Kapitel II

Minimum-Distanz-Schätzfolgen

Reinhard Höpfner

Vorlesung Mathematische Statistik

Wintersemester 2004/2005 und Wintersemester 2007/2008

Institut für Mathematik, Johannes Gutenberg Universität Mainz

01.12.04 und 03.01.08

Übersicht zu Kapitel II :

A. Stochastische Prozesse mit Pfaden in $L^p(T, \mathcal{T}, \mu)$

B. Minimum-Distanz-Schätzfolgen

C. Gaußprozesse und asymptotische Normalität für MD-Schätzfolgen

D. Beispiel: Lokations- und Skalenmodelle für iid Beobachtungen

A. Stochastische Prozesse mit Pfaden in $L^p(T, \mathcal{T}, \mu)$

Als Vorbereitung auf die Diskussion der Asymptotik von MD-Schätzfolgen charakterisieren wir schwache Konvergenz in L^p-Pfadräumen; der wichtigste Fall ist dabei $p = 2$. Die Darstellung folgt Cremers und Kadelka (1986), siehe auch Grinblat (1976). Diesem Teilkapitel liegen stets – ohne daß es in den einzelnen Resultaten erwähnt wird – die folgenden Voraussetzungen zugrunde.

2.1 Bezeichnungen und Voraussetzungen für Teilkapitel A: a) μ ist ein endliches Maß auf einem meßbaren Raum $(T, \mathcal{T})$ mit abzählbar erzeugter σ-Algebra $\mathcal{T}$.

b) Für $1 \le p < \infty$ betrachten wir den Raum $L^p(\mu) = L^p(T, \mathcal{T}, \mu)$ aller (μ-Äquivalenzklassen von) p-fach integrierbaren Funktionen $(T, \mathcal{T}) \to (I\!R, \mathcal{B}(I\!R))$ mit Norm

$$\|f\| \;=\; \|f\|_{L^p(T,\mathcal{T},\mu)} \;=\; \left(\int_T |f(t)|^p \mu(dt) \right)^{\frac{1}{p}} \;,$$

und versehen $L^p(\mu)$ mit seiner Borel-σ-Algebra $\mathcal{B}(L^p(\mu))$.

c) Man weiß: da $\mathcal{T}$ abzählbar erzeugt ist, gibt es eine abzählbare Teilmenge $S \subset L^p(\mu)$, welche in $L^p(T, \mathcal{T}, \mu)$ dicht liegt (*Separabilität*, zum Beweis siehe Witting (1985), S. 138). Folglich wird die Borel-σ-Algebra $\mathcal{B}(L^p(\mu))$ von den offenen Kugeln in $L^p(\mu)$ erzeugt, etwa vom System

$$B_r(g) \;:=\; \{\, f \in L^p(\mu) : \|f - g\|_{L^p(\mu)} < r \,\} \;, \quad r > 0 \text{ rational}, \;\; g \in S.$$

d) Ein reellwertiger stochastischer Prozess $X = (X_t)_{t \in T}$, definiert auf einem Raum $(\Omega, \mathcal{A}, P)$, indiziert durch die Parametermenge T, heißt *meßbar* falls

$$T \times \Omega \;\ni\; (t, \omega) \;\to\; X(t, \omega) \;\in I\!R$$

eine meßbare Abbildung von $(T \times \Omega, \mathcal{T} \otimes \mathcal{A})$ nach $(I\!R, \mathcal{B}(I\!R))$ ist. In einem meßbaren stochastischen Prozeß $X = (X_t)_{t \in T}$ ist jeder ω–Pfad $X_\bullet(\omega)$, $\omega \in \Omega$:

$$X_\bullet(\omega) : \; T \ni t \;\to\; X(t, \omega) \;\in I\!R$$

eine meßbare Abbildung von $(T, \mathcal{T})$ nach $(I\!R, \mathcal{B}(I\!R))$.

2.1' Beispiel: Betrachte iid ZV $Y_1, Y_2, \ldots$ auf $(\Omega, \mathcal{A}, P)$ mit Werten in $(T, \mathcal{T}) := (I\!R^m, \mathcal{B}(I\!R^m))$.

a) Sei $n \ge 1$. Definiere zu den ersten n Beobachtungen die *empirische Verteilungsfunktion*

$$\widehat{F}_n(t, \omega) \;:=\; \frac{1}{n} \sum_{i=1}^{n} 1_{(-\infty, t]}(Y_i(\omega)) \;=\; \frac{1}{n} \sum_{i=1}^{n} 1_{\{Y_i \le t\}}(\omega) \;, \quad t \in I\!R^m \,, \omega \in \Omega$$

(mit Schreibweise $(-\infty, t] := \bigtimes_{j=1}^{m} (-\infty, t_j]$ für $t = (t_1, ..., t_m)$, und mit $(t_1, ..., t_m) \leq (t'_1, ..., t'_m)$ falls $t_j \leq t'_j$ für alle $1 \leq j \leq m$). Dann ist $\widehat{F}_n = (\widehat{F}_n(t))_{t \in \mathbb{R}^m}$ ein stochastischer Prozeß auf $(\Omega, \mathcal{A})$. Wir zeigen, daß $\widehat{F}_n$ ein meßbarer Prozeß im Sinne von 2.1 d) ist.

Definiere für $k \geq 1$ und $l = (l_1, ..., l_m) \in \mathbb{Z}^m$

$$t_l(k) := \left(\frac{l_1+1}{2^k}, \ldots, \frac{l_m+1}{2^k}\right) , \quad A_l(k) := \bigtimes_{j=1}^{m} \left[\frac{l_j}{2^k}, \frac{l_j+1}{2^k}\right) .$$

Für jedes feste k ist

$$(t, \omega) \longrightarrow \sum_{l \in \mathbb{Z}^m} 1_{A_l(k)}(t) \; \widehat{F}_n(t_l(k), \omega)$$

eine meßbare Abbildung $(\mathbb{R}^m \times \Omega, \mathcal{B}(\mathbb{R}^m) \otimes \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Die Pfade von $\widehat{F}_n$ sind in jeder Komponente des Arguments $t \in \mathbb{R}^m$ rechtsstetig, stückweise konstant, nichtfallend; folglich ist die empirische Verteilungsfunktion $\widehat{F}_n$ als punktweiser Limes meßbarer Abbildungen

$$(t, \omega) \longrightarrow \widehat{F}_n(t, \omega) = \lim_{k \to \infty} \sum_{l \in \mathbb{Z}^m} 1_{A_l(k)}(t) \; \widehat{F}_n(t_l(k), \omega)$$

ein meßbarer stochastischer Prozess.

b) Die Verteilungsfunktion $F(t) = P(Y_1 \leq t)$, $t \in \mathbb{R}^m$, der Variable Y_1 unter P ist in jeder Komponente des Arguments $t \in \mathbb{R}^m$ rechtsstetig, stückweise konstant, nichtfallend. Mit demselben Argument wie oben ist also auch $\sqrt{n}(\widehat{F}_n - F)$

$$(t, \omega) \longrightarrow \sqrt{n} \left(\widehat{F}_n(t, \omega) - F(t)\right) , \quad t \in \mathbb{R}^m , \omega \in \Omega$$

ein meßbarer stochastischer Prozess. □

2.2 Hilfssatz: Genügt ein meßbarer stochastischer Prozeß $X = (X_t)_{t \in T}$ auf $(\Omega, \mathcal{A}, P)$ der Bedingung

$$\int_T |X(t, \omega)|^p \mu(dt) < \infty \quad \text{für alle } \omega \in \Omega ,$$

so ist der *Pfad* $X_\bullet$

$$X_\bullet : \Omega \ni \omega \to X_\bullet(\omega) \in L^p(\mu)$$

eine wohldefinierte Zufallsvariable auf $(\Omega, \mathcal{A})$ mit Werten in $(L^p(T, \mathcal{T}, \mu), \mathcal{B}(L^p(T, \mathcal{T}, \mu)))$.
Wir nennen X dann kurz einen *meßbaren stochastischen Prozeß mit Pfaden in* $L^p(\mu)$.

Beweis: Für beliebiges $g \in L^p(\mu) = L^p(T, \mathcal{T}, \mu)$ schreibe

$$Z_g : \quad \omega \to \|X_\bullet(\omega) - g\| = \left(\int_T |X(t, \omega) - g(t)|^p \mu(dt)\right)^{\frac{1}{p}} .$$

Wegen Meßbarkeit des Prozesses $X = (X_t)_{t \in T}$ ist Z_g eine meßbare Abbildung $(\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Für jede Kugel $B_r(g)$ wie in 2.1 c) gilt also

$$X_\bullet^{-1}(B_r(g)) \;=\; \{\omega : X_\bullet(\omega) \in B_g(r)\} \;=\; \{Z_g < r\} \;\in \mathcal{A}\,.$$

Da die Borelsche σ-Algebra $\mathcal{B}(L^p(\mu))$ nach 2.1 c) von den offenen Kugeln erzeugt wird, ist der Pfad $X_\bullet$ eine Zufallsvariable auf $(\Omega, \mathcal{A})$ mit Werten in $(L^p(\mu), \mathcal{B}(L^p(\mu)))$. □

2.3 Satz: Seien $(X_t^n)_{t \in T}$, $n \geq 1$, $(X_t)_{t \in T}$ meßbare stochastische Prozesse mit Pfaden in $L^p(T, \mathcal{T}, \mu)$. Dann sind die folgenden Aussagen i) und ii) äquivalent:

i) es gilt

$$X_\bullet^n \longrightarrow X_\bullet \quad \text{(schwache Konvergenz in } L^p(T, \mathcal{T}, \mu)\text{, für } n \to \infty\text{);}$$

ii) für beliebiges $l \geq 1$ und beliebige g_1, ..., g_l in $L^p(T, \mathcal{T}, \mu)$ gilt

$$(\|X_\bullet^n - g_1\|, \ldots, \|X_\bullet^n - g_l\|) \;\longrightarrow\; (\|X_\bullet - g_1\|, \ldots, \|X_\bullet - g_l\|)$$
$$\text{(schwache Konvergenz in } \mathbb{R}^l\text{, für } n \to \infty\text{).}$$

Beweis: Für schwache Konvergenz in metrischen Räumen siehe z.B. Billingsley (1968). Schreibe kurz $L^p = L^p(T, \mathcal{T}, \mu)$. Die Richtung i)$\Longrightarrow$ii) folgt aus dem Satz über stetige Abbildungen, anzuwenden auf

$$h: \quad L^p \ni f \;\to\; (\|f - g_1\|, \ldots, \|f - g_l\|) \;\in \mathbb{R}^l\,.$$

Zeige also ii)$\Longrightarrow$i). Seien X^n definiert auf irgendwelchen $(\Omega_n, \mathcal{A}_n, P_n)$, $n \geq 1$, und X auf $(\Omega, \mathcal{A}, P)$. Schreibe Q^n für die Wahrscheinlichkeitsmaße $\mathcal{L}(X_\bullet^n | P_n)$ auf $(L^p, \mathcal{B}(L^p))$, $n \geq 1$, und Q für $\mathcal{L}(X_\bullet | P)$.

Nach dem Portmanteau-Theorem genügt es zu zeigen, daß für alle offenen Mengen G in L^p gilt

$$(+) \qquad \liminf_{n \to \infty} \; Q^n(G) \quad \geq \quad Q(G)\,.$$

Wegen Separabilität des L^p nach 2.1 c) kann man jede offene Menge als abzählbare Vereinigung geeigneter offener Kugeln schreiben

$$G \;=\; \bigcup_{i=1}^{\infty} B_{r_i}(g_i) \quad \text{mit geeigneten } g_i \in S,\; r_i > 0 \text{ rational.}$$

Zu beliebigem $\varepsilon > 0$ existiert dann ein $l = l(\varepsilon)$ so daß

$$Q(G) \;\leq\; Q\left(\bigcup_{i=1}^{l} B_{r_i}(g_i)\right) + \varepsilon.$$

Damit folgt für $n \to \infty$

$$
\begin{aligned}
\liminf_n\ Q^n(G) \ &\geq\ \liminf_n\ Q^n\left(\bigcup_{i=1}^l B_{r_i}(g_i)\right) \quad = \quad \liminf_n\ P_n\left(X^n_\bullet \in \bigcup_{i=1}^l B_{r_i}(g_i)\right) \\
&=\ \liminf_n\ P_n\,(\text{es existiert ein } 1 \leq i \leq l \text{ mit } \|X^n_\bullet - g_i\| < r_i) \\
&=\ \liminf_n\ \left(1\ -\ P_n\left((\|X^n_\bullet - g_1\|, \ldots, \|X^n_\bullet - g_l\|)\ \in\ \mathop{\times}_{i=1}^l [r_i, \infty)\right)\right)
\end{aligned}
$$

was unter ii) – mit schwacher Konvergenz in $\mathbb{R}^l$ und der Aussage des Portmanteau-Theorems für abgeschlossene Mengen in $\mathbb{R}^l$ – weiter abzuschätzen ist durch

$$
\begin{aligned}
&\geq\ 1\ -\ P\left((\|X_\bullet - g_1\|, \ldots, \|X_\bullet - g_l\|)\ \in\ \mathop{\times}_{i=1}^l [r_i, \infty)\right) \\
&=\ P\,(\text{es existiert ein } 1 \leq i \leq l \text{ mit } \|X_\bullet - g_i\| < r_i) \\
&=\ P\left(X_\bullet \in \bigcup_{i=1}^l B_{r_i}(g_i)\right) \quad = \quad Q\left(\bigcup_{i=1}^l B_{r_i}(g_i)\right) \\
&\geq\ Q(G) - \varepsilon
\end{aligned}
$$

nach Wahl von l. Da $\varepsilon > 0$ beliebig war, ist (+) gezeigt und damit die Behauptung bewiesen. □

2.4 Hauptsatz (Cremers und Kadelka 1986): Betrachte meßbare stochastische Prozesse

$$(X_t^n)_{t\in T},\ \text{definiert auf } (\Omega_n, \mathcal{A}_n, P_n),\ n \geq 1, \text{ und } (X_t)_{t\in T},\ \text{definiert auf } (\Omega, \mathcal{A}, P),$$

alle mit Pfaden in demselben $L^p(T, \mathcal{T}, \mu)$. Für

$$X^n_\bullet \longrightarrow X_\bullet \quad (\text{schwache Konvergenz in } L^p(T, \mathcal{T}, \mu), \text{ für } n \to \infty)$$

sind die folgenden Bedingungen i) + ii) hinreichend:

i) es gibt $N \in \mathcal{T}$ mit $\mu(N) = 0$, so daß für beliebiges $l \geq 1$ und beliebige $t_1, \ldots, t_l$ in $T \setminus N$ gilt

$$\mathcal{L}\left((X^n_{t_1}, \ldots, X^n_{t_l}) \mid P_n\right)\ \longrightarrow\ \mathcal{L}\left((X_{t_1}, \ldots, X_{t_l}) \mid P\right) \quad (\text{schwach in } \mathbb{R}^l, \text{ für } n \to \infty)$$

(Konvergenz der endlichdimensionalen Randverteilungen, μ-fast überall);

ii) für jedes $\varepsilon > 0$ gibt es ein $K = K(\varepsilon) < \infty$ so daß (mit $X^0 := X$ und $P_0 := P$)

$$\sup_{n\geq 0} \int_{T\times\Omega_n} 1_{\{|X^n|>K\}}\, |X^n|^p\, d(\mu\otimes P_n) \quad < \quad \varepsilon$$

(*gleichgradige Integrierbarkeit* der $\{|X^n(\cdot,\cdot)|^p : n \geq 0\}$ als Familie meßbarer Abbildungen

$$X^n\ :\quad (T\times\Omega_n, \mathcal{T}\otimes\mathcal{A}_n, \mu\otimes P_n)\ \longrightarrow\ (\mathbb{R}, \mathcal{B}(\mathbb{R}))\ , \quad n \geq 0\ ,$$

die auf verschiedenen Wahrscheinlichkeitsräumen leben).

Vor dem Beweis von Satz 2.4 erwähnen wir eine nützliche Konsequenz:

2.5 Folgerung: Unter den Voraussetzungen i) + ii) des Satzes 2.4 gilt insbesondere

$$\int_T g(s)\, X_s^n\, \mu(ds) \;\longrightarrow\; \int_T g(s)\, X_s\, \mu(ds) \quad \text{(schwache Konvergenz in } I\!R\text{, für } n \to \infty)$$

für jedes $g \in L^q(\mu)$, wobei q durch $\frac{1}{p} + \frac{1}{q} = 1$ gegeben ist, und

$$\|X_\bullet^n\|_{L^p(\mu)} \;\longrightarrow\; \|X_\bullet\|_{L^p(\mu)} \quad \text{(schwache Konvergenz in } I\!R\text{, für } n \to \infty).$$

Beweis: $L^q(\mu)$ ist der Dualraum zu $L^p(\mu)$, insbesondere ist für jedes feste $g \in L^q(\mu)$ die Abbildung $f \to \int f g\, d\mu$ eine stetige Abbildung von $L^p(\mu) = L^p(T, \mathcal{T}, \mu)$ nach $I\!R$. Auch $f \to \|f\|_{L^p(\mu)}$ ist stetig auf $L^p(\mu)$. Beide Behauptungen folgen also aus

$$X_\bullet^n \;\longrightarrow\; X_\bullet \quad \text{(schwache Konvergenz in } L^p(\mu)\text{, für } n \to \infty)$$

mit dem Satz über stetige Abbildungen. □

2.5' Beweis von Satz 2.4: Für den Beweis normieren wir das endliche Maß μ auf $\mu(T) = 1$. Bezeichne $\mathcal{H}$ die Klasse aller meßbaren und beschränkten Funktionen $\varphi : T \times I\!R \to I\!R$ mit der Eigenschaft, daß für jedes feste $t \in T$ die Abbildung $\varphi(t, \cdot) : x \to \varphi(t, x)$ stetig ist.
Für jeden meßbaren reellwertigen Prozeß $(X'_t)_{t \in T}$ auf einem $(\Omega', \mathcal{A}', P')$ ist $(t, \omega) \to (t, X'(t, \omega))$ $\mathcal{T} \otimes \mathcal{A}$–$\mathcal{T} \otimes \mathcal{B}(I\!R)$–meßbar, damit ist $(t, \omega) \to \varphi(t, X'(t, \omega))$ $\mathcal{T} \otimes \mathcal{A}$–$\mathcal{B}(I\!R)$–meßbar, folglich ist

$$\omega \;\longrightarrow\; \int_T \varphi(s, X'(s, \omega))\, \mu(ds)$$

eine ordentliche reellwertige Zufallsvariable auf $(\Omega', \mathcal{A}')$.
1) Wir zeigen zuerst: Bedingung i) aus 2.4 impliziert

$$(+) \qquad \int_T \varphi(s, X_s^n)\, \mu(ds) \;\longrightarrow\; \int_T \varphi(s, X_s)\, \mu(ds) \quad \text{(schwach in } I\!R\text{, für } n \to \infty)$$

für jedes $\varphi \in \mathcal{H}$.
Bezeichne $\mathcal{C}_b(I\!R)$ die Klasse der stetigen und beschränkten Funktionen $I\!R \to I\!R$. Zu zeigen ist

$$(\diamond) \qquad E\left(g\left(\int_T \varphi(s, X_s^n)\, \mu(ds)\right)\right) \;\longrightarrow\; E\left(g\left(\int_T \varphi(s, X_s)\, \mu(ds)\right)\right)\ , \quad g \in \mathcal{C}_b(I\!R) \text{ beliebig}$$

für $n \to \infty$. Wähle eine Konstante $M < \infty$ so daß $\sup_{T \times I\!R} |\varphi| \leq M$. Wegen $\mu(T) = 1$ spielt in $(\diamond)$ nur die Restriktion von g auf das Intervall $[-M, +M]$ eine Rolle. Nach Weierstraß wird die stetige Funktion g gleichmäßig auf dem Kompaktum $[-M, +M]$ durch Polynome approximiert. Folglich reicht es, $(\diamond)$ nur für Polynome g bzw. sogar nur für die Funktionen $g(x) := x^l$ nachzuweisen, $l \in I\!N$ beliebig. Wir beweisen nun für festes $l \in I\!N$

$$E\left(\left(\int_T \varphi(s, X_s^n)\,\mu(ds)\right)^l\right) \;\longrightarrow\; E\left(\left(\int_T \varphi(s, X_s)\,\mu(ds)\right)^l\right) \,, \quad n \to \infty \,.$$

Mit $X^0 := X$ schreibt man rechte bzw. linke Seite in der Form

$$\begin{aligned} E\left(\left(\int_T \varphi(s, X_s^n)\,\mu(ds)\right)^l\right) &= \int_T \cdots \int_T E\left(\prod_{i=1}^l \varphi(s_i, X_{s_i}^n)\right) \mu(ds_1)\ldots\mu(ds_l) \\ &= \int_T \cdots \int_T E\left(\,\psi_{(s_1,\ldots,s_l)}(X_{s_1}^n, \ldots, X_{s_l}^n)\,\right) \mu(ds_1)\ldots\mu(ds_l) \end{aligned}$$

wobei für festes $(s_1, \ldots, s_l)$ nach Voraussetzung über $\varphi \in \mathcal{H}$ gilt

$$\psi(x_1, \ldots, x_l) \;=\; \psi_{(s_1,\ldots,s_l)}(x_1, \ldots, x_l) \;:=\; \prod_{i=1}^l \varphi(s_i, x_i) \quad \in\; \mathcal{C}_b(I\!R^l) \,.$$

Die in i) vorausgesetzte Konvergenz der endlichdimensionalen Randverteilungen von X^n gegen die von X, μ–fast überall, impliziert daher für jedes feste $(s_1, \ldots, s_l)$ mit $s_1, \ldots, s_l \in T \setminus N$

$$E\left(\psi(X_{s_1}^n, \ldots, X_{s_l}^n)\right) \;\longrightarrow\; E\left(\psi(X_{s_1}, \ldots, X_{s_l})\right) \,, \quad n \to \infty \,.$$

Unabhängig von $(s_1, \ldots, s_l)$ sind alle diese Terme beschränkt durch M^l, also strebt

$$\int_T \cdots \int_T E\left(\,\psi_{(s_1,\ldots,s_l)}(X_{s_1}^n, \ldots, X_{s_l}^n)\,\right) \mu(ds_1)\ldots\mu(ds_l)$$

mit dominierter Konvergenz für $n \to \infty$ gegen

$$\int_T \cdots \int_T E\left(\,\psi_{(s_1,\ldots,s_l)}(X_{s_1}, \ldots, X_{s_l})\,\right) \mu(ds_1)\ldots\mu(ds_l) \,.$$

Dies ist die Aussage $(\diamond)$ für $g(x) = x^l$. Folglich ist $(+)$ bewiesen.

2) Wir zeigen nun: wegen Voraussetzung ii) kann für jedes feste $g \in L^p(T, \mathcal{T}, \mu)$ die Familie

$$Z^n(g) \;:=\; \|X_\bullet^n - g\|^p \,, \quad n \geq 0$$

gleichmäßig in $n \in I\!N_0$ durch Zufallsvariable des in 1) betrachteten Typs

$$\int_T \varphi(s, X_s^n)\,\mu(ds) \,, \quad \varphi \in \mathcal{H} \quad \text{geeignet} \,,\; n \geq 0$$

approximiert werden, genauer formuliert (mit $X^0 = X$ und $P_0 = P$ wie oben): für jedes $g \in L^p(\mu)$ und jedes $\delta > 0$ gibt es eine Funktion $\varphi = \varphi_{g,\delta} \in \mathcal{H}$ so daß gilt

$$(*) \qquad \sup_{n\geq 0} P_n \left(\left| \|X^n_\bullet - g\|^p - \int_T \varphi(s, X^n_s)\mu(ds) \right| > \delta \right) \quad < \quad \delta \, .$$

Zum Beweis von $(*)$ seien $g \in L^p(T, \mathcal{T}, \mu)$ und $\delta > 0$ fest. Mit Voraussetzung ii) aus 2.4 wählt man zu δ ein $1 < C < \infty$ so daß

$$\sup_{n\geq 0} \int_{T\times\Omega_n} 1_{\{|X^n|>C\}} \, |X^n|^p \, d(\mu\otimes P_n) \quad < \quad 3^{-2}\, 2^{-p}\, \delta^2 \, ;$$

eventuell notwendiges Vergrößern von C stellt zugleich sicher

$$\int_T 1_{\{|g|>C\}} \, |g|^p \, d\mu \quad < \quad 3^{-2}\, 2^{-p}\, \delta^2 \, .$$

Schreibe $h(x) \; = \; (-C) \vee x \wedge (+C)$ und definiere $\varphi = \varphi_{g,\delta} : T \times I\!\!R \to I\!\!R$ durch

$$\varphi(s,x) \; := \; |h(x) - h(g(s))|^p \, , \quad s \in T \, , \; x \in I\!\!R \, .$$

Klar gilt $\varphi \in \mathcal{H}$, und aus $|a+b|^p \leq (|a|+|b|)^p \leq 2^p(|a|^p + |b|^p)$ erhält man

$$\begin{aligned}
& \left| \|X^n_\bullet(\omega) - g\|^p - \int_T \varphi(s, X^n_s(\omega)) \, \mu(ds) \right| \\
\leq \; & \int_T 1_{\{|X^n(s,\omega)|>C\} \cup \{|g(s)|>C\}} \, \left| \, |X^n_s(\omega) - g(s)|^p - \varphi(s, X^n_s(\omega)) \, \right| \; \mu(ds) \\
\leq \; & \int_T 1_{\{|X^n(s,\omega)|>C\} \cup \{|g(s)|>C\}} \, (\, 2^p |X^n_s(\omega)|^p + 2^p |g(s)|^p + 2^p C^p \,) \; \mu(ds) \, .
\end{aligned}$$

Daraus ergibt sich wegen $\mu(T) = 1$

$$\begin{aligned}
& E \quad \left(\left| \|X^n_\bullet(\omega) - g\|^p - \int_T \varphi(s, X^n_s(\omega)) \, \mu(ds) \right| \right) \\
= \; & 2^p \int_{T\times\Omega} 1_{\{|X^n|>C\} \cup \{|g|>C\}} \, (|X^n|^p + |g|^p + C^p) \; d(\mu\otimes P_n) \\
\leq \; & 2^p \int_{T\times\Omega} 1_{\{|X^n|>C \, , \, |g|>C\}} \left(\frac{3}{2}|X^n|^p + \frac{3}{2}|g|^p \right) \; d(\mu\otimes P_n) \\
& + 2^p \int_{T\times\Omega} 1_{\{|X^n|>C \, , \, |g|\leq C\}} \, (\, 3 \, |X^n|^p \,) \; d(\mu\otimes P_n) \\
& + 2^p \int_{T\times\Omega} 1_{\{|g|>C \, , \, |X^n|\leq C\}} \, (\, 3 \, |g|^p \,) \; d(\mu\otimes P_n) \\
\leq \; & \frac{9}{2} \cdot 2^p \cdot \left(\int_{T\times\Omega} 1_{\{|X^n|>C\}} \, |X^n|^p \, d(\mu\otimes P_n) \; + \; \int_T 1_{\{|g|>C\}} \, |g|^p \; d\mu \right) \\
< \; & \delta^2
\end{aligned}$$

nach Wahl von C. Eine Anwendung der Markov-Ungleichung schließt den Beweis dieses Teilschritts ab und liefert $(*)$.

3) Zum Nachweis der Behauptung

$$X^n_\bullet \longrightarrow X_\bullet \quad \text{(schwache Konvergenz in } L^p(T,\mathcal{T},\mu)\text{, für } n \to \infty)$$

genügt es nach 2.3 zu zeigen, daß für beliebiges $l \geq 1$ und beliebige g_1, ..., g_l in $L^p(T,\mathcal{T},\mu)$ gilt

$$(\|X^n_\bullet - g_1\|, \ldots, \|X^n_\bullet - g_l\|) \longrightarrow (\|X_\bullet - g_1\|, \ldots, \|X_\bullet - g_l\|)$$
$$\text{(schwache Konvergenz in } I\!R^l\text{, für } n \to \infty),$$

bzw. nach Dazwischenschalten einer stetigen Bijektion auf $(0,\infty)^l$ die äquivalente Behauptung

$$(\|X^n_\bullet - g_1\|^p, \ldots, \|X^n_\bullet - g_l\|^p) \longrightarrow (\|X_\bullet - g_1\|^p, \ldots, \|X_\bullet - g_l\|^p)$$
$$\text{(schwache Konvergenz in } I\!R^l\text{, für } n \to \infty)$$

zu beweisen. Nach Cramér-Wold (bzw. nach dem Stetigkeitssatz von P. Lévy) hat man also für beliebiges $\alpha \in I\!R^l$ zu zeigen

$$(++) \qquad Y^n := \sum_{i=1}^{l} \alpha_i \|X^n_\bullet - g_i\|^p \longrightarrow \sum_{i=1}^{l} \alpha_i \|X_\bullet - g_i\|^p =: Y \quad \text{(schwach in } I\!R,\ n \to \infty)\ .$$

Für den Beweis von (++) reicht es, nur den Fall $\alpha_i \neq 0$ für $1 \leq i \leq l$ zu betrachten, und solche Tupel $(\alpha_1, \ldots, \alpha_l)$ durch Multiplikation mit einer Konstanten auf $\frac{1}{l}\sum_{i=1}^{l}\frac{1}{\alpha_i} = 1$ zu normieren.

4) Zum Nachweis von (++) werden wir mit 'begleitenden Folgen' argumentieren:
Bezeichne $\mathcal{C}_{b,u}(I\!R)$ die Klasse der gleichmäßig stetigen und beschränkten Funktionen auf $I\!R$. Eine Folge $(Z^n)_n$ heißt *δ-begleitend für* $(Y^n)_n$ wenn gilt

$$\sup_{n\geq 0} P_n\left(|Y^n - Z^n| > \delta\right) < \delta\ .$$

Sei $f \in \mathcal{C}_{b,u}(I\!R)$ fest, setze $M := 2\sup|f|$. Wählt man zu beliebig kleinem $\eta > 0$ ein $\delta > 0$ mit

$$\sup_{|x-x'|<\delta} |f(x) - f(x')| < \eta\ ,$$

so gilt für jede δ-begleitende Folge $(Z^n)_n$ die Abschätzung

$$(\circ) \qquad \left|E(f(Y^n)) - E(f(Y^0))\right| \leq \left|E(f(Z^n)) - E(f(Z^0))\right| + 2\eta + 2M\delta\ .$$

Kann man also bei gegebener Folge $(Y^n)_n$ *für jedes $\delta > 0$ eine δ-begleitende Folge $(Z_n)_n$ assoziieren, welche schwach konvergiert*, so impliziert $(\circ)$ die schwache Konvergenz von $(Y^n)_n$.

5) Abschluß des Beweises: sei $(Y^n)_n$ die Folge aus (++), $Y^0 := Y$. Zu zeigen ist

$$\lim_{n\to\infty} E(f(Y^n)) \;=\; E(f(Y^0)) \quad \text{für alle } f \in \mathcal{C}_{b,u}(I\!\!R)\ .$$

Sei $\delta > 0$ beliebig klein. Mit $(\alpha_1, \ldots, \alpha_l)$ und $g_1, \ldots, g_l$ aus (++) betrachte

$$\varphi_{\alpha,g_1,\ldots,g_l,\delta} \;:=\; \sum_{i=1}^{l} \alpha_i\, \varphi_{g_i, \frac{\delta}{l\alpha_i}} \;\in\; \mathcal{H}\ , \quad Z^n \;:=\; \int_T \varphi_{\alpha,g_1,\ldots,g_l,\delta}(s, X^n_s)\, \mu(ds)\ , \quad n \in I\!\!N_0\ .$$

Dann folgt aus $(*)$ in Schritt 2) (benutze Dreiecksungleichung und Normierung $\frac{1}{l}\sum_{i=1}^{l} \frac{1}{\alpha_i} = 1$)

$$P_n\left(|Y^n - Z^n| > \delta \right) \;\leq\; \sum_{i=1}^{l} P_n\left(\left| \alpha_i \|X^n_\bullet - g_i\|^p - \alpha_i \int_T \varphi_{g_i,\frac{\delta}{l\alpha_i}}(s, X^n_s)\mu(ds) \right| > \frac{\delta}{l} \right) \;<\; \delta$$

unabhängig von $n \in I\!\!N_0$. Damit ist aber eine δ-begleitende Folge zu $(Y^n)_{n\in I\!\!N_0}$ gefunden:

$$(**) \qquad \sup_{n\geq 0} P_n\left(\left| Y^n - \int_T \varphi_{\alpha,g_1,\ldots,g_l,\delta}(s, X^n_s)\, \mu(ds) \right| > \delta \right) \;<\; \delta\ ,$$

für die man nach (+) in Schritt 1) bereits weiß:

$$\int_T \varphi_{\alpha,g_1,\ldots,g_l,\delta}(s, X^n_s)\, \mu(ds) \;\longrightarrow\; \int_T \varphi_{\alpha,g_1,\ldots,g_l,\delta}(s, X_s)\, \mu(ds) \quad (\text{schwach in } I\!\!R,\ n \to \infty)\ .$$

Dies gilt für beliebig kleines $\delta > 0$. Aus dem Argument $(\circ)$ folgt daher (++)

$$\lim_{n\to\infty} E(f(Y^n)) \;=\; E(f(Y)) \quad \text{für alle } f \in \mathcal{C}_{b,u}(I\!\!R)\ ,$$

und der Beweis des Satzes 2.4 ist vollständig. □

Anstelle der der gleichgradigen Integrierbarkeitsbedingung ii) aus Satz 2.4 benutzt man häufig gröbere Varianten, die aber sehr einfach zu verifizieren sind.

2.6 Satz: Betrachte meßbare stochastische Prozesse X^n, $n \geq 1$, X mit Pfaden in $L^p(T, \mathcal{T}, \mu)$. Konvergieren die endlich-dimensionalen Randverteilungen μ-fast überall, so impliziert jede der folgenden Aussagen iia) oder iib) die Bedingung ii) aus 2.4:

iia) es gilt

$$X^n(.,.)\,,\ n \geq 1\,,\ X(.,.) \quad \in \quad L^p(T \times \Omega_n, \mathcal{T} \otimes \mathcal{A}_n, \mu \otimes P_n)\ ,$$
$$\limsup_{n\geq 1} \int_{T\times\Omega_n} |X^n|^p\, d(\mu \otimes P_n) \quad \leq \quad \int_{T\times\Omega} |X|^p\, d(\mu \otimes P)\ ;$$

iib) es gibt eine Funktion $f \in L^1(T,\mathcal{T},\mu)$ so daß

$$\begin{aligned} \sup_{n\geq 1} E\left(|X_t^n|^p\right) &\leq f(t) \quad \text{für } \mu\text{-fast alle } t \in T\,, \\ \lim_{n\to\infty} E\left(|X_t^n|^p\right) &= E\left(|X_t|^p\right)\,. \end{aligned}$$

Beweis: Normiere wieder μ zu $\mu(T)=1$.

1) Es gilt iib)$\Longrightarrow$iia), denn mit dominierter Konvergenz bezüglich μ liefert iib) sogar

$$\lim_{n\to\infty} \int_{T\times\Omega_n} |X^n|^p \, d(\mu\otimes P_n) \;=\; \int_{T\times\Omega} |X|^p \, d(\mu\otimes P)\,.$$

2) Wir zeigen, daß iia) zusammen mit Bedingung i) aus Satz 2.4 (Konvergenz der endlichdimensionaler Randverteilungen, μ-fast überall) die Bedingung ii) aus Satz 2.4 impliziert.

Unter iia) gibt es wegen $X(.,.) \in L^p(T\times\Omega,\mathcal{T}\otimes\mathcal{A},\mu\otimes P)$ zu jedem $\varepsilon>0$ ein $C=C(\varepsilon)<\infty$ mit

$$\int_{T\times\Omega} 1_{\{|X|>C\}}\, |X|^p \, d(\mu\otimes P) \quad < \quad \varepsilon\,.$$

Zerlege die Funktion $f_0(x) := |x|^p$ auf $I\!R$ in eine Summe $f_0 = f_1 + f_2$ mit stetigen nichtnegativen Summanden f_1, f_2. Dabei stimme $f_1 \in \mathcal{C}_b(I\!R)$ auf $\{|x|\leq C\}$ mit der Funktion f_0 überein, und verschwinde auf $\{|x|>2C\}$. Für f_2 gilt dann $f_2 \leq f_0$, f_2 stimmt auf $\{|x|>2C\}$ mit f_0 überein, f_2 verschwindet auf $\{|x|\leq C\}$; insbesondere nach Wahl von $C=C(\varepsilon)$

$$E\left(\int_T f_2(X_s)\,\mu(ds)\right) \;<\; \varepsilon\,.$$

Weiter gehört $\phi(s,x) := f_1(x)$ zur Klasse $\mathcal{H}$ wie in Schritt 1) des Beweises von 2.4, also

$$\int_T f_1(X_s^n)\,\mu(ds) \;\longrightarrow\; \int_T f_1(X_s)\,\mu(ds) \quad \text{(schwache Konvergenz in } I\!R\text{, für } n\to\infty)\,.$$

An dieser Stelle hat man die Bedingung i) aus Satz 2.4 (Konvergenz der endlichdimensionalen Randverteilungen, μ-fast überall) benutzt. Setze $M := \sup|f_1|$ und wähle ein $g \in \mathcal{C}_b(I\!R)$, welches auf $[-M,+M]$ mit der Identität übereinstimmt. Dann

$$E\left(g\left(\int_T f_1(X_s^n)\,\mu(ds)\right)\right) \;\longrightarrow\; E\left(g\left(\int_T f_1(X_s^n)\,\mu(ds)\right)\right)\;, \quad n\geq 0\,,$$

und wegen $g=\text{id}$ in Einschränkung auf $[-M,+M]$

$$E\left(\int_T f_1(X_s^n)\,\mu(ds)\right) \;\longrightarrow\; E\left(\int_T f_1(X_s)\,\mu(ds)\right)\;, \quad n\to\infty\,.$$

Unter Voraussetzung iia) – beachte $f_1 + f_2 = f_0 = |\cdot|^p$ – impliziert dies

$$\limsup_{n\to\infty} E\left(\int_T f_2(X_s^n)\,\mu(ds)\right) \;\le\; E\left(\int_T f_2(X_s)\,\mu(ds)\right) \;<\; \varepsilon\,.$$

Da f_2 auf $\{|x| > 2C\}$ mit f_0 übereinstimmt, hat man insbesondere

$$\int_{T\times\Omega_n} 1_{\{|X^n|>2C\}}\,|X^n|^p\,d(\mu\otimes P_n) \;<\; \varepsilon \quad \text{für hinreichend große } n.$$

Unter iia) gilt aber $X^n(.,.) \in L^p(T\times\Omega_n, \mathcal{T}\otimes\mathcal{A}_n, \mu\otimes P_n)$ für jedes $n \ge 0$, also kann man stets die Konstante $2C$ zu einem $K = K(\varepsilon)$ vergrößern, welches die Bedingung ii) aus Satz 2.4

$$\sup_{n\ge 0} \int_{T\times\Omega_n} 1_{\{|X^n|>K\}}\,|X^n|^p\,d(\mu\otimes P_n) \;<\; \varepsilon$$

erfüllt. □

B. Minimum-Distanz-Schätzfolgen

Minimum-Distanz(MD)-Schätzfolgen sind, wie Millar (1984) gezeigt hat, unter sehr allgemeinen Voraussetzungen konsistent und asymptotisch normal. In einem ersten Abschnitt beweisen wir eine Strukturformel für die Grenzverteilung reskalierter Schätzfehler, aus der wir im nächsten Teilkapitel die asymptotische Normalität von MD-Schätzfolgen herleiten werden. Die hier gegebene Darstellung folgt Millar (1984).

2.7 Motivation: Gegeben seien iid ZV $Y_1, Y_2, \ldots, Y_n$, deren Verteilung von einem unbekannten Parameter $\vartheta \in \Theta$ abhängt, $\Theta \subset \mathbb{R}^d$ offen. Aus den Daten $Y_1, Y_2, \ldots, Y_n$ soll der unbekannte Parameter geschätzt werden.

Die ZV $Y_1, Y_2, \ldots$ seien definiert auf einem $(\Omega, \mathcal{A})$, mit Werten in $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$, mit stetiger Verteilungsfunktion $F_\vartheta : \mathbb{R}^m \to [0,1]$ unter P_ϑ. Wir betrachten die Familie $\{F_\vartheta : \vartheta \in \Theta\}$ als Teilmenge des Raumes $\mathcal{C}_b(\mathbb{R}^m)$, versehen mit gleichmäßiger Konvergenz, und setzen stetige Parametrisierung voraus:

$$(+) \qquad \Theta \ni \;\vartheta \;\longrightarrow\; F_\vartheta(\cdot) \;\in \mathcal{C}_b(\mathbb{R}^m) \quad \text{ist stetig.}$$

Sei $\widehat{F}_n(\cdot,\omega)$ die empirische Verteilungsfunktion

$$\widehat{F}_n(t,\omega) \;:=\; \frac{1}{n}\sum_{i=1}^n 1_{(-\infty,t]}(Y_i(\omega)) \;=\; \frac{1}{n}\sum_{i=1}^n 1_{\{Y_i\le t\}}(\omega)\,, \quad t \in \mathbb{R}^m\,.$$

Der Pfad $\omega \to \widehat{F}_n(\cdot,\omega)$ ist nach 2.1' und 2.2 eine wohldefinierte Zufallsvariable auf $(\Omega, \mathcal{A})$ mit Werten in $(L^2(\mu), \mathcal{B}(L^2(\mu)))$, für jede Wahl eines endlichen Maßes μ auf $(I\!R^m, \mathcal{B}(I\!R^m))$, mit Kurzschreibweise $L^2(\mu)$ für $L^2(I\!R^m, \mathcal{B}(I\!R^m), \mu)$. Wegen (+) ist für festes ω die Abbildung

$$(++) \qquad \xi \;\to\; \|\widehat{F}_n(\cdot,\omega) - F_\xi(\cdot)\|_{L^2(\mu)}$$

stetig auf Θ. Stellt man sich (++) als zufallsabhängige Fläche über Θ vor, so liegt es nahe, den unbekannten Parameter $\vartheta \in \Theta$ durch eine Minimalstelle von (++) zu schätzen

$$\vartheta_n^*(\omega) \;=\; \operatorname*{arginf}_{\xi\in\Theta} \;\|\widehat{F}_n(\cdot,\omega) - F_\xi(\cdot)\|_{L^2(\mu)} \;.$$

Jede Minimalstelle von (++) entspricht einer bestmöglichen Anpassung der aus der Beobachtung gewonnenen Größe $\widehat{F}_n(\cdot,\omega)$ an die Schar der Funktionen $F_\xi(\cdot)$, $\xi \in \Theta$, im Hilbertraum $L^2(\mu)$. Nach dem Satz von Glivenko-Cantelli

$$(*) \qquad \lim_{n\to\infty} \sup_{t\in I\!R^m} |\widehat{F}_n(t) - F_\vartheta(t)| \;=\; 0 \quad P_\vartheta\text{-fast sicher, unter jedem } \vartheta\in\Theta$$

stabilisieren sich die zufallsabhängigen Flächen (++) für $n \to \infty$ P_ϑ-fast sicher; der Limes unter ϑ ist die deterministische Fläche

$$(+++) \qquad \Theta \ni \quad \xi \;\longrightarrow\; \|F_\vartheta(\cdot) - F_\xi(\cdot)\|_{L^2(\mu)} \quad \in I\!R \;.$$

Im Gegensatz zu Maximum-Likelhood-Schätzern ist es für MD-Schätzfolgen $(\vartheta_n^*)_n$ möglich, unter sehr einfachen Voraussetzungen die Konsistenz zu sichern (vergleiche mit den in 1.20 benutzten Voraussetzungen für ML-Schätzfolgen). Wir skizzieren den Kern des Argumentes (über Existenz und geeignete $\sigma(Y_1, ..., Y_n)$-meßbare Festlegungen von Minimalstellen von (++) wird in 2.10–2.11 nachgedacht werden): aus

$$\left\{ \min_{\xi:|\xi-\vartheta|\le\varepsilon} \|\widehat{F}_n - F_\xi\|_{L^2(\mu)} \;<\; \inf_{\xi:|\xi-\vartheta|>\varepsilon} \|\widehat{F}_n - F_\xi\|_{L^2(\mu)} \right\} \;\subset\; \{\, |\vartheta_n^* - \vartheta| \le \varepsilon \,\}$$

folgt nach Übergang zu den Komplementen

$$\begin{aligned}
\{\, |\vartheta_n^* - \vartheta| > \varepsilon \,\} \;&\subset\; \left\{ \min_{\xi:|\xi-\vartheta|\le\varepsilon} \|\widehat{F}_n - F_\xi\|_{L^2(\mu)} \;\ge\; \inf_{\xi:|\xi-\vartheta|>\varepsilon} \|\widehat{F}_n - F_\xi\|_{L^2(\mu)} \right\} \\
&\subset\; \left\{ \|\widehat{F}_n - F_\vartheta\|_{L^2(\mu)} \;\ge\; \inf_{\xi:|\xi-\vartheta|>\varepsilon} \|\widehat{F}_n - F_\xi\|_{L^2(\mu)} \right\} \\
&\subset\; \left\{ 2\,\|\widehat{F}_n - F_\vartheta\|_{L^2(\mu)} \;\ge\; \inf_{\xi:|\xi-\vartheta|>\varepsilon} \|F_\vartheta - F_\xi\|_{L^2(\mu)} \right\}
\end{aligned}$$

mit umgekehrter Dreiecksungleichung. Wegen Glivenko-Cantelli $(*)$ reicht nun die folgende einfache Bedingung

$$(**) \qquad \text{für jedes } \vartheta \in \Theta \text{ und jedes } \varepsilon > 0 \ : \quad \inf_{\xi : |\xi - \vartheta| > \varepsilon} \|F_\xi - F_\vartheta\|_{L^2(\mu)} \quad > \quad 0 \ ,$$

um Konsistenz der Schätzfolge $(\vartheta_n^*)_n$ zu erzielen. Wegen $(*)$ und $(**)$ hat man unter P_ϑ sogar fast sichere Konvergenz für $n \to \infty$ der Folge $(\vartheta_n^*)_n$ gegen ϑ, für jedes $\vartheta \in \Theta$.

Die Identifizierbarkeitsbedingung $(**)$ ist nur minimal mehr als als die unvermeidliche Bedingung, das Maß μ auf $(T, \mathcal{T})$ müsse geeignet sein, jede Verteilungsfunktion F_ϑ von anderen Möglichkeiten F_ξ, $\xi \neq \vartheta$, zu unterscheiden. $\square$

2.8 Voraussetzungen und Bezeichnungen für Teilkapitel B:

Wir betrachten eine Familie von Wahrscheinlichkeitsmaßen $\{P_\vartheta : \vartheta \in \Theta\}$ auf einem Raum $(\Omega, \mathcal{A})$, mit Parameterraum $\Theta \subset I\!R^d$ offen. Sei $(\mathcal{F}_n)_n$ eine aufsteigende Folge von Sub-σ-Algebren von $\mathcal{A}$; dabei steht $\mathcal{F}_n$ für den Informationsstand eines Beobachters im Stadium n einer Versuchsserie, etwa wie in 1.5 b) nach n-facher unabhängiger Wiederholung desselben Einzelversuchs. Schreibe $P_{n,\vartheta} := P_\vartheta | \mathcal{F}_n$ für die Restriktion des Wahrscheinlichkeitsmaßes P_ϑ auf $\mathcal{F}_n$. Wir betrachten die Folge von Experimenten

$$\mathcal{E}_n \ := \ (\Omega, \mathcal{F}_n, \{P_{n,\vartheta} : \vartheta \in \Theta\}) \ , \quad n \geq 1 \ .$$

I) Sei H ein Hilbertraum mit Skalarprodukt $\langle \cdot, \cdot \rangle_H$ und Norm $\| \cdot \|_H$ und Borel-σ-Algebra $\mathcal{B}(H)$. Sei $\widehat{\Psi}_n$ eine Folge von $\mathcal{F}_n$-meßbaren H-wertigen Zufallsvariablen

$$\widehat{\Psi}_n : \quad (\Omega, \mathcal{F}_n) \ \longrightarrow \ (H, \mathcal{B}(H)) \ , \quad n \geq 1 \ .$$

Sei $\{\Psi_\xi : \xi \in \Theta\}$ eine Familie von Elementen aus H in stetiger Parametrisierung

$$\Theta \ni \ \xi \ \to \ \Psi_\xi \ \in H \text{ ist stetig.}$$

$\widehat{\Psi}_n$ spielt die Rolle einer im Experiment $\mathcal{E}_n$ relevanten empirischen Größe; Ψ_ξ ist eine unter wahrem Wert ξ des unbekannten Parameters verfügbare Vergleichsgröße für $\widehat{\Psi}_n$, die durch das statistische Modell festgelegt ist.

II) Wegen der vorausgesetzten Stetigkeit von $\Theta \ni \xi \to \Psi_\xi \in H$ ist für festes ω auch

$$\Theta \ni \ \xi \ \to \ \left\| \widehat{\Psi}_n(\omega) - \Psi_\xi \right\|_H \ \in [0, \infty)$$

stetig, also sind für G offen und F kompakt, $G, F \subset \Theta$, stets

$$\inf_{\xi \in G} \left\| \widehat{\Psi}_n - \Psi_\xi \right\|_H \ , \quad \min_{\xi \in F} \left\| \widehat{\Psi}_n - \Psi_\xi \right\|_H$$

wohldefinierte reellwertige $\mathcal{F}_n$-meßbare Zufallsvariable.

III) Wir stellen für die Objekte aus I) eine Liste von Voraussetzungen zusammen.

Man sagt, an der Stelle ϑ gilt

a) das *starke Gesetz der großen Zahlen* $\mathbf{SLLN}(\vartheta)$ falls

$$P_\vartheta\text{-fast sicher:} \quad \|\widehat{\Psi}_n - \Psi_\vartheta\|_H \to 0 \ \text{ für } n \to \infty;$$

b) die *Identifizierungsbedingung* $\mathbf{I}(\vartheta)$ falls für beliebig kleines $\delta > 0$

$$\inf_{\xi \in \Theta \, , \, |\xi - \vartheta| > \delta} \|\Psi_\xi - \Psi_\vartheta\|_H \ > \ 0 \ ;$$

c) die *Straffheitsbedingung* $\mathbf{T}(\vartheta)$ falls es eine Folge $\varphi_n = \varphi_n(\vartheta) \uparrow \infty$ gibt so daß die Familie

$$\mathcal{L}\left(\varphi_n \|\widehat{\Psi}_n - \Psi_\vartheta\|_H \mid P_\vartheta \right) \ , \quad n \geq 1$$

(Wahrscheinlichkeitsmaße in $(I\!R, \mathcal{B}(I\!R))$) straff in $I\!R$ ist;

d) die *Differenzierbarkeitsbedingung* $\mathbf{D}(\vartheta)$ falls die Abbildung $\Theta \ni \xi \to \Psi_\xi \in H$ an der Stelle $\xi = \vartheta$ differenzierbar (Fréchét) ist: es existiere

$$D\Psi_\vartheta \ := \begin{pmatrix} D_1 \Psi_\vartheta \\ \dots \\ D_d \Psi_\vartheta \end{pmatrix} \ , \quad D_j \Psi_\vartheta \in H \ , \quad 1 \leq j \leq d$$

so daß gilt

$$\frac{1}{|\xi - \vartheta|} \left\| \Psi_\xi - \Psi_\vartheta - (\xi - \vartheta)^\top D\Psi_\vartheta \right\|_H \ \longrightarrow \ 0 \quad \text{für } \ |\xi - \vartheta| \to 0 \ .$$

Wir schließen in die Differenzierbarkeitsbedingung eine *Nichtsingularitätsbedingung* für die Ableitung $D\Psi_\vartheta$ mit ein:

$$\text{die Komponenten} \quad D_j \Psi_\vartheta \ , \ 1 \leq j \leq d \ , \quad \text{seien linear unabhängig in } H.$$

Die in 2.8.I) gemachten Voraussetzungen werden in allen Resultaten dieses Teilkapitels benutzt, ohne daß dies eigens erwähnt wird; diejenigen aus der Liste 2.8.III) werden jeweils explizit benannt werden.

2.9 Definition: Man nennt eine Folge von Schätzern $\vartheta_n^* : (\Omega, \mathcal{F}_n) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ für den unbekannten Parameter $\vartheta \in \Theta$ eine *Minimum-Distanz(MD)-Schätzfolge*, wenn es eine Folge von Ereignissen $A_n \in \mathcal{F}_n$ mit der Eigenschaft

$$\text{für jedes } \vartheta \in \Theta : \quad P_\vartheta \left(\liminf_{n\to\infty} A_n \right) \; = \; 1$$

gibt, so daß für jedes feste $n \geq 1$ gilt

$$\left\| \widehat{\Psi}_n(\omega) - \Psi_{\vartheta_n^*(\omega)} \right\|_H \; = \; \inf_{\xi\in\Theta} \left\| \widehat{\Psi}_n(\omega) - \Psi_\xi \right\|_H \; , \quad \omega \in A_n \; .$$

Für Minimum-Distanz-Schätzfolgen ist die symbolische Schreibweise

$$\vartheta_n^* \; = \; \operatorname*{arginf}_{\xi\in\Theta} \left\| \widehat{\Psi}_n - \Psi_\xi \right\|_H \; , \quad n \to \infty$$

üblich, womit stillschweigend alle Voraussetzungen aus 2.9 als erfüllt angesehen werden.

2.10 Satz: Betrachte eine kompakte Ausschöpfung von Θ

$$K_n \text{ kompakt in } \mathbb{R}^d \; , \quad K_n \subset \operatorname{int}(K_{n+1}) \; , \quad K_n \uparrow \Theta$$

und zu $(K_n)_n$ die Folge von Ereignissen

$$A_n \; := \; \left\{ \min_{\xi\in K_n} \left\| \widehat{\Psi}_n - \Psi_\xi \right\|_H = \inf_{\xi\in\Theta} \left\| \widehat{\Psi}_n - \Psi_\xi \right\|_H \right\} \quad \in \mathcal{F}_n \; .$$

Dann kann man konstruktiv eine Folge $(T_n)_n$ angeben, so daß für jedes $n \geq 1$ gilt

$$T_n : (\Omega, \mathcal{F}_n) \longrightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) \quad \text{ist meßbar},$$

$$\forall\, \omega \in A_n \; : \quad T_n(\omega) \in K_n \; , \quad \left\| \widehat{\Psi}_n(\omega) - \Psi_{T_n(\omega)} \right\|_H \; = \; \inf_{\xi\in\Theta} \left\| \widehat{\Psi}_n(\omega) - \Psi_\xi \right\|_H \; .$$

Beweis: 1) Fixiere eine kompakte Ausschöpfung $(K_n)_n$ von Θ. Definiere für jedes n eine zufällige abgeschlossene Teilmenge M_n des Kompaktums K_n:

$$M_n(\omega) \; := \; \left\{ \zeta \in K_n : \left\| \widehat{\Psi}_n(\omega) - \Psi_\zeta \right\|_H \; = \; \inf_{\xi\in\Theta} \left\| \widehat{\Psi}_n(\omega) - \Psi_\xi \right\|_H \right\} \; , \quad \omega \in \Omega \; .$$

Dies ist die Menge aller globalen Minimalstellen von $\xi \to \left\| \widehat{\Psi}_n(\omega) - \Psi_\xi \right\|_H$, welche in K_n zu finden sind. Für $\omega \in A_n$ ist $M_n(\omega)$ ein nichtleeres Kompaktum, also kann man aus Folgen in $M_n(\omega)$ stets konvergente Teilfolgen und Limiten in $M_n(\omega)$ auswählen.

2) Für festes n und $\omega \in A_n$ wählen wir einen Punkt $\alpha(\omega) \in M_n(\omega)$ wie folgt. Setze

$$\alpha_1(\omega) \ := \ \inf \{\zeta_1 : \text{ es gibt Punkte } \zeta = (\zeta_1, \zeta_2, ..., \zeta_d) \in M_n(\omega)\} \ .$$

Durch Auswahl konvergenter Teilfolgen im Kompaktum $M_n(\omega)$ sieht man, daß $M_n(\omega)$ Punkte der Bauart $(\alpha_1(\omega), \zeta_2, ..., \zeta_d)$ enthält. Danach betrachte man

$$\alpha_2(\omega) \ := \ \inf\{\zeta_2 : \text{ es gibt Punkte } \zeta = (\alpha_1(\omega), \zeta_2, ..., \zeta_d) \in M_n(\omega)\} \ .$$

Durch Auswahl konvergenter Teilfolgen sieht man, daß das Kompaktum $M_n(\omega)$ Punkte der Bauart $(\alpha_1(\omega), \alpha_2(\omega), \zeta_3, ..., \zeta_d)$ enthält. Dies fortsetzend findet man $\alpha(\omega) = (\alpha_1(\omega), \alpha_2(\omega), ..., \alpha_d(\omega))$ in $M_n(\omega)$ mit der Eigenschaft

$$\alpha_j(\omega) \ := \ \min \{\zeta_j : \text{ es gibt Punkte } \zeta = (\alpha_1(\omega), ..., \alpha_{j-1}(\omega), \zeta_j, ..., \zeta_d) \in M_n(\omega)\} \ , \quad 1 \leq j \leq d \ .$$

3) Schreibe nun genauer $\alpha^{(n)}(\omega) := \alpha(\omega)$ und definiere mit $\vartheta_0 \in \Theta$ beliebig

$$T_n(\omega) \ := \ \vartheta_0 \, 1_{A_n^c}(\omega) + \alpha^{(n)}(\omega) \, 1_{A_n}(\omega) \ , \quad \omega \in \Omega \ .$$

Für $\omega \in A_n$ ist $T_n(\omega) \in K_n$ *eine* Minimalstelle von $\xi \to \left\| \widehat{\Psi}_n(\omega) - \Psi_\xi \right\|_H$ auf Θ.
Wir werden in Schritt 5) unten zeigen, daß T_n eine $\mathcal{F}_n$-meßbare Zufallsvariable ist.
4) Zuerst betrachten wir sukzessiv für $j = 1, 2, ..., d$ die Komponenten $\alpha_j^{(n)}$ von $\alpha^{(n)}$ als Abbildungen $A_n \to I\!R$ und weisen deren Meßbarkeit bezüglich der Spur von $\mathcal{F}_n$ auf A_n nach:

$$(\circ) \qquad \alpha_j^{(n)} \ : \ (A_n, \{A_n \cap F : F \in \mathcal{F}_n\}) \longrightarrow (I\!R, \mathcal{B}(I\!R)) \quad \text{ist meßbar, } j = 1, 2, ..., d \ .$$

Schreibe kurz $M := \inf\limits_{\xi \in \Theta} \|\widehat{\Psi}_n - \Psi_\xi\|_H$ und fasse $\{\alpha_j^{(n)} \leq b\} = \{\omega \in A_n : \alpha_j^{(n)}(\omega) \leq b\}$ stets als Teilmenge von A_n auf. Für $i = 1$ gilt für jedes $b \in I\!R$

$$\{\alpha_1^{(n)} \leq b\} \quad = \quad \bigcap_m \bigcup_{\left\{ \substack{\xi \in \mathbb{Q}^d \cap K_n \\ \xi_1 < b + \frac{1}{m}} \right.} \left\{ \|\widehat{\Psi}_n - \Psi_\xi\|_H < M + \frac{1}{m} \right\} \quad \in \quad \mathcal{F}_n$$

nach Konstruktion von $\alpha_1^{(n)}$, und mit Auswahl konvergenter Teilfolgen im Kompaktum K_n (beachte dabei, daß auf der rechten Seite eine Teilmenge von A_n steht). Das ist $(\circ)$ für $i = 1$. Sei nun schon $(\circ)$ für $i = 1, 2, \ldots, j{-}1$ bewiesen. Nach Konstruktion von $\alpha_j^{(n)}$ (und wieder mit Auswahl konvergenter Teilfolgen im Kompaktum K_n) kann dann $\{\alpha_j^{(n)} \leq b\}$ dargestellt werden als

$$\bigcap_m \bigcup_{\left\{ \substack{\xi \in \mathbb{Q}^d \cap K_n \\ \xi_j < b + \frac{1}{m}} \right.} \left[\left\{ |\alpha_i^{(n)} - \xi| < \frac{1}{m} \, , \, i = 1, 2, \ldots, j{-}1 \right\} \bigcap \left\{ \|\widehat{\Psi}_n - \Psi_\xi\|_H < M + \frac{1}{m} \right\} \right] \ .$$

Dies aber ist wegen ($\circ$) für $i = 1, 2, \ldots, j{-}1$ ein Ereignis in $\mathcal{F}_n$, und eine Teilmenge von $A_n \in \mathcal{F}_n$. So erhält man für alle $j = 1, \ldots, d$ die Aussage ($\circ$).

5) Abschluß des Beweises: wegen $A_n \in \mathcal{F}_n$ liefert ($\circ$) für $\beta \in \mathbb{R}^d$ beliebig

$$A_n \cap \{T_n \leq \beta\} \;=\; \{\alpha^{(n)} \leq \beta\} \;\in\; \mathcal{F}_n \;,$$

sicher gilt aber auch

$$A_n^c \cap \{T_n \leq \beta\} \;\in\; \mathcal{F}_n$$

denn das letzte Ereignis ist stimmt entweder mit A_n^c überein, oder ist leer. Damit ist die $\mathcal{F}_n$-Meßbarkeit von T_n bewiesen. □

Das eben in Beweisschritt 4) von Satz 2.10 gegebene Argument löst im hier betrachteten Fall ein 'measurable selection' Problem, das in der asymptotischen Statistik immer dann auftaucht, wenn Schätzer durch Minimalstellen, Maximalstellen oder Nullstellen geeigneter Abbildungen $\xi \to H(\omega, \xi)$ definiert werden sollen. Ein allgemein und zugleich konstruktiv formuliertes 'measurable selection' Theorem findet man in Hammer (2005, Theorem A2 in Anhang A).

2.11 Satz: Unter den Voraussetzungen $\mathbf{SLLN}(\vartheta)$ und $\mathbf{I}(\vartheta)$ für alle $\vartheta \in \Theta$

a) liefert $(T_n)_n$ wie in 2.10 konstruiert eine MD-Schätzfolge für den unbekannten Parameter,

b) sind *beliebige* MD-Schätzfolgen $(\vartheta_n^*)_n$ gemäß 2.9 stark konsistent:

$$\forall\, \vartheta \in \Theta : \quad \vartheta_n^* \;\longrightarrow\; \vartheta \quad P_\vartheta\text{-fast sicher für } n \to \infty.$$

Beweis: Sei $\vartheta \in \Theta$ fest.

a) Es ist zu zeigen, daß die in 2.10 zur Konstruktion von $(T_n)_n$ benutzte Folge

$$A_n \;:=\; \left\{ \min_{\xi \in K_n} \left\| \widehat{\Psi}_n - \Psi_\xi \right\|_H = \inf_{\xi \in \Theta} \left\| \widehat{\Psi}_n - \Psi_\xi \right\|_H \right\} \;\in \mathcal{F}_n$$

unter den Voraussetzungen $\mathbf{SLLN}(\vartheta)$ und $\mathbf{I}(\vartheta)$ die Eigenschaft

$$(+) \qquad P_\vartheta \left(\liminf_{n \to \infty} \, A_n \right) \;=\; 1$$

besitzt. Nach Wahl von $(K_n)_n$ als kompakter Ausschöpfung der offenen Menge Θ gibt es ein n_0 und ein $\varepsilon_0 > 0$ so daß $B_{2\varepsilon_0}(\vartheta) \subset K_n$ für alle $n \geq n_0$. Sei $0 < \varepsilon < \varepsilon_0$ beliebig. Nach Definition von A_n und von T_n in 2.10 muß dann für $n \geq n_0$ gelten

$$\left\{ \min_{\xi : |\xi - \vartheta| \leq \varepsilon} \left\| \widehat{\Psi}_n - \Psi_\xi \right\|_H < \inf_{\xi : |\xi - \vartheta| > \varepsilon} \left\| \widehat{\Psi}_n - \Psi_\xi \right\|_H \right\} \quad \subset \quad \left(A_n \cap \{ |T_n - \vartheta| \leq \varepsilon \} \right) \;.$$

Nach Übergang zu Komplementen heißt das

$$\begin{aligned}\{|T_n-\vartheta|>\varepsilon\}\cup A_n^c \quad&\subset\quad \left\{\min_{\xi:|\xi-\vartheta|\le\varepsilon}\left\|\widehat{\Psi}_n-\Psi_\xi\right\|_H \ge \inf_{\xi:|\xi-\vartheta|>\varepsilon}\left\|\widehat{\Psi}_n-\Psi_\xi\right\|_H\right\}\\ &\subset\quad \left\{\left\|\widehat{\Psi}_n-\Psi_\vartheta\right\|_H \ge \inf_{\xi:|\xi-\vartheta|>\varepsilon}\left\|\widehat{\Psi}_n-\Psi_\xi\right\|_H\right\}\\ &\subset\quad \left\{2\cdot\left\|\widehat{\Psi}_n-\Psi_\vartheta\right\|_H \ge \inf_{\xi:|\xi-\vartheta|>\varepsilon}\left\|\Psi_\xi-\Psi_\vartheta\right\|_H\right\}\end{aligned}$$

für $n\ge n_0$, wobei die umgekehrte Dreiecksungleichung benutzt wurde. Schreibe kurz

$$C_n \;:=\; \left\{2\cdot\left\|\widehat{\Psi}_n-\Psi_\vartheta\right\|_H \ge \inf_{\xi:|\xi-\vartheta|>\varepsilon}\left\|\Psi_\xi-\Psi_\vartheta\right\|_H\right\}\;.$$

Wegen $\mathbf{SLLN}(\vartheta)$ und $\mathbf{I}(\vartheta)$ sieht man sofort

$$P_\vartheta\left(\limsup_{n\to\infty}\,C_n\right)\;=\;P_\vartheta\left(\{\omega:\omega\in C_n \text{ für unendlich viele } n\}\right)\;=\;0\;.$$

Wegen $A_n^c\subset C_n$ für $n\ge n_0$ (das steht insbesondere in der letzten Kette von Inklusionen) hat man also $P_\vartheta\left(\limsup_{n\to\infty}\,A_n^c\right)=0$ und damit (+). Aussage a) ist bewiesen.

b) Sei nun $(\vartheta_n^*)_n$ irgendeine Festlegung einer MD-Schätzfolge für den unbekannten Parameter wie in 2.9, seien $B_n\in\mathcal{F}_n$ Ereignisse, so daß für $\omega\in B_n$ die Abbildung $\Theta\ni\xi\to\left\|\widehat{\Psi}_n(\omega)-\Psi_\xi\right\|_H$ eine Minimalstelle in $\vartheta_n^*(\omega)$ besitzt, $n\ge1$, und so daß gilt

$$(++)\qquad\qquad P_\vartheta\left(\liminf_{n\to\infty}\,B_n\right)\;=\;1\;.$$

Nach Definition von ϑ_n^* und B_n in 2.9 muß gelten

$$\left\{\min_{\xi:|\xi-\vartheta|\le\varepsilon}\left\|\widehat{\Psi}_n-\Psi_\xi\right\|_H < \inf_{\xi:|\xi-\vartheta|>\varepsilon}\left\|\widehat{\Psi}_n-\Psi_\xi\right\|_H\right\}\quad\subset\quad\left(B_n^c\cup\{|\vartheta_n^*-\vartheta|\le\varepsilon\}\right)\;,$$

woraus man mit denselben Ereignissen C_n wie in a) zuerst die Inklusion

$$\{|\vartheta_n^*-\vartheta|>\varepsilon\}\cap B_n\quad\subset\quad C_n$$

und damit die Aussage

$$\{|\vartheta_n^*-\vartheta|>\varepsilon\}\quad\subset\quad C_n\cup B_n^c\;,\quad n\ge1$$

erhält. Aber $\mathbf{SLLN}(\vartheta)$ und $\mathbf{I}(\vartheta)$ garantieren $P_\vartheta\left(\limsup_{n\to\infty}C_n\right)=0$, woraus wegen (++)

$$P_\vartheta\left(\limsup_{n\to\infty}\{|\vartheta_n^*-\vartheta|>\varepsilon\}\right)\;=\;0$$

folgt. Für P_ϑ-fast alle $\omega\in\Omega$ ist also die Folge der Schätzwerte $\vartheta_n^*(\omega)$ für schließlich alle n im Abschluß der Kugel $B_\varepsilon(\vartheta)$ zu finden. Nach Wahl von $\varepsilon>0$ ist das die in b) behauptete P_ϑ-fast sichere Konvergenz der Folge $(\vartheta_n^*)_n$ gegen ϑ. □

Zur Vorbereitung auf den nächsten Satz überlegt man sich

2.12 Hilfssatz: Unter $\mathbf{I}(\vartheta)$ und $\mathbf{D}(\vartheta)$ gilt für jede Folge $\varphi_n \uparrow \infty$ reeller Zahlen

$$\lim_{c\uparrow\infty} \liminf_{n\to\infty} \inf_{|h|>c} \left\| \varphi_n \left(\Psi_{\vartheta+h/\varphi_n} - \Psi_\vartheta \right) \right\|_H \;=\; \infty \,.$$

Beweis: Wegen $\mathbf{I}(\vartheta)$ gilt für jedes $\varepsilon > 0$ zunächst

$$(+) \qquad \inf_{h:|h/\varphi_n|\geq\varepsilon} \left\| \varphi_n \left(\Psi_{\vartheta+h/\varphi_n} - \Psi_\vartheta \right) \right\|_H \;=\; \varphi_n \cdot \inf_{\xi:|\xi-\vartheta|\geq\varepsilon} \|\Psi_\xi - \Psi_\vartheta\|_H \quad \longrightarrow \; \infty$$

für $n \to \infty$. Wegen der in $\mathbf{D}(\vartheta)$ eingeschlossenen Nichtsingularitätsvoraussetzung an die Ableitung $D\Psi_\vartheta$ folgt aus der Stetigkeit von $u \to \|u^\top D\Psi_\vartheta\|_H$

$$(++) \qquad \Delta \;:=\; \min_{|u|=1} \|u^\top D\Psi_\vartheta\|_H \;>\; 0\,.$$

Weiter gilt nach Differenzierbarkeitsbedingung $\mathbf{D}(\vartheta)$

$$(+++) \qquad \sup_{\xi:|\xi-\vartheta|<\varepsilon} \frac{1}{|\xi-\vartheta|} \left\| \Psi_\xi - \Psi_\vartheta - (\xi-\vartheta)^\top D\Psi_\vartheta \right\|_H \;<\; \frac{\Delta}{2}$$

für hinreichend kleines $\varepsilon > 0$. Für festes $c < \infty$ bleibt von

$$\inf_{|h|>c} \left\| \varphi_n \left(\Psi_{\vartheta+h/\varphi_n} - \Psi_\vartheta \right) \right\|_H$$

wegen (+) nur der folgende Anteil (mit hinreichend kleinem $\varepsilon > 0$) zu betrachten:

$$\begin{aligned}
&\inf_{|h|>c\,,\,|h/\varphi_n|<\varepsilon} \left\| \varphi_n \left(\Psi_{\vartheta+h/\varphi_n} - \Psi_\vartheta \right) \right\|_H \;=\; \inf_{|h|>c\,,\,|h/\varphi_n|<\varepsilon} |h| \, \frac{\left\| \left(\Psi_{\vartheta+h/\varphi_n} - \Psi_\vartheta \right) \right\|_H}{|h/\varphi_n|} \\
&\geq\; c \,\cdot\, \inf_{\xi:|\xi-\vartheta|<\varepsilon} \frac{\|(\Psi_\xi - \Psi_\vartheta)\|_H}{|\xi-\vartheta|} \\
&\geq\; c \,\cdot\, \left(\inf_{\xi:|\xi-\vartheta|<\varepsilon} \frac{\left\|(\xi-\vartheta)^\top D\Psi_\vartheta\right\|_H}{|\xi-\vartheta|} \;-\; \sup_{\xi:|\xi-\vartheta|<\varepsilon} \frac{\left\|\Psi_\xi - \Psi_\vartheta - (\xi-\vartheta)^\top D\Psi_\vartheta\right\|_H}{|\xi-\vartheta|} \right) \\
&\geq\; c \,\cdot\, (\Delta - \frac{\Delta}{2}) \quad=\; c \,\cdot\, \frac{\Delta}{2}\,,
\end{aligned}$$

nach (++) und (+++). Das aber strebt für $c \to \infty$ gegen $+\infty$. □

2.13 Hilfssatz: Unter $\mathbf{I}(\vartheta)$, $\mathbf{D}(\vartheta)$, $\mathbf{T}(\vartheta)$ gilt mit $\varphi_n = \varphi_n(\vartheta) \uparrow \infty$ wie in 2.8 d): die Familie

$$\mathcal{L}\left(\varphi_n(\vartheta_n^* - \vartheta) \mid P_\vartheta \right)\,, \quad n \geq 1$$

(Wahrscheinlichkeitsmaße in $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$) ist straff in $\mathbb{R}^d$.

Beweis: Zu zeigen ist: für jedes $\varepsilon > 0$ gibt es eine Konstante $K = K(\varepsilon)$ so daß

$$(+) \qquad \limsup_{n\to\infty} P_\vartheta \left(|\varphi_n (\vartheta_n^* - \vartheta)| > K \right) \; < \; \varepsilon \, .$$

1) Man hat für beliebiges K mit den üblichen Argumenten (wie im Beweis von 2.11)

$$\begin{aligned} D_n(K) \quad &:= \left\{ \min_{\xi : |\xi - \vartheta| \le K/\varphi_n} \left\| \widehat{\Psi}_n - \Psi_\xi \right\|_H \; \ge \inf_{\xi : |\xi - \vartheta| > K/\varphi_n} \left\| \widehat{\Psi}_n - \Psi_\xi \right\|_H \right\} \\ &\subset \quad \left\{ 2 \cdot \left\| \widehat{\Psi}_n - \Psi_\vartheta \right\|_H \; \ge \inf_{\xi : |\xi - \vartheta| > K/\varphi_n} \left\| \Psi_\xi - \Psi_\vartheta \right\|_H \right\} \\ &\subset \quad \left\{ 2 \cdot \left\| \varphi_n \left(\widehat{\Psi}_n - \Psi_\vartheta \right) \right\|_H \; \ge \; \inf_{|h| > K} \left\| \varphi_n \left(\Psi_{\vartheta + h/\varphi_n} - \Psi_\vartheta \right) \right\|_H \right\} \, . \end{aligned}$$

2) Nach Hilfssatz 2.12 kann im Limes für $n \to \infty$ die deterministische Größe

$$\inf_{|h| > K} \left\| \varphi_n \left(\Psi_{\vartheta + h/\varphi_n} - \Psi_\vartheta \right) \right\|_H$$

am Ende von Schritt 1) durch geeignete Wahl von K beliebig groß gemacht werden, wegen $\mathbf{I}(\vartheta)$ und $\mathbf{D}(\vartheta)$. Zugleich ist wegen Voraussetzung $\mathbf{T}(\vartheta)$ die Familie

$$\mathcal{L}\left(\left\| \varphi_n \left(\widehat{\Psi}_n - \Psi_\vartheta \right) \right\|_H \mid P_\vartheta \right) \;, \qquad n \ge 1$$

straff in $\mathbb{R}$, so daß es in der Kette von Inklusionen aus Schritt 1) für jedes $\varepsilon > 0$ eine Konstante $K_0 = K_0(\varepsilon)$ gibt mit

$$(++) \qquad \limsup_{n\to\infty} P_\vartheta \left(D_n(K_0) \right) \; < \; \varepsilon \, .$$

3) Wählt man nun eine Folge $(B_n)_n$ von Ereignissen wie in 2.9, so daß auf B_n der Schätzer ϑ_n^* ein globales Minimum der Abbildung $\xi \to \left\| \widehat{\Psi}_n - \Psi_\xi \right\|_H$ liefert, so erhält man

$$(+++) \qquad \{ |\varphi_n(\vartheta_n^* - \vartheta)| > K_0 \} \quad \subset \quad B_n^c \cup D_n(K_0) \, .$$

Gleichzeitig erfüllt $(B_n)_n$ nach 2.9 aber $P_\vartheta \left(\liminf_{n\to\infty} B_n \right) = 1$, und damit erst recht

$$\lim_{n\to\infty} P_\vartheta \left(B_n \right) \; = \; 1$$

(beachte $\liminf_{n\to\infty} B_n = \bigcup_n \bigcap_{m \ge n} B_m$; dabei ist $\bigcap_{m \ge n} B_m$ aufsteigend in n). Wegen (++) und (+++) ist die Behauptung (+) bewiesen. □

2.14 Hauptsatz (Millar 1984) : Es gelte **SLLN**(ϑ), **I**(ϑ), **D**(ϑ), **T**(ϑ), mit einer Folge von Normierungskonstanten $\varphi_n = \varphi_n(\vartheta) \uparrow \infty$ wie in 2.8.III.c). Sei $(\vartheta_n^*)_n$ eine beliebige Festlegung einer MD-Schätzfolge gemäß 2.9. Dann gilt: die reskalierten Schätzfehler können als

$$\varphi_n(\vartheta_n^* - \vartheta) \;=\; \Pi_\vartheta\left(\varphi_n(\widehat{\Psi}_n - \Psi_\vartheta)\right) \;+\; o_{P_\vartheta}(1) \quad \text{für } \; n \to \infty$$

dargestellt werden, mit einer linearen Abbildung

$$\Pi_\vartheta(f) \;:=\; \Lambda_\vartheta^{-1} \begin{pmatrix} \langle D_1\Psi_\vartheta, f\rangle \\ \ldots \\ \langle D_d\Psi_\vartheta, f\rangle \end{pmatrix} \;, \quad f \in H$$

von H nach $I\!R^d$ mit

$$\Lambda_\vartheta \;:=\; \left(\langle D_i\Psi_\vartheta, D_j\Psi_\vartheta\rangle\right)_{1\le i,j\le d} \;.$$

Beweis: 0) Beachte: nach Voraussetzung **D**(ϑ) sind die Komponenten $D_1\Psi_\vartheta$, .., $D_d\Psi_\vartheta$ der Ableitung $D\Psi_\vartheta$ linear unabhängig in H.

1) Für $f \in H$ und $h \in I\!R^d$ gilt $\Pi_\vartheta(f) = h$ genau dann, wenn die Projektion von f auf den linearen Unterraum

$$V_\vartheta \;:=\; \text{span}(D_i\Psi_\vartheta : 1 \le i \le d)$$

durch $\sum_{i=1}^n h_i\, D_i\Psi_\vartheta$ gegeben ist:

Sei $f \in H$; dann gibt es ein eindeutig bestimmtes $h \in I\!R^d$ so daß

$$\left(f - h^\top D\Psi_\vartheta\right) \;=\; \left(f - \sum_{i=1}^d h_i\, D_i\Psi_\vartheta\right) \quad \perp \quad V_\vartheta$$

Dabei ist die Aussage

$$0 \;=\; \langle f - \sum_{i=1}^d h_i\, D_i\Psi_\vartheta, D_j\Psi_\vartheta\rangle \;=\; \langle f, D_j\Psi_\vartheta\rangle \;-\; \sum_{i=1}^d h_i\,(\Lambda_\vartheta)_{i,j} \;, \quad 1 \le j \le d$$

gleichwertig mit $(\langle D_1\Psi_\vartheta, f\rangle, \ldots, \langle D_d\Psi_\vartheta, f\rangle) \;=\; h^\top\, \Lambda_\vartheta$, und folglich wegen Invertierbarkeit von Λ_ϑ gleichwertig mit

$$h \;=\; \Lambda_\vartheta^{-1} \begin{pmatrix} \langle D_1\Psi_\vartheta, f\rangle \\ \ldots \\ \langle D_d\Psi_\vartheta, f\rangle \end{pmatrix} \;=\; \Pi_\vartheta(f) \;.$$

2) Nach Schritt 1) besitzt die Abbildung

$$(*) \qquad I\!R^d \ni \quad h \;\longrightarrow\; \left\| \varphi_n\left(\widehat{\Psi}_n - \Psi_\vartheta\right) \;-\; h^\top D\Psi_\vartheta \right\| \quad =: \; d_{n,\vartheta}(h) \quad \in \; [0,\infty)$$

eine eindeutige Minimalstelle $\widehat{h}_n$ gegeben durch

$$\widehat{h}_n \;:=\; \Pi_\vartheta\left(\varphi_n(\widehat{\Psi}_n - \Psi_\vartheta)\right)\;;$$

für diese gilt

$$\left(\varphi_n(\widehat{\Psi}_n - \Psi_\vartheta) \;-\; \widehat{h}_n^\top D\Psi_\vartheta\right) \;\perp\; V_\vartheta$$

und damit nach Pythagoras

$$d^2_{n,\vartheta}(h) \;=\; \left\|(h - \widehat{h}_n)^\top D\Psi_\vartheta\right\|^2 \;+\; d^2_{n,\vartheta}(\widehat{h}_n)\;.$$

Wegen $\Delta := \min\limits_{|u|=1} \|u^\top D\Psi_\vartheta\| > 0$ nach $\mathbf{D}(\vartheta)$ bedeutet das

$$(+) \qquad \forall\, h \in I\!R^d\,, \forall n\;: \quad d^2_{n,\vartheta}(h) \;\geq\; d^2_{n,\vartheta}(\widehat{h}_n) \;+\; \left|h - \widehat{h}_n\right|^2 \Delta^2\;.$$

3) Π_ϑ ist eine lineare Abbildung von H nach $I\!R^d$, und wegen $\widehat{h}_n = \Pi_\vartheta\left(\varphi_n(\widehat{\Psi}_n - \Psi_\vartheta)\right)$ zeigt die Straffheitsvoraussetzung $\mathbf{T}(\vartheta)$:

$$\text{die Familie } \mathcal{L}(\,\widehat{h}_n \mid P_\vartheta),\; n \geq 1\,, \text{ ist straff in } I\!R^d.$$

4) Sei nun $(\vartheta^*_n)_n$ eine MD-Schätzfolge und $(A_n)_n$ eine Folge von Ereignissen $A_n \in \mathcal{F}_n$ wie in Definition 2.9. Dann gilt (insbesondere) $\lim\limits_{n\to\infty} P_\vartheta(A_n) = 1$, und für jedes $n \geq 1$:

$$\text{auf } A_n \text{ ist } \vartheta^*_n \text{ eine Minimalstelle der Abbildung } \;\Theta \ni \xi \;\to\; \left\|\widehat{\Psi}_n - \Psi_\xi\right\| \in [0,\infty)\;.$$

Schreibe $h^*_n := \varphi_n(\vartheta^*_n - \vartheta)$. Ersetzt man in der letzten Aussage ξ durch $\vartheta + h/\varphi_n$, so folgt

$$(**) \qquad \text{auf } A_n \text{ ist } h^*_n \text{ Minimalstelle von } \quad \Theta_{\vartheta,n} \ni\; h \;\to\; \left\|\varphi_n(\widehat{\Psi}_n - \Psi_{\vartheta+h/\varphi_n})\right\| \;=:\; \delta_{n,\vartheta}(h)$$

wobei (da Θ offen in $I\!R^d$)

$$\Theta_{\vartheta,n} \;:=\; \{h \in I\!R^d : \vartheta + h/\varphi_n \in \Theta\} \quad\uparrow\; I\!R^d\,,\quad n \to \infty\;.$$

5) Nach Hilfssatz 2.13 ist bereits bewiesen:

$$\text{die Familie } \mathcal{L}(\,h^*_n \mid P_\vartheta),\; n \geq 1\,, \text{ ist straff in } I\!R^d.$$

6) Die Funktion $\delta_{n,\vartheta}(\cdot)$ in $(**)$ zerlegt man in

$$\begin{aligned}
\delta_{n,\vartheta}(h) \;&=\; \left\|\,\varphi_n(\widehat{\Psi}_n - \Psi_\vartheta) - \varphi_n(\Psi_{\vartheta+h/\varphi_n} - \Psi_\vartheta)\,\right\| \\
&=\; \left\|\,\varphi_n(\widehat{\Psi}_n - \Psi_\vartheta) \;-\; h^\top D\Psi_\vartheta \;-\; \varphi_n\left(\Psi_{\vartheta+h/\varphi_n} - \Psi_\vartheta - (h/\varphi_n)^\top D\Psi_\vartheta\right)\,\right\|
\end{aligned}$$

so daß nach Voraussetzung $\mathbf{D}(\vartheta)$ und wegen $(*)$ für jedes Kompaktum K in $\mathbb{R}^d$ gilt

$$(++) \qquad \sup_{h\in K} |\,\delta_{n,\vartheta}(h) - d_{n,\vartheta}(h)\,| \quad \longrightarrow \quad 0 \quad \text{für } n\to\infty.$$

Zugleich hat man aber nach $(+)$ für dieses K

$$\begin{aligned} P_\vartheta\left(|h_n^* - \widehat{h}_n| > \varepsilon\right) \;&\le\; P_\vartheta\left(\widehat{h}_n \notin K\right) + P_\vartheta\left(h_n^* \notin K\right) + P_\vartheta\left(A_n^c\right) \\ &\quad + \; P_\vartheta\left(\{\widehat{h}_n \in K\} \cap \{h_n^* \in K\} \cap A_n \cap \{|h_n^* - \widehat{h}_n| > \varepsilon\}\right) \\ &\le\; P_\vartheta\left(\widehat{h}_n \notin K\right) + P_\vartheta\left(h_n^* \notin K\right) + P_\vartheta\left(A_n^c\right) \\ &\quad + \; P_\vartheta\left(\{\widehat{h}_n \in K\} \cap \{h_n^* \in K\} \cap A_n \cap \{d^2_{n,\vartheta}(h_n^*) \ge d^2_{n,\vartheta}(\widehat{h}_n) + \varepsilon^2\Delta^2\}\right) \;. \end{aligned}$$

7) Sei nun $\varepsilon > 0$ beliebig klein. Wähle ein Kompaktum $K = K(\varepsilon)$ in $\mathbb{R}^d$ so daß

$$\sup_{n\ge 1} P_\vartheta\left(\widehat{h}_n \notin K\right) \;<\; \frac{\varepsilon}{2}\,, \quad \sup_{n\ge 1} P_\vartheta\left(h_n^* \notin K\right) \;<\; \frac{\varepsilon}{2}\,,$$

dann zeigt Schritt 6) wegen $\lim\limits_{n\to\infty} P_\vartheta(A_n) = 1$:

$$\begin{aligned} &\limsup_{n\to\infty} \; P_\vartheta\left(|h_n^* - \widehat{h}_n| > \varepsilon\right) \\ &< \quad \varepsilon \;+\; \limsup_{n\to\infty} \; P_\vartheta\left(\{\widehat{h}_n \in K\} \cap \{h_n^* \in K\} \cap A_n \,\cap \left\{d^2_{n,\vartheta}(h_n^*) \ge d^2_{n,\vartheta}(\widehat{h}_n) + \varepsilon^2\Delta^2\right\}\right) \;. \end{aligned}$$

Wegen $(++)$ verschwindet die Differenz $|\delta^2_{n,\vartheta} - d^2_{n,\vartheta}|$ für $n\to\infty$ gleichmäßig auf K, daher

$$\begin{aligned} &\limsup_{n\to\infty} \; P_\vartheta\left(|h_n^* - \widehat{h}_n| > \varepsilon\right) \\ &< \quad \varepsilon \;+\; \limsup_{n\to\infty} \; P_\vartheta\left(A_n \,\cap \{\widehat{h}_n \in K\} \cap \left\{\delta^2_{n,\vartheta}(h_n^*) \ge \delta^2_{n,\vartheta}(\widehat{h}_n) + \frac{\varepsilon^2}{2}\Delta^2\right\}\right) \;. \end{aligned}$$

Für hinreichend großes n gilt $\Theta_{\vartheta,n} \supset K$, und auf A_n ist nach Schritt 4) h_n^* eine Minimalstelle der Abbildung $\Theta_{\vartheta,n} \ni h \to \delta^2_{n,\vartheta}(h)$. Folglich erscheint auf der rechten Seite der letzten Ungleichung für großes n die Wahrscheinlichkeit einer leeren Menge, und man erhält

$$\limsup_{n\to\infty} \; P_\vartheta\left(|h_n^* - \widehat{h}_n| > \varepsilon\right) \;<\; \varepsilon \;.$$

Da $\varepsilon > 0$ beliebig war, sind die Folgen $(h_n^*)_n$ und $(\widehat{h}_n)_n$ unter P_ϑ asymptotisch äquivalent, und aus

$$\widehat{h}_n \;=\; \Pi_\vartheta\left(\varphi_n(\widehat{\Psi}_n - \Psi_\vartheta)\right)$$

folgt die Behauptung

$$h_n^* \;=\; \Pi_\vartheta\left(\varphi_n(\widehat{\Psi}_n - \Psi_\vartheta)\right) \;+\; o_{P_\vartheta}(1)\,, \quad n\to\infty\,.$$

Der Beweis von 2.14 ist damit abgeschlossen. □

2.15 Bemerkung: Will man sich in Satz 2.11 anstelle der starken Konsistenz der MD-Schätzfolge $(\vartheta_n^*)_n$ mit der üblichen (schwachen) Konsistenz (wie in 1.9) begnügen, so reicht es, in der Definition 2.9 einer MD-Schätzfolge anstelle von $P_\vartheta(\liminf_{n\to\infty} A_n) = 1$ nur $\lim_{n\to\infty} P_\vartheta(A_n) = 1$ zu fordern, und Voraussetzung **SLLN**(ϑ) in 2.8.II a) zu einem *schwachen Gesetz der großen Zahlen* **WLLN**(ϑ) herabzustufen, d.h. in 2.8.II.a) nur P_ϑ-stochastische Konvergenz von $\left\|\widehat{\Psi}_n - \Psi_\vartheta\right\|_H$ gegen 0 zu fordern. Die Ergebnisse aus 2.12–2.13 bleiben mit dieser Modifikation ohne jede Änderung in den Beweisen gültig. Die Aussage des Hauptsatzes 2.14 bleibt (mit schwacher statt starker Konsistenz der MD Schätzfolge) dieselbe.

C. Gaußprozesse und asymptotische Normalität für Minimum-Distanz-Schätzfolgen

2.16 Voraussetzungen und Bezeichnungen für Teilkapitel C: Sei $k \geq 1$, sei T ein k-dimensionales Intervall versehen mit seiner Borel-σ-Algebra $\mathcal{T} = \mathcal{B}(T)$, der Spur von $\mathcal{B}(\mathbb{R}^k)$ auf T. Sei μ ein endliches Maß auf $(T, \mathcal{T})$, absolutstetig bezüglich des auf T eingeschränkten Lebesgue-Maßes $\lambda\!\!\lambda$. Für den Hilbertraum

$$H \ := \ L^2(T, \mathcal{T}, \mu)$$

führen wir die Ergebnisse aus den Teilkapiteln A und B zusammen: wie dort ist

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

ein statistisches Experiment, $(\mathcal{F}_n)_n$ eine aufsteigende Familie von Sub-σ-Algebren von $\mathcal{A}$, und

$$\{\Psi_\xi : \xi \in \Theta\} \ \subset \ L^2(T, \mathcal{T}, \mu)$$

eine stetig parametrisierte Schar von Funktionen $\Psi_\xi : T \ni t \to \Psi_\xi(t) \in \mathbb{R}$. Neu setzen wir voraus, daß die Familie $(\widehat{\Psi}_n)_n$ in 2.8 aus einer Familie reellwertiger Prozesse

$$X^n \ : \ (T \times \Omega, \mathcal{T} \otimes \mathcal{F}_n) \ \longrightarrow \ (\mathbb{R}, \mathcal{B}(\mathbb{R})) \quad \text{meßbar, mit Pfaden in } L^2(T, \mathcal{T}, \mu) \ , \ n \geq 1 \ ,$$

als Pfad von X^n entstehe:

$$\text{für alle } \omega \in \Omega \ : \quad \widehat{\Psi}_n(\omega) \ := \ X^n_\bullet(\omega) \ = \ X^n(\cdot, \omega) \ , \quad n \geq 1 \ .$$

Damit ist $\widehat{\Psi}_n$ nach 2.2 eine wohldefinierte $(H, \mathcal{B}(H))$–wertige Zufallsvariable auf $(\Omega, \mathcal{F}_n)$, für jedes $n \geq 1$, und alle in 2.1 und in 2.8 I) gemachten Voraussetzungen sind erfüllt.

Wir beschreiben zuerst eine Klasse von Limesobjekten für schwach in $L^2(T, \mathcal{T}, \mu)$ konvergierende Teilfolgen der $\varphi_n(\widehat{\Psi}_n - \Psi_\vartheta)$ unter P_ϑ aus 2.14.

2.17 Definition: Ein reellwertiger meßbarer Prozeß $(X_t)_{t \in T}$ heißt (zentrierter) *Gaußprozeß mit Kovarianzkern* $K(\cdot, \cdot)$ falls es eine Teilmenge $N \in \mathcal{T}$ mit $\lambda\!\!\lambda(N) = 0$ gibt so daß

$$\mathcal{L}\,(X_{t_1}, \ldots, X_{t_l}) \;=\; \mathcal{N}\,(\,0_l\,,\,(K(t_i, t_j)_{i,j=1,\ldots,l}\,)$$

für beliebige Punkte $t_1, \ldots, t_l \in T \setminus N$, $l \geq 1$.

Zuerst nennen wir ein einfaches (aber wichtiges) Beispiel.

2.18 Beispiele: a) Sei $T = [0, \infty)$ und sei $(B_t)_{t \geq 0}$ eine Standard-Brownsche Bewegung mit $B_0 \equiv 0$. Wegen der Stetigkeit aller Pfade ist B, definiert auf irgendeinem Raum $(\Omega', \mathcal{A}', P')$, ein meßbarer stochastischer Prozeß. Da Zuwächse von B unabhängig und über Zeitintervalle $[t_1, t_2]$ nach $\mathcal{N}(0, t_2 - t_1)$ verteilt sind, gilt $E(B_t) = 0$ für alle t, und

$$E(B_{t_1} B_{t_2}) \;=\; E(B_{t_1}^2) + E(B_{t_1}(B_{t_2} - B_{t_1})) \;=\; t_1 \quad \text{falls } t_1 < t_2$$

und damit $E(B_{t_1} B_{t_2}) = t_1 \wedge t_2$ für alle t_1, t_2 in $[0, \infty)$. Also ist B im Sinn von 2.17 ein Gaußprozeß mit Kovarianzkern

$$K(t_1, t_2) = t_1 \wedge t_2\,, \quad t_1, t_2 \geq 0\,.$$

b) Sei $T = [0, 1]$ und sei $B^0 = (B_t^0)_{0 \leq t \leq 1}$ eine *Brownsche Brücke*, d.h.

$$B_t^0 \;=\; B_t \;-\; t \cdot B_1\,, \quad 0 \leq t \leq 1\,.$$

Nach den üblichen Transformationseigenschaften von Normalverteilungen erhält man mit

$$\begin{pmatrix} B_{t_1}^0 \\ \cdots \\ B_{t_l}^0 \end{pmatrix} = \widetilde{A} \cdot \begin{pmatrix} B_{t_1} \\ \cdots \\ B_{t_l} \\ B_1 \end{pmatrix} \quad , \quad \widetilde{A} := \begin{pmatrix} 1 & 0 & \ldots & 0 & -t_1 \\ 0 & 1 & \ldots & 0 & -t_2 \\ \cdots & & & & \\ 0 & 0 & \ldots & 1 & -t_l \end{pmatrix}$$

die endlichdimensionalen Randverteilungen: B^0 ist ein Gaußprozeß mit Kovarianzkern

$$K(t_1, t_2) = t_1 \wedge t_2 \;-\; t_1\, t_2 \;, \quad t_1, t_2 \in [0,1] \;.$$

Da jeder Pfad der Brownschen Brücke nach Konstruktion auf $[0,1]$ beschränkt ist, ist B^0 für jedes endliche Maß μ auf $[0,1]$ nach 2.2 ein Prozeß mit Pfaden in $L^2([0,1], \mathcal{B}([0,1]), \mu)$.
c) Sei F eine beliebige Verteilungsfunktion auf $\mathbb{R}$. Die *durch F zeittransformierte Brownsche Brücke* $B^{0,F}$

$$B_t^{0,F} \;:=\; B^0_{F(t)} \;, \quad t \in \mathbb{R}$$

ist wegen der Rechtsstetigkeit aller Pfade ein meßbarer stochastischer Prozeß. Nach b) ist $B^{0,F}$ ein Gaußprozeß mit Kovarianzkern

$$K(t_1, t_2) \;=\; F(t_1) \wedge F(t_2) \;-\; F(t_1) F(t_2) \;, \quad t_1, t_2 \in \mathbb{R} \;.$$

Mit demselben Argument wie in b) ist für jede Verteilungsfunktion F und für jedes endliche Maß μ auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ $B^{0,F}$ ein Prozeß mit Pfaden in $L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$.

Die Existenz von Gaußprozessen wurde – zusammen mit expliziten 'orthogonalen' Darstellungen – von Loève um 1945 bewiesen, siehe Loève (1978, Vol.2, Sec. 36-37; 1963, S. 478).

2.19 Satz: Die Abbildung $K(\cdot,\cdot) : T \times T \to \mathbb{R}$ sei symmetrisch, nichtnegativ definit (d.h.

$$\sum_{i,j=1,\ldots,\ell} \alpha_i\, K(t_i, t_j)\, \alpha_j \;\;\geq\;\; 0$$

für beliebige $\ell \geq 1$, $t_1, \ldots, t_\ell \in T$, $\alpha_1, \ldots, \alpha_\ell \in \mathbb{R}$), beschränkt und stetig.
Dann gibt es einen reellwertigen Gaußprozeß $(X_t)_{t \in T}$ mit Kovarianzkern $K(\cdot,\cdot)$ gemäß 2.17.

Beweis: 1) Da $K(.,.)$ symmetrisch und nichtnegativ definit, gibt es für beliebige $\ell \geq 1$ und $t_1, \ldots, t_\ell$ in T eine zentrierte Normalverteilung $P_{t_1,\ldots,t_\ell}$ mit Kovarianzmatrix $(K(t_i, t_j))_{i,j=1,\ldots,\ell}$, mit charakteristischer Funktion

$$\mathbb{R}^\ell \;\ni\; \lambda \;\;\longrightarrow\;\; e^{-\frac{1}{2}\,\lambda^\top \Sigma\, \lambda} \;, \quad \Sigma \;:=\; (K(t_i, t_j))_{i,j=1,\ldots,\ell} \;.$$

Damit ist eine konsistente Familie von Wahrscheinlichkeitsmaßen

$$P_{t_1,\ldots,t_\ell} \quad \text{Ws-Maß auf} \quad \left(\mathop{\times}_{i=1}^{\ell} \mathbb{R} \,,\; \bigotimes_{i=1}^{\ell} \mathcal{B}(\mathbb{R}) \right) \;, \quad t_1, \ldots, t_\ell \in T \;, \quad \ell \geq 1$$

festgelegt. Zu dieser gibt es mit dem Konsistenzsatz von Kolmogorov genau ein Wahrscheinlichkeitsmaß

$$P \quad \text{auf} \quad \left(\bigtimes_{t\in T} I\!R\,,\ \bigotimes_{t\in T} \mathcal{B}(I\!R) \right) \,,$$

so daß unter P der kanonische Prozeß $X = (X_t)_{t\in T}$ auf dem Produktraum

$$(\Omega, \mathcal{A}) \ := \ \left(\bigtimes_{t\in T} I\!R\,,\ \bigotimes_{t\in T} \mathcal{B}(I\!R) \right)$$

die richtigen endlichdimensionalen Randverteilungen besitzt: für $\ell \geq 1$ und $t_1, \ldots, t_\ell$ in T gilt

$$(+) \qquad \mathcal{L}\left((X_{t_1}, \ldots, X_{t_\ell}) \mid P\right) \ = \ \mathcal{N}\left((0)_{i=1,\ldots,\ell}\,,\ (K(t_i,t_j))_{i,j=1,\ldots,\ell} \right) \,.$$

2) Wegen Stetigkeit von $K(.,.)$ strebt für konvergente Folgen $t_n \to t$

$$E_P\left((X_{t_n} - X_t)^2\right) \ = \ K(t_n,t_n) - 2K(t_n,t) + K(t,t)$$

gegen 0. Insbesondere hat man stochastische Konvergenz, also für konvergente Folgen $t_n \to t$

$$(++) \qquad X_{t_n} \ = \ X_t + o_P(1)\,, \quad n \to \infty \,.$$

3) Zu X definiert man eine Schar $\mathcal{T}\otimes\mathcal{A}$–meßbarer Approximationen $(X^m)_m$.

i) Wir pflastern $I\!R^k$ wie in 2.1' mit halboffenen Würfeln $A_l(m)$ der Kantenlänge 2^{-m}

$$A_l(m) := \bigtimes_{j=1}^{k} \left[\frac{l_j}{2^m}, \frac{l_j+1}{2^m} \right) \,, \quad l = (l_1, ..., l_k) \in Z\!\!\!Z^k$$

und definieren für das k-dimensionale Intervall T $\Lambda(m,T)$ als Menge aller Indices $l \in Z\!\!\!Z^k$ mit

$$A_l(m) \,\cap\, T \,\neq\, \emptyset \quad \text{und} \quad \max_{i=1,\ldots,k} |l_i| < m \,.$$

Dann wählen wir für jedes $l \in \Lambda(m,T)$ einen Punkt $t_l(m)$ in $\overline{A_l(m)} \cap T$ (für alle Indices l mit $\overline{A_l(m)} \subset T$ kann man also wie in 2.1'

$$t_l(m) := \left(\frac{l_1+1}{2^m}, \ldots, \frac{l_m+1}{2^m} \right)$$

festlegen). Für jedes $t \in T$ gibt es für schließlich alle m genau ein l mit $t \in A_l(m)$. Definiere

$$X^m(t,\omega) \ := \ \sum_{l\in\Lambda(m,T)} 1_{A_l(m)}(t)\ X(t_l(m),\omega)\,, \quad t \in T\,,\, \omega \in \Omega \,.$$

Dann ist X^m ein $\mathcal{T}\otimes\mathcal{A}$–meßbarer Prozess. Wegen Endlichkeit von $\Lambda(m,T)$ und wegen (+) in Schritt 1) sind die endlichdimensionalen Randverteilungen von X^m Normalverteilungen. Also ist für jedes $m \geq 1$ X^m ein Gaußprozeß im Sinne von 2.17. Nach Konstruktion von X^m und mit stochastischen Entwicklungen (++) konvergieren die endlichdimensionalen Randverteilungen der X^m gegen die von X:

$$(*) \qquad \mathcal{L}\left((X^m_{t_1}, \ldots, X^m_{t_\ell}) \mid P\right) \longrightarrow \mathcal{L}\left((X_{t_1}, \ldots, X_{t_\ell}) \mid P\right) \quad (\text{schwach in } \mathbb{R}^\ell, \text{ für } m \to \infty)$$

für beliebige $\ell \geq 1$ und $t_1, \ldots, t_\ell \in T$.

ii) Fixiere ein Wahrscheinlichkeitsmaß $\widetilde{\mu}$ auf $(T, \mathcal{T})$ mit strikt positiver Lebesgue-Dichte. Wir zeigen: $(X^m)_m$ ist eine Cauchyfolge in $L^2(T\times\Omega, \mathcal{T}\otimes\mathcal{A}, \widetilde{\mu}\otimes P)$.

Wegen Beschränktheit von $K(.,.)$ gilt für jedes $m \geq 1$ zunächst

$$\int_{T\times\Omega} (X^m(t,\omega))^2 \ (\widetilde{\mu}\otimes P)(dt, d\omega) \ = \ \int_T E\left((X^m_t)^2\right) \widetilde{\mu}(dt) \ = \ \int_T [\sup_{t\in T} K(t,t)]\, \widetilde{\mu}(dt) \ < \ \infty \, .$$

Wegen Endlichkeit von $\Lambda(m', T)$ verschwinden für $m \to \infty$ und $m' > m$ die Integrale

$$\int_{T\times\Omega} |X^{m'} - X^m|^2 \, d(\widetilde{\mu}\otimes P) \ = \ \int_T E\left((X^{m'}_t - X^m_t)^2\right) \widetilde{\mu}(dt) \, ,$$

wegen dominierter Konvergenz, denn der Integrand auf der rechten Seite ist

$$\sum_{\substack{l'\in\Lambda(m',T)\\ l\in\Lambda(m,T)}} 1_{A_{l'}(m')\cap A_l(m)}(t) \ \left(K(t_{l'}(m'), t_{l'}(m')) - 2K(t_{l'}(m'), t_l(m)) + K(t_l(m), t_l(m))\right)$$

mit stetigem und beschränktem $(t_1, t_2) \to K(t_1, t_2)$.

iii) Zur Cauchyfolge $(X^m)_m$ in $L^2(T\times\Omega, \mathcal{T}\otimes\mathcal{A}, \widetilde{\mu}\otimes P)$ gibt es eine Teilfolge $(m_k)_k$ und ein $\widetilde{X}$ in $L^2(T\times\Omega, \mathcal{T}\otimes\mathcal{A}, \widetilde{\mu}\otimes P)$ so daß gilt

$$X^{m_k} \ \longrightarrow \ \widetilde{X} \qquad \widetilde{\mu}\otimes P\text{–fast sicher auf } (T\times\Omega, \mathcal{T}\otimes\mathcal{A}) \, , \quad k \to \infty \, .$$

Als fast sicherer Limes $\mathcal{T}\otimes\mathcal{A}$–meßbarer Prozesse ist $\widetilde{X}$ ein $\mathcal{T}\otimes\mathcal{A}$–meßbarer Prozeß, und es gibt eine $\widetilde{\mu}\otimes P$–Nullmenge $M \in \mathcal{T}\otimes\mathcal{A}$ so daß

$$(**) \qquad [1 - 1_M]\, X^{m_k} \ \longrightarrow \ [1 - 1_M]\, \widetilde{X} \quad \text{punktweise auf } T\times\Omega \text{ für } k \to \infty \, .$$

4) Mit Schreibweise M_t für t-Schnitte durch M

$$M_t \ = \ \{\omega \in \Omega : (t,\omega) \in M\} \ \in \ \mathcal{A} \, , \quad t \in T$$

folgt aus $0 = (\widetilde{\mu}\otimes P)(M) = \int_T P(M_t)\, \widetilde{\mu}(dt)$ die Existenz einer $\widetilde{\mu}$-Nullmenge $N \in \mathcal{T}$ so daß $P(M_t) = 0$ für alle t in $T \setminus N$. Vorausgesetzt wurde $\widetilde{\mu} \sim \lambda\!\!\lambda$, also ist die Ausnahmemenge $N \in \mathcal{T}$ auch eine $\lambda\!\!\lambda$-Nullmenge. Gezeigt ist

$$(***) \qquad \text{es gibt } N \in \mathcal{T} \text{ mit } \lambda\!\!\lambda(N) = 0 \text{ so daß} \quad P(M_t) \ = \ 0 \quad \text{für alle } t \text{ in } T \setminus N \, .$$

5) Abschluss des Beweises: für beliebige $\ell \geq 1$ und $t_1, \ldots, t_\ell$ in $T \setminus N$ konvergieren nach $(**)$

$$(\diamond) \qquad \Big([1-1_{M_{t_1}}]X_{t_1}^{m_k}(\omega), \ldots, [1-1_{M_{t_\ell}}]X_{t_\ell}^{m_k}(\omega)\Big)$$

für $k \to \infty$ gegen

$$\Big([1-1_{M_{t_1}}]\widetilde{X}_{t_1}(\omega), \ldots, [1-1_{M_{t_\ell}}]\widetilde{X}_{t_\ell}(\omega))\Big) \ .$$

Nach $(*)$ und $(***)$ streben aber endlichdimensionale Randverteilungen zu $(\diamond)$ gegen die entsprechenden endlichdimensionalen Randverteilungen des Prozesses X aus Schritt 1). Also hat – mit Ausnahme einer $\lambda\!\!\lambda$–Nullmenge von Punkten $t \in T$ – der meßbare Prozess $[1-1_M]\,\widetilde{X}$ dieselben endlichdimensionalen Randverteilungen wie X aus Schritt 1). Mit $[1-1_M]\,\widetilde{X}$ ist also zum Kovarianzkern $K(.,.)$ ein Gaußprozeß mit den in 2.17 geforderten Eigenschaften konstruiert. $\square$

2.20 Satz: Sei $(X_t)_{t\in T}$ ein Gaußprozeß mit stetigem und beschränktem Kovarianzkern $K(\cdot,\cdot)$, definiert auf irgendeinem $(\Omega', \mathcal{A}', P')$. Dann ist X (gegebenenfalls nach Abänderung der Pfade auf einer P'-Nullmenge $N' \in \mathcal{A}'$) ein Prozeß mit Pfaden in $L^2(T, \mathcal{T}, \mu)$, und es gilt

$$\int_T g(t)\, X_t\, \mu(dt) \ \sim\ \mathcal{N}(0, \tau^2) \quad , \quad \tau^2 \ := \ \int_T \int_T g(t_1)\, K(t_1, t_2)\, g(t_2)\ \mu(dt_1)\mu(dt_2)$$

für jedes $g \in L^2(T, \mathcal{T}, \mu)$.

Beweis: 1) Die Meßbarkeit des Prozesses X impliziert

$$N' \ := \ \left\{\omega \in \Omega' : \int_T X^2(t,\omega)\ \mu(dt) = +\infty\right\} \ \in\ \mathcal{A}' \ .$$

Betrachte X' definiert durch $X'(t,\omega) := 1_{\Omega'\setminus N'}(\omega)X(t,\omega)$. Wegen Beschränktheit von $K(.,.)$ muß wie in Beweisschritt 3)ii) von 2.19 N' eine P'-Nullmenge sein. Mit X ist auch X' ein meßbarer stochastischer Prozeß, sowie ein Gaußprozess, dessen Pfade sämtlich in $L^2(T, \mathcal{T}, \mu)$ liegen. Die P'-ununterscheidbaren Prozesse X und X' werden im folgenden stets identifiziert.
2) Sei nun $g \in L^2(T, \mathcal{T}, \mu)$. Wir betrachten wieder die Prozesse X^m aus dem Beweis von 2.19. Nach Voraussetzung über $K(.,.)$ ist für diese die Bedingung iib) aus Satz 2.6 erfüllt:

$$\sup_m E(|X_t^m|^2) \ \leq\ \sup_{t\in T} K(t,t) \ <\ \infty\ , \quad t \in T\ ,$$

$$\lim_{m\to\infty} E(|X_t^m|^2) \ =\ E(|X_t|^2)\ , \quad t \in T\ ,$$

dabei ergibt sich die zweite Aussage aus Beweisschritt 2) von 2.19. Zugleich konvergieren nach $(*)$ in Beweisschritt 3) von 2.19 die endlichdimensionalen Randverteilungen der X^m für $m \to \infty$

gegen die von X. Nach 2.6 sind damit alle Voraussetzungen des Satzes 2.4 erfüllt, und 2.5 liefert

$$\int_T g(t)\, X_t^m\, \mu(dt) \quad \longrightarrow \quad \int_T g(t)\, X_t\, \mu(dt) \quad \text{(schwache Konvergenz in } I\!R\text{, für } m \to \infty) .$$

3) Mit den Notationen aus Beweisschritt 3) von 2.19 ist leicht zu sehen, daß für jedes feste m

$$\int_T g(t)\, X_t^m\, \mu(dt) \;=\; \sum_{l \in \Lambda(m,T)} X_{t_l(m)} \int g(t)\, 1_{A_l(m)}(t)\, \mu(dt)$$

normalverteilt ist, wobei die Varianz dieser Normalverteilung

$$\tau_m^2 \;:=\; \sum_{l,l' \in \Lambda(m,T)} \int g(t)\, 1_{A_l(m)}(t)\, \mu(dt)\;\; K\left(t_l(m), t_{l'}(m)\right) \int g(s)\, 1_{A_{l'}(m)}(s)\, \mu(ds)$$

wegen Stetigkeit und Beschränktheit von $(t_1, t_2) \to K(t_1, t_2)$ für $m \to \infty$ gegen die Varianz

$$\tau^2 \;=\; \int_T \int_T g(t_1)\, K(t_1, t_2)\, g(t_2)\; \mu(dt_1)\mu(dt_2)$$

der in der Behauptung des Satzes angegebenen Normalverteilung konvergiert. □

2.21 Bemerkung: Für kompaktes $T \subset I\!R^k$ konstruiert man Gaußprozesse – unter allen in 2.19 gemachten Voraussetzungen an $K(.,.)$ – explizit als Summen

$$(*) \qquad \sum_{j=1}^{\infty} \lambda_j^{1/2}\, \xi_j(\omega)\, \psi_j(t)\,, \quad t \in T\,,\ \omega \in \Omega$$

mit iid normalverteilten (u.U. komplexwertigen) ZV $\xi_1, \xi_2, \ldots$, Eigenwerten $\lambda_1, \lambda_2, \ldots$, und orthonormalen Eigenfunktionen $\psi_1, \psi_2, \ldots$ des Operators

$$I\!K \;:\quad f \;\longrightarrow\; \int_T K(s,t)\, f(t)\, dt \quad , \quad f \in L^2(T, \mathcal{T}, \lambda\!\!\lambda)\,,$$

wobei die Summen $(*)$ in $L^2(T \times \Omega, \mathcal{T} \otimes \mathcal{A}, \lambda\!\!\lambda \otimes P)$ konvergieren (auch hier ist die in den endlichdimensionalen Randverteilungen in 2.17 erscheinende Ausnahme-Nullmenge $N \in \mathcal{T}$ nicht zu vermeiden). Man findet diese Konstruktion in Loève (1978, Vol. 2, S. 144), Gikhman und Skorohod (1974, Ch. IV.3, Thm. 5), siehe auch Zabreyko (1975).

Nun können wir eine Voraussetzung formulieren, die für $H = L^2(T, \mathcal{T}, \mu)$ die asymptotische Normalität der MD-Schätzfolgen für den unbekannten Parameter in Satz 2.14 sicherstellt.

2.22 Voraussetzung AN(ϑ): Sei $(X^n)_n$ wie in 2.16 (beachte $\widehat{\Psi}_n = X^n_\bullet$, $n \geq 1$), sei $(\varphi_n)_n$ wie in $\mathbf{T}(\vartheta)$ und wie in Satz 2.14. Wir setzen voraus, daß es einen Kern

$$K(\cdot,\cdot) : \; T \times T \to I\!R \quad \text{symmetrisch, nichtnegativ definit, stetig, beschränkt,}$$

eine $\lambda\!\!\lambda$-Nullmenge $N \in \mathcal{T}$ und eine Funktion $f \in L^1(T, \mathcal{T}, \mu)$ gibt so daß die durch Reskalierung entstehende Folge meßbarer Prozesse

$$W^n := \varphi_n \left(X^n - \Psi_\vartheta \right) \ , \quad n \geq 1 \ , \quad \text{unter } P_\vartheta$$

(beachte $W^n = W^n(\vartheta)$, $\varphi_n = \varphi_n(\vartheta)$, $n \geq 1$) die folgenden Eigenschaften i)–iii) besitzt:

i) $$\sup_{n \geq 1} E_\vartheta \left((W_t^n)^2 \right) \leq f(t) \ , \quad t \in T \setminus N \ ;$$

ii) $$\lim_{n \to \infty} E_\vartheta \left((W_t^n)^2 \right) = K(t,t) \ , \quad t \in T \setminus N \ ;$$

iii) für beliebige $l \geq 1$, $\alpha_1, \ldots, \alpha_l \in \mathbb{R}$, $t_1, \ldots, t_l \in T \setminus N$ gilt

$$\mathcal{L} \left(\sum_{i=1}^{l} \alpha_i \, W_{t_i}^n \mid P_\vartheta \right) \longrightarrow \mathcal{N} \left(0 \, , \, \sum_{i,j=1}^{l} \alpha_i \, K(t_i, t_j) \, \alpha_j \right)$$

(schwache Konvergenz in $\mathbb{R}$, für $n \to \infty$).

Der folgende Satz ist das zentrale Ergebnis über MD–Schätzfolgen.

2.23 Hauptsatz: An jeder Stelle $\vartheta \in \Theta$ sei der Satz von Voraussetzungen

$$\mathbf{SLLN}(\vartheta),\ \mathbf{I}(\vartheta),\ \mathbf{D}(\vartheta),\ \mathbf{T}(\vartheta),\ \mathbf{AN}(\vartheta),\ \text{mit } \varphi_n = \varphi_n(\vartheta) \uparrow \infty$$

erfüllt. Dann ist jede Festlegung $(\vartheta_n^*)_n$ einer MD-Schätzfolge für den unbekannten Parameter $\vartheta \in \Theta$ (stark konsistent und) asymptotisch normal: an jeder Stelle $\vartheta \in \Theta$ gilt

$$\mathcal{L} \left(\varphi_n (\vartheta_n^* - \vartheta) \mid P_\vartheta \right) \longrightarrow \mathcal{N} \left(0_d \ , \ \Lambda_\vartheta^{-1} \Xi_\vartheta \Lambda_\vartheta^{-1} \right)$$

(schwache Konvergenz in $\mathbb{R}^d$, für $n \to \infty$), mit $d \times d$-Matrizen mit Einträgen

$$\begin{aligned} (\Lambda_\vartheta)_{i,j} &= \langle D_i \Psi_\vartheta, D_j \Psi_\vartheta \rangle \ , \quad 1 \leq i, j \leq d \ , \\ (\Xi_\vartheta)_{i,j} &= \int_T \int_T D_i \Psi_\vartheta(t_1) \, K(t_1, t_2) \, D_j \Psi_\vartheta(t_2) \, \mu(dt_1) \mu(dt_2) \ . \end{aligned}$$

Beweis: Sei $\vartheta \in \Theta$ fest. Sei $K(\cdot, \cdot)$ wie in $\mathbf{AN}(\vartheta)$. Sei $W = W(\vartheta)$ ein Gaußprozeß mit Kovarianzkern $K(\cdot, \cdot)$ und Pfaden in $L^2(T, \mathcal{T}, \mu)$, definiert auf irgendeinem $(\Omega', \mathcal{A}', P')$.
Schreibe kurz $W^0 := W$, $P_0 := P'$, und $P_n := P_\vartheta$ für $n \geq 1$.

1) Analog zu Beweisschritt 1) von 2.20 stellt Voraussetzung $\mathbf{AN}(\vartheta)$ i) sicher, daß alle Prozesse W^n, $n \geq 0$, Pfade in $L^2(T, \mathcal{T}, \mu)$ besitzen (gegebenenfalls jeweils nach Abänderung der Pfade auf einer P_ϑ-Nullmenge). $\mathbf{AN}(\vartheta)$ iii) liefert dann nach Cramér-Wold die Konvergenz der endlichdimensionalen Randverteilungen von W^n gegen die von W, μ-fast überall:

$$(*) \qquad \mathcal{L}\left((W^n_{t_1}, \ldots, W^n_{t_l}) \mid P_\vartheta\right) \longrightarrow \mathcal{L}\left((W_{t_1}, \ldots, W_{t_l}) \mid P'\right)$$

(schwache Konvergenz in $\mathbb{R}^l$, $n \to \infty$), für beliebige $l \geq 1$ und $t_1, \ldots, t_l$ in $T \setminus N$.
Zusammen mit $(*)$ implizieren die Voraussetzungen $\mathbf{AN}(\vartheta)$ i)+ii) nach Satz 2.6:
für jedes $\varepsilon > 0$ gibt es ein $K = K(\varepsilon) < \infty$ so daß

$$(**) \qquad \sup_{n \geq 0} \int_{T \times \Omega} 1_{\{|W^n| > K\}} \, |W^n|^2 \, d(\mu \otimes P_n) \; < \; \varepsilon \, .$$

Mit Hauptsatz 2.4 erhält man aus $(*)$ und $(**)$ nun

$$W^n_\bullet \longrightarrow W_\bullet \quad \text{(schwache Konvergenz in } L^2(T, \mathcal{T}, \mu)\text{, für } n \to \infty)$$

woraus für jedes $g \in L^2(T, \mathcal{T}, \mu)$ nach 2.5 folgt (mit $\langle ., . \rangle$ Skalarprodukt in $L^2(T, \mathcal{T}, \mu)$)

$$(\diamond) \qquad \langle g, W^n_\bullet \rangle \longrightarrow \langle g, W_\bullet \rangle \quad \text{(schwache Konvergenz in } \mathbb{R}\text{, für } n \to \infty) \, .$$

Für jedes $g \in L^2(T, \mathcal{T}, \mu)$ gilt nach 2.20

$$(\diamond\diamond) \qquad \mathcal{L}\left(\langle g, W_\bullet \rangle \mid P'\right) \;=\; \mathcal{N}\left(0\,, \int_T \int_T g(s)\, K(s,t)\, g(t)\, \mu(ds)\mu(dt)\right)$$

2) Betrachte nun die Komponenten $D_1\Psi_\vartheta, \ldots, D_d\Psi_\vartheta$ der Ableitung $D\Psi_\vartheta$ in Voraussetzung $\mathbf{D}(\vartheta)$. Setzt man $g := h^\top D\Psi_\vartheta$ mit beliebigem $h \in \mathbb{R}^d$, so folgt mit Cramér-Wold aus $(\diamond)$ und $(\diamond\diamond)$

$$\mathcal{L}\left(\begin{pmatrix} \langle D_1\Psi_\vartheta, W^n_\bullet \rangle \\ \ldots \\ \langle D_d\Psi_\vartheta, W^n_\bullet \rangle \end{pmatrix} \mid P_\vartheta\right) \longrightarrow \mathcal{L}\left(\begin{pmatrix} \langle D_1\Psi_\vartheta, W_\bullet \rangle \\ \ldots \\ \langle D_d\Psi_\vartheta, W_\bullet \rangle \end{pmatrix} \mid P'\right) \;=:\; \mathcal{N}\,(\,0\,, \,\Xi_\vartheta\,)$$

(schwache Konvergenz in $\mathbb{R}^d$, für $n \to \infty$); die Kovarianzmatrix der Grenzverteilung hat Einträge

$$\begin{aligned} (\Xi_\vartheta)_{i,j} \;&=\; E\,(\,\langle D_i\Psi_\vartheta, W_\bullet \rangle\,,\,\langle D_j\Psi_\vartheta, W_\bullet \rangle\,) \\ &=\; E\left(\int_T \int_T D_i\Psi_\vartheta(s)\, W_s\, W_t\, D_j\Psi_\vartheta(t)\, \mu(ds)\mu(dt)\right) \\ &=\; \int_T \int_T D_i\Psi_\vartheta(s)\, K(s,t)\, D_j\Psi_\vartheta(t)\, \mu(ds)\mu(dt)\,, \quad i,j = 1, \ldots, d\,. \end{aligned}$$

3) Wegen Hauptsatz 2.14 aus Teilkapitel B besitzen

$$\mathcal{L}\,(\,\varphi(\vartheta^*_n - \vartheta) \mid P_\vartheta) \quad , \quad \mathcal{L}\,(\,\Pi_\vartheta\,(W^n_\bullet) \mid P_\vartheta) \quad , \quad n \to \infty$$

dieselbe Limesverteilung. Die lineare Abbildung $\Pi_\vartheta : L^2(T, \mathcal{T}, \mu) \to I\!R^d$ ist gegeben durch

$$f \quad \longrightarrow \quad \Lambda_\vartheta^{-1} \begin{pmatrix} \langle D_1 \Psi_\vartheta, f \rangle \\ \ldots \\ \langle D_d \Psi_\vartheta, f \rangle \end{pmatrix} .$$

Wegen Schritt 2) ist damit die Konvergenz der reskalierten MD-Schätzfehler unter P_ϑ gegen

$$\mathcal{N}\left(0 \, , \, \Lambda_\vartheta^{-1} \, \Xi_\vartheta \, \Lambda_\vartheta^{-1} \right)$$

nachgewiesen. Dies schließt den Beweis des Satzes 2.23 ab. □

D. Beispiel: Lokations- und Skalenmodelle für iid Beobachtungen

In Fortführung des zu Beginn von Teilkapitel B andiskutierten Beispiels (siehe 2.7) schätzen wir in Lokations- und Skalenmodellen für reellwertige iid Beobachtungen den unbekannten Parameter durch Anpassen der empirischen Verteilungsfunktion an die Schar der im Modell möglichen Verteilungsfunktionen $\{F_\xi : \xi \in \Theta\}$, und geben die Grenzverteilungen für Minimum-Distanz-Schätzfolgen explizit an.

2.24 Satz: Seien $Y_1, Y_2, \ldots$ iid ZV, $I\!R$-wertig, mit Verteilungsfunktion $F : I\!R \to [0,1]$. Sei $\widehat{F}_n$ die empirische Verteilungsfunktion zu den ersten n Beobachtungen, und μ ein endliches Maß auf $I\!R$. Dann gilt

$$\sqrt{n} \left(\widehat{F}_n - F \right) \; \longrightarrow \; B^{0,F}_{\bullet} \quad \text{(schwache Konvergenz in } L^2(I\!R, \mathcal{B}(I\!R), \mu) \text{, für } n \to \infty)$$

wobei $B^{0,F}$ wie in 2.18 c) die durch $t \to F(t)$ zeittransformierte Brownsche Brücke bezeichnet.

Beweis: Seien $Y_1, Y_2, \ldots$ definiert auf $(\Omega, \mathcal{A}, P)$, und $B^{0,F}$ auf $(\Omega', \mathcal{A}', P')$. Zuerst sind

$$(t, \omega) \; \longrightarrow \; X^n(t, \omega) \; := \; \sqrt{n} \left(\widehat{F}_n(t, \omega) - F(t) \right) \quad , \quad (t, \omega) \; \longrightarrow \; B^{0,F}(t, \omega) = B^0(F(t), \omega)$$

meßbare stochastische Prozesse, nach 2.1' und 2.18 c). Wir zeigen, daß $(X^n)_n$ und $B^{0,F}$ die Bedingungen des Satzes 2.4 erfüllen. Zum Nachweis der Konvergenz der endlichdimensionalen Randverteilungen fixiere $t_1, \ldots, t_l$ in $I\!R$. Für den Gaußprozeß $B^{0,F}$ gilt

$$\mathcal{L}\left((B^{0,F}_{t_1}, \ldots, B^{0,F}_{t_l}) \mid P' \right) \; = \; \mathcal{N}(0, \Sigma) \; , \quad \Sigma_{i,j} \; := \; F(t_i) \wedge F(t_j) - F(t_i) F(t_j) \; .$$

Für $X^n = \sqrt{n}\left(\widehat{F}_n - F\right)$ ist zu zeigen

$$\mathcal{L}\left((X^n_{t_1}, \ldots, X^n_{t_l}) \mid P\right) \longrightarrow \mathcal{N}(0,\Sigma) \quad \text{(schwache Konvergenz in } I\!R^l\text{, für } n\to\infty) \,,$$

bzw. nach Cramér-Wold für beliebiges $\alpha \in I\!R^l$:

$$\mathcal{L}\left(\sum_{i=1}^{l} \alpha_i \, X^n_{t_i} \mid P\right) \longrightarrow \mathcal{N}(\,0\,,\ \alpha^\top \Sigma\, \alpha\,) \quad \text{(schwache Konvergenz in } I\!R\text{, für } n\to\infty) \,.$$

Schreibe nun

$$\sum_{i=1}^{l} \alpha_i X^n_{t_i} \;=\; \frac{1}{\sqrt{n}} \sum_{j=1}^{n} R_j \quad , \quad R_j \;:=\; \left(\sum_{i=1}^{l} \alpha_i 1_{(-\infty,t_i]}\right)(Y_j) \;-\; \sum_{i=1}^{l} \alpha_i F(t_i)$$

mit iid ZV $R_1, R_2, \ldots$, zentriert und mit endlicher Varianz. Für diese gilt der zentrale Grenzwertsatz, also reduziert sich der Nachweis der Konvergenz der endlichdimensionalen Randverteilungen darauf, die Varianz der Einzelbeobachtung zu betrachten:

$$\begin{aligned} Var(R_1) \;&=\; \sum_{i=1}^{l}\sum_{i'=1}^{l} \alpha_i\,\alpha_{i'}\, E\left(\left[1_{(-\infty,t_i]}(Y_1) - F(t_i)\right]\left[1_{(-\infty,t_{i'}]}(Y_1) - F(t_{i'})\right]\right) \\ &=\; \sum_{i=1}^{l}\sum_{i'=1}^{l} \alpha_i\,\alpha_{i'}\; \left(F(t_i \wedge t_{i'}) - F(t_i)F(t_{i'})\right) \\ &=\; \sum_{i=1}^{l}\sum_{i'=1}^{l} \alpha_i\,\alpha_{i'}\; \left(F(t_i) \wedge F(t_{i'}) - F(t_i)F(t_{i'})\right) \quad = \quad \alpha^\top \Sigma\, \alpha \,. \end{aligned}$$

Damit ist Bedingung i) in Satz 2.4 erfüllt. Als einfachen Spezialfall erhält man

$$(+) \qquad E\left((X^n_t)^2\right) \;=\; F(t)(1-F(t)) \;=:\; f(t) \quad \text{für alle } t\in I\!R\text{, alle } n\geq 0\text{, mit } X^0 := B^{0,F} \;\;.$$

Klar gilt $f \in L^1(I\!R, \mathcal{B}(I\!R), \mu)$, und wegen Hilfssatz 2.6 ist mit (+) die Bedingung ii) aus Satz 2.4 verifiziert. Mit 2.4 folgt schwache Konvergenz der Pfade $X^n_\bullet \to B^{0,F}_\bullet$ in $L^2(I\!R, \mathcal{B}(I\!R), \mu)$. □

2.25 Beispiel: Sei F ein Wahrscheinlichkeitsmaß auf $(I\!R, \mathcal{B}(I\!R))$ mit stetiger und beschränkter Dichte f bezüglich des Lebesgue-Maßes $\lambda\!\!\lambda$ auf $(I\!R, \mathcal{B}(I\!R))$. Sei durch $\{f > 0\} = (a,b)$ ein Intervall $I := (a,b)$ als Träger von F gegeben, $-\infty \leq a < b \leq +\infty$. Betrachte das Lokationsmodell

$$(I\!R, \mathcal{B}(I\!R), \{F_\vartheta : \vartheta\in\Theta\}) \;, \quad \Theta = I\!R \;, \quad F_\vartheta := F(\cdot - \vartheta) \;,\; f_\vartheta := f(\cdot - \vartheta) \;.$$

Wähle ein endliches Maß $\mu \sim \lambda\!\!\lambda$. Sei $\widehat{F}_n$ die empirische Verteilungsfunktion zu den ersten n Beobachtungen. Dann ist jede MD-Schätzfolge im Sinne von 2.9

$$\vartheta^*_n \;=\; \operatorname*{arginf}_{\xi\in\Theta} \left\|\widehat{F}_n - F_\xi\right\|_{L^2(I\!R,\mathcal{B}(I\!R),\mu)} \;, \quad n\geq 1$$

an jeder Stelle $\vartheta \in \Theta$ stark konsistent und asymptotisch normal. Genauer: an jeder Stelle $\vartheta \in \Theta$ hat man eine Entwicklung

$$(+) \qquad \sqrt{n}(\vartheta_n^* - \vartheta) \;=\; \frac{\left\langle -f_\vartheta \,,\, \sqrt{n}\left(\widehat{F}_n - F_\vartheta\right)\right\rangle_{L^2(T,\mathcal{T},\mu)}}{\|f_\vartheta\|^2_{L^2(T,\mathcal{T},\mu)}} \;+\; o_{P_\vartheta}(1) \quad \text{für } n \to \infty$$

der reskalierten Schätzfehler, und es gilt

$$\mathcal{L}\left(\sqrt{n}(\vartheta_n^* - \vartheta) \mid P_\vartheta\right) \;\longrightarrow\; \mathcal{N}\left(0\,,\, \Sigma(\vartheta,\mu)\right)$$

wobei

$$\Sigma(\vartheta,\mu) \;:=\; \frac{\int_{I\!R}\int_{I\!R} f_\vartheta(s)\left(F_\vartheta(s\wedge t) - F_\vartheta(s)F_\vartheta(t)\right) f_\vartheta(t)\; \mu(ds)\mu(dt)}{\left(\, \int_{I\!R} f_\vartheta^2(t)\; \mu(dt)\,\right)^2}\;.$$

Zum Beweis der gemachten Behauptungen fixieren wir $\vartheta \in \Theta$. Wir betrachten $\widehat{\Psi}_n(\omega) = \widehat{F}_n(\cdot,\omega)$, $\Psi_\xi = F_\xi$ in $L^2(I\!R, \mathcal{B}(I\!R), \mu)$, und $\varphi_n = \sqrt{n}$. Aufgrund der Monotonieeigenschaften des (einparametrigen) Lokationsmodells (für $\xi \in \Theta$ und $h > 0$ gilt

$$F_{\xi+h} \;\leq\; F_\xi \;\text{ auf }\; I\!R\,, \quad F_{\xi+h} \;<\; F_\xi \;\text{ auf }\; (a+\xi, b+\xi)\,,$$

wobei der Träger $(a+\xi, b+\xi)$ von F_ξ für jeden Wert $\xi \in I\!R$ strikt positives μ-Maß besitzt) ist die Identifizierungsbedingung $\mathbf{I}(\vartheta)$ erfüllt. Das starke Gesetz der großen Zahlen $\mathbf{SLLN}(\vartheta)$ folgt aus dem Satz von Glivenko-Cantelli. Die Differenzierungsbedingung $\mathbf{D}(\vartheta)$ gilt mit

$$D\Psi_\vartheta \;=\; -f_\vartheta\,, \quad \text{wobei} \quad 0 \;<\; \int_{I\!R} f_\vartheta^2(t)\,\mu(dt) \;<\; \infty\,,$$

nach Voraussetzung über $f(\cdot)$. Die Entwicklung (+) der reskalierten Schätzfehler folgt nun sofort aus 2.14. Nach 2.24 hat man schwache Konvergenz in $L^2(I\!R, \mathcal{B}(I\!R), \mu)$ der Pfade

$$(++) \qquad \varphi_n\left(\widehat{\Psi}_n - \Psi_\vartheta\right) \;=\; \sqrt{n}\left(\widehat{F}_n(\cdot) - F_\vartheta\right) \quad\longrightarrow\quad W_\bullet\,, \quad n \to \infty$$

mit $W := B^{0,F_\vartheta}$, unter P_ϑ; der Kovarianzkern des Limesprozesses in (++) ist

$$K_\vartheta(s,t) \;:=\; E\left(B_s^{0,F_\vartheta}\; B_t^{0,F_\vartheta}\right) \;=\; F_\vartheta(s\wedge t) - F_\vartheta(s)F_\vartheta(t)\,, \quad s,t \in I\!R\,.$$

Aus (+) und (++) ergibt sich die Grenzverteilung für reskalierte Schätzfehler.

Der Beweis von 2.24 zeigt, daß in der Konvergenz (++) alle in $\mathbf{AN}(\vartheta)$ gemachten Voraussetzungen erfüllt sind; auch ist durch (++) die Straffheitsbedingung $\mathbf{T}(\vartheta)$ miterfüllt, vgl. Folgerung 2.5. Damit sind alle Voraussetzungen des Hauptsatzes 2.23 erfüllt. $\square$

2.26 Beispiel: Zu F und f, die den in 2.25 gemachten Voraussetzungen genügen, betrachte das Lokations- und Skalenmodell

$$(\mathbb{R}, \mathcal{B}(\mathbb{R}), \{F_\vartheta : \vartheta \in \Theta\}) \ , \quad \vartheta = (\sigma, \lambda) \ , \quad F_\vartheta = F(\sigma(\cdot - \lambda)) \ , \ f_\vartheta = \sigma f(\sigma(\cdot - \lambda))$$

mit Parameterraum

$$\Theta \text{ offen in } \mathbb{R}^2, \text{ beschränkt, mit Abschluß } \overline{\Theta} \subset (0,\infty)\times\mathbb{R} .$$

Insbesondere hat F_ϑ dann den Träger $I_\vartheta := (\frac{a}{\sigma} + \lambda, \frac{b}{\sigma} + \lambda)$, und in Einschränkung auf eine beliebig kleine offene Kugel in I_ϑ kann man F_ϑ von allen F_ξ, $\xi \in \overline{\Theta}\backslash\{\vartheta\}$, unterscheiden. Sei T ein Kompaktum in $\mathbb{R}$ mit der Eigenschaft

$$\mathrm{int}(T) \ \cap \ I_\vartheta \ \neq \ \emptyset \quad \text{für jedes } \vartheta \in \Theta \ ,$$

sei μ ein endliches Maß auf $(T, \mathcal{T})$, äquivalent zu dem auf T eingeschränkten Lebesgue-Maß $\lambda\!\!\lambda$. Dann sind gemäß 2.9 gebildete MD-Schätzfolgen für den unbekannten Parameter $\vartheta = (\sigma, \lambda)$

$$\vartheta_n^* \ = \ \underset{\xi\in\Theta}{\mathrm{arginf}} \ \left\| \widehat{F}_n - F_\xi \right\|_{L^2(T,\mathcal{T},\mu)} \ , \quad n \geq 1$$

an jeder Stelle $\vartheta \in \Theta$ stark konsistent und asymptotisch normal. Genauer: an jeder Stelle $\vartheta \in \Theta$ hat man mit

$$DF_\vartheta \ = \ \begin{pmatrix} D_1 F_\vartheta \\ D_2 F_\vartheta \end{pmatrix} \ , \quad D_1 F_\vartheta = \frac{(\cdot - \lambda)}{\sigma} f_\vartheta \ , \ D_2 F_\vartheta = -f_\vartheta$$

eine Entwicklung der reskalierten Schätzfehler

$$(*) \qquad \sqrt{n}(\vartheta_n^* - \vartheta) \ = \ \Pi_\vartheta\left(\sqrt{n}(\widehat{F}_n - F_\vartheta)\right) \ + \ o_{P_\vartheta}(1) \ , \quad n \to \infty$$

mit der in 2.14 gegebenen linearen Abbildung $\Pi_\vartheta : L^2(T, \mathcal{T}, \mu) \to \mathbb{R}^2$; mit Kurzschreibweise

$$\widetilde{m}_\vartheta(dt_1, dt_2) \ := \ f_\vartheta(t_1)\,(F_\vartheta(t_1 \wedge t_2) - F_\vartheta(t_1)F_\vartheta(t_2))\, f_\vartheta(t_2)\, \mu(dt_1)\mu(dt_2)$$

und mit Matrizen

$$\Xi_\vartheta \ = \ \begin{pmatrix} \int_T \int_T \frac{t_1-\lambda}{\sigma} \frac{t_2-\lambda}{\sigma}\, \widetilde{m}_\vartheta(dt_1, dt_2) & -\int_T \int_T \frac{t_2-\lambda}{\sigma}\, \widetilde{m}_\vartheta(dt_1, dt_2) \\ -\int_T \int_T \frac{t_1-\lambda}{\sigma}\, \widetilde{m}_\vartheta(dt_1, dt_2) & \int_T \int_T \widetilde{m}_\vartheta(dt_1, dt_2) \end{pmatrix}$$

$$\Lambda_\vartheta \ = \ \begin{pmatrix} \int_T \left(\frac{t-\lambda}{\sigma}\right)^2 f_\vartheta^2(t)\,\mu(dt) & -\int_T \frac{t-\lambda}{\sigma}\, f_\vartheta^2(t)\,\mu(dt) \\ -\int_T \frac{t-\lambda}{\sigma}\, f_\vartheta^2(t)\,\mu(dt) & \int_T f_\vartheta^2(t)\,\mu(dt) \end{pmatrix}$$

erhält man als Grenzverteilung der reskalierten Schätzfehler

$$\mathcal{L}\left(\sqrt{n}(\vartheta_n^* - \vartheta) \mid P_\vartheta\right) \longrightarrow \mathcal{N}\left(0\,,\ \Lambda_\vartheta^{-1}\Xi_\vartheta\,\Lambda_\vartheta^{-1}\right)$$

(schwache Konvergenz in $I\!R^2$, für $n \to \infty$).

Zum Beweis dieser Aussagen fixieren wir $\vartheta \in \Theta$.

Wir betrachten unter P_ϑ die empirische Verteilungsfunktion $\widehat{\Psi}_n(\omega) = \widehat{F}_n(\cdot, \omega)$ zu den ersten n Beobachtungen, $\Psi_\xi = F_\xi$ in $L^2(T, \mathcal{T}, \mu)$, $\xi \in \Theta \subset I\!R^2$, und $\varphi_n = \sqrt{n}$.

Zunächst verifiziert man die Identifizierbarkeitsbedingung $\mathbf{I}(\vartheta)$: für $\xi \neq \vartheta$ gilt $F_\xi \neq F_\vartheta$ insbesondere auf einer kleinen Kugel von positivem μ-Maß in $\mathrm{int}(T) \cap I_\vartheta$. Also verschwindet die stetige Funktion $\xi \to \|F_\xi - F_\vartheta\|_{L^2(T,\mathcal{T},\mu)}$ nur an der Stelle $\xi = \vartheta$. Da $\overline{\Theta}$ kompakt ist, hat man

$$\inf_{\xi\in\Theta\,,\,|\xi-\vartheta|>\varepsilon} \|F_\xi - F_\vartheta\|_{L^2(T,\mathcal{T},\mu)} \;\geq\; \min_{\xi\in\overline{\Theta}\,,\,|\xi-\vartheta|\geq\varepsilon} \|F_\xi - F_\vartheta\|_{L^2(T,\mathcal{T},\mu)} \;>\; 0\,.$$

Wir verifizieren die Differenzierbarkeitsbedingung $\mathbf{D}(\vartheta)$, mit DF_ϑ wie oben angegeben. Wegen Kompaktheit von T und Beschränktheit von f gilt für hinreichend kleines $\varepsilon > 0$ sogar

$$(+) \qquad \sup_{\xi\in B_\varepsilon(\vartheta)} |D_i F_\xi| \;\in\; L^2(T, \mathcal{T}, \mu)\,, \quad i = 1, 2\,.$$

Die Komponenten D_1F_ϑ, D_2F_ϑ von DF_ϑ sind linear unabhängig in $L^2(T, \mathcal{T}, \mu)$, da das zu $\lambda\!\!\lambda$ äquivalente Maß μ die Funktionen $\frac{\cdot-\lambda}{\sigma} f_\vartheta$ und $-f_\vartheta$ unterscheiden kann (genauer: ... bereits auf einer beliebig kleinen Kugel in $\mathrm{int}(T) \cap I_\vartheta$ unterscheiden kann). Nun schreibt man

$$F_\xi(t) - F_\vartheta(t) \;=\; \int_0^{|\xi-\vartheta|} dr\; u^\top DF_{\vartheta+ru}(t) \;=\; |\xi - \vartheta| \int_0^1 dr\; u^\top DF_{\vartheta+r|\xi-\vartheta|u}(t)$$

für $\xi \in B_\varepsilon(\vartheta)\,,\ \xi \neq \vartheta\,$, und $u := \frac{\xi-\vartheta}{|\xi-\vartheta|}$, also

$$(\diamond) \qquad F_\xi(t) - F_\vartheta(t) - (\xi - \vartheta)^\top DF_\vartheta(t) \;=\; |\xi - \vartheta| \int_0^1 dr\; u^\top \left(DF_{\vartheta+r|\xi-\vartheta|u}(t) - DF_\vartheta(t)\right)\,.$$

Mit Jensen–Ungleichung hat man

$$\left\| \int_0^1 dr\; u^\top \left(DF_{\vartheta+r|\xi-\vartheta|u} - DF_\vartheta\right) \right\|^2_{L^2(T,\mathcal{T},\mu)} \;\leq\; \int_0^1 dr\; \left\| u^\top \left(DF_{\vartheta+r|\xi-\vartheta|u} - DF_\vartheta\right) \right\|^2_{L^2(T,\mathcal{T},\mu)}$$

und schätzt dies weiter ab durch

$$(++) \qquad 2 \sup_{\xi\in B_\varepsilon(\vartheta)} \left(\sum_{i=1}^2 \|D_iF_\xi - D_iF_\vartheta\|^2_{L^2(T,\mathcal{T},\mu)} \right)\,.$$

Wegen (+) oben und wegen Stetigkeit von

$$\Theta \ni \xi \longrightarrow D_i F_\xi \in L^2(T, \mathcal{T}, \mu) , \quad i = 1, 2$$

verschwinden mit dominierter Konvergenz die Ausdrücke (++) für $\varepsilon \to 0$. Mit ($\diamond$) und dem Verschwinden von (++) für $\varepsilon \to 0$ ist die Differenzierbarkeitsbedingung $\mathbf{D}(\vartheta)$ nachgewiesen. Für die Bedingungen $\mathbf{SLLN}(\vartheta)$, $\mathbf{T}(\vartheta)$ und $\mathbf{AN}(\vartheta)$ bleiben die Argumente unverändert dieselben wie im Beweis von 2.25. Insbesondere hat man nach 2.24 unter P_ϑ schwache Konvergenz in $L^2(I\!R, \mathcal{B}(I\!R), \mu)$ der Pfade

$$(**) \qquad \varphi_n\left(\widehat{\Psi}_n - \Psi_\vartheta\right) = \sqrt{n}\left(\widehat{F}_n(\cdot) - F_\vartheta\right) \longrightarrow W_\bullet , \quad n \to \infty$$

mit $W := \left(B_t^{0,F_\vartheta}\right)_{t \in T}$; der Kovarianzkern des Limesprozesses in $(**)$ ist

$$K_\vartheta(s,t) := E\left(B_s^{0,F_\vartheta}\, B_t^{0,F_\vartheta}\right) = F_\vartheta(s \wedge t) - F_\vartheta(s) F_\vartheta(t) , \quad s, t \in T .$$

Wieder folgt aus $(*)$ und $(**)$ die Aussage über die Grenzverteilung reskalierter Schätzfehler. Auch hier zeigt der Beweis von 2.24, daß wie im vorhergehenden Beispiel alle Voraussetzungen des Hauptsatzes 2.23 erfüllt sind.

Will man die empirische Verteilungsfunktion zu $I\!R^k$-wertigen iid Beobachtungen an eine Familie $\{F_\xi : \xi \in \Theta\}$ von Verteilungsfunktionen auf $I\!R^k$ anpassen, braucht man anstelle von 2.24 den folgenden Satz.

2.27 Satz: Sei $k \geq 1$, sei $(T, \mathcal{T})$ ein k-dimensionales Intervall versehen mit seiner Borel-σ-Algebra, sei $\mu \ll \lambda\!\!\lambda$ ein endliches Maß auf $(T, \mathcal{T})$. Für $I\!R^k$–wertige iid ZV $Y_1, Y_2, \ldots$ mit stetiger Verteilungsfunktion bezeichne $\widehat{F}_n$ die empirische Verteilungsfunktion zu den ersten n Beobachtungen. Dann gibt es einen Gaußprozeß $W = (W_t)_{t \in T}$ mit Kovarianzfunktion

$$K(\cdot, \cdot) : \quad T \times T \ni (s,t) \longrightarrow F(s \wedge t) - F(s) F(t) \quad \in [0,1]$$

(mit komponentenweise definiertem Minimum in $s \wedge t$ da $s, t \in I\!R^k$), und es gilt

$$\sqrt{n}\left(\widehat{F}_n - F\right) \longrightarrow W_\bullet \quad \text{(schwache Konvergenz in } L^2(T, \mathcal{T}, \mu)\text{, für } n \to \infty) .$$

Beweis: Wegen der vorausgesetzten Stetigkeit von F gibt es nach Satz 2.19 einen Gaußprozeß $W = (W_t)_{t \in T}$ mit der angegebenen Kovarianzfunktion $K(\cdot, \cdot)$; beachte, daß hier gemäß 2.17

und 2.19 die endlichdimensionalen Randverteilungen von W nur bis auf eine Ausnahmemenge $N \in \mathcal{T}$ mit $\lambda\!\!\lambda(N) = 0$ bestimmt sind. Man wiederholt nun den Beweis von 2.24 mit W anstelle von $B^{0,F}$, um Konvergenz der endlichdimensionalen Randverteilungen $\lambda\!\!\lambda$-fast sicher – damit auch μ-fast sicher, nach Voraussetzung über μ – sowie

$$E\left(\left[\sqrt{n}\left(\widehat{F}_n - F\right)\right]_t^2\right) \;=\; K(t,t) \;=\; F(t)(1-F(t)) \quad \text{für alle } t \text{ in } T$$

zu erhalten, und wendet 2.6 und 2.4 an. □

2.28 Ausblicke: In den beiden iid Beispielen 2.25+2.26 hat man gesehen, daß unter minimalen Voraussetzungen MD-Schätzfolgen die Eigenschaften der starken Konsistenz und der asymptotischen Normalität besitzen. Es gibt viele Variationsmöglichkeiten für das im Teilkapitel B geschilderte Grundschema.

a) Eine Reihe 'klassischer' Anwendungsbeispiele gibt Millar (1984). In statistischen Modellen für $\mathbb{R}^k$-wertige iid Beobachtungen $Y_1, Y_2, \ldots$, in denen die charakteristische Funktion der Einzelbeobachtung eine einfache explizite Gestalt hat, kann man z.B. versuchen, empirische charakterische Funktionen $\widehat{\Psi}_n$

$$\widehat{\Psi}_n(t) \;:=\; \frac{1}{n}\sum_{j=1}^{n}(\cos(t^\top Y_j) + i\sin(t^\top Y_j)) \;, \quad t \in \mathbb{R}^k$$

in einem L^2-Raum komplexwertiger Funktionen auf $\mathbb{R}^k$ an die Schar der im Modell möglichen charakteristischen Funktionen $\{\Psi_\xi : \xi \in \Theta\}$

$$\Psi_\xi(t) := E_\xi\left(e^{i\,t^\top Y_1}\right) \;, \quad t \in \mathbb{R}^k$$

anzupassen. Ein Beispiel, in dem Stabilitäts- und Skalenparameter von symmetrisch stabil verteilten Zufallsvariablen zu schätzen sind, findet man – zusammen mit einer Diskussion älterer Referenzen – in Höpfner und Rüschendorf (1999). Analog kann man auch Laplace-Transformierte oder Quantilfunktionen anpassen.

b) Die Einzelbeobachtungen in iid Modellen müssen auch nicht $\mathbb{R}^k$-wertig sein. Ohne jede Änderung kann man Zufallsvariable betrachten, die Werte in polnischen Räumen (z.B. kanonischen Pfadräumen für stochastische Prozesse eines festgelegten Typs) annehmen: so können Einzelbeobachtungen – wie in Kutoyants (1998) – unabhängige Realisierungen desselben Punktprozesses mit geeignet parametrisierter deterministischer Intensität sein.

c) Über die Welt der iid Versuchswiederholung hinaus kann man mit dem hier geschilderten Vorgehen ohne jede Änderung $\mathbb{R}^k$-wertige *stationäre und ergodische* Markovprozesse behandeln:

nach Beobachtung einer Trajektorie des Markovprozesses $X = (X_t)_{t \geq 0}$ unter unbekanntem $\vartheta \in \Theta$ über das Zeitintervall $[0, n]$ kann man versuchen, das Okkupationszeitmaß $\widehat{\Psi}_n$

$$\widehat{\Psi}_n(v) \;=\; \frac{1}{n} \int_0^n 1_{(-\infty, v]}(X_s)\, ds \;, \quad v \in T$$

auf einem geeigneten Kompaktum T in $I\!R^k$ an die Familie der im Modell möglichen invarianten Maße $\{\Psi_\xi : \xi \in \Theta\}$

$$\Psi_\xi(v) \;=\; P_\xi \left\{ X_1 \in (-\infty, v] \right\} \;, \quad v \in T$$

anzupassen. Für stationäre und ergodische Diffusionsprozesse findet man viele interessante auf diesem Ansatz beruhende MD-Schätzfolgen in dem Buch Kutoyants (2003).

d) Verallgemeinerungen des Vorgehens in Teilkapitel B sind möglich, welche erlauben, Stationaritätsvoraussetzungen in ergodischen Markovprozessen fallenzulassen. In diesem Fall vergleicht man – manchmal auch zu geeigneten zufälligen Zeiten τ_n anstelle der deterministischen Zeiten n, für $n \to \infty$ – den Zustand eines aus der Beobachtung extrahierten wachsenden Prozesses mit dem Zustand des Kompensators (im Sinne der Martingaltheorie) dieses wachsenden Prozesses unter den möglichen Werten des Parameters $\xi \in \Theta$. Für Markov-Sprungprozesse findet man ein Beispiel für dieses Vorgehen in Höpfner und Kutoyants (1997).