Folgen $(h_n)_n$ auch

$$n\, P_\vartheta \left(\left| \sqrt{L^{\xi_n/\vartheta}} - 1 - \frac{1}{2}(\xi_n - \vartheta)^\top V_\vartheta \right| > \frac{\delta}{2} \right)$$

für $n \to \infty$ verschwindet. Mit den Bezeichnungen aus dem Beweis von 4.12 schreibt sich der Term zwischen Betragsstrichen als

$$\frac{|h_n|}{\sqrt{n}} |r_\vartheta(n, h_n)| \;\le\; \frac{C}{\sqrt{n}} |r_\vartheta(n, h_n)|$$

für geeignete Wahl von C, wobei Markov-Ungleichung und ($\diamond$) aus dem Beweis von 4.12 liefern

$$n\, P_\vartheta \left(|r_\vartheta(n, h_n)| \;>\; \sqrt{n}\, \frac{\delta}{2\,C} \right) \;\;\le\;\; \left(\frac{2\,C}{\delta} \right)^2 \sup_{|h|<C} E_\vartheta \left(r_\vartheta(n,h)^2 \right) \;\;\longrightarrow\;\; 0$$

für $n \to \infty$. Damit ist für Paare (γ, δ) wie oben mit beliebig kleinen Einträgen gezeigt worden

$$P_\vartheta^n \left(\sum_{j=1}^n \varrho(Z_{n,j})\, 1_{\{Z_{n,j} > \delta\}} \;>\; \gamma \right) \quad \longrightarrow \quad 0\,, \quad n \to \infty\,.$$

Ganz analog behandelt man negative betragsmäßig 'große' Werte der $Z_{n,j}$ und erhält

$$P_\vartheta^n \left(\sum_{j=1}^n \varrho(Z_{n,j})\, 1_{\{Z_{n,j} < -\delta\}} \;>\; \gamma \right) \quad \longrightarrow \quad 0\,, \quad n \to \infty\,.$$

Nun ist (+) bewiesen und damit die zweite Approximation in Schritt 1) gerechfertigt. □

4.15 Beweis des Zweiten Le Cam'schen Lemmas 4.11: Kombiniert man die Aussagen der Hilfssätze 4.14, 4.12 und 4.13, ergibt sich sofort die Aussage von Satz 4.11. □