

Empirical-Likelihood-Schätzung der Fehlerverteilung im nichtparametrischen
Regressionsmodell

Dissertation
zur Erlangung des Doktorgrades
der Fakultät für Mathematik, Informatik
und Naturwissenschaften
der Universität Hamburg

vorgelegt
im Fachbereich Mathematik

von
Sebastian Kiwitt

Hamburg

2010

Als Dissertation angenommen vom Fachbereich
Mathematik der Universität Hamburg
Auf Grund der Gutachten von Prof. Dr. Natalie Neumeyer
Prof. Dr. Holger Dette

Hamburg, den 04.Oktober 2010

Prof. Dr. Vincente Cortés
Leiter des Fachbereichs Mathematik

Inhaltsverzeichnis

1	Einleitung	5
2	Grundlegendes	9
2.1	Allgemeines	9
2.2	Kern(dichte-)schätzer	26
2.3	Empirical Likelihood	31
2.4	Bezug zu anderen Arbeiten	36
3	Modell und Annahmen	39
3.1	Modell und grundsätzliche Annahmen	39
3.2	Annahmen an die dem Schätzer zugrundeliegende Stichprobe	40
3.3	Annahmen für den Einbezug der die Fehlerverteilung betreffenden Zusatzinformation	41
4	Schätzansatz	45
4.1	Empirical-Likelihood-Schätzer	45
4.2	Empirical-Likelihood-Schätzer unter Nebenbedingungen	47
4.3	Entwicklung von Bestandteilen des Empirical-Likelihood Schätzers	51
5	Entwicklung des Empirical-Likelihood-Schätzers	69
5.1	Zerlegung in drei Summanden	69
5.2	Entwicklung der drei Summanden	69
6	Erwartungswert, Kovarianz und schwache Konvergenz	85
6.1	Erwartungswert	87
6.2	Angabe des Prozesses	92
6.3	Kovarianzfunktion	95
6.4	Beweis der schwachen Konvergenz der Prozesse	107
7	Beispiele	137
7.1	Beispiel 1	137
7.2	Beispiel 2	140
7.3	Zusammenfassung der Ergebnisse	144

8 Simulationen	145
8.1 Einleitung	145
8.2 Berechnung der Schätzer	146
8.3 Datenerzeugung	146
8.4 Simulationsbeispiel 1 (Normalverteilungsmodell)	146
8.5 Simulationsbeispiel 2 (schiefe Verteilung)	158
8.6 Simulationsbeispiel 3 (Cauchy-verteilte Fehler)	161
8.7 Simulationsbeispiel 4 (geschätzte Nebenbedingungen)	164
8.8 Simulationsbeispiel 5 (skew normal distribution)	166
8.9 Zusammenfassung der Simulationsergebnisse	168
8.10 Informationen zum Simulationsprogramm	169
9 Vorhersageintervalle und Testverfahren	171
9.1 Vorhersageintervalle	171
9.2 Testverfahren	174
9.2.1 Erstes Testverfahren	176
9.2.2 Zweites Testverfahren	177
9.2.3 Drittes Testverfahren	178
10 Zusammenfassung der Arbeit und Ausblick	181
10.1 Zusammenfassung	181
10.2 Ausblick	181
Symbolverzeichnis	183
Literaturverzeichnis	185

1 Einleitung

Einführung

In der angewandten Statistik spielt die Modellierung von Zusammenhängen zwischen Zufallsgrößen X und Y eine wichtige Rolle. Das Regressionsmodell hat allgemein die Form

$$(1.1) \quad Y = m(X) + \varepsilon,$$

dabei beschreibt die zufällige Störgröße ε Messfehler oder andere, nicht berücksichtigte Einflussfaktoren. Meist wird vorausgesetzt, dass ε und X unabhängig sind. Sind Zusatzinformationen zu der den Zusammenhang zwischen X und Y beschreibenden Funktion m gegeben kann ggf. ein parametrisches Modell $\{m_\vartheta, \vartheta \in \Theta\}$ angesetzt werden. Ein allgemeiner Zugang ist aber ein nichtparametrisches Modell, d. h. die den Zusammenhang beschreibende Funktion m werde mittels eines nichtparametrischen Kurvenschätzers geschätzt, was in dieser Arbeit Grundlage sein soll. Ein dafür in der Praxis häufig genutzter Schätzer ist der Nadaraya-Watson-Schätzer (vgl. Definition 2.2).

In dieser Arbeit soll aber nicht m , sondern die Verteilung der Fehler ε im Mittelpunkt stehen. Es sei \hat{m} ein Schätzer für m , \hat{F} ein Schätzer für F_ε . Die Schätzung der Verteilung F_ε der Fehler ist von Interesse, da damit zum Beispiel (im Falle der Unabhängigkeit von ε und X) Vorhersageintervalle für Y mit asymptotischem Niveau $(1 - \alpha)$ der Form

$$[\hat{m}(x) - \hat{F}_\varepsilon^{-1}(\alpha/2), \hat{m}(x) + \hat{F}_\varepsilon^{-1}(1 - \alpha/2)]$$

angegeben oder auch Tests auf Gültigkeit von Nebeninformationen konstruiert werden können. Auch ist F_ε wegen des engen Zusammenhangs zur Verteilung von Y gegeben $X = x$ von Interesse. Im Vordergrund dieser Arbeit soll aber der Beweis der schwachen Konvergenz des im Weiteren vorgestellten, geeignet standardisierten Verteilungsfunktions-Schätzers gegen einen Gaußprozess sein.

Das Problem der Schätzung der Verteilung der Fehler ist die Unbeobachtbarkeit derselben, so dass zur Schätzung der Verteilungsfunktion nicht ε selbst, sondern nur $\hat{\varepsilon} := Y - \hat{m}(X)$ genutzt werden kann. Hat man eine Stichprobe $(X_1, Y_1), \dots, (X_n, Y_n)$ unabhängig und identisch verteilt gegeben, welche dem obigen Modell (1.1) genügt, so wird die Grundlage der Schätzung der Verteilungsfunktion der Fehler ε die daraus (mittels des nichtparametrischen Kurvenschätzers \hat{m}) erstellte Stichprobe $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n$ sein. Dies ist aber keine Stichprobe unabhängiger Zufallsvariablen

mehr, vielmehr hängt jede Beobachtung $\hat{\varepsilon}_i$ von allen anderen ab, was die Analyse von darauf basierenden Schätzern deutlich erschwert.

In dem hier betrachteten Modell sei aber noch Zusatzinformation zur Verteilung der Fehler gegeben, zum Beispiel bekannter Erwartungswert, bekannte Quantile oder Ähnliches. Ziel ist es einen Schätzer zu entwickeln, welcher diese Zusatzinformation miteinbezieht und diesen mit einem herkömmlichen nichtparametrischen Schätzer, speziell einem Kernschätzer theoretisch und in Simulationen zu vergleichen.

Kiwitt, Nagel und Neumeyer (2009) betrachten einen ebensolchen Schätzer der Verteilungsfunktion der Fehler auf Basis von $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n$ unter Einbezug von Nebeninformationen, setzen aber voraus, dass im Modell $Y = m(X) + \varepsilon$ der Fehler ε unabhängig vom Regressor X ist. In dieser Arbeit werden wir diese recht strikte Voraussetzung aufheben, denn diese Voraussetzung ist im allgemeinen nicht gegeben. Betrachtet man zum Beispiel die Wuchshöhe Y einer Pflanze in Abhängigkeit der Sonneneinstrahlung X o. ä., so ist anzunehmen, dass in einer Stichprobe von Pflanzen mit geringer Sonneneinstrahlung die Variabilität in der Größe kleiner ist als in einer vergleichbaren Stichprobe mit normaler Sonneneinstrahlung und somit das beschreibende Modell nicht die Unabhängigkeit von X und ε voraussetzen sollte.

Diese Arbeit verallgemeinert das Modell in dieser Hinsicht, stellt einen nichtparametrischen Schätzer für die bedingte Verteilungsfunktion der Fehler gegeben X , d.h. $F_{\varepsilon|X}$, auf Basis von $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n$ unter Einbezug von Nebeninformationen vor und vergleicht diesen mit einem die Nebeninformation nicht nutzenden Kernschätzer. Das Hauptziel dieser Arbeit ist der Beweis der schwachen Konvergenz der beiden im weiteren vorgestellten Schätzer. Zum Beispiel können die oben schon erwähnten Vorhersageintervalle für Y mit asymptotischem Niveau $(1 - \alpha)$

$$[\hat{m}(x) - \hat{F}_{\varepsilon|X}^{-1}(\alpha/2|x), \hat{m}(x) + \hat{F}_{\varepsilon|X}^{-1}(1 - \alpha/2|x)]$$

mittels dieser Schätzer auch für dieses deutlich allgemeinere Modell angegeben und ggf. verbessert werden, vgl. für einen Ausblick Kapitel 9, S. 171.

Die Grundlage des Schätzers $\hat{F}_{\varepsilon|X}$, welcher Nebeninformationen einbezieht ist dabei die im Weiteren vorgestellte Empirical-Likelihood-Methode. Für die beiden in dieser Arbeit betrachteten Schätzer wird (geeignet standardisiert) im Verlauf der Arbeit jeweils die schwache Konvergenz gegen einen Gaußprozess bewiesen, was das Hauptresultat dieser Arbeit darstellt. Die Schwierigkeiten resultieren dabei hauptsächlich aus den Abhängigkeiten in der zugrundeliegenden Stichprobe.

Die Voraussetzungen, welche im Weiteren an das Modell gestellt werden (siehe Kapitel 3), sind, bedingt durch den nichtparametrischen Ansatz, hauptsächlich Glattheitsannahmen. Der Nutzen ist meist nicht direkt ersichtlich; häufig dienen diese Annahmen nur der Abschätzung von Resttermen in verschiedensten Teilen der Entwicklung des Schätzers.

Aufbau der Arbeit

Im folgenden Kapitel wird neben stochastischen Grundlagen und spezielleren Beweistechniken zur schwachen Konvergenz der im Weiteren genutzte nichtparametrische Kurvenschätzer \hat{m} im Modell $Y = m(X) + \varepsilon$ eingeführt, nützliche Eigenschaften zusammengestellt und auf das Problem der Bandbreitenwahl eingegangen. Darauf wird die Empirical-Likelihood-Methode in Grundzügen vorgestellt, so dass in folgenden Kapiteln neben einem Kernschätzer für $F_{\varepsilon|X}$ der in dieser Arbeit im Vordergrund stehende Empirical-Likelihood-Schätzer für $F_{\varepsilon|X}$ motiviert werden kann.

Das Kapitel wird abgeschlossen durch eine kurze Einordnung im Hinblick auf verwandte Arbeiten.

Das dritte Kapitel stellt das Modell und seine Annahmen vor. Die Schätzer der bedingten Verteilungsfunktion der Fehler gegeben den Regressor X , $F_{\varepsilon|X}$, speziell der Empirical-Likelihood-Schätzer (ELE) und ein Kernschätzer (KE) werden im darauf folgenden Kapitel eingeführt. Es werden Entwicklungen für Bauteile des ELE hergeleitet bevor dann im fünften Kapitel damit begonnen werden kann den Schätzer selbst so zu entwickeln, dass er als Summe unabhängiger Zufallsvariablen zuzüglich eines Resttermes von hinreichend schneller Rate darstellbar ist. Es wird eine Entwicklung angegeben, welche beide Schätzer in Verbindung bringt.

Mit diesen recht länglichen Vorarbeiten kann im sechsten Kapitel der Erwartungswert der Schätzer berechnet und können (geeignet standardisiert) Aussagen zur schwachen Konvergenz des Prozesses $t \mapsto \hat{F}_{\varepsilon|X}(t|x)$ getroffen werden. Die asymptotische Kovarianzstruktur wird angegeben.

Beispiele im Hinblick auf die Gültigkeit der Voraussetzungen und die Berechnung der asymptotischen Varianz und des Bias liefert das siebte Kapitel.

Das achte Kapitel beinhaltet hauptsächlich die Simulationsergebnisse des Vergleichs des Empirical-Likelihood-Schätzers (ELE) und des Kernschätzers (KE) der Verteilungsfunktion $F_{\varepsilon|X}$, insbesondere für die in Kapitel sieben betrachteten Beispiele. Approximiert werden der erwartete quadratische Fehler, $MSE(y, x) = E[(\hat{F}_{\varepsilon|X}(y|x) - E[\hat{F}_{\varepsilon|X}(y|x)])^2]$, und der erwartete integrierte quadratische Fehler, $MISE(x) = E[\int (\hat{F}_{\varepsilon|X}(y|x) - E[\hat{F}_{\varepsilon|X}(y|x)])^2 dy]$. Es zeigt sich auch hier, ebenso wie in den theoretischen Berechnungen keine gleichmäßige Verbesserung des MSE , für bestimmte Bereiche kann aber eine Verbesserung ausgemacht werden. Ebenso zeigt

der approximierte *MISE* Verbesserungen auf. Im neunten Kapitel wird ein kurzer Einblick in Konfidenzintervalle für m und Tests auf die Gültigkeit der Nebeninformation auf Basis der zuvor betrachteten Schätzer für $F_{\varepsilon|X}$ gegeben, dabei wird die Theorie nicht genauer ausgeführt. Abschließend fasst das zehnte Kapitel die Ergebnisse übersichtlich zusammen und gibt einen Ausblick.

2 Grundlegendes

2.1 Allgemeines

Im Weiteren sei X eine Zufallsvariable auf einem abstrakten Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ nach $(\Omega', \mathcal{A}', P)$, dabei bezeichnet P das Bildmaß, d.h. $P = \mathbb{P}^X = \mathbb{P} \circ X^{-1}$. Meist ist in dieser Arbeit X reellwertige Zufallsvariable mit Dichte f , so dass $(\Omega', \mathcal{A}') = (\mathbb{R}, \mathcal{B})$ und $P(A) = \int_A dP = \int_A f(x) dx$ für alle A aus \mathcal{B} , der Borel- σ -Algebra.

Landau-Symbole

Es seien a_n, b_n Zahlenfolgen. Schreibe $a_n = o(b_n)$, falls $\frac{a_n}{b_n} \rightarrow 0$ und $a_n = O(b_n)$ falls $\frac{a_n}{b_n}$ eine beschränkte Folge ist. Auf eine stochastische Folge $(X_n)_{n \in \mathbb{N}}$ übertragen definieren wir:

$X_n = o_p(a_n)$ falls $\lim_{n \rightarrow \infty} P\left(\left|\frac{X_n}{a_n}\right| \geq \varepsilon\right) = 0$ für alle $\varepsilon > 0$ und $X_n = O_p(a_n)$ falls für alle $\varepsilon > 0$ ein $M := M(\varepsilon) > 0$ und ein $N := N(\varepsilon) \in \mathbb{N}$ existiert mit $P\left(\left|\frac{X_n}{a_n}\right| > M\right) \leq \varepsilon$ für alle $n \geq N$.

Konvergenz

Betrachtet man eine Folge $(X_n)_{n \in \mathbb{N}}$ von Zufallsvariablen, so können verschiedene Arten von Konvergenz definiert und betrachtet werden. Wir schreiben $X_n \xrightarrow{f.s.} X$ (X_n konvergiert fast sicher gegen X), falls $P(\lim_{n \rightarrow \infty} X_n = X) = 1$ und $X_n \xrightarrow{P} X$ (X_n konvergiert stochastisch gegen X), falls $\lim_{n \rightarrow \infty} P(|X_n - X| \geq \varepsilon) = 0$ für alle $\varepsilon > 0$.

Sei (Ω, d) ein metrischer Raum. Als dritte Konvergenz-Art definiere $X_n \xrightarrow{D} X$, X_n konvergiert in Verteilung gegen X falls die dazugehörigen Bildmaße P_n schwach gegen das zu X gehörige Bildmaß P konvergieren. Man sagt die Maße P_n konvergieren schwach gegen das Maß P , falls für alle beschränkten und bezüglich der Metrik d stetigen Funktionen $h : \Omega \rightarrow \mathbb{R}$ gilt, dass $\lim_{n \rightarrow \infty} \int h dP_n = \int h dP^X$.

Schwache Konvergenz (für Folgen stochastischer Prozesse) wird im folgenden noch definiert werden, diese ist die in dieser Arbeit maßgeblich genutzte Art der Konvergenz.

Stochastische Prozesse und schwache Konvergenz stochastischer Prozesse

Die im Weiteren in diesem Abschnitt auftretenden Zufallsvariablen Z_1, \dots, Z_n seien immer unabhängig und identisch verteilt.

Definition 2.1 (stochastischer Prozess) Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, (Ω', \mathcal{A}') ein mit einer σ -Algebra versehener Raum und T beliebige Indexmenge. Ein stochastischer Prozess X ist dann eine Familie von Zufallsvariablen $X(t) := X(\cdot, t) : \Omega \rightarrow \Omega', t \in T$, so dass $X(\cdot, t)$ für alle $t \in T$ \mathcal{A}/\mathcal{A}' -messbar ist. Für jedes $\omega \in \Omega$ heißt die Abbildung $X(\omega, \cdot), t \mapsto X(\omega, t)$ Pfad. Die Menge aller Pfade nennt man Pfadraum. Gilt $E[X(t)] = 0$ für alle $t \in T$, so heißt der Prozess zentriert.

Bemerkung 2.2 Die obige Definition hält T allgemein, so betrachtet man zum Beispiel häufig $T = [0, 1]$. Es sind aber noch viel abstraktere Indexmengen möglich, zum Beispiel Funktionenklassen. Diese werden uns im Weiteren noch interessieren, denn auf der Basis von mit Funktionenklassen indizierten stochastischen Prozessen gibt es einige in dieser Arbeit genutzte Resultate zur oben schon erwähnten, aber noch nicht definierten schwachen Konvergenz.

Bemerkung 2.3 Meist fordert man in der Stochastik von Zufallsvariablen $(X_n)_{n \in \mathbb{N}}$ die Borel-messbarkeit für jedes n , d.h. die Messbarkeit bezüglich der durch die offenen Mengen erzeugten σ -Algebra. Stochastische Prozesse erfüllen diese Forderung mitunter nicht, ein Beispiel findet sich in Billingsley (1968, Seite 152). Die Messbarkeit ist aber nötig, wollen wir Erwartungswerte etc. von Prozessen bestimmen. Es gibt zwei grundsätzliche Methoden dieses Problem anzugehen, einmal die Definition der äußeren Erwartung und den der Beschränkung auf eine kleinere σ -Algebra und passende Norm, in diesem Fall die σ -Algebra erzeugt durch die offenen Bälle anstelle derjenigen erzeugt durch die offenen Mengen versehen mit der Supremumsnorm. Beide Vorgehensweisen sind kompatibel, vergleiche dazu van der Vaart und Wellner (2000), S. 34 und 82 und Bemerkung 2.12.

Definition 2.4 Ein stochastischer Prozess $(X(t))_{t \in T}$ heißt Gaußprozess, falls für jede Teilmenge $\{t_1, \dots, t_k\}$ von T , $k \in \mathbb{N}$ beliebig, der Vektor $(X(t_1), \dots, X(t_k))^t$ k -dimensional normalverteilt ist, wobei Erwartungswert und Varianz von t_1, \dots, t_k abhängen können. Gilt für einen solchen Gaußprozess $\text{Cov}(X(s), X(t)) = \min(s, t) - s \cdot t$, so nennt man den Prozess eine Brownsche Brücke.

Definition 2.5 Sei $a, b \in \bar{\mathbb{R}}$ mit $a < b$ und $f : [a, b] \rightarrow \mathbb{R}$ eine Abbildung für die in jedem Punkt die links und rechtsseitigen Grenzwerte existieren und $f(x) = \lim_{x_n \searrow x} f(x_n)$ gilt, dann heißt f Cadlag-Funktion. Der Raum aller Cadlag-Funktionen auf $[a, b]$ wird mit $D([a, b])$, bezeichnet

und Skorokhod-Raum genannt. Mit $C([0, 1])$ bezeichne im Weiteren den Raum der auf $[0, 1]$ stetigen Funktionen, $C([0, 1]) \subset D([0, 1])$.

Definition 2.6 Es seien X_1, \dots, X_n reellwertige Zufallsvariablen, $t \in (-\infty, \infty)$. Dann heißt

$$F_n(t) := \frac{1}{n} \sum_{i=1}^n I\{X_i \leq t\}$$

empirische Verteilungsfunktion (EDF) von X_1, \dots, X_n .

Die empirische Verteilungsfunktion erfüllt einige wünschenswerte Eigenschaften, so ist sie zum Beispiel erwartungstreu. Als Schätzer für die Verteilungsfunktion von X_i , falls X_1, \dots, X_n unabhängig und identisch verteilt sind und es gilt das Glivenko-Cantelli Theorem:

Proposition 2.7 (Glivenko-Cantelli) Seien X_1, \dots, X_n unabhängig und identisch verteilt mit Verteilungsfunktion F . Dann gilt:

$$P\left(\lim_{n \rightarrow \infty} \sup_{t \in \mathbb{R}} |F_n(t) - F(t)| = 0\right) = 1$$

Zum Beweis siehe zum Beispiel Rohatgi und Saleh (2001).

Wir nutzen im Verlauf der Arbeit häufiger eine dem Glivenko-Cantelli Theorem ähnliche Aussage für allgemeinere stochastische Prozesse, vgl. Theorem 2.39.

Bemerkung 2.8 Als Beispiel eines speziellen stochastischen Prozesses mit $T = [0, 1]$ nutzen wir die empirische Verteilungsfunktion. Betrachte auf $[0, 1]$ unabhängig gleichverteilte Zufallsvariablen X_1, \dots, X_n . Dann hat der Prozess $(G_n(t))_{t \in [0, 1]} := (\sqrt{n}(F_n(t) - t))_{t \in [0, 1]}$ Pfade in $D([0, 1])$ und ist zentriert. Für auf ebensolche Weise aus X_1, \dots, X_n mit stetiger Verteilungsfunktion F konstruierte Prozesse, welche Pfade in $D([-\infty, \infty])$ besitzen, kann wieder ein Prozess mit Pfaden in $D([0, 1])$ erzeugt werden (vgl. z.B. Billingsley 1968), da $F(X_i)$ gleichverteilt auf $[0, 1]$ ist.

Definition 2.9 Eine Folge von stochastischen Prozessen $(X_n(\cdot))_{n \in \mathbb{N}}$ im Raum $D([0, 1])$ konvergiert schwach gegen einen stochastischen Prozess $X(\cdot)$ in $D([0, 1])$ falls

$$\lim_{n \rightarrow \infty} E[H(X_n)] = E[H(X)]$$

für alle beschränkten Funktionale $H : D([0, 1]) \rightarrow \mathbb{R}$ gilt, welche bezüglich der Supremumsnorm stetig sind, und falls X messbar bezüglich der durch die offenen Bälle in $D([0, 1])$ erzeugten σ -Algebra ist.

Der Satz von Donsker besagt nun, dass $G_n(t)$ aus Bemerkung 2.8 schwach gegen einen Gaußprozess $B^0(t)$, mit $\text{Cov}(B^0(s), B^0(t)) = \min(s, t) - s \cdot t$, also eine Brownsche Brücke, konvergiert. Für beliebige Zufallsvariablen X_1, \dots, X_n unabhängig mit stetiger Verteilungsfunktion F kann man zeigen, dass $G_n^F(t) := (\sqrt{n}(F_n(t) - F(t)))_{t \in \mathbb{R}}$ schwach gegen $B^0 \circ F := B^0(F(t))_{t \in [-\infty, \infty]}$ konvergiert (Billingsley 1968, Theorem 16.4), dabei sei $F(-\infty) := 0, F(\infty) := 1$ gesetzt und es wird ausgenutzt, dass $F(X_i)$ gleichverteilt auf $[0, 1]$ ist.

Proposition 2.10 (Continuous Mapping Theorem)

Sei $(Z_n)_{n \in \mathbb{N}}$ eine Folge stochastischer Prozesse, welche schwach gegen einen stochastischen Prozess Z mit $P(Z \in C([0, 1])) = 1$ konvergiert und bezeichne $\mathcal{B}^\circ(D)$ die von den offenen Bällen in $D[0, 1]$ erzeugte σ -Algebra. Die Abbildung $\psi : D([0, 1]) \rightarrow \mathbb{R}$ sei $\mathcal{B}^\circ(D)/\mathcal{B}(\mathbb{R})$ messbar und stetig in $C([0, 1])$. Dann konvergiert $\psi(Z_n)$ schwach gegen $\psi(Z)$.

zum Beweis siehe Pollard (1984), S. 70.

Wir benötigen in einigen Beweisen Aussagen zur schwachen Konvergenz. Im folgenden wird ein Vorgehen vorgestellt, welches in dieser Arbeit häufiger angewandt wird. Dabei betrachtet man stochastische Prozesse nicht mehr als Prozesse mit einem Indexbereich, welcher Teilmenge der reellen Zahlen ist, sondern betrachte Prozesse mit einer Funktionenklasse \mathcal{F} als Indexbereich:

$$(2.1) \quad G_n^P(\varphi) = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \varphi(Z_i) - E[\varphi(Z_1)] \right), \quad \varphi \in \mathcal{F},$$

wobei Z_1, \dots, Z_n χ -wertig und unabhängig und identisch verteilt mit Verteilung P sind. Zu klären ist nun für welche Funktionenklassen \mathcal{F} für den Prozess schwache Konvergenz vorliegt. Für den

Prozess aus Beispiel 2.8

$$G_n^F(y) := \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n I\{Z_i \leq t\} - P(Z_1 \leq t) \right), \quad t \in \mathbb{R}$$

ist mit $\mathcal{F} = \{ \varphi : \mathcal{X} \rightarrow \mathbb{R} \mid z \mapsto I\{z \leq t\}, t \in \mathbb{R} \}$ der Prozess wie in (2.1) darstellbar.

Zu bemerken ist, dass insbesondere für diese Arbeit Funktionenklassen von Interesse sind, welche vom Stichprobenumfang n abhängen, also die Folge der Funktionenklassen mit n variiert; passende Sätze werden zum Ende des Kapitels noch angegeben.

Im Weiteren werden grundlegenden Definitionen zusammengetragen und Theoreme angegeben, welche es ermöglichen auf schwache Konvergenz (vgl. Definition 2.14) des Prozesses (2.1) zu schließen.

Definition 2.11 (i) *Der Raum der gleichmäßig beschränkten, reellwertigen Funktionen auf \mathcal{F} , d.h. $\{ h : \mathcal{F} \rightarrow \mathbb{R} \mid \sup_{\varphi \in \mathcal{F}} |h(\varphi)| < \infty \}$, versehen mit der Supremumsnorm und der von den offenen Bällen erzeugten σ -Algebra $\mathcal{B}^\circ(D)$ bezeichne mit $l^\infty(\mathcal{F})$.*

(ii) *Ein Element G aus $l^\infty(\mathcal{F})$ heißt straff, falls zu jedem $\epsilon > 0$ eine kompakte Menge K existiert, so dass $P(G \in K) \geq 1 - \epsilon$.*

(iii) *Mit $\| \cdot \|_{r,P}$ bezeichne die $L_r(P)$ -Norm, d.h. die L_r -Norm bezüglich des Wahrscheinlichkeitsmaßes P , $\| \varphi \|_{r,P} := (\int |\varphi|^r dP)^{1/r}$.*

(iv) *Sei $\mathcal{F} = \{ \varphi : \mathcal{X} \rightarrow \mathbb{R} \}$ Funktionenklasse. Dann heißt Φ Einhüllende von \mathcal{F} , falls für jedes $\varphi \in \mathcal{F}$ und $x \in \mathcal{X}$ $|\varphi(x)| \leq \Phi(x)$. Φ muss nicht notwendigerweise in \mathcal{F} liegen.*

Bemerkung 2.12 *Betrachtet man den empirischen Prozess $G_n(t)$ aus Bemerkung 2.8 als Abbildung von $[0, 1]^n$ nach $D[0, 1]$, so ist der Prozess nicht Borel-messbar, die σ -Algebra ist zu groß (vgl. Billingsley (1986), S. 150 - 153). Um die Messbarkeit wieder herzustellen beschränken wir uns auf eine kleinere σ -Algebra, die σ -Algebra der offenen Bälle.*

Ein anderer Weg ist die Definition der äußeren Erwartung, definiert für beliebige Funktionen, wobei man auf die Messbarkeit verzichten kann (vgl. van der Vaart und Wellner (2000), S. 2 und 3), bzw. dies nur noch für den Grenzprozess gefordert wird.

Die schwache Konvergenz eines Prozesses in $l^\infty(\mathcal{F})$ definiert man aber analog zu Definition 2.9, verzichtet aber auf die Messbarkeitsannahme und nutzt die äußere Erwartung:

Definition 2.13 Gegeben (Ω, \mathcal{A}, P) und $Z : \Omega \rightarrow \bar{\mathbb{R}}$ definiere den äußeren Erwartungswert von Z als

$$E^*[Z] := \inf \{ E[U] : U \geq Z, U : \Omega \rightarrow \bar{\mathbb{R}} \text{ messbar und } E_P[U] \text{ existent} \}$$

Diese Konstruktion nutzen wir in der folgenden allgemeineren Definition der schwachen Konvergenz stochastischer Prozesse, die im weiteren berechneten Erwartungswerte existieren aber jeweils und lassen sich direkt berechnen.

Definition 2.14 Eine Folge von stochastischen Prozessen $(X_n(\cdot))_{n \in \mathbb{N}}$ im Raum $l^\infty(\mathcal{F})$ konvergiert schwach gegen einen stochastischen Prozess $X(\cdot)$ in $l^\infty(\mathcal{F})$ falls

$$\lim_{n \rightarrow \infty} E^*[H(X_n)] = E[H(X)]$$

für alle beschränkten Funktionale $H : l^\infty(\mathcal{F}) \rightarrow \mathbb{R}$ gilt, welche bezüglich der Supremumsnorm stetig sind und X messbar bezüglich der Borel- σ -Algebra ist.

Definition 2.15 (Glivenko-Cantelliklasse) Es seien Z_1, Z_2, \dots unabhängig, identisch verteilt. Eine Funktionenklasse $\mathcal{F} = \{ \varphi : \mathcal{X} \rightarrow \mathbb{R} \}$ heißt P -Glivenko-Cantelli-Klasse (oder kurz Glivenko-Cantelli) f.s. (oder in Wahrscheinlichkeit), falls die folgende Konvergenz

$$\sup_{\varphi \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n \varphi(Z_i) - E[\varphi(Z_1)] \right| \rightarrow 0 \text{ für } n \rightarrow \infty$$

f.s. (oder in Wahrscheinlichkeit) gilt.

Definition 2.16 (Donskerklasse) Es seien Z_1, Z_2, \dots unabhängig, identisch verteilt mit Maß P . Eine Funktionenklasse $\mathcal{F} = \{ \varphi : \mathcal{X} \rightarrow \mathbb{R} \}$ heißt P -Donsker-Klasse (oder kurz Donsker), falls schwache Konvergenz von

$$G_n(\varphi) = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \varphi(Z_i) - E[\varphi(Z_1)] \right), \varphi \in \mathcal{F}$$

gegen ein ein straffes, Borel-messbares Element G aus $l^\infty(\mathcal{F})$ gilt.

Zum Beispiel ist die oben betrachtete Funktionenklasse $\mathcal{F} = \{\varphi \mid \varphi(z) = I\{z \leq y\}, y \in \mathbb{R}\}$ Donskerklasse.

Proposition 2.17

Vereinigungen von Donskerklassen und die Summe von Donskerklassen $\mathcal{F} + \mathcal{G}$, definiert als paarweise Summe von Elementen zweier Donskerklassen sind wieder Donsker, falls $\sup_{\varphi \in \mathcal{F} \cup \mathcal{G}} |\int \varphi dP| < \infty$.

Zum Beweis vgl. van der Vaart und Wellner (2000), S. 192.

Im Weiteren werden nun Voraussetzungen angegeben, die die Eigenschaft Donsker bzw. die Eigenschaft Glivenko-Cantelli zu sein folgern lassen.

Definition 2.18 Die Überdeckungszahl $\mathcal{N}(\epsilon, \mathcal{F}, \|\cdot\|)$ bezeichne die minimale Anzahl von Bällen mit Radius ϵ zur Norm $\|\cdot\|$, also Mengen derart $\{\psi \in \mathcal{F} \mid \|\psi - \varphi\| < \epsilon\}$, welche nötig sind um \mathcal{F} zu überdecken. Die Zentren der Bälle müssen $\|\varphi\| < \infty$ erfüllen, aber nicht notwendig in \mathcal{F} liegen.

$\log \mathcal{N}(\epsilon, \mathcal{F}, \|\cdot\|)$ bezeichnen wir als Entropie ohne Klammerung,

$$\int_0^\infty \sqrt{\log \mathcal{N}(\epsilon, \mathcal{F}, \|\cdot\|)} d\epsilon$$

als Überdeckungsintegral.

Definition 2.19 Sei $\|\cdot\|$ eine Norm und l und u zwei reellwertige Funktionen mit endlicher Norm. Dann bezeichne mit einer Klammer $[l, u]$ die Menge aller Funktionen φ mit $l \leq \varphi \leq u$. Eine ϵ -Klammer sei eine Klammer $[l, u]$ mit $\|l - u\| < \epsilon$.

Mit der Klammerungszahl $\mathcal{N}_{[\cdot]}(\epsilon, \mathcal{F}, \|\cdot\|)$ bezeichne dann die minimale Anzahl von ϵ -Klammern, welche nötig sind um \mathcal{F} zu überdecken. l und u müssen dabei nicht notwendig in \mathcal{F} liegen.

$\log \mathcal{N}_{[\cdot]}(\epsilon, \mathcal{F}, \|\cdot\|)$ bezeichnen wir als Entropie mit Klammerung,

$$\int_0^\infty \sqrt{\log \mathcal{N}_{[\cdot]}(\epsilon, \mathcal{F}, \|\cdot\|)} d\epsilon$$

als Klammerungsintegral.

Proposition 2.20 *Es seien \mathcal{F} und \mathcal{G} Funktionenklassen. Dann gilt für jedes Wahrscheinlichkeitsmaß Q und jedes $1 \leq p < \infty$*

$$(i) \mathcal{N}_{[]} (2\epsilon, \mathcal{F} + \mathcal{G}, L_p(Q)) \leq \mathcal{N}_{[]} (\epsilon, \mathcal{F}, L_p(Q)) \cdot \mathcal{N}_{[]} (\epsilon, \mathcal{G}, L_p(Q))$$

(ii) *Vorausgesetzt, dass die Klassen \mathcal{F} und \mathcal{G} eine Einhüllende $\Phi \leq 1$ besitzen gilt:*

$$\mathcal{N}_{[]} (2\epsilon, \mathcal{F} \cdot \mathcal{G}, L_p(Q)) \leq \mathcal{N}_{[]} (\epsilon, \mathcal{F}, L_p(Q)) \cdot \mathcal{N}_{[]} (\epsilon, \mathcal{G}, L_p(Q))$$

vgl. Kosorok (2008), S. 169, Lemma 9.25.

Definition 2.21 *Eine Klasse von Funktionen \mathcal{F} heißt euklidisch zur Einhüllenden Φ , falls Konstanten A und V existieren, so dass $\mathcal{N}(\epsilon, \mathcal{F}, L_1(Q)) \leq A \cdot \epsilon^{-V} (\int \Phi dQ)^V$ für alle $\epsilon > 0$ und alle Wahrscheinlichkeitsmaße Q mit $\int \Phi dQ < \infty$.*

Wir bezeichnen eine Folge \mathcal{F}_n von Funktionenklassen als euklidisch zur Einhüllenden Φ_n , falls von n unabhängige Konstanten A und V existieren, so dass $\mathcal{N}(\epsilon, \mathcal{F}_n, L_1(Q)) \leq A \cdot \epsilon^{-V} (\int \Phi_n dQ)^V$ für alle $\epsilon > 0$ und alle Wahrscheinlichkeitsmaße Q mit $\int \Phi_n dQ < \infty$.

Proposition 2.22 *Sei \mathcal{F} euklidisch zur Einhüllenden Φ und $p \geq 1$. Dann existieren Konstanten B und W , so dass $\mathcal{N}(\epsilon, \mathcal{F}, L_p(Q)) \leq A \cdot \epsilon^{-W} (\int \Phi^p dQ)^{W/p}$ für alle $\epsilon > 0$ und alle Wahrscheinlichkeitsmaße Q mit $\int \Phi^p dQ < \infty$.*

vgl. Neumeyer (2006) S. 150, Lemma A.10.

Im Verlauf dieser Arbeit ist es häufiger nötig zu zeigen, dass eine auftretende Funktionenklasse euklidisch ist. Die Beweise dazu sind in Kapitel 6, Lemma 6.6 ausgelagert.

Nun kann mittels obiger Konstruktionen Rückschluss auf die Donsker-Eigenschaft einer Funktionenklasse \mathcal{F} gezogen werden:

Proposition 2.23 *Eine Funktionenklasse \mathcal{F} von messbaren Funktionen ist P -Donsker, falls das dazugehörige Klammerungsintegral, versehen mit der $L_2(P)$ -Norm,*

$$\int_0^\infty \sqrt{\log \mathcal{N}_{[]} (\epsilon, \mathcal{F}, L_2(P))} d\epsilon$$

endlich ist.

zum Beweis siehe van der Vaart und Wellner (2000) S. 85.

Diese Forderung der Endlichkeit des Integrales ist sehr stark und das Theorem häufig nicht anwendbar, da es in vielen interessanten Fällen schwer zu zeigen ist, dass das Integral endlich ist. Diese Forderung ist noch abschwächbar. Im folgenden werden weitere Möglichkeiten aufgezeigt die Donsker-Eigenschaft zu beweisen:

Definition 2.24 (*VC-Index und VC-Klasse*)

Sei \mathcal{M} eine Menge und \mathcal{M}_1 Teilmenge von \mathcal{M} bestehend aus n Elementen. \mathcal{C} sei eine Menge von Teilmengen von \mathcal{M} . \mathcal{C} heißt \mathcal{M}_1 -zerschlagend, falls durch Schneiden von \mathcal{M}_1 mit Elementen $C \in \mathcal{C}$ alle möglichen 2^n Teilmengen von \mathcal{M}_1 erzeugbar sind.

Der VC-Index $V(\mathcal{C})$ einer Menge \mathcal{C} von Teilmengen von \mathcal{M} ist nun definiert als das kleinste n , für das keine aus n Elementen bestehende Teilmenge \mathcal{M}_1 von \mathcal{M} durch \mathcal{C} zerschlagen wird.

Eine Menge \mathcal{C} heißt VC-Klasse von Mengen, falls der dazugehörige VC-Index $V(\mathcal{C})$ endlich ist.

Siehe zu dieser Konstruktion auch Vapnik und Cervonenkis (1971).

Über die folgende Definition schlagen wir eine Verbindung zu Funktionenklassen:

Definition 2.25 (*Subgraph*)

Mit dem Subgraphen einer Funktion $\varphi : \mathcal{X} \rightarrow \mathbb{R}$ bezeichne im Weiteren die Menge $\{(x, t) \mid t < \varphi(x)\}$, welche Teilmenge von $\mathcal{X} \times \mathbb{R}$ ist.

Definition 2.26 Eine Klasse von reellwertigen Funktionen \mathcal{F} auf einem Messraum \mathcal{X} heißt VC-Klasse, falls die Menge aller Subgraphen eine VC-Klasse von Mengen in $\mathcal{X} \times \mathbb{R}$ bildet, ebenso wird der VC-Index einer Klasse von Mengen auf Funktionenklassen übertragen.

Beispiel 2.27 Jeder endlichdimensionale Vektorraum \mathcal{F} aus messbaren Funktionen $\varphi : \mathcal{X} \rightarrow \mathbb{R}$ ist VC-Klasse mit VC-Index $\leq \dim(\mathcal{F}) + 2$.

vgl. van der Vaart und Wellner (2000), Lemma 2.6.15 und die Verwendung im Beweis des Satzes 3.12.

Beispiel 2.28 Die Funktionenklasse $\mathcal{F} := \{\varphi_y : \mathbb{R} \rightarrow \mathbb{R} \mid \varphi_y(x) = I\{x \leq y\}, y \in \mathbb{R}\}$ ist VC-Klasse, denn jede Menge der Form $M_1 := \{(a, c_1), (b, c_2)\}$ mit $a < b$, c_1, c_2 beliebig wird von den Subgraphen $\mathcal{C} := \{((-\infty, y] \times \mathbb{R}^{\leq 1}) \cup (\mathbb{R} \times \mathbb{R}^{\leq 0}), y \in \mathbb{R}\}$ nicht zerschlagen, denn, falls man (a, c_1) darstellen kann, so kann man den Punkt (b, c_2) nicht durch Schnitte von M_1 mit Elementen aus \mathcal{C} darstellen ohne den Punkt (a, c_1) auch zu erhalten. Damit ist \mathcal{F} VC-Klasse mit VC-Index 2.

Die beiden folgenden Propositionen aus Kosorok (2008), S. 161, Lemma 9.9 werden in späteren Kapitel häufiger angewendet:

Proposition 2.29 Sei \mathcal{F} VC-Klasse von reellwertigen Funktionen auf einem Messraum \mathcal{X} und sei $g : \mathcal{X} \rightarrow \mathbb{R}$ eine feste messbare Funktion, dann ist $\mathcal{F} \cdot g := \{f \cdot g \mid f \in \mathcal{F}\}$ VC-Klasse mit VC-Index $\leq 2V(\mathcal{F}) - 1$.

Proposition 2.30 Sei \mathcal{F} VC-Klasse von reellwertigen Funktionen auf einem Messraum \mathcal{X} und sei $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ eine feste monotone Funktion, dann ist $\varphi \circ \mathcal{F} := \{\varphi \circ f \mid f \in \mathcal{F}\}$ VC-Klasse mit VC-Index $\leq V(\mathcal{F})$.

Proposition 2.31 Sei \mathcal{F} euklidische Klasse von reellwertigen Funktionen auf einem Messraum \mathcal{X} und sei $g : \mathcal{X} \rightarrow \mathbb{R}$ eine feste messbare Funktion, dann ist $\mathcal{F} \cdot g := \{f \cdot g \mid f \in \mathcal{F}\}$ euklidisch.

vgl. Kosorok (2008), Lemma 9.9 (vi).

Proposition 2.32 Sei \mathcal{F} Klasse von reellwertigen monotonen Funktionen $f : \mathbb{R} \rightarrow [-1, 1]$, dann ist \mathcal{F} euklidisch.

vgl. Kosorok (2008), Lemma 9.11.

Proposition 2.33 *Jede VC-Klasse \mathcal{F} von Funktionen mit VC-Index $V(\mathcal{F})$ und Einhüllender Φ , $r \geq 1$ ist euklidisch und es existiert eine Konstante A und $\epsilon \in (0, 1)$, so dass gilt:*

$$\sup_{P: \|\Phi\|_{r,P} > 0} \mathcal{N}(\epsilon \|\mathcal{F}\|_{r,P}, \mathcal{F}, L_r(P)) \leq A \cdot V(\mathcal{F}) (16e)^{V(\mathcal{F})} \epsilon^{-r(V(\mathcal{F})-1)}$$

vgl. van der Vaart und Wellner (2000), S. 141, Theorem 2.6.7.

Im Weiteren begegnet uns häufiger die folgende Funktionenklasse, für welche es Abschätzungen der Entropie mit und ohne Klammerung gibt (vgl. dazu auch van der Vaart und Wellner (2000), S. 154):

Definition 2.34 $(C_\delta^{1+\alpha}[0, 1])$

Die Menge der differenzierbaren Funktionen $h : [0, 1] \rightarrow \mathbb{R}$ mit

$$\max\left(\sup_{x \in [0,1]} |h(x)|, \sup_{x \in [0,1]} |h'(x)|\right) + \sup_{x,y \in [0,1]} \frac{|h'(x) - h'(y)|}{|x - y|^\alpha} \leq \delta$$

für ein $\alpha \in (0, 1]$ bezeichne im Weiteren mit $C_\delta^{1+\alpha}[0, 1]$.

Proposition 2.35 *Betrachte die Funktionenklasse $C_\delta^{1+\alpha}[0, 1]$ mit $\delta = 1$. Dann existiert ein $K_1 = K_1(\alpha)$, so dass für jedes $\epsilon > 0$ gilt:*

$$\log \mathcal{N}(\epsilon, C_\delta^{1+\alpha}[0, 1], \|\cdot\|_\infty) \leq K_1 \left(\frac{1}{\epsilon}\right)^{1/(1+\alpha)}.$$

Es existiert ebenso ein $K_2 = K_2(\alpha)$, so dass für jedes $\epsilon > 0$, $r \geq 1$ und Wahrscheinlichkeitsmaß Q auf $[0, 1]$ gilt:

$$\log \mathcal{N}_{[\cdot]}(\epsilon, C_\delta^{1+\alpha}[0, 1], L_r(Q)) \leq K_2 \left(\frac{1}{\epsilon}\right)^{1/(1+\alpha)}$$

vgl. van der Vaart und Wellner (2000), S. 154 und Nagel (2006), S. 13.

Hat man für eine Klasse von Funktionen \mathcal{F} die Donskereigenschaft nachgewiesen so kann die folgende Proposition nützlich sein:

Proposition 2.36 *Seien Z_1, Z_2, \dots iid $\sim P$ und sei \mathcal{F} eine P -Donsker-Klasse von messbaren Funktionen, und bezeichnet $(\hat{\varphi}_n)_{n \in \mathbb{N}}$ eine Folge zufälliger Funktionen, die ihre Werte in \mathcal{F} annehmen, so dass $\int (\hat{\varphi}_n - \varphi_0)^2 dP$ für ein $\varphi_0 \in L_2(P)$ in Wahrscheinlichkeit gegen Null konvergiert. Dann konvergiert*

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n (\hat{\varphi}_n(Z_i) - \varphi_0(Z_i)) - \int (\hat{\varphi}_n - \varphi_0) dP \right)$$

in Wahrscheinlichkeit gegen Null, und damit gilt insbesondere auch Verteilungskonvergenz gegen Null.

vgl. dazu van der Vaart (1998), S. 280, Lemma 19.24.

Prozesse auf Basis von von n abhängenden Funktionenklassen

In dieser Arbeit sind insbesondere Funktionenklassen von Interesse, welche von n abhängen. Für diese gibt es auch Resultate zur schwachen Konvergenz etc. Zuerst ist zu klären, was es für von n abhängende Funktionenklassen bedeutet Donsker zu sein, die Definition 2.16 greift hier nicht:

Definition 2.37 *Seien Z_1, Z_2, \dots iid $\sim P$. Um die Formulierungen beizubehalten bezeichne im Weiteren eine Folge von von n abhängenden Funktionenklassen $(\mathcal{F}_n)_{n \in \mathbb{N}} := (\{ \varphi_{n,t} \mid t \in T \})_{n \in \mathbb{N}}$ als P -Donsker, falls der dazugehörige Prozess*

$$(2.2) \quad \{ G_n(t) \mid t \in T \} := \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n (\varphi_{n,t}(Z_i) - E[\varphi_{n,t}(Z_1)]) \right\}$$

schwach gegen einen Gaußprozess G aus $l^\infty(T)$ konvergiert.

Analog übertrage auch die Definition der Glivenko-Cantelli Eigenschaft:

Definition 2.38 Seien Z_1, Z_2, \dots iid $\sim P$. Eine Folge von von n abhängenden Funktionenklassen $(\mathcal{F}_n)_{n \in \mathbb{N}} := (\{\varphi_{n,t} \mid t \in T\})_{n \in \mathbb{N}}$ heißt P -Glivenko-Cantelli f.s. (oder in Wahrscheinlichkeit), falls die folgende Konvergenz

$$\sup_{\varphi_{n,t} \in \mathcal{F}_n} \left| \frac{1}{n} \sum_{i=1}^n \varphi_{n,t}(Z_i) - E[\varphi_{n,t}(Z_1)] \right| \rightarrow 0 \text{ für } n \rightarrow \infty$$

f.s (oder in Wahrscheinlichkeit) gilt.

Wir beginnen nach den Definitionen mit einem Glivenko-Cantelli-Theorem:

Proposition 2.39 Es seien Z_1, Z_2, \dots unabhängig identisch verteilte \mathcal{X} -wertige Zufallsvariablen. Mit \mathcal{Q} bezeichne die Menge aller Wahrscheinlichkeitsmaße auf \mathcal{X} . Für jedes n sei \mathcal{F}_n Funktionenklasse \mathcal{X} -wertiger Funktionen deren Überdeckungsanzahl

$$(2.3) \quad \sup_{Q \in \mathcal{Q}} \mathcal{N}(\epsilon, \mathcal{F}_n, L_1(Q)) \leq A\epsilon^{-W} \quad \text{mit } \epsilon \in (0, 1)$$

erfüllt, wobei A, W reellwertige Konstanten, unabhängig von n . Es existiere ein $c < \infty$, so dass $|\varphi| < c$ für alle $\varphi \in \mathcal{F}_n$ und $E[\varphi^2(Z_1)] \leq \delta_n^2$ für alle $\varphi \in \mathcal{F}_n$. α_n sei eine monoton fallende reellwertige Folge, so dass $\frac{\log(n)}{n\delta_n^2\alpha_n^2} \rightarrow 0$, dann gilt:

$$\sup_{\varphi \in \mathcal{F}_n} \left| \frac{1}{n} \sum_{i=1}^n \varphi(Z_i) - E[\varphi(Z_1)] \right| = o(\delta_n^2\alpha_n) \quad \text{f.s.}$$

Zum Beweis siehe Pollard (1984) Seite 34, Theorem 37.

Bemerkung 2.40 Die geforderte Eigenschaft, dass $|\varphi| < c$ kann noch abgeschwächt werden. Liegt nicht wie im obigen Satz gefordert $|\varphi| < c$, also keine konstante Einhüllende Φ vor, so ist der Satz allgemeiner gültig, es muss dann aber gelten:

(i) Es existiere ein $p \geq 1$, eine Einhüllende Φ für \mathcal{F}_n für alle $n \in \mathbb{N}$ und A, W wie in Proposition 2.39, so dass für jedes Wahrscheinlichkeitsmaß Q mit $\int \Phi^p dQ < \infty$ gilt, dass

$$\mathcal{N}(\epsilon, \mathcal{F}_n, L_1(Q)) \leq A\epsilon^{-W} \left(\int \Phi^p dQ \right)^{W/p}$$

(ii) $P\left(\sup_{\varphi \in \mathcal{F}_n} \frac{1}{n} \sum_{i=1}^n \varphi^2(Z_i) > \delta_n^2\right) = o(1)$

(iii) Es existiere eine Konstante \tilde{c} , so dass $\sup_{n \in \mathbb{N}} \frac{1}{n} \sum_{i=1}^n \Phi^p(Z_i) \leq \tilde{c}$ f.s.

vgl. dazu Neumeyer (2006) S. 151, remark A.12.

Für den Nachweis der Donsker-Klassen-Eigenschaft nutzen wir die folgenden Theoreme für von n abhängige Funktionenklassen:

Proposition 2.41 Sei Z \mathcal{X} -wertige Zufallsvariable und P das dazugehörige Wahrscheinlichkeitsmaß. Zusätzlich seien Z_1, Z_2, \dots iid. $\sim P$. Für jedes n sei $\mathcal{F}_n = \{\varphi_{n,t} \mid t \in T\}$ Funktionenklasse \mathcal{X} -wertiger Funktionen mit konstanter Einhüllender $\Phi = 1$ und (T, ρ) totalbeschränkter semimetrischer Raum, so dass für jede reelle Folge $\delta_n \rightarrow 0$ gilt:

$$(2.4) \quad \int_0^{\delta_n} \sqrt{\log \mathcal{N}_{[]}(\epsilon, \mathcal{F}_n, L_2(P))} d\epsilon \rightarrow 0$$

$$(2.5) \quad \sup_{\rho(s,t) < \delta_n} E[(\varphi_{n,t}(Z) - \varphi_{n,s}(Z))^2] \rightarrow 0.$$

Zusätzlich konvergiere $E[\varphi_{n,t}(Z) \cdot \varphi_{n,s}(Z)] - E[\varphi_{n,t}(Z)] \cdot E[\varphi_{n,s}(Z)]$ punktweise in $T \times T$. Dann konvergiert der Prozess $\{G_n(t) \mid t \in T\} := \{\sqrt{n}(\frac{1}{n} \sum_{i=1}^n \varphi_{n,t}(Z_i) - E[\varphi_{n,t}(Z_1)]) \mid t \in T\}$ gegen einen straffen Gaußprozess G .

siehe dazu Neumeyer (2006), S. 150 bzw. eine allgemeinere Aussage in van der Vaart und Wellner (2000), S. 221, Theorem 2.11.23.

In den meisten Fällen dieser Arbeit liegt keine beschränkte Einhüllende Φ vor. Wende in diesen Fällen die folgende Verallgemeinerung von Proposition 2.41 an:

Bemerkung 2.42 Seien Z, Z_1, Z_2, \dots \mathcal{X} -wertige Zufallsvariablen, iid. $\sim P$. Mit \mathcal{Q} bezeichne die Menge aller Wahrscheinlichkeitsmaße auf \mathcal{X} . und es gelten die Voraussetzungen von Proposition 2.41, nur existiere für die Funktionenklassen \mathcal{F}_n und deren jeweilige Einhüllende Φ_n keine gleichmäßig beschränkte Einhüllende Φ mit $\Phi_n \leq \Phi = 1$. Dann sind die Bedingungen (2.4) und $|\Phi_n| \leq 1$ durch die folgenden drei Bedingungen zu ersetzen und die Gültigkeit des Satzes 2.41 bleibt erhalten:

$$(2.6) \quad E[\Phi_n^2(Z)] = O(1)$$

$$(2.7) \quad E[\Phi_n^2(Z) I\{\Phi_n(Z) > \beta \sqrt{n}\}] \rightarrow 0 \quad \text{für jedes } \beta > 0$$

$$(2.8) \quad \sup_{Q \in \mathcal{Q}} \int_0^{\delta_n} \sqrt{\log \mathcal{N}(\epsilon \|\Phi_n\|_{2,Q}, \mathcal{F}_n, L_2(Q))} d\epsilon \rightarrow 0 \quad \text{für jedes } \delta_n \rightarrow 0$$

Die Bedingung (2.8) kann auch ersetzt werden durch

$$(2.9) \quad \int_0^{\delta_n} \sqrt{\log \mathcal{N}_{[]}(\epsilon \|\Phi_n\|_{2,P}, \mathcal{F}_n, L_2(P))} d\epsilon \rightarrow 0 \quad \text{für jedes } \delta_n \rightarrow 0$$

Siehe zu dieser Verallgemeinerung van der Vaart und Wellner (2000), S. 220, Theorem 2.11.22 und 2.11.23. Einige Beispiele für in dieser Arbeit benötigte Donskerklassen folgen in Lemma 6.7.

Proposition 2.43 Ist \mathcal{F}_n VC-Klasse mit Einhüllender Φ_n und VC-Index unabhängig von n oder euklidisch mit Konstanten unabhängig von n , so ist die Bedingung (2.8) aus 2.42 erfüllt.

Siehe dazu van der Vaart und Wellner (2000), S. 221, Beispiel 2.11.24.

Definition 2.44 Sei $(X_n)_{n \in \mathbb{N}}$ Folge stochastischer Prozesse in $l^\infty(T)$ und (T, ρ) sei totalbeschränkter metrischer Raum. Bezeichne $(X_n)_{n \in \mathbb{N}}$ als asymptotisch stochastisch gleich-stetig (asympt. stoch. equicontinuos) bezüglich ρ , falls

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} P\left(\sup_{\substack{s, t \in T: \\ \rho(s, t) < \delta}} |X_n(s) - X_n(t)| > \epsilon \right) = 0$$

für alle $\epsilon > 0$.

Die folgende Proposition ist ein Analogon zu Proposition 2.36 im Falle der Vorliegens n abhängiger Funktionenklassen. Diese Aussage wird an einigen Stellen der Arbeit zum Abschätzen von Resttermen genutzt.

Proposition 2.45 *Sei $\mathcal{C} = l^\infty(\mathcal{X})$, \mathcal{X} der Wertebereich von Z_1, \dots, Z_n , $T = \mathbb{R} \times \mathcal{C}$, $0_{\mathcal{C}}$ die Null in \mathcal{C} und $(G_n(t))_{t \in T}$ stochastischer Prozess der Form*

$$\{G_n(y, c) \mid (y, c) \in T\} := \left\{ \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \psi_{n,y,c}(Z_i) - \int \psi_{n,y,c} dP \right) \mid (y, c) \in T \right\},$$

welcher schwach gegen einen Gaußprozess G_P mit $G_P(\cdot, 0_{\mathcal{C}}) = 0$ konvergiere. Zusätzlich sei (T, ρ) totalbeschränkter semimetrischer Raum, so dass G_n asymptotisch stochastisch gleich-stetig bezüglich ρ (vgl. Definition 2.44). $(\hat{c}_n)_{n \in \mathbb{N}}$ sei eine Folge, welche ihre Werte in \mathcal{C} annimmt, so dass $\|\hat{c}_n\|_\infty = o(1)$ in Wahrscheinlichkeit. Für alle Folgen $(c_n)_{n \in \mathbb{N}}$ in \mathcal{C} mit $\|c_n\|_\infty \rightarrow 0$, für alle $\delta > 0$ existiere ein $n_0 := n_0(\delta) \in \mathbb{N}$, so dass gilt $\rho((y, c_n), (y, 0_{\mathcal{C}})) < \delta$ für alle $n \geq n_0$. Dann konvergiert der Prozess

$$\{G_n(y, \hat{c}_n) \mid y \in \mathbb{R}\} := \left\{ \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \psi_{n,y,\hat{c}_n}(Z_i) - \int \psi_{n,y,\hat{c}_n} dP \right) \mid y \in \mathbb{R} \right\}$$

schwach gegen 0.

Beweis:

Mit Theorem 7.15 aus Kosorok (2008), S. 112 und den Voraussetzungen folgt, dass $X_n := (G_n(\cdot, \cdot), \hat{c}_n)$ schwach gegen $X := (G_P(\cdot, \cdot), 0)$ konvergiert, dabei bezeichnet $0 = 0_{\mathcal{C}}$ und schwache Konvergenz von \hat{c}_n gegen 0 folgt nach Voraussetzung.

Bezeichne mit $UC(T, \rho)$ im Weiteren die Teilmenge von $l^\infty(T)$ der bezüglich ρ gleichmäßig stetigen Abbildungen, d.h.

$$\lim_{\delta \rightarrow 0} \sup_{\substack{t_1, t_2 \in T \\ \rho(t_1, t_2) < \delta}} |f(t_1) - f(t_2)| = 0 \text{ für alle } f \in UC(T, \rho).$$

Es gilt nach Voraussetzung, dass X_n stochastisch asymptotisch gleichstetig bzgl. ρ ist. Setze $\mathcal{D} := l^\infty(T) \times \mathcal{C}$ und $\mathcal{D}_0 := UC(T, \rho) \times \{0_{\mathcal{C}}\}$. Nach Theorem 7.19, S. 114 aus Kosorok (2008) folgt aus der asymptotischen stoch. gleich-Stetigkeit von X_n , dass $P(X \in \mathcal{D}_0) = 1$. Nun möchten wir das Continuous-Mapping-Theorem, Theorem 7.7, S. 109 aus Kosorok (2008) anwenden, dazu sind die Voraussetzungen zu prüfen. Definiere $g : \mathcal{D} \rightarrow l^\infty(\mathbb{R})$ mit $g(x(\cdot, \cdot), c) = x(\cdot, c)$. Um die

Aussage des obigen Satzes zu erhalten ist noch zu prüfen, ob g stetig in \mathcal{D}_0 ist, was aber aus der Konstruktion folgt:

Sei $(x_n(\cdot, \cdot), c_n)_{n \in \mathbb{N}}$ Folge in \mathcal{D} , die (bezüglich $\|\cdot\|_\infty \times \|\cdot\|_\infty$ in $l^\infty(T) \times \mathcal{C}$) gegen $(x(\cdot, \cdot), 0) \in \mathcal{D}_0$ konvergiert, d.h. $\sup_{(y,c) \in T} |x_n(y, c) - x(y, c)| \rightarrow 0$ und $\|c_n\|_\infty \rightarrow 0$. Für $\delta \rightarrow 0$ sei n_0 wie in der Voraussetzung. Dann gilt:

$$\begin{aligned}
\|g(x_n(\cdot, \cdot), c_n) - g(x(\cdot, \cdot), 0)\|_\infty &= \|x_n(\cdot, c_n) - x(\cdot, 0)\|_\infty \\
&\leq \|x_n(\cdot, c_n) - x(\cdot, c_n)\|_\infty + \sup_{y \in \mathbb{R}} |x(y, c_n) - x(y, 0)| \\
&\stackrel{(1)}{=} o(1) + \sup_{y \in \mathbb{R}} |x(y, c_n) - x(y, 0)| \\
&\stackrel{n \geq n_0}{\leq} o(1) + \sup_{\substack{y \in \mathbb{R}, c \in \mathcal{C}: \\ \rho((y,c), (y,0)) < \delta}} |x(y, c) - x(y, 0)| \\
&= o(1),
\end{aligned}$$

für $\delta \rightarrow 0$, da $X \in UC(T, \rho)$. (1) nutzt aus, dass $x_n \rightarrow x$ gleichmäßig, was aus den Eigenschaften von $l^\infty(T)$ folgt.

Mit Theorem 7.7 aus Kosorok (2008), S. 109 folgt dann die schwache Konvergenz von $G_n(\cdot, \hat{c}_n) = g(G_n(\cdot, \cdot), \hat{c}_n)$ gegen $g(G_P(\cdot, \cdot), 0_C) = G_P(\cdot, 0_C)$ mit $G_P(\cdot, 0_C) = 0$ nach Voraussetzung.

□

2.2 Kern(dichte-)schätzer

Kern(dichte-)Schätzer sind nichtparametrische Kurvenschätzer. Einen Einblick in nichtparametrische Kurvenschätzung findet man in Härdle (1993), im Bezug auf Dichteschätzung in Silverman (1986) und Scott (1992). Frühe wichtige Arbeiten zur Kern(dichte-)schätzung sind zum Beispiel Rosenblatt (1956), Nadaraya (1964), Watson (1964) und Epanechnikov (1969).

Dieser Abschnitt stellt die in dieser Arbeit benötigten Kern(dichte-)schätzer vor und fasst wichtige und häufig in dieser Arbeit genutzte Resultate zusammen.

Wir unterscheiden zu Beginn zwei Arten von Kernschätzern, einmal den herkömmlichen Kerndichteschätzer, das ist ein nichtparametrischer Kurvenschätzer der Dichte f_X einer Zufallsvariablen X auf Basis einer Stichprobe von unabhängig und identisch verteilten Beobachtungen X_1, \dots, X_n der Zufallsvariablen X und den Nadaraya-Watson Schätzer \hat{m} der Funktion m im Modell $Y = m(X) + \varepsilon$.

Im folgenden werden wir zuerst diese beiden Schätzer definieren und darauf Eigenschaften der Schätzer in Propositionen zusammenfassen:

Definition 2.46 $K : D \subseteq \mathbb{R} \rightarrow \mathbb{R}$ heißt (eindimensionale) Kernfunktion (oder Kern) der Ordnung 2, falls K die folgenden Eigenschaften besitzt:

- (i) $\int_D K(u) du = 1$
- (ii) K stetig, symmetrisch und beschränkt
- (iii) $\mu^K := \int_D u K(u) du = 0$ und $\mu_2^K := \int_D u^2 K(u) du \neq 0$

Einige Beispiele für (eindimensionale) Kerne sind in der folgenden Tabelle dargestellt, diese Kerne werden wir zum Teil im Weiteren für die Simulationen nutzen:

Kern	$K(u)$	Ordnung
Epanechnikov	$\frac{3}{4}(1 - u^2) I\{ u \leq 1\}$	2
Uniform	$\frac{1}{2} I\{ u \leq 1\}$	2
Quadric	$\frac{15}{16}(1 - u^2)^2 I\{ u \leq 1\}$	2
Gauss	$(2\pi)^{-\frac{1}{2}} \exp(-\frac{u^2}{2})$	2

Tab.1: Kernfunktionen

Definition 2.47 Sei $(b_n)_{n \in \mathbb{N}}$ mit $b_n > 0$ und $b_n \rightarrow 0$ für $n \rightarrow \infty$ eine reellwertige Zahlenfolge. Dann bezeichnen wir b_n im Weiteren als Bandbreite. Der Übersicht dienend schreibe oft nur b anstelle von b_n .

Definition 2.48 Es seien X_1, \dots, X_n reellwertige Zufallsvariablen und \tilde{K} eine Kernfunktion 2. Ordnung, $b := b_n$ eine Bandbreite. Dann heißt

$$\hat{f}_X(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{b} \tilde{K}\left(\frac{X_i - x}{b}\right) = \frac{1}{n} \sum_{i=1}^n \frac{1}{b} \tilde{K}_b(X_i - x)$$

Kerndichteschätzer der Dichte $f_X(x)$. Hierbei und im folgenden verwenden wir die Notation $\tilde{K}_b(\cdot) = \tilde{K}(\frac{\cdot}{b})$.

Hat man eine Stichprobe $(X_1, Y_1), \dots, (X_n, Y_n)$ unabhängig und identisch verteilt mit funktionalem Zusammenhang $Y = m(X) + \varepsilon$ gegeben und möchte man die den Zusammenhang beschreibende Funktion m schätzen, so bietet sich folgender nichtparametrischer Ansatz an:

Definition 2.49 Es seien $(X_1, Y_1), \dots, (X_n, Y_n)$ unabhängig und identisch verteilte \mathbb{R}^2 -wertige Zufallsvariablen und K eine Kernfunktion 2. Ordnung, b eine Bandbreite. Dann heißt

$$\hat{m}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{b} \tilde{K}\left(\frac{X_i - x}{b}\right) Y_i \frac{1}{\hat{f}_X(x)} =: \frac{1}{n} \sum_{i=1}^n \frac{1}{b} \tilde{K}_b(X_i - x) Y_i \frac{1}{\hat{f}_X(x)}$$

der Nadaraya-Watson Schätzer [Nadaraya (1964), Watson (1964)] für die Modellfunktion m .

Bemerkung 2.50 Die obigen Definitionen der Kernschätzer sind diejenigen für eindimensionale Regressoren X_1, \dots, X_n . Analog lassen sich auch Kernschätzer mit $X_1, \dots, X_n \in \mathbb{R}^d$, $d \geq 2$ definieren, die Bandbreiten sind in diesen Fällen Matrizen. Wir nutzen in dieser Arbeit nur die Schätzer für $d = 1$, die gesamte in dieser Arbeit vorgestellte Theorie wäre aber auch auf Kernschätzer höherer Dimension übertragbar.

Bemerkung 2.51 Die in die Kernschätzer eingehende Bandbreite $b := b_n$ ist wesentlich für ihr Verhalten. Kleine Bandbreiten liefern sehr stark oszillierende Schätzkurven, große Bandbreiten

liefern einen stark geglätteten Schätzer. Es gibt optimale, von n (dem zugrundeliegenden Stichprobenumfang) abhängige Raten für die Bandbreite b . Diese Bandbreitenraten sind aber abhängig von der jeweils zugrunde gelegten Definition der Optimalität. So kann man zum Beispiel eine exakte optimale, den $MISE(\hat{f}_X) := E[\int (\hat{f}_X(x) - E[\hat{f}_X(x)])^2 dx]$ minimierende Rate für die Bandbreite des Kerndichteschätzers $\hat{f}_X(x)$ angeben, diese hängt aber noch von der zweiten Ableitung der zu schätzenden Dichte f_X ab, was die direkte Nutzung meist unmöglich macht. Genutzt werden kann aber die Rate der Bandbreite im Bezug auf den Stichprobenumfang n . Ein Hypothesentest, dessen Statistik auf eben diesem Kerndichteschätzer aufbaut, wird i.A. aber eine optimale Bandbreite b von anderer Rate benötigen. Schon der Schätzer der Verteilungsfunktion von X , $\hat{F}_X(x)$, definiert als Integral über den Kerndichteschätzer liefert für die (den MISE minimierende) Bandbreite b eine andere Rate als es für $\hat{f}_X(x)$ der Fall ist.

Bemerkung 2.52 Die im folgenden angegebenen Aussagen gelten unter den Annahmen an die Bandbreite b , die Kernfunktion \tilde{K} , die Modellfunktion m , die Stichprobe und an die Verteilung von (X, Y) wie sie in dieser Arbeit unter (K), (H1) und (M1) und gegeben sind (vgl. dazu Kapitel 3, Modell und Annahmen), sie sind aber zum Teil auch unter schwächeren Annahmen gültig. Zusätzlich ist zu beachten, dass die in den folgenden beiden Propositionen gegebenen Aussagen nur für $x \in [b, 1 - b]$ gelten, um die Aussage auf $[0, 1]$ auszuweiten muss der Schätzer am Rand modifiziert werden, darauf werden wir aber in den Propositionen nicht näher eingehen, diese modifizierten Schätzer würden die theoretischen Betrachtungen stark verkomplizieren.

Proposition 2.53 Unter den Voraussetzungen (M1), (H1) und (K) gilt:

$$\sup_{x \in [0, 1]} |\hat{m}(x) - m(x)| = O\left(\sqrt{\frac{\log(b^{-1})}{nb}}\right) \quad f.s.$$

Zum Beweis siehe AvK (2001) Proposition 3.

Die Schwachstelle der meisten Kernschätzer ist die Schätzung der Funktionswerte an den Rändern des Definitionsbereiches der zu schätzenden Funktion, da dort zu wenig Beobachtungen vorliegen. Eine Verbesserung der Schätzung kann man durch Modifizieren der Schätzer im Randbereich erzielen, siehe bezüglich verschiedener Ansätze zum Beispiel Karunamuni und Alberts (2005).

Die folgende Proposition greift schon etwas vor, sie benutzt die im folgenden Kapitel angegebenen und grundlegenden Voraussetzungen, gehört aber inhaltlich in diesen Abschnitt. Es sei dazu $(\varepsilon, X) \sim F_{\varepsilon, X}$, $F_{\varepsilon|X}$ die dazugehörige bedingte Verteilungsfunktion von ε gegeben X und f_X die Dichte von X . Zusätzlich sei $\ddot{f}_{\varepsilon, X}(\varepsilon, x) := \frac{\partial^2 f_{\varepsilon, X}(\varepsilon, x)}{\partial^2 x}$ bzw. $f_X''(x) := \frac{\partial^2 f_X(x)}{\partial^2 x}$. $(\varepsilon_1, X_1), \dots, (\varepsilon_n, X_n)$ sei eine iid. Stichprobe von (ε, X) :

Proposition 2.54 *Im Modell (3.1) gilt unter den Voraussetzungen (M1), (M2), den Annahmen (K) an die Kernfunktion und den Annahmen (H1), (H2) an die Bandbreiten h und b :*

- (a) $E[\hat{f}_X(x)] - f_X(x) = \frac{b^2}{2!} \cdot \mu_2^{\tilde{K}} \cdot f_X''(x) + o(b^2)$
- (b) $E[\frac{1}{nh} \sum_{i=1}^n K(\frac{X_i - x}{h}) (I\{\varepsilon_i \leq y\} - F_{\varepsilon|X}(y|x))] = \frac{h^2}{2!} \cdot \mu_2^K \cdot (\int_{(-\infty, y]} \ddot{f}_{\varepsilon, X}(\varepsilon, x) d\varepsilon - F_{\varepsilon|X}(y|x) f_X''(x)) + o(h^2)$

Der Beweis zu (a) folgt direkt und ohne Schwierigkeiten aus der Berechnung des Erwartungswertes des Kerndichteschätzers. Der Beweis zu (b) folgt mittels Transformation, Taylorentwicklung:

$$\begin{aligned} & E[\frac{1}{nh} \sum_{i=1}^n K(\frac{X_i - x}{h}) (I\{\varepsilon_i \leq y\} - F_{\varepsilon|X}(y|x))] \\ &= \int \int K(u) I\{\varepsilon \leq y\} f_{\varepsilon, X}(\varepsilon, x + uh) d\varepsilon du \\ &\quad - \int \int K(u) F_{\varepsilon|X}(y|x) f_X(x + uh) d du \\ &= \frac{h^2}{2} \mu_2^K \left(\int_{(-\infty, y]} \ddot{f}_{\varepsilon, X}(\varepsilon, x) d\varepsilon - F_{\varepsilon|X}(y|x) f_X''(x) \right) + o(h^2) \end{aligned}$$

Die Abschätzung der Restterme benutzt dabei die in den Voraussetzungen (M1), (M2), (K), (H1), (H2) genannten Annahmen an die auftretenden Funktionen, insbesondere deren gleichmäßige Beschränktheit.

□

Proposition 2.55 *Unter den Voraussetzungen (M1), (H1), (K) gilt:*

- (a) $\sup_{x \in [0, 1]} |\hat{f}_X(x) - f_X(x)| = O(\sqrt{\frac{\log(b^{-1})}{nb}}) f.s.$
- (a*) $\hat{f}_X(x) - f_X(x) = O_p(b^2)$

$$(b) \frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) (I\{\varepsilon_i \leq y\} - F_{\varepsilon|X}(y|x)) = O_p\left(\frac{1}{\sqrt{nh}}\right)$$

Beweis: (a) siehe Neumeyer (2006), Seite 161.

Beweis von (a*) und (b): Man berechne die Varianz, dabei nutzte Substitution und Taylorentwicklung; im weiteren Verlauf folgen solche Rechnungen noch häufig.

□

2.3 Empirical Likelihood

Empirical-Likelihood ist die Bezeichnung einer nichtparametrischen statistischen Rückschlussmethode. Diese Methode wird im Weiteren vorgestellt. Sie benötigt im Gegensatz zu parametrischen Likelihood-Verfahren der Stochastik nicht die Annahme einer speziellen der den Daten zugrundeliegenden Klasse von Wahrscheinlichkeitsmaßen $\{P_\vartheta \mid \vartheta \in \Theta\}$. Diese parametrischen Modelle haben den Nachteil bei Mißspezifikation unbrauchbare Schätzer, Konfidenzintervalle etc. zu liefern. Die Empirical-Likelihood-Methode hat diesen Nachteil nicht. Es lassen sich mittels dieser Methode meist aber nur Schätzer, Konfidenzintervalle etc. mit asymptotischen Eigenschaften bestimmen, was in parametrischen Modellen zum Teil auch exakt zu festem Stichprobenumfang n möglich ist. Ebenso liefern die parametrischen Methoden i. A. bei (nahezu) passender Wahl des Modells $\{P_\vartheta \mid \vartheta \in \Theta\}$ bessere Schätzer.

Im Vergleich zu anderen nichtparametrischen Methoden der Statistik - wie zum Beispiel dem Bootstrap - kann die Empirical-Likelihood-Methode auch Nebeninformationen, die zur Verteilung der Daten gegeben sind, mit einbeziehen, ein Nachteil kann aber im konkreten Fall die aufwändige Optimierung der auftretenden Empirical-Likelihood-Funktion sein.

Grundlage der Empirical-Likelihood-Methode ist die empirische Verteilungsfunktion F_n , vgl. Definition 2.6.

Im parametrischen Modell sei (X_1, \dots, X_n) ein Zufallsvektor mit Dichte $h_\vartheta(x_1, \dots, x_n)$. Dann definiert

$$L(X_1, \dots, X_n, \vartheta) := f_{X_1, \dots, X_n, \vartheta}(x_1, \dots, x_n) = h_\vartheta(x_1, \dots, x_n),$$

gesehen als Funktion in ϑ , die Likelihood-Funktion. Im Falle, dass X_1, \dots, X_n unabhängig und identisch verteilt mit Zähl- oder Lebesgue-Dichte aus der Klasse $\{f_\vartheta \mid \vartheta \in \Theta\}$ sind, hat die Likelihood-Funktion die Gestalt

$$L(X_1, \dots, X_n, \vartheta) := f_{X_1, \dots, X_n, \vartheta}(x_1, \dots, x_n) = \prod_{i=1}^n f_\vartheta(x_i)$$

Analog definieren wir:

Definition 2.56 *Es seien X_1, \dots, X_n unabhängig und identisch verteilte Zufallsvariablen mit derselben Verteilungsfunktion wie X , $X \sim F$ und es sei $F(x-) := P(X < x)$.*

Dann heißt

$$(2.10) \quad L(F) := \prod_{i=1}^n (F(x_i) - F(x_{i-}))$$

nichtparametrische Likelihood-Funktion der Verteilungsfunktion F .

Die Struktur ist nahezu identisch mit der im parametrischen Fall, so ist hier im diskreten Fall

$$F_{\vartheta}(x_i) - F_{\vartheta}(x_{i-}) = P_{\vartheta}(X = x_i) = f_{\vartheta}(x_i).$$

Ist $X \sim F$, so ist $L(F)$ die Wahrscheinlichkeit mit unabhängigen Kopien von X genau x_1, \dots, x_n zu beobachten und $L(F) > 0$ ist äquivalent zu $P(X = x_i) \neq 0$ für alle i . Mit diesen beiden Definitionen lässt sich schon eine erste Proposition aufstellen:

Proposition 2.57 *Es seien X_1, \dots, X_n unabhängig und identisch verteilte Zufallsvariablen. Dann ist die empirische Verteilungsfunktion Maximalstelle von $L(F)$ maximiert über alle Verteilungsfunktionen F , das heißt die empirische Verteilungsfunktion ist der nichtparametrische Maximum-Likelihood-Schätzer (NPML) von F .*

Zum Beweis siehe Owen (2001), S. 8.

Ist ein allgemeiner Parameter ϑ der Verteilung F der zugrundeliegenden Daten von Interesse, welcher auf irgendeine Weise als Funktion der Verteilungsfunktion F darstellbar ist, also $\vartheta = T(F)$, zum Beispiel der Erwartungswert $\vartheta = \int x dF(x)$, so ist $\hat{\vartheta} := T(F_n)$ der nichtparametrische Maximum-Likelihood-Schätzer (NPML) von ϑ , und damit um im Beispiel zu bleiben für den Erwartungswert der NPML $\hat{\vartheta} = \int x dF_n(x) = \bar{X}_n$.

Ein anderes Gebiet, in dem die Empirical-Likelihood-Methode angewendet werden kann sind Konfidenzintervalle und Hypothesentests eines Parameters ϑ der den Daten zugrundeliegenden Verteilung. In der parametrischen Statistik wird dazu zum Beispiel der Likelihood-Quotient genutzt um (asymptotische) Tests der Form $H_0 : \vartheta = T(F_0) = \vartheta_0$, $H_1 : \vartheta \neq \vartheta_0$ zum Niveau α herzuleiten. Dieser Quotient wird definiert durch

$$R = \frac{L(X_1, \dots, X_n, \vartheta_0)}{\sup_{\vartheta \in \Theta} L(X_1, \dots, X_n, \vartheta)} = \frac{L(X_1, \dots, X_n, \vartheta_0)}{L(X_1, \dots, X_n, \hat{\vartheta})}$$

mit $\hat{\vartheta}$ der MLE für ϑ . Die Heuristik der Testkonstruktion ist die, dass kleine Werte des Quotienten gegen die Nullhypothese sprechen, große dafür. Eine Möglichkeit der Konstruktion eines asymptotischen Tests liefert das Theorem von Wilk (vgl. z.B. Rohatgi (2001)). Wilks Theorem besagt, dass unter bestimmten Voraussetzungen $-2 \log(R)$ in Verteilung unter H_0 gegen eine Chiquadrat-Verteilung mit Freiheitsgraden gleich der Dimension des Parameterraumes Θ konvergiert.

Ebenso lässt sich zum Testproblem $H_0 : F = F_0, H_1 : F \neq F_0$ der nichtparametrische Likelihood-Quotient definieren durch

$$R(F_0) = \frac{L(F_0)}{L(F_n)}$$

Die Analogie zum parametrischen Likelihood-Quotienten wird leicht deutlich: Ist zum Beispiel ein Parameter ϑ einer Klasse von Verteilungsfunktionen \mathcal{F} von Interesse, welcher sich über die Verteilungsfunktion darstellen lässt, d.h. $\vartheta = T(F)$, so betrachte die Statistik $R(\vartheta) := \sup\{R(F) \mid T(F) = \vartheta, F \in \mathcal{F}\}$. Definiert man $\vartheta_0 := T(F_0)$ und betrachtet man damit die Hypothese $H_0 : \vartheta = \vartheta_0$, so verwerfe man $H_0 : \vartheta = \vartheta_0$, falls $R(\vartheta_0) < \tau_0$. Es lässt sich in vielen Fällen ein Analogon zu Wilks Theorem aufstellen, mit dem τ_0 so zu bestimmen ist, dass sich ein asymptotisches Niveau von α ergibt. Zum Beispiel im Falle des Erwartungswertes μ mit $T(F) = \mu$. Die Statistik $R(\mu_0)$ ist unter $H_0 : \mu = \mu_0$ und weiteren Voraussetzungen ebenfalls asymptotisch Chiquadrat-verteilt mit Freiheitsgraden gleich der Dimension des Parameterraumes Θ , vgl. Owen Kapitel 11.2.

Es lassen sich auch in beiden Fällen (parametrisch und nichtparametrischer Likelihood-Quotient) unter geeigneten Voraussetzungen die asymptotischen Verteilungen unter H_1 berechnen (nicht-zentrierte Chiquadratverteilungen), was im konkreten Fall zum Vergleich der Power des jeweiligen Tests hilfreich ist.

Diese Aussage zur asymptotischen Verteilung des nichtparametrischen Empirical-Likelihood-Quotienten kann noch erweitert werden auf Empirical Likelihood-Schätzer, welche sich unter Einbeziehung von Nebenbedingungen ergeben. Dazu wollen wir diese zuerst vorstellen:

Einbeziehen von Zusatzinformationen

Diese Arbeit betrachtet einen Empirical-Likelihood-Schätzer einer Verteilungsfunktion, welcher Nebeninformationen ausnutzt.

Wir haben im letzten Abschnitt gesehen, dass die empirische Verteilungsfunktion F_n der NPMLE der Verteilungsfunktion F im Modell X_1, \dots, X_n unabhängig und identisch verteilt mit Vertei-

lungsfunktion F ist. Dabei lag keine einbezogene Zusatzinformation vor. F_n ist als NPMLE unter Nebeninformationen i.A. nicht zu erwarten. Allgemein hat der Schätzer der Verteilungsfunktion, welcher die Likelihoodfunktion (2.10) maximiert, die Form

$$\tilde{F}_n(x) := \sum_{i=1}^n p_i I\{X_i \leq x\},$$

dabei sind die Gewichte notwendigerweise positiv und addieren sich zu 1 um der Definition der Verteilungsfunktion zu genügen. Masse auf andere Punkte x als die Beobachteten x_1, \dots, x_n würde den Wert der Likelihood-Funktion verringern.

Es sei eine Nebeninformation in Form einer Funktion g der Zufallsvariablen X von der Art

$$E[g(X)] = \int g(x) dF(x) = 0$$

gegeben. Das kann zum Beispiel bekanntes zweites Moment sein, d.h. $g(X) = X^2 - \tau^2$, weitere Beispiele folgen später.

Sei nun $F_n^*(x) = \sum_{i=1}^n p_i^* I\{X_i \leq x\}$ der Empirical-Likelihood-Schätzer in diesem Modell mit Nebeninformationen. Der Empirical-Likelihood-Schätzer für $E[g(X)]$ ist dann

$$\int g(x) d\hat{F}_n^*(x) = \sum_{i=1}^n p_i^* g(X_i)$$

und somit erhält man wegen $E[g(X)] = 0$ als zu erfüllende Nebenbedingung, dass

$$0 = \sum_{i=1}^n p_i^* g(X_i)$$

gelten muss. Jetzt ist also zur Bestimmung des Empirical-Likelihood-Schätzers die Funktion

$$L(\tilde{F}_n) = L((p_1, \dots, p_n)) = \prod_{i=1}^n p_i$$

unter den Nebenbedingungen

$$\sum_{i=1}^n p_i = 1 \quad \text{und} \quad \sum_{i=1}^n p_i g(X_i) = 0$$

über die Menge $p = (p_1, \dots, p_n) \in [0, 1]^n$ zu maximieren, was je nach der Gestalt von g recht

aufwändig sein kann, da zuvor auch noch geklärt werden muss, ob überhaupt ein eindeutiges Maximum existiert.

Für den nichtparametrischen Likelihood-Quotienten unter Einbezug von Nebeninformationen $R(\vartheta) := \sup\{R(F) \mid T(F) = \vartheta, E_F[g(X)] = 0, F \in \mathcal{F}\}$ kann ebenfalls eine asymptotische Chiquadrat-Verteilung bewiesen werden, vgl. Owen (2001), Theorem 3.6. Dieses Theorem stellt eine große Anzahl von Voraussetzungen an die die Nebeninformation beschreibende Funktion g , speziell Stetigkeits- und Differenzierbarkeitsannahmen, so dass so wichtige Nebeninformationen wie $g(X) = I\{X \leq \vartheta_1\} - \vartheta_2$, d.h. Informationen über Quantile etc. nicht von diesem Theorem abgedeckt werden. Für asymptotische Theorie des Likelihood-Quotienten unter Nebenbedingungen dieser Art siehe Molanez-Lopez, van Keilegom und Veraverbeke (2009).

2.4 Bezug zu anderen Arbeiten

Die Fehlerverteilung der Fehler im Regressionsmodell $y = m(X) + \varepsilon$ ist in der nichtparametrischen Statistik schon seit längerem von Interesse, speziell das homoskedastische Modell, so zum Beispiel in den Arbeiten von Loynes (1980), Buckley, Eagleson & Silverman (1988) und Seifert, Gasser & Wolf (1993). Auf einige Arbeiten wollen wir noch genauer eingehen:

Stute (1986)

In dieser Arbeit wird unter anderem die schwache Konvergenz des (geeignet standardisierten) Nadaraya-Watson-Schätzers der bedingten Verteilungsfunktion $\hat{F}_{Y|X}(t|x)$ gegeben $(X_1, Y_1), \dots, (X_n, Y_n)$ als Prozess in t und x fest bewiesen.

Hall, Wolff & Yao (1999)

Diese Arbeit betrachtet verschiedene Schätzer für die bedingte Verteilungsfunktion $F_{Y|X}$ auf Basis einer unabhängig und identisch verteilten Stichprobe $(Y_1, X_1), \dots, (Y_n, X_n)$. Im Vergleich zur vorliegenden Arbeit ist die erste Variable beobachtbar, was der Fehler ε in unserem Modell nicht ist, er wird geschätzt, so dass also eine Stichprobe der Art $(\hat{\varepsilon}_1, X_1), \dots, (\hat{\varepsilon}_n, X_n)$ vorliegt, welche nicht mehr aus unabhängigen Beobachtungen besteht.

Speziell die Struktur des genutzten (modifizierten) Nadaraya-Watson-Schätzers übertragen wir auf die Schätzer der vorliegenden Arbeit (vgl. Definition 4.1). Auch stellen Hall, Wolff & Yao (1999) einen Empirical-Likelihood-Schätzer für $F_{Y|X}(y|x)$ vor (vgl. den Empirical-Likelihood-Schätzer dieser Arbeit, Definition 4.6), ohne diesen genauer über die Empirical-Likelihood-Methode zu motivieren. Hall, Wolff & Yao (1999) zeigen für festes x und y für die oben genannten und weitere Schätzer unter einigen technischen Annahmen, die asymptotische Normalität ihrer (geeignet standardisierten) Schätzer $\hat{F}_{Y|X}(y|x)$ und geben jeweils Varianz und Bias an.

Nagel (2006) und Kiwitt, Nagel & Neumeier (2009)

Die Arbeit von Nagel (2006) (und zusammenfassend die Arbeit von Kiwitt, Nagel & Neumeier (2009)) betrachtet, wie schon in der Einleitung erwähnt, ein ähnliches Regressionsmodell wie es dieser Arbeit zugrunde liegt. Im Mittelpunkt steht die Verteilung der Fehler ε im Modell $Y = m(X) + \varepsilon$. Der grundlegende Unterschied zu dieser Arbeit ist die vorausgesetzte Unabhängig-

keit von X und ε . Es wird ein Empirical-Likelihood-Schätzer entwickelt, welcher Nebeninformation der Form $E[g(\varepsilon)] = 0$ miteinbeziehen kann und mit einem herkömmlichen Kernschätzer der Verteilungsfunktion der Fehler verglichen. Zusätzlich wird die Kovarianzstruktur des hergeleiteten Empirical-Likelihood-Schätzers $\hat{F}_{EL}(\cdot)$ berechnet und mit der des herkömmlichen Schätzers verglichen. Es zeigt sich, dass keiner der Schätzer gleichmäßig besser ist. In Simulationen und in der Theorie werden Varianz, Erwartungswert und MSE für verschiedenste Fehler-Verteilungen und Nebeninformationen verglichen. In vielen Fällen, besonders für "gängige" Fehlerverteilungen, z.B. im Falle normalverteilter Fehler, liefert der Empirical-Likelihood-Schätzer bessere Resultate. Auf diese Simulationsergebnisse und Beispiele wird die vorliegende Arbeit in Simulationen zu den neuen Schätzern dieser Arbeit noch Stellung nehmen.

Die Arbeit von Nagel (2006) rückt mittels eines etwas allgemeineren Modelles $Y = m(X) + \varepsilon$ mit $\varepsilon = \sigma(X)\tilde{\varepsilon}$, X und $\tilde{\varepsilon}$ unabhängig von der Unabhängigkeit des Regressors X und des Fehlers ε ab, was schon etwas allgemeiner ist und auch häufiger genutzt wird. Es sind damit aber auch weitere Annahmen, speziell an die Funktion σ nötig und es stellt nicht die allgemeinste mögliche Form der Abhängigkeit dar. Die vorliegende Arbeit verzichtet von vornherein auf die Unabhängigkeit von X und ε , die Beweise von Nagel sind darum nur in einigen Teilen nutzbar. Viele der nötigen Entwicklungen bei Nagel, insbesondere die Abschätzungen nutzen gezielt die Produktdarstellung der gemeinsamen Dichte von X und ε , $f_{X,\varepsilon} = f_X \cdot f_\varepsilon$ und die damit verbundene Möglichkeit der Trennung der Integrale, was in dieser Arbeit nicht möglich ist und die Entwicklungen deutlich umständlicher macht.

Im letzten Abschnitt der Arbeit von Nagel wird auf Testverfahren auf Basis der Schätzers der Verteilungsfunktion der Fehler ε eingegangen und es werden verschiedene Verfahren vorgestellt. An diese Struktur wird sich die vorliegende Arbeit auch halten und ggf. an späterer Stelle Bezug zu den Ergebnissen bei Nagel (2006) nehmen.

Qin & Lawless (1994)

Qin und Lawless betrachten in ihrer Arbeit einen Verteilungsfunktionsschätzer für beobachtbare identisch verteilte Zufallsvariablen X_1, \dots, X_n . Ihr Schätzer ist ein Empirical Likelihood-Schätzer unter Nebenbedingung $E[g(\varepsilon)|X] = 0$ und sie zeigen für diesen asymptotische Normalität. Die Arbeit von Nagel (2006) baut in Grundzügen darauf auf.

Akritis & van Keilegom (2001)

Die Arbeit von Akritis & van Keilegom (2001) betrachtet das Regressionsmodell $Y = m(X) + \sigma(X)\varepsilon$, konstruiert einen nichtparametrischen Schätzer für die Verteilungsfunktion der Residuen und beweist schwache Konvergenz des Schätzers. Sie ist eine grundlegende Arbeit für Nagel (2006). Viele der Entwicklungen können aber, wegen der im vorherigen Abschnitt beschriebenen Problematik im Bezug auf die in dem Modell dieser Arbeit nicht mehr vorausgesetzte Unabhängigkeit von X und ε , nicht für diese Arbeit genutzt werden. Auf genutzte Aussagen und Beweisideen wird an passender Stelle hingewiesen.

Einmahl & McKeague (2003)

Einmahl & McKeague (2003) konstruieren Tests für nichtparametrische Hypothesen auf Basis der Empirical Likelihood Methode, unter anderem Tests auf Symmetrie, Wechsel der Verteilungsfunktion und Unabhängigkeit (der Komponenten einer bivariaten Zufallsvariable). Grundlage der Tests ist der Likelihood-Quotient R , vergleiche dazu den vorherigen Abschnitt. Die Autoren standardisieren den Likelihood-Quotient, jeweils für jedes ihrer betrachteten Testprobleme passend, und geben die Asymptotik der Statistik an. Auf ähnliche Weise werden die Teststatistiken dieser Arbeit konstruiert.

3 Modell und Annahmen

3.1 Modell und grundsätzliche Annahmen

Wie in der Einleitung dargestellt betrachten wir in dieser Arbeit ein nichtparametrisches Regressionsmodell der Art

$$(3.1) \quad Y = m(X) + \varepsilon,$$

Dabei sind X, ε Zufallsvariablen, die nicht als unabhängig vorausgesetzt sind. Gegeben seien n Beobachtungen $(X_1, Y_1), \dots, (X_n, Y_n)$ iid. $\stackrel{D}{=} (X, Y)$. Die dazugehörigen Fehler $\varepsilon_1, \dots, \varepsilon_n$ sind nicht gegeben, m ist unbekannt. Das Interesse dieser Arbeit liegt in der Schätzung der Verteilung der Fehler ε gegeben $X = x$. Weitere Annahmen sind:

(M1) Es gelte das Modell (3.1). (ε, X) sei verteilt mit Dichte $f_{\varepsilon, X}$ und Verteilungsfunktion $F_{\varepsilon, X}$. f_X bezeichne die Dichte von X und habe kompakten Träger $D := [0, 1]$, F_X sei die dazugehörige Verteilungsfunktion. Die bedingte Dichte von ε gegeben X werde im Weiteren mit $f_{\varepsilon|X}$, die dazugehörige Verteilungsfunktion mit $F_{\varepsilon|X}$ bezeichnet. Es existiere ein $r_0 > 0$, mit $E[|\varepsilon|^{2+r_0}] < \infty$.

m sei zweimal stetig differenzierbar, ebenso $f_{\varepsilon, X}(\varepsilon, x)$, $f_{\varepsilon|X}(\varepsilon|x)$ und $f_X(x)$ in x , es gelte $f_X(x) > 0$ für alle $x \in D$ und $E[\varepsilon|X] = 0$.

Zusätzlich sei noch $E[\varepsilon^2|X = x] =: \sigma^2(x)$ zweimal stetig differenzierbar in $x \in (0, 1)$.

Diese Ableitung sei gleichmäßig in x beschränkt.

Es sei ebenfalls noch angenommen, dass $B(x) := (m \cdot f_X)''(x) - m(x) f_X''(x)$ für alle $x \in (0, 1)$ Lipschitzstetig mit Lipschitzkonstante L ist.

(M2) Annahmen an die Ableitungen:

Für alle $x \in D$ existieren die Ableitungen $\dot{f}_{\varepsilon|X}(y|x) := \frac{\partial f_{\varepsilon|X}(y|x)}{\partial x}$, $\ddot{f}_{\varepsilon|X}(y|x) := \frac{\partial^2 f_{\varepsilon|X}(y|x)}{\partial x^2}$ und $f'_{\varepsilon|X}(y|x) := \frac{\partial f_{\varepsilon|X}(y|x)}{\partial y}$ und es existiere ein $0 < c < \infty$, so dass

$$(i) \quad \sup_{x \in D, y \in \mathbb{R}} |f_{\varepsilon|X}(y|x)| \leq c$$

$$(ii) \quad \sup_{x \in D, y \in \mathbb{R}} |\dot{f}_{\varepsilon|X}(y|x)| \leq c$$

$$(iii) \quad \sup_{x \in D, y \in \mathbb{R}} |\ddot{f}_{\varepsilon|X}(y|x)| \leq c$$

$$(iv) \quad \sup_{x \in D, y \in \mathbb{R}} |f'_{\varepsilon|X}(y|x)| := \sup_{x \in D, y \in \mathbb{R}} \left| \frac{\partial f_{\varepsilon|X}(y|x)}{\partial y} \right| \leq c$$

Diese, wie auch viele der folgenden Annahmen sind nötig um Restterme von passender Rate abzuschätzen.

3.2 Annahmen an die dem Schätzer zugrundeliegende Stichprobe

Die Fehler $\varepsilon_1, \dots, \varepsilon_n$ im Modell (3.1) sind aus der Stichprobe $(X_1, Y_1), \dots, (X_n, Y_n)$ nicht direkt beobachtbar. Um die bedingte Verteilung $F_{\varepsilon|X}$ der unbeobachtbaren Fehler zu schätzen nutzen wir deshalb nichtparametrisch geschätzte Residuen

$$\hat{\varepsilon}_i = Y_i - \hat{m}(X_i), \quad i = 1, \dots, n,$$

wobei \hat{m} den Nadaraya-Watson Schätzer für die Modellfunktion m bezeichne (Nadaraya (1964), Watson (1964), vgl. Definition 2.49 und Definition 2.48).

An die obigen Schätzer sind noch die folgenden Annahmen nötig:

- (K) Die Kernfunktion \tilde{K} sei zweimal stetig differenzierbar, nichtnegativ, symmetrisch mit Träger $[-1, 1]$, von beschränkter Totalvariation und es gelte $\tilde{K}(-1) = \tilde{K}(1) = 0$ und $\int \tilde{K}(u) du = 1$. Im Weiteren setze $\mu_j^{\tilde{K}} := \int u^j \tilde{K}(u) du$ und $\mu_j^{\tilde{K}^2} := \int u^j \tilde{K}^2(u) du$.
- (H1) $b := b_n$ sei eine Folge von Bandbreiten und es existiere ein $\alpha > 0$, so dass für $n \rightarrow \infty$ gilt: $nb^5 = O(1)$ und $nb^{3+2\alpha}(\log b^{-1})^{-1} \rightarrow \infty$.

Im weiteren bezeichnen \tilde{K} und K zwei (ggf.) unterschiedliche, die Voraussetzungen (K) erfüllende Kerne.

Bemerkung 3.1 Die Bandbreitenannahmen (H1) dieser Arbeit sind so angepasst, dass für die Differenz $\hat{m} - m$ gilt, dass $P(\hat{m} - m \in \mathcal{H}) \rightarrow 1$ für $n \rightarrow \infty$, dabei ist $\mathcal{H} := C_\delta^{1+\alpha}[0, 1]$ (vgl. Definition 2.34 und für α bzw. δ die Annahmen (H1) und (G2)). Siehe dazu und zu den Bandbreitenannahmen auch Nagel (2006), S. 17 bzw. Akritas und van Keilegom (2001).

- (H2) $h = h_n$ sei Folge von Bandbreiten mit $nh^5 = O(1)$ und $(nh)^{1/2}(\log h^{-1})^{-1/2} \rightarrow \infty$ für $n \rightarrow \infty$. Zusätzlich gelte $\lim_{n \rightarrow \infty} \frac{b}{h} = \lambda$, $\lambda \in [0, \infty)$ $\sqrt{\frac{h \log(b^{-1})^{(\beta+1)/2}}{(nb)^{\beta/2}}} \rightarrow 0$, dabei sei $\beta > 0$ wie in (G1) vorausgesetzt, $\frac{\log(b^{-1})}{\sqrt{nhb}} \rightarrow 0$ und es existiere ein $\gamma > 0$, so dass $\frac{\log(n)^{2+\gamma}}{h^3 n} \rightarrow 0$.

Diese Bandbreite wird im weiteren Verlauf der Arbeit zur Definition des Kernschätzers bzw. des Empirical-Likelihood-Schätzers für $F_{\varepsilon|X}(\varepsilon|x)$ (vgl. Definition 4.1 bzw. 4.6) benötigt.

3.3 Annahmen für den Einbezug der die Fehlerverteilung betreffenden Zusatzinformation

Im Weiteren nehmen wir an, dass zusätzliche Information zur Fehlerverteilung gegeben ist:

(A) $E[g(\varepsilon, X) | X = x] = \int g(y, x) f_{\varepsilon|X}(y|x) dy = 0$ für alle $x \in D$, wobei $g = (g_1, \dots, g_k)^T : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R}^k$ bekannt. Es sei für jedes $x \in D$ die Matrix $\Sigma(x) = f_X(x) \mu_0^{K^2} E[g(\varepsilon, X)g^t(\varepsilon, X) | X = x]$ positiv definit. Zusätzlich sei noch $E[g(\varepsilon, X) \cdot \varepsilon | X = x]$ und $E[g(\varepsilon, X) | X = x]$ zweimal stetig differenzierbar in $x \in (0, 1)$ und diese Ableitung sei beschränkt.

(M3) g_j sei stetig differenzierbar in der ersten Komponente (setze $g'_j(y, x) = \frac{\partial g_j(y, x)}{\partial y}$, analog $g''_j(y, x)$) und es gelten für $j = 1, \dots, k$ die folgenden Existenz-Annahmen an Integrale:

- (i) $\sup_{z \in [0, 1]} \int |g'_j(y, x)| f_{\varepsilon, X}(y, z) dy < \infty,$
- (ii) $\sup_{z \in [0, 1]} \int (g'_j(y, x))^2 |\dot{f}_{\varepsilon, X}(y, z)| dy < \infty,$
- (iii) $\sup_{z \in [0, 1]} \int |g'_j(y, x)| |\ddot{f}_{\varepsilon, X}(y, z)| dy < \infty.$
- (iv) $\sup_{z \in [0, 1]} \int |g_j(y, x)| |\dot{f}_{\varepsilon, X}(y, z)| dy < \infty,$
- (v) $\sup_{z \in [0, 1]} \int g_j^2(y, x) |\ddot{f}_{\varepsilon, X}(y, z)| dy < \infty,$
- (vi) $\sup_{x \in [0, 1]} \left| \frac{\partial E[g''_j(\varepsilon, X) | X=x]}{\partial x} \right| < \infty$
- (vii) $\sup_{x \in [0, 1]} \int (g'_j(y, x))^2 f_{\varepsilon, X}(y, x) dy < \infty,$ und $\sup_{x \in [0, 1]} \left| \frac{\partial E[g'_j(\varepsilon, X) | X=x]}{\partial x} \right| < \infty$
- (viii) Es existiere ein $m > 0$, so dass für $j = 1, \dots, k$ gilt:

$$E\left[\frac{1}{h} \left(K\left(\frac{X-x}{h}\right) g_j(\varepsilon, x)\right)^{m+2}\right] = O(1).$$

Aus (A) folgt auch, dass $E[g(\varepsilon, X)] = 0$ (it. Erwartung).

Eindimensionale Beispiele für Funktionen g ohne X -Abhängigkeit sind bekannte Varianz, d.h. $g(\varepsilon, x) = g(\varepsilon) = \varepsilon^2 - \sigma^2$, andere Momente wie Schiefe etc. oder auch bekannte Quantile, d.h. $g(\varepsilon, x) = g(\varepsilon) = I\{\varepsilon \leq a\} - d$. Für diese ist (M3) nicht erfüllt, aber man kann die Theorie

auch für Nebeninformationen von dieser Art durchführen. Mehrdimensionale Nebeninformationen können durch Kombination der eindimensionalen Nebeninformationen erstellt werden. Die Quantilsfunktionen werden im Weiteren immer gesondert behandelt, da sie nicht um ε in einer Taylorreihe entwickelt werden können etc.

Für $g(\varepsilon, x) = I\{\varepsilon \leq a(x)\} - d(x)$ kann das zum Beispiel für $d(x) = \frac{1}{2}$ eine durch die X -Abhängigkeit erzeugte Verschiebung des Medians bedeuten. In den später folgenden Beispielen ist meist entweder $d(x)$ oder $a(x)$ konstant. Ebenso sind X -abhängige Momente gut interpretierbar, speziell eine von X abhängige Varianz, woraus eine Nebeninformation der Art $g(\varepsilon, x) = \varepsilon^2 - \sigma^2(x)$ resultiert.

Bemerkung 3.2 *Es ist ein noch allgemeineres Modell mit X -Abhängigkeit möglich, welches Parameter-Abhängigkeit miteinbezieht, d.h. $g(\varepsilon, x) = \varepsilon^k - c_{\vartheta}(x)$, $\vartheta \in \Theta$ oder auch $g(\varepsilon, x) = I\{\varepsilon \leq a(x)\} - d_{\vartheta}(x)$, $\vartheta \in \Theta$ etc. So könnten zum Beispiel auch geschätzte Nebeninformationen einbezogen werden.*

*Insbesondere Nebeninformationen der Form $g(\varepsilon, x) = \varepsilon^2 - \sigma^2(x)$ mit $\sigma(x) = \sigma_{\vartheta}(x)$, so dass also nur noch der Parameter ϑ unbekannt ist können von Interesse sein. Mit recht allgemeinen Voraussetzungen kann man Entwicklungen für Parameterschätzer $\hat{\vartheta}_n$ für ϑ angeben (vgl. dazu z.B. *M-Schätzer, van der Vaart (1998), S. 68*) und mit weiteren (Differenzierbarkeits-)Annahmen an g dann $g_{\hat{\vartheta}_n}(\varepsilon, x) = g_{\vartheta}(\varepsilon, x) + \hat{R}_n$ entwickeln. Ist $g_{\vartheta}(\varepsilon, x)$ um ϑ in einer Taylorreihe entwickelbar und gilt für den Term der ersten Ableitung $\hat{R}_n = g'_{\varepsilon}(\varepsilon, x) \cdot (\hat{\vartheta}_n - \vartheta)$ die Rate $o_p(\frac{1}{\sqrt{nh}})$ so erhält man dieselben Resultate wie sie in dieser Arbeit für bekannten Parameter ϑ , d.h. für bekannte Funktionen g , hergeleitet werden. Im Falle $\hat{R}_n = O_p(\frac{1}{\sqrt{nh}})$ wäre der Term der führenden Rate der Differenz mit in Entwicklungen und die Aussage zur schwachen Konvergenz einzubeziehen, was prinzipiell möglich ist, aber deutlich umfangreichere Entwicklungen nötig machen würde. Die Entwicklungen unseres Schätzers der bedingten Verteilungsfunktion der Residuen gegeben X unter Einbezug von Nebeninformationen werden im folgenden aber nur für das schon sehr komplexe Modell mit vollständig bekannten Nebeninformationen durchgeführt. Die Simulationen betrachten auch ein Modell, welches geschätzte Nebeninformationen nutzt. In folgenden betrachten wir dazu noch ein Beispiel:*

Beispiel 3.3 *Es gelte das Modell dieser Arbeit mit*

$$(3.2) \quad \text{Var}(\varepsilon|X) = E[\varepsilon^2 | X] = \sigma^2(X) = \alpha_0 + \alpha_1 \cdot X, \quad \alpha_0, \alpha_1 \in \mathbb{R}$$

zusätzlich existieren die vierten Momente von ε . Zur Schätzung von $\vartheta = (\alpha_0, \alpha_1)^t$ nutzen wir $(\hat{\varepsilon}_1^2, X_1), \dots, (\hat{\varepsilon}_n^2, X_n)$, da $(\varepsilon_1^2, X_1), \dots, (\varepsilon_n^2, X_n)$ nicht beobachtbar ist. Setze $\hat{\varepsilon}^2 := (\hat{\varepsilon}_1^2, \dots, \hat{\varepsilon}_n^2)^t$, $\tilde{\varepsilon}^2 := (\varepsilon_1^2, \dots, \varepsilon_n^2)^t$ und \tilde{X} eine $(n \times 2)$ -Matrix mit $\tilde{X}_{i,1} = 1$ und $\tilde{X}_{i,2} = X_i$ für $i = 1, \dots, n$. Als Schätzer für ϑ nutzen wir den Kleinst-Quadrate-Schätzer-Ansatz linearer Modelle:

$$(\hat{\alpha}_0, \hat{\alpha}_1)^t = \hat{\vartheta} = (\tilde{X}^t \tilde{X})^{-1} \tilde{X}^t \hat{\varepsilon}^2,$$

und

$$\begin{aligned} \hat{\alpha}_0 &= \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i^2 - \hat{\alpha}_1 \frac{1}{n} \sum_{i=1}^n X_i =: \hat{\varepsilon}^2_n - \hat{\alpha}_1 \bar{X}_n \\ \hat{\alpha}_1 &= \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)(\hat{\varepsilon}_i^2 - \hat{\varepsilon}^2_n)}{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2} = \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n) \hat{\varepsilon}_i^2}{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2}. \end{aligned}$$

Es gilt:

$$\begin{aligned} \hat{\vartheta} &= (\tilde{X}^t \tilde{X})^{-1} \tilde{X}^t \hat{\varepsilon}^2 \\ &= (\tilde{X}^t \tilde{X})^{-1} \tilde{X}^t (\hat{\varepsilon}^2 - E[\hat{\varepsilon}^2 | \tilde{X}]) + (\tilde{X}^t \tilde{X})^{-1} \tilde{X}^t E[\hat{\varepsilon}^2 | \tilde{X}] \\ &= O_p\left(\frac{1}{\sqrt{n}}\right) + \vartheta \\ &= \vartheta + o_p\left(\frac{1}{\sqrt{nh}}\right) \end{aligned}$$

Der zweite Summand der zweiten Zeile liefert direkt ϑ , denn $E[\hat{\varepsilon}^2 | \tilde{X}] = \tilde{X} \vartheta$, vgl. Gleichung (3.2). Für den ersten Teil ist zu begründen, warum die Bandbreite b sich nicht in der Rate niederschlägt. Lemma 3.10 aus Nagel (2006) liefert für Summanden der Art $\frac{1}{n} \sum_{i=1}^n g(\hat{\varepsilon}_i)$ mit $E[g(\varepsilon_i) | X_i] = 0$ die Rate $O_p\left(\frac{1}{\sqrt{n}}\right)$. Beweise die Rate $O_p\left(\frac{1}{\sqrt{n}}\right)$ mit ähnlichen Techniken wie das Lemma 3.10 aus Nagel (2006), was in diesem speziellen Fall sehr aufwändig ist.

Nach diesem Beispiel wollen wir noch einige weitere Voraussetzungen zusammentragen:

An die Komponenten g_j , $j = 1, \dots, k$ von g werden noch die folgenden Annahmen gestellt:

- (G1)** Es existieren Konstanten γ, C und $\beta > 0$, so dass für $j = 1, \dots, k$ und für jedes $z \in C_\delta^{1+\alpha}[0, 1]$

$$\left| \int \frac{1}{h} K\left(\frac{u-x}{h}\right) (g_j(y+z(u), x) - g_j(y, x) - z(u)g'_j(y)) f_{\varepsilon, X}(y, u) d(y, u) \right|$$

$$\leq C \cdot \int \frac{1}{h} K\left(\frac{u-x}{h}\right) \cdot |z(u)|^{1+\beta} f_X(u) du,$$

wobei α wie in Annahme (H1), δ wie in Annahme (G2).

(G2) Es existieren Konstanten δ, C , positives $\kappa < 2(1 + \alpha)$ und $r_1 > 0$, so dass für $j = 1, \dots, k$ gilt

$$\sup_{u \in [0,1]} \left(E \left[\sup_{\substack{z, \tilde{z} \in \mathbb{R}: |z| \leq \delta, \\ |\tilde{z}| \leq \delta, |z - \tilde{z}| \leq \xi}} (g_j(\varepsilon + z, x) - g_j(\varepsilon + \tilde{z}, x))^{2+r_1} \middle| X = u \right] \right)^{\frac{1}{2+r_1}} \leq C \xi^{1/\kappa},$$

wobei α in Annahme (H1) definiert wurde. Zusätzlich gelte

$$\sup_{u \in [0,1]} \left(E \left[\sup_{z, \tilde{z} \in \mathbb{R}: |z| \leq \delta} (g_j(\varepsilon + z, x) - g_j(\varepsilon, x))^4 \middle| X = u \right] \right) < \infty.$$

Beispiel 3.4 Nebeninformationen $g(\varepsilon, x) = \varepsilon^k - c(x)$, welche aus Annahmen an die Momente folgen, erfüllen (G1) und (G2) falls $\sup_{x \in [0,1]} E[\varepsilon_1^{4k} | X = x] < \infty$. Das gleiche gilt für Polynome die Abhängigkeiten zwischen Momenten darstellen.

Für ein g_j , welches (G2) erfüllt oder für g_j eine Indikatorfunktion derart, dass $g_j(\varepsilon, x) = I\{\varepsilon \leq a(x)\} - d(x)$ und (ε, X) mit Dichte $f_{\varepsilon, X}$, so gilt:

$$(3.3) \quad \exists \delta > 0 \text{ so dass } \sup_{u \in [0,1]} E \left[\sup_{y \in \mathbb{R}: |y| \leq \delta} (g_j(\varepsilon + y, x) - g_j(\varepsilon, x))^2 \middle| X = u \right] < \infty$$

Dies nutzen wir im Weiteren, weisen dann aber explizit darauf hin.

Bemerkung 3.5 Die Annahmen an g und $f_{\varepsilon, X}$ sind für viele interessante Beispiele erfüllt, zum Beispiel für $\{f_{\varepsilon|X}(y|x)\}$ die Klasse Normalverteilungen, wobei die (gleichzeitig die Nebeninformation darstellende) x -Abhängigkeit der Fehler ε gegeben durch die Varianz der Fehler $E[\varepsilon^2 | X = x] = \sigma^2(x)$ mit $\sigma^2 : [0, 1] \rightarrow \mathbb{R}$ stetig und beschränkt. Weitere Beispiele finden sich im Kapitel Simulationen.

Nach den meist technischen Annahmen wollen wir nun im nächsten Abschnitt die in dieser Arbeit betrachteten Schätzer vorstellen.

4 Schätzansatz

Dieser Abschnitt stellt die im Weiteren nötigen Schätzer, respektive den im Mittelpunkt dieser Arbeit stehenden Empirical-Likelihood-Schätzer der bedingten Verteilungsfunktion der Residuen ε gegeben den Regressor X , $F_{\varepsilon|X}(y|x)$, vor. Genutzt werden jeweils die Annahmen und Bezeichnungen des vorherigen Kapitels. Wir beginnen mit der Motivation der grundlegenden Schätzer dieser Arbeit.

4.1 Empirical-Likelihood-Schätzer

Im Weiteren betrachte als Auswahlklasse für die Schätzer dieser Arbeit die folgende Klasse $\mathcal{F}_{\varepsilon|X}$ von Verteilungsfunktionen $F_{\varepsilon|X}$:

$$(4.1) \quad \mathcal{F}_{\varepsilon|X} := \left\{ F_n^*(y|x) : F_n^*(y|x) := \sum_{i=1}^n \frac{\frac{1}{h} K_h(X_i - x) p_i}{\tilde{f}_X(x)} \cdot I\{\varepsilon_i \leq y\}, \right. \\ \left. \tilde{f}_X(x) = \sum_{j=1}^n \frac{1}{h} K_h(X_j - x) p_j, p_i > 0, i = 1, \dots, n, \sum_{i=1}^n p_i = 1 \right\}.$$

Ziel der Empirical-Likelihood-Methode ist es nun aus obiger Klasse von nichtparametrischen Verteilungsfunktionen diejenige auszuwählen, welche die Wahrscheinlichkeit des Auftretens der Stichprobe $(\varepsilon_1, X_1), \dots, (\varepsilon_n, X_n)$ maximiert:

$$\begin{aligned} \max_{F \in \mathcal{F}_{\varepsilon|X}} \prod_{i=1}^n P_F((\varepsilon, X) = (\varepsilon_i, X_i)) &= \max_{F \in \mathcal{F}_{\varepsilon|X}} \prod_{i=1}^n P_F(\varepsilon = \varepsilon_i | X = X_i) \cdot P(X = X_i) \\ &= \max_{F \in \mathcal{F}_{\varepsilon|X}} \prod_{i=1}^n (F(\varepsilon_i | X_i) - F(\varepsilon_i - | X_i)) \cdot P(X = X_i) \end{aligned}$$

Nach den Voraussetzungen dieser Arbeit ist X als stetig mit Dichte f_X angenommen.

In obiger Gleichung ersetzen wir deshalb $P(X = x)$ durch den für $f_X(x)$ asymptotisch erwartungstreuen, konsistenten Schätzer $\tilde{f}_X(x) = \sum_{j=1}^n \frac{1}{h} K_h(X_j - x) p_j$ und maximieren damit:

$$\prod_{i=1}^n \frac{1}{nh} K_h(X_i - X_i) p_i = \left(\frac{1}{nh} K_h(0) \right)^n \prod_{i=1}^n p_i$$

über p_1, \dots, p_n .

Um die im weiteren deutlich werdende Abhängigkeit der Gewichte von x zu verdeutlichen schreiben wir im weiteren $p_i(x)$ anstelle von p_i . Liegen keine Nebeninformationen vor so ist das Produkt maximal für $p_i = \frac{1}{n}$, $i = 1, \dots, n$. Dies führt zu unserem ersten Schätzansatz:

Definition 4.1 Sei K eine nichtnegative Kernfunktion und h eine Bandbreite. Dann ist die empirische Verteilungsfunktion von ε gegeben $X = x$ auf Basis der geschätzten Residuen $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n$ definiert als

$$\hat{F}_n(y|x) := \sum_{i=1}^n \frac{\frac{1}{nh} K_h(X_i - x)}{\sum_{j=1}^n \frac{1}{nh} K_h(X_j - x)} \cdot I\{\hat{\varepsilon}_i \leq y\}.$$

Wir verwenden hierbei die Notation $K_h(\cdot) := K(\frac{\cdot}{h})$.

Zur Unterscheidung wird in dieser Arbeit obiger Schätzer als Kernschätzer(-ansatz) (KE) bezeichnet.

Nicht aus den Daten bestimmbar, da $\varepsilon_1, \dots, \varepsilon_n$ nicht direkt beobachtbar, aber für die Entwicklung nötig ist:

Definition 4.2 Sei K eine nichtnegative Kernfunktion und h eine Bandbreite. Dann ist die empirische Verteilungsfunktion von ε gegeben $X = x$ auf Basis der (unbeobachtbaren) original Residuen $\varepsilon_1, \dots, \varepsilon_n$ definiert als

$$F_n(y|x) := \sum_{i=1}^n \frac{\frac{1}{nh} K_h(X_i - x)}{\sum_{j=1}^n \frac{1}{nh} K_h(X_j - x)} \cdot I\{\varepsilon_i \leq y\}.$$

Die in den obigen Definitionen auftretende Bandbreite h ist eine andere als die Bandbreite b des Nadaraya-Watson-Schätzers \hat{m} der Modellfunktion m , vorgestellt im vorherigen Kapitel. Gleiches gelte für den genutzten Kern, dieser erfülle aber ebenfalls die Voraussetzungen (K). h erfülle die Bandbreitenannahme (H2).

4.2 Empirical-Likelihood-Schätzer unter Nebenbedingungen

Im Weiteren wollen wir nun einen weiteren Schätzer der bedingten Verteilungsfunktion $F_{\varepsilon|X}$ herleiten. Wir wählen als Empirical-Likelihood-Schätzer denjenigen aus obiger Klasse (4.1), welcher das Produkt $\prod_{i=1}^n p_i(x)$ maximiert unter den Nebenbedingungen

$$(4.2) \quad \sum_{i=1}^n p_i(x) = 1 \quad \text{und} \quad \sum_{i=1}^n p_i(x) g(\hat{\varepsilon}_i, x) K_h(X_i - x) = 0_{k \times 1}.$$

Mittels der ersten Bedingung werden die p_i eindeutig festgelegt, eine andere Konstante als 1 ist aber ebenso möglich, denn die Gewichte $\frac{\frac{1}{h} K_h(X_i - x) p_i(x)}{\sum_{j=1}^n \frac{1}{h} K_h(X_j - x) p_j(x)}$ summieren sich unabhängig davon über alle i nach Konstruktion zu 1, was nötig ist, damit der Schätzer der Definition der Verteilungsfunktion genügt. Die letztere der beiden Bedingungen ist sinnvoll, denn nach Voraussetzung gilt

$$E[g(\varepsilon, X) | X = x] = \int g(y, x) F_{\varepsilon|X}(dy|x) = \int g(y, x) f_{\varepsilon|X}(y|x) dy = 0_{k \times 1}$$

und damit fordern wir für den Schätzer

$$0_{k \times 1} = \int g(y, x) \hat{F}_n^*(dy|x) = \frac{1}{nh} \sum_{i=1}^n \frac{p_i(x) g(\hat{\varepsilon}_i, x) K\left(\frac{X_i - x}{h}\right)}{\frac{1}{nh} \sum_{j=1}^n K\left(\frac{X_j - x}{h}\right) p_j(x)},$$

also den zweiten Teil der Nebenbedingung (4.2). Im Weiteren werde die Gültigkeit der folgenden Gleichung

$$(4.3) \quad \min_{1 \leq i \leq n} g_j(\hat{\varepsilon}_i, x) \frac{1}{h} K\left(\frac{X_i - x}{h}\right) < 0 < \max_{1 \leq i \leq n} g_j(\hat{\varepsilon}_i, x) \frac{1}{h} K\left(\frac{X_i - x}{h}\right)$$

und die Positiv-Definitheit von

$$\frac{1}{n} \sum_{i=1}^n \frac{1}{h} K^2\left(\frac{X_i - x}{h}\right) g(\hat{\varepsilon}_i, x) g^t(\hat{\varepsilon}_i, x).$$

angenommen. Diese ist zusätzlich nötig um die Existenz einer eindeutigen Lösung zu garantieren. Die technischen Voraussetzungen für die asymptotische Gültigkeit der Ungleichung in

Wahrscheinlichkeit und ebenso die asymptotische Positiv-Definitheit in Wahrscheinlichkeit werden in Nagel (2006), Abschnitt 3.5, speziell Lemma 3.18, S. 67, bewiesen und sind auch unter den Voraussetzungen dieser Arbeit gültig. Beides ist nötig für die folgende Proposition:

Proposition 4.3 *Existenz und Eindeutigkeit der Empirical-Likelihood-Lösung: $\prod_{i=1}^n p_i(x)$ besitzt für jedes $x \in (0, 1)$ in $B_n = \{(p_1, \dots, p_n)(x) \in [0, 1]^n \mid \sum_{i=1}^n p_i(x) = 1, \sum_{i=1}^n p_i(x) g(\hat{\varepsilon}_i, x) K_h(X_i - x) = 0\}$ unter der Bedingung (4.3) genau eine Lösung $(p_1(x), \dots, p_n(x))$.*

Beweis:

Mit Gleichung (4.3) können wir nun zum Beweis der Existenz und Eindeutigkeit den Beweis von Lemma 2.1 aus Nagel (2006) nutzen. Dort betrachtet man als Funktion $g_j(\hat{\varepsilon}_i)$. Der Beweis kann auf Funktionen derart $g(\hat{\varepsilon}_i, x) K_h(X_i - x)$ wie sie in unserem Fall vorliegen übertragen werden.

□

Proposition 4.4 *Der Schätzer $\hat{p}_i(x)$, $i = 1, \dots, n$ welcher $\prod_{i=1}^n p_i(x)$ unter den Nebenbedingungen $\sum_{i=1}^n p_i(x) = 1$, $\sum_{i=1}^n p_i(x) g(\hat{\varepsilon}_i, x) K_h(X_i - x) = 0_{k \times 1}$ maximiert ist:*

$$(4.4) \quad p_i(x) = \hat{p}_i(x) := \frac{1}{n} \frac{1}{1 + \hat{\eta}_n(x)^t g(\hat{\varepsilon}_i, x) K_h(X_i - x)},$$

wobei $\hat{\eta}_n(x)$ definiert ist als die Lösung der Gleichung

$$(4.5) \quad \sum_{i=1}^n p_i(x) g(\hat{\varepsilon}_i, x) K_h(X_i - x) = \sum_{i=1}^n \frac{g(\hat{\varepsilon}_i, x) K_h(X_i - x)}{1 + \hat{\eta}_n(x)^t g(\hat{\varepsilon}_i, x) K_h(X_i - x)} = 0_{k \times 1}.$$

Beweis:

Zur Bestimmung der Maximalstelle benutzen wir den Lagrange-Multiplikatorenansatz, welcher besagt, dass die Extrema einer Funktion $f(p)$ gegeben die \mathbb{R}^q -wertige Nebenbedingung $\psi(p) = 0$ nur unter denjenigen Werten p zu suchen ist, für welche Lagrange-Multiplikatoren $\alpha_1, \dots, \alpha_q$ existieren, so dass $\frac{\partial L(p)}{\partial p} = \nabla f(p) + \sum_{i=1}^q \alpha_i \nabla \psi_i(p) = 0$.

In unserem Fall gibt es genau $q = 1 + k$ Nebenbedingungen, wobei wir die erste noch in die passende Form umstellen müssen. Gesucht wird die Maximalstelle $p := p(x) = (p_1(x), \dots, p_n(x)) \in \mathbb{R}^n$ von $f(p)$ unter Nebenbedingungen. Die Funktion f sowie die Nebenbedingungen sind reellwertig. Zur Vereinfachung der weiteren Rechnungen logarithmieren wir f noch, was wegen der strikten

Monotonie keine Auswirkungen auf die Position der Extremstellen hat. Somit hat die Lagrange-Funktion die Gestalt:

$$L(p_1(x), \dots, p_n(x)) = \sum_{i=1}^n \log(p_i(x)) - \beta^t \sum_{i=1}^n p_i(x) g(\hat{\varepsilon}_i, x) K_h(X_i - x) - \lambda \left(\sum_{i=1}^n p_i(x) - 1 \right),$$

mit λ reellwertig, $\beta \mathbb{R}^k$ -wertig. Im Weiteren ist $\nabla L(p(x)) = \left(\frac{\partial L(p(x))}{\partial p_1(x)}, \dots, \frac{\partial L(p(x))}{\partial p_n(x)} \right)^t = 0_{n \times 1}$ zu lösen, woraus im Weiteren für alle $i = 1, \dots, n$ folgt:

$$(4.6) \quad \frac{\partial L(p(x))}{\partial p_i(x)} = \frac{1}{p_i(x)} - \beta^t g(\hat{\varepsilon}_i, x) K_h(X_i - x) - \lambda = 0$$

Wir multiplizieren die Gleichung mit $p_i(x)$ und summieren über alle i :

$$0 = \sum_{i=1}^n p_i(x) \cdot \frac{\partial L(p(x))}{\partial p_i(x)} = n - \lambda \Rightarrow \lambda = n$$

Einsetzen in (4.6) liefert

$$(4.7) \quad 0 = \frac{1}{p_i(x)} - \beta^t g(\hat{\varepsilon}_i, x) K_h(X_i - x) - n$$

was äquivalent ist zu:

$$(4.8) \quad p_i(x) = \hat{p}_i(x) = \frac{1}{n + \beta^t g(\hat{\varepsilon}_i, x) K_h(X_i - x)} = \frac{1}{n} \cdot \frac{1}{1 + \hat{\eta}_n(x)^t g(\hat{\varepsilon}_i, x) K_h(X_i - x)}$$

Schreibe $\hat{\eta}_n(x)$ anstelle von $\frac{\beta}{n}$ um die später in der Entwicklung deutlich werdende Abhängigkeit η 's von $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n$ und x aufzuzeigen.

□

Jetzt beweisen wir noch eine Proposition zur Existenz und Eindeutigkeit der Lösung. Da $p_i(x) \in (0, 1)$, folgt aus (4.8):

$$(4.9) \quad 1 + \hat{\eta}_n(x)^t g(\hat{\varepsilon}_i, x) K_h(X_i - x) > \frac{1}{n},$$

also wird eine Lösung in der Menge

$$(4.10) \quad D_k^x := \{\hat{\eta}_n(x) \in \mathbb{R}^k : 1 + \hat{\eta}_n(x)^t g(\hat{\varepsilon}_i, x) K_h(X_i - x) > \frac{1}{n} \text{ für } 1 \leq i \leq n\}$$

gesucht. Wir benutzen diese Menge (in der die gesuchte Lösung liegt) und zeigen die folgende Proposition:

Proposition 4.5 *Unter der Annahme (4.3) und der Annahme der Positiv-Definitheit von $\frac{1}{n} \sum_{i=1}^n \frac{1}{h} K^2(\frac{X_i-x}{h}) g(\hat{\varepsilon}_i, x) g^t(\hat{\varepsilon}_i, x)$ existiert für jedes $x \in (0, 1)$ genau eine Lösung $\hat{\eta}_n(x)$ von Gleichung (4.5) in D_k^x .*

Beweis:

Unter der Annahme der Gültigkeit der Gleichung (4.3) können wir zum Beweis der Existenz und Eindeutigkeit vorgehen wie im Beweis Lemma 2.2 aus Nagel (2006).

□

Der in dieser Arbeit grundlegende Schätzer kann nun definiert werden:

Definition 4.6 *Der Empirical-Likelihood-Schätzer (ELE) der Verteilung der Residuen gegeben $X = x$ auf Basis der geschätzten Fehler $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n$ ist gegeben durch:*

$$\hat{F}_n^*(y|x) := \sum_{i=1}^n \frac{\frac{1}{nh} K_h(X_i - x) \hat{p}_i(x)}{\frac{1}{nh} \sum_{j=1}^n K_h(X_j - x) \hat{p}_j(x)} \cdot I\{\hat{\varepsilon}_i \leq y\},$$

dabei bezeichne K eine nichtnegative Kernfunktion, welche (K) erfüllt und h eine Bandbreite, welche die Bedingung (H2) erfüllt.

Bemerkung 4.7 *Es ist zu bemerken, dass die Wahl der Raten der Bandbreiten b und h in (H1) und (H2) nicht kanonisch ist, sondern angepasst ist an Wechselwirkungen und das Ziel dieser*

Arbeit, der Nachweis der schwachen Konvergenz der oben vorgestellten Schätzer. Dies ist jetzt noch nicht ersichtlich, wird später aber während der Entwicklung deutlich. Der Übersicht halber werden wir die in den einzelnen Abschnitten gegebenen Wechselwirkungen von h und b und die Vereinbarkeit der genutzten Relationen mit den Voraussetzungen (H1) und (H2) in Bemerkungen zum Schluss der Abschnitte zusammenfassen. Es beschreibe im Weiteren λ das Verhältnis zwischen b und h , d.h. $\lambda := \lim_{n \rightarrow \infty} \frac{b}{h}$. Es zeigt sich, dass durch die Wahl $\lambda = 0$ Bias-Terme entfallen und sich die Varianzstruktur vereinfacht. Um den Zusammenhang von b und h besser verfolgen zu können wird ein auftretendes λ in den noch folgenden Rechnungen nicht durch Null ersetzt, dies geschieht erst ganz zu Ende der Rechnungen, so dass die Gründe der Wahl der Raten der Bandbreiten deutlicher zu erkennen sind. Erfüllt im Weiteren eine Bandbreite (H2) so schreiben wir h für diese Bandbreite, ist (H1) erfüllt, so schreiben wir b .

Bemerkung 4.8 Die hier auftretende Bandbreite h ist (wie schon zuvor angemerkt) eine andere als die Bandbreite b zum Nadaraya-Watson-Schätzer \hat{m} , vorgestellt im vorherigen Kapitel. Vergleiche dazu auch (H1) und (H2), bzw. Bemerkung 4.7.

Für diesen Schätzer wollen wir im den weiteren Kapiteln die Differenz $\hat{F}_n^*(y|x) - F_{\varepsilon|X}(y|x)$ entwickeln und $\hat{F}_n^*(y|x)$ asymptotisch mit dem Schätzer aus Definition 4.1, $\hat{F}_n(y|x)$, welcher ebenfalls zu entwickeln ist, vergleichen.

4.3 Entwicklung von Bestandteilen des Empirical-Likelihood Schätzers

Das bei der weiteren Untersuchung des Schätzers $\hat{F}_n^*(y|x)$ maßgebliche Problem ist das Auftreten von $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n$ als Grundlage der Schätzung, es können also nicht ohne weiteres Erwartungswert etc. dieses Schätzers berechnet werden. Sukzessive ist der Schätzer $\hat{F}_n^*(y|x)$ zu entwickeln; wir beginnen dabei mit der Entwicklung von $\hat{\eta}_n^t(x)$, welches als Bauteil im Weiteren häufiger auftritt. Das nachstehende Lemma liefert zentrale Aussagen zu $\hat{\eta}_n^t(x)$ und dessen Bauteilen, so dass darauf eine Entwicklung für $\hat{\eta}_n^t(x)$ als Summe unabhängiger Zufallsvariablen zzgl. eines Resttermes von geeigneter Rate hergeleitet werden kann. Diese Entwicklungen werden dann im nächsten Kapitel Bauteile der weiteren Entwicklung des Schätzers $\hat{F}_n^*(y|x)$ im Hinblick auf die Berechnung des Erwartungswerts und der Varianz sein. Der Abschnitt mündet in der Entwicklung von $\hat{\eta}_n$ in einer Summe unabhängiger Zufallsvariablen zzgl. eines Resttermes von geeigneter Rate, siehe Proposition 4.12.

Proposition 4.9 *Im Modell (3.1) gilt unter den Voraussetzungen (M1), (M2), (K), (A), (H1), (H2), und (G1), (G2), (M3) bzw. g von der Form $g(\varepsilon, x) = I\{\varepsilon \leq a(x)\} - d(x)$:*

$$(i) \max_{1 \leq i \leq n} |g_j(\hat{\varepsilon}_i, x)K_h(X_i - x)| = o_p(\sqrt{nh}) \quad (j = 1, \dots, k)$$

$$(ii) \hat{\eta}_n(x) = O_p\left(\frac{1}{\sqrt{nh}}\right)$$

$$(iii) \max_{1 \leq i \leq n} \left| \frac{1}{1 + \hat{\eta}_n^t(x)g(\hat{\varepsilon}_i, x)K_h(X_i - x)} \right| = O_p(1)$$

$$(iv) \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K_h^2(X_i - x) (g_j(\hat{\varepsilon}_i, x) - g_j(\varepsilon_i, x))^2 = o_p(1) \text{ f\"ur } j = 1, \dots, k$$

(v)

$$\begin{aligned} \frac{1}{nh} \sum_{i=1}^n g(\hat{\varepsilon}_i, x)g^t(\hat{\varepsilon}_i, x)K_h^2(X_i - x) &= f_X(x) E[g(\varepsilon_1, X_1)g^t(\varepsilon_1, X_1) | X_1 = x] \mu_0^{K^2} + o_p(1) \\ &=: \Sigma(x) + o_p(1) \end{aligned}$$

Beweise zu (i) - (v), beginnend mit (i):

Es gilt

$$\max_{1 \leq i \leq n} |g_j(\hat{\varepsilon}_i, x)K_h(X_i - x)| \leq \max_{1 \leq i \leq n} |g_j(\varepsilon_i, x)K_h(X_i - x)| + \max_{1 \leq i \leq n} |(g_j(\hat{\varepsilon}_i, x) - g_j(\varepsilon_i, x))K_h(X_i - x)|$$

Für den ersten Summanden gilt:

$$\begin{aligned} P\left(\max_{1 \leq i \leq n} |g_j(\varepsilon_i, x)K_h(X_i - x)| > \rho\sqrt{nh}\right) &\leq n E[I\{|g_j(\varepsilon, x)K_h(X - x)| > \rho\sqrt{nh}\}] \\ &\leq \frac{1}{\rho^2} E[g_j^2(\varepsilon, x)\frac{1}{h}K_h^2(X - x) I\{|g_j(\varepsilon, x)K_h(X - x)| > \rho\sqrt{nh}\}] \\ &= o(1) \end{aligned}$$

Die letzte Gleichheit folgt, da

$$\begin{aligned} E[g_j^2(\varepsilon, x)\frac{1}{h}K_h^2(X - x)] &= \int E[g_j^2(\varepsilon, x) | X = z]\frac{1}{h}K_h^2(z - x) f_X(z) dz \\ &= \int E[g_j^2(\varepsilon, x) | X = x + uh]K^2(u) f_X(x + uh) du \\ &= E[g_j^2(\varepsilon, X) | X = x] f_X(x) \mu_0^{K^2} + o(1) < \infty, \end{aligned}$$

wobei die Substitution $u = \frac{z-x}{h}$ und die Annahme (A) genutzt wird. Für den zweiten Summanden gilt:

$$\begin{aligned} & P\left(\left|\max_{1 \leq i \leq n} |(g_j(\hat{\varepsilon}_i, x) - g_j(\varepsilon_i, x)) K_h(X_i - x)| > \rho\sqrt{nh}\right.\right) \\ & \leq P\left(\left|\max_{1 \leq i \leq n} |(g_j(\hat{\varepsilon}_i, x) - g_j(\varepsilon_i, x)) K_h(X_i - x)| > \rho\sqrt{nh}, \max_{1 \leq i \leq n} |\hat{\varepsilon}_i - \varepsilon_i| \leq \delta\right.\right) \\ & \quad + P\left(\max_{1 \leq i \leq n} |\hat{\varepsilon}_i - \varepsilon_i| > \delta\right) \end{aligned}$$

Die zweite Wahrscheinlichkeit konvergiert für alle $\delta > 0$ gegen 0, denn mit Proposition 2.53 (i) (Eigenschaften der Regressionsfunktion \hat{m}) gilt:

$$\max_{1 \leq i \leq n} |\hat{\varepsilon}_i - \varepsilon_i| \leq \sup_{z \in [0,1]} |\hat{m}(z) - m(z)| = o_p(1)$$

Für den anderen Teil folgt mit $h_j(z, u, x) := \sup_{y \in \mathbb{R} |y| \leq \delta} |(g_j(z+y, x) - g_j(z, x))K(\frac{u-x}{h})|$:

$$\begin{aligned} & P\left(\left|\max_{1 \leq i \leq n} |(g_j(\hat{\varepsilon}_i, x) - g_j(\varepsilon_i, x)) K_h(X_i - x)| > \rho\sqrt{nh}, \max_{1 \leq i \leq n} |\hat{\varepsilon}_i - \varepsilon_i| \leq \delta\right.\right) \\ & \leq P\left(\left|\max_{1 \leq i \leq n} |h_j(\varepsilon_i, X_i, x)| > \rho\sqrt{nh}\right.\right) \\ & \leq n E[I\{|h_j(\varepsilon_1, X_1, x)| > \rho\sqrt{nh}\}] \\ & \leq \frac{1}{\rho^2 h} E[h_j^2(\varepsilon_1, X_1, x) I\{|h_j(\varepsilon_1, X_1, x)| > \rho\sqrt{nh}\}] \\ & \stackrel{(G2)}{=} o(1), \end{aligned}$$

dabei ergibt sich die Rate mit einer ähnlichen Rechnung wie zum vorletzten Summanden, substituiere dazu, so dass noch $\frac{1}{h}$ entfällt und nutze dabei die Folgerung (3.3) mit dem dazugehörigen δ .

Beweis zu (ii):

Nach Gleichung (4.5) und $\frac{1}{1+x} = 1 - \frac{x}{1+x}$ gilt:

$$\begin{aligned} 0 &= \left\| \frac{1}{n} \sum_{i=1}^n \frac{g(\hat{\varepsilon}_i, x) K_h(X_i - x)}{1 + \hat{\eta}_n(x)^t g(\hat{\varepsilon}_i, x) K_h(X_i - x)} \right\| \\ &= \left\| \frac{1}{n} \sum_{i=1}^n g(\hat{\varepsilon}_i, x) K_h(X_i - x) - \frac{1}{n} \sum_{i=1}^n \hat{\eta}_n(x)^t g(\hat{\varepsilon}_i, x) K_h(X_i - x) \frac{g(\hat{\varepsilon}_i, x) K_h(X_i - x)}{1 + \hat{\eta}_n(x)^t g(\hat{\varepsilon}_i, x) K_h(X_i - x)} \right\| \end{aligned}$$

$$\geq \left\| \frac{1}{n} \sum_{i=1}^n \hat{\eta}_n(x)^t g(\hat{\varepsilon}_i, x) K_h(X_i - x) \frac{g(\hat{\varepsilon}_i, x) K_h(X_i - x)}{1 + \hat{\eta}_n(x)^t g(\hat{\varepsilon}_i, x) K_h(X_i - x)} \right\| - \left\| \frac{1}{n} \sum_{i=1}^n g(\hat{\varepsilon}_i, x) K_h(X_i - x) \right\|$$

Es sei nun $\tilde{\eta}_n^t(x) := \frac{\hat{\eta}_n(x)^t}{\|\hat{\eta}_n(x)^t\|}$. Dann folgt weiter:

$$\begin{aligned} 0 &\geq \|\tilde{\eta}_n^t(x)\| \cdot \left\| \frac{1}{n} \sum_{i=1}^n \hat{\eta}_n(x)^t g(\hat{\varepsilon}_i, x) K_h(X_i - x) \frac{g(\hat{\varepsilon}_i, x) K_h(X_i - x)}{1 + \hat{\eta}_n(x)^t g(\hat{\varepsilon}_i, x) K_h(X_i - x)} \right\| - \left\| \frac{1}{n} \sum_{i=1}^n g(\hat{\varepsilon}_i, x) K_h(X_i - x) \right\| \\ &\stackrel{(1)}{\geq} \cdot \left\| \tilde{\eta}_n^t(x) \frac{1}{n} \sum_{i=1}^n \frac{g(\hat{\varepsilon}_i, x) g^t(\hat{\varepsilon}_i, x) K_h^2(X_i - x) \hat{\eta}_n(x)}{1 + \hat{\eta}_n(x)^t g(\hat{\varepsilon}_i, x) K_h(X_i - x)} \right\| - \left\| \frac{1}{n} \sum_{i=1}^n g(\hat{\varepsilon}_i, x) K_h(X_i - x) \right\| \\ &= \|\hat{\eta}_n(x)\| \cdot \left\| \frac{1}{n} \sum_{i=1}^n \frac{K_h^2(X_i - x) \tilde{\eta}_n^t(x) g(\hat{\varepsilon}_i, x) g^t(\hat{\varepsilon}_i, x) \tilde{\eta}_n(x)}{1 + \hat{\eta}_n(x)^t g(\hat{\varepsilon}_i, x) K_h(X_i - x)} \right\| - \left\| \frac{1}{n} \sum_{i=1}^n g(\hat{\varepsilon}_i, x) K_h(X_i - x) \right\| \\ &\stackrel{(2)}{\geq} \|\hat{\eta}_n(x)\| \cdot \left\| \frac{1}{n} \sum_{i=1}^n \frac{K_h^2(X_i - x) \tilde{\eta}_n^t(x) g(\hat{\varepsilon}_i, x) g^t(\hat{\varepsilon}_i, x) \tilde{\eta}_n(x)}{1 + \|\hat{\eta}_n(x)^t\| \sqrt{k} \max_{1 \leq i \leq n, 1 \leq j \leq k} |g_j(\hat{\varepsilon}_i, x) K_h(X_i - x)|} \right\| - \left\| \frac{1}{n} \sum_{i=1}^n g(\hat{\varepsilon}_i, x) K_h(X_i - x) \right\| \\ &= \|\hat{\eta}_n(x)\| \cdot \frac{1}{1 + \|\hat{\eta}_n(x)^t\| \sqrt{k} \max_{1 \leq i \leq n, 1 \leq j \leq k} |g_j(\hat{\varepsilon}_i, x) K_h(X_i - x)|} \\ &\quad \cdot \frac{1}{n} \sum_{i=1}^n \tilde{\eta}_n^t(x) K_h^2(X_i - x) g(\hat{\varepsilon}_i, x) g^t(\hat{\varepsilon}_i, x) \tilde{\eta}_n(x) - \left\| \frac{1}{n} \sum_{i=1}^n g(\hat{\varepsilon}_i, x) K_h(X_i - x) \right\| \end{aligned}$$

Dabei folgt (1) mittels der Cauchy-Schwarz-Ungleichung und (2) mittels der Voraussetzung $1 + \hat{\eta}_n(x)^t g(\hat{\varepsilon}_i, x) K_h(X_i - x) > 0$, vgl. Gleichung (4.9).

Umgeformt ergibt sich unter Ausnutzung von $\frac{1}{1+x} = 1 - \frac{x}{1+x}$, $x^t A x \geq \min_{\|x\|=1} x^t A x =: \lambda_{\min}(A) >$

0, (i) und $\left\| \frac{1}{\sqrt{nh}} \sum_{i=1}^n g(\hat{\varepsilon}_i, x) K_h(X_i - x) \right\| \leq O_p(1)$ (Proposition 4.11 + ZGWS):

$$\begin{aligned} \|\hat{\eta}_n(x)\| &\leq \left\| \frac{1}{nh} \sum_{i=1}^n g(\hat{\varepsilon}_i, x) K_h(X_i - x) \right\| \cdot \left[\frac{1}{nh} \sum_{i=1}^n K_h^2(X_i - x) \tilde{\eta}_n^t(x) g(\hat{\varepsilon}_i, x) g^t(\hat{\varepsilon}_i, x) \tilde{\eta}_n(x) \right. \\ &\quad \left. - \frac{1}{nh} \sum_{i=1}^n g(\hat{\varepsilon}_i, x) K_h(X_i - x) \sqrt{k} \max_{1 \leq i \leq n, 1 \leq j \leq k} |g_j(\hat{\varepsilon}_i, x) K_h(X_i - x)| \right]^{-1} \\ &\leq \left\| \frac{1}{nh} \sum_{i=1}^n g(\hat{\varepsilon}_i, x) K_h(X_i - x) \right\| (\lambda_{\min}(A) + o_p(1))^{-1} \\ &\leq O_p\left(\frac{1}{\sqrt{nh}}\right) \end{aligned}$$

Die Rate der letzten Zeile ergibt sich dabei wiederum mittels Proposition 4.11 und dem ZGWS.

Beweis zu (iii):

Wir werden im weiteren zeigen, dass: $\max_{1 \leq i \leq n} |1 - (1 + \hat{\eta}_n^t(x)g(\hat{\varepsilon}_i, x)K_h(X_i - x))^{-1}| = o_p(1)$.

Nach (i) und (ii) gilt $P(\|\hat{\eta}_n^t(x)\| \max_{1 \leq i \leq n, 1 \leq j \leq k} |g_j(\hat{\varepsilon}_i, x)K_h(X_i - x)| \geq 1) = o(1)$. Für $\rho > 0$ und $0 < z < 1$ gilt $\frac{z}{1-z} > \rho \Leftrightarrow z + \rho z > \rho \Leftrightarrow z > \frac{\rho}{1+\rho}$ und damit:

$$\begin{aligned}
& P\left(\max_{1 \leq i \leq n} \left|1 - \frac{1}{1 + \hat{\eta}_n^t(x)g(\hat{\varepsilon}_i, x)K_h(X_i - x)}\right| > \rho\right) \\
&= P\left(\max_{1 \leq i \leq n} \left|\frac{\hat{\eta}_n^t(x)g(\hat{\varepsilon}_i, x)K_h(X_i - x)}{1 + \hat{\eta}_n^t(x)g(\hat{\varepsilon}_i, x)K_h(X_i - x)}\right| > \rho, \|\hat{\eta}_n^t(x)\| \max_{1 \leq i \leq n, 1 \leq j \leq k} |g_j(\hat{\varepsilon}_i, x)K_h(X_i - x)| < 1\right) + o(1) \\
&\leq P\left(\frac{\|\hat{\eta}_n^t(x)\| \max_{1 \leq i \leq n, 1 \leq j \leq k} |g_j(\hat{\varepsilon}_i, x)K_h(X_i - x)|}{1 - \|\hat{\eta}_n^t(x)\| \max_{1 \leq i \leq n, 1 \leq j \leq k} |g_j(\hat{\varepsilon}_i, x)K_h(X_i - x)|} > \rho\right) + o(1) \\
&= P\left(\|\hat{\eta}_n^t(x)\| \max_{1 \leq i \leq n, 1 \leq j \leq k} |g_j(\hat{\varepsilon}_i, x)K_h(X_i - x)| \geq \frac{\rho}{1 + \rho}\right) + o(1) \\
&= o(1).
\end{aligned}$$

Damit folgt (iii). Die beiden noch ausstehenden Beweise nutzen keine der Aussagen (i) - (iii), wir beginnen mit dem Beweis zu (iv). Wir zeigen zuerst, dass:

$$\begin{aligned}
& \left| \frac{1}{n} \sum_{i=1}^n \frac{1}{h} \left((g_j(\varepsilon_i + (\hat{m} - m)(X_i), x) - g_j(\varepsilon_i, x)) K\left(\frac{X_i - x}{h}\right) \right)^2 \right. \\
& \quad \left. - \int \frac{1}{h} \left((g_j(\varepsilon + (\hat{m} - m)(z), x) - g_j(\varepsilon, x)) K\left(\frac{z - x}{h}\right) \right)^2 f_{\varepsilon, X}(\varepsilon, z) d(\varepsilon, z) \right| = o_p(1).
\end{aligned}$$

Betrachte die Funktionenklasse

$$\mathcal{G}_{j,n} := \left\{ \tilde{g}_{j,n} : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R}, \tilde{g}_{j,n}(\varepsilon, z) = \frac{1}{h} \left[g_j(\varepsilon + \tilde{h}(z), x) - g_j(\varepsilon, x) \right]^2 K^2\left(\frac{z - x}{h}\right) \mid \tilde{h} \in \mathcal{H} \right\},$$

mit $\mathcal{H} := C_\delta^{1+\alpha}[0, 1]$, siehe zu dieser Klasse Definition 2.34. Für diese zeigen wir im weiteren, dass gleichmäßig in n für jedes $\epsilon > 0$ gilt, dass $\mathcal{N}_{[]}(\epsilon, \mathcal{G}_{j,n}, L_1(\mathcal{P})) < \infty$. Die Einhüllende ist gegeben durch

$$\Phi_n(\varepsilon, z) := \Phi_{j,n}(\varepsilon, z) := \sup_{|y| \leq \delta} \frac{1}{h} \left[g_j(\varepsilon + y, x) - g_j(\varepsilon, x) \right]^2 K^2\left(\frac{z - x}{h}\right).$$

Für $(\hat{m} - m)$ gilt nach Akritas und van Keilegom (2001), dass $P(\hat{m} - m \in \mathcal{H}) \rightarrow 1$ für $n \rightarrow \infty$.
Es sei im Weiteren

$$\tilde{\mathcal{G}}_{j,n} := \left\{ \tilde{g}_{j,n} : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R}, \tilde{g}_{j,n}(\varepsilon, z) = \frac{1}{\sqrt{h}} \left(g_j(\varepsilon + \tilde{h}(z), x) - g_j(\varepsilon, x) \right) K\left(\frac{z-x}{h}\right) \mid \tilde{h} \in \mathcal{H} \right\},$$

die Einhüllende ist gegeben durch $\tilde{\Phi}_n(\varepsilon, z) := \sqrt{\Phi_n(\varepsilon, z)}$. Nach Proposition 6.7 ist $\tilde{\mathcal{G}}_{j,n} := \sqrt{\mathcal{G}_{j,n}}$ Donskerklasse, dabei ist die dortige Bezeichnung $\tilde{\mathcal{F}}_{1,n}$.

Dort wird gezeigt, dass gleichmäßig in n für jedes $\epsilon > 0$ für die Klammerungszahl gilt, dass $\mathcal{N}_{[\cdot]}(\epsilon \|\tilde{\Phi}_n\|_{2,\mathcal{P}}, \tilde{\mathcal{G}}_{j,n}, L_2(\mathcal{P})) < \infty$. Es folgt, dass für die Klammerungszahl von $\mathcal{G}_{j,n}$ gilt, dass $\sup_n \mathcal{N}_{[\cdot]}(\epsilon \|\Phi_n\|_{1,\mathcal{P}}, \mathcal{G}_{j,n}, L_1(\mathcal{P})) < \infty$, denn es ist $\mathcal{G}_{j,n} = \tilde{\mathcal{G}}_{j,n}^2$ und für $[\tilde{l}, \tilde{u}]$ eine $L_2(\mathcal{P})$ ϵ -Klammer setze $\tilde{l}_a^2(z) := \min(\tilde{l}^2(z), \tilde{u}^2(z)) \cdot I\{\tilde{u}(z) \geq 0, \tilde{l}(z) \geq 0 \text{ oder } \tilde{u}(z) \leq 0, \tilde{l}(z) \leq 0\}$ sonst identisch 0 und $\tilde{u}_a^2 := \max(\tilde{l}^2(z), \tilde{u}^2(z))$. $[\tilde{l}_a^2, \tilde{u}_a^2]$ ist eine $L_1(\mathcal{P})$ $\epsilon \cdot c$ -Klammer für $\mathcal{G}_{j,n}$, denn

$$\begin{aligned} E[\tilde{u}_a^2 - \tilde{l}_a^2] &\leq E[|\tilde{u}^2 - \tilde{l}^2|] \\ &= E[|\tilde{u} - \tilde{l}| |\tilde{u} + \tilde{l}|] \\ &\leq \sqrt{E[(\tilde{u} - \tilde{l})^2] E[(\tilde{u} + \tilde{l})^2]} \\ &\leq 2\|\tilde{\Phi}_n\|_{2,\mathcal{P}} \sqrt{E[(\tilde{u} - \tilde{l})^2]} \\ &< 2\sqrt{\|\Phi_n\|_{1,\mathcal{P}}} \cdot \epsilon \\ &=: c \cdot \epsilon, \end{aligned}$$

vorausgesetzt, dass $c_1 < \|\Phi_n\|_{1,\mathcal{P}} < c_2$ für alle n . Dabei ist noch zu klären, warum die erste Ungleichung gilt. Für den Fall, dass $\tilde{l}_a^2 = \min(\tilde{l}^2(z), \tilde{u}^2(z))$ gilt Gleichheit nach Konstruktion. Interessant sind nur Fälle mit $u(z) > 0$ und $l(z) < 0$, also $\tilde{l}_a^2 = 0$. In diesem Fall folgt obige Ungleichung aus der Dreiecksungleichung. Diese Klammern überdecken $\mathcal{G}_{j,n}$ nach Konstruktion. Nun ist zu klären, warum daraus die Glivenko-Cantelli Eigenschaft für die Klasse $\mathcal{G}_{j,n}$ folgt. Dazu betrachte Theorem 2.4.1 aus van der Vaart und Wellner (2000), S. 122. Dort wird gezeigt, dass Klassen, welche nicht von n abhängen und welche $\mathcal{N}(\epsilon, \mathcal{G}, L_1(\mathcal{P})) < \infty$ erfüllen Glivenko-Cantelli sind. Der Beweis ist nicht direkt auf von n abhängende Klassen zu übertragen. Trifft man aber die zusätzliche Voraussetzung, dass für alle $\tilde{g}_{j,n}$ aus $\tilde{\mathcal{G}}_{j,n}$ gilt $E[h \cdot (\tilde{g}_{j,n})^2] < \infty$ (d.h. Existenz der Varianz), so kann der Beweis übertragen werden. Dies ist in unserem Modell vorausgesetzt (vgl. (G2)) und damit folgt mit Theorem 2.4.1 aus van der Vaart und Wellner (2000), S. 122, die Glivenko-Cantelli Eigenschaft der Klasse $\mathcal{G}_{j,n}$.

Es gilt damit insgesamt:

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{h}} \left((g_j(\varepsilon_i + (m - \hat{m})(X_i), x) - g_j(\varepsilon_i, x)) K\left(\frac{X_i - x}{h}\right) \right)^2 \\ &= \int \frac{1}{\sqrt{h}} \left((g_j(\varepsilon + (m - \hat{m})(z), x) - g_j(\varepsilon, x)) K\left(\frac{z - x}{h}\right) \right)^2 f_{\varepsilon, X}(\varepsilon, z) d(\varepsilon, z) + o_p(\sqrt{h}). \end{aligned}$$

Nun ist für das jeweilige g_j zu zeigen, dass der Term auf der rechten Seite ebenfalls von der Rate $o_p(\sqrt{h})$ ist. Im Falle einer Funktion, die (G1) und (G2) erfüllt, gilt

$$(4.11) \quad \int \int (g_j(y + (m - \hat{m})(z), x) - g_j(y, x))^2 \frac{1}{h} K^2\left(\frac{z - x}{h}\right) f_{\varepsilon, X}(y, z) dz dy = o_p(1),$$

denn dies folgt mittels Taylorentwicklung, Substitution, der Tatsache, dass $(m - \hat{m})(x)$ gleichmäßig in x fast sicher gegen Null konvergiert, der Voraussetzung, dass $g_j^{\prime 2}$ stetig in der ersten Komponente und damit in einer Umgebung jedes festen y beschränkt ist und der Anwendung des Satzes zur majorisierenden Konvergenz (die Existenz der integrierbaren Majorante folgt dabei aus Gleichung (3.3)).

Betrachte hier den Fall $g_j(\varepsilon, x) = I\{\varepsilon \leq a(x)\} - d(x)$ Indikatorfunktion:

$$\begin{aligned} & \int \frac{1}{\sqrt{h}} \left((g_j(\varepsilon + (m - \hat{m})(z), x) - g_j(\varepsilon, x)) K\left(\frac{z - x}{h}\right) \right)^2 f_{\varepsilon, X}(\varepsilon, z) d(\varepsilon, z) \\ &= \int \frac{1}{\sqrt{h}} |F_{\varepsilon|X}(a(x) + (m - \hat{m})(z)|z) - F_{\varepsilon|X}(a(x)|z)| K^2\left(\frac{z - x}{h}\right) f_X(z) dz \\ &= O_p\left(\frac{\sqrt{h \log(b^{-1})}}{\sqrt{nb}}\right) = o_p(\sqrt{h}) \end{aligned}$$

Der letzte Schritt folgt mittels Taylorentwicklung, Substitution, Proposition 2.55 (i) zur Konvergenzrate von $\sup_{z \in [0, 1]} |m(z) - \hat{m}(z)|$ und den Eigenschaften der Bandbreiten h und b , dazu siehe (H1) und (H2).

Der Beweis zu (v) ist wie der Beweis zu (iv) zu führen. Die nötige Funktionenklassenfolge $(\mathcal{G}_{j,n})_{n \in \mathbb{N}}$, welche die Donsker-Eigenschaft erfüllt, ist hier gegeben durch $\mathcal{G}_{j,n} = \{\tilde{g}_{j,n} : \mathbb{R} \times [0, 1] \times [0, 1] \rightarrow \mathbb{R}, \tilde{g}_{j,n}(\varepsilon, z, x) = \frac{1}{\sqrt{h}} K\left(\frac{z-x}{h}\right) g_j(\varepsilon + \tilde{h}(z), x) \mid \tilde{h} \in \mathcal{H}\}$, siehe dazu den Beweis zu Lemma 6.7 (ii).

□

Jetzt kann mit dieser Proposition eine Entwicklung für $\hat{\eta}_n(x)$ hergeleitet werden:

Proposition 4.10 *Im Modell (3.1) unter den Annahmen (M1) – (M3), (K), (H1), (H2) und für Funktionen g , die (G1), (G2) erfüllen, gilt die Entwicklung*

$$(4.12) \quad \hat{\eta}_n(x) = \Sigma^{-1}(x) \frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) g(\hat{\varepsilon}_i, x) + o_p\left(\frac{1}{\sqrt{nh}}\right).$$

Dabei bezeichnet $\Sigma(x) = f_X(x) \mu_0^{K^2} E[g(\varepsilon, X)g^t(\varepsilon, X) | X = x]$ und ist nach Voraussetzung positiv definit. Zu diesem Term siehe auch Proposition 4.9.

Beweis:

Der Beweis der Aussage benötigt viele der Hilfsaussagen, die im vorherigen Hilfslemma zusammengestellt wurden. Dabei benutzen wir, dass $\frac{1}{1+x} = 1 - x + \frac{x^2}{1+x}$:

$\hat{\eta}_n(x)$ wurde definiert über die Gleichung

$$\begin{aligned} 0_{k \times 1} &= \frac{1}{n} \sum_{i=1}^n p_i(x) g(\hat{\varepsilon}_i, x) \frac{1}{h} K_h(X_i - x) \\ &= \frac{1}{n} \sum_{i=1}^n \frac{g(\hat{\varepsilon}_i, x) \frac{1}{h} K_h(X_i - x)}{1 + \hat{\eta}_n^t(x) g(\hat{\varepsilon}_i, x) K_h(X_i - x)} \\ &= \frac{1}{nh} \sum_{i=1}^n g(\hat{\varepsilon}_i, x) K_h(X_i - x) - \frac{1}{nh} \sum_{i=1}^n g(\hat{\varepsilon}_i, x) g^t(\hat{\varepsilon}_i, x) K_h^2(X_i - x) \hat{\eta}_n(x) \\ &\quad + \frac{1}{nh} \sum_{i=1}^n \frac{g(\hat{\varepsilon}_i, x) K_h(X_i - x) (\hat{\eta}_n^t(x) g(\hat{\varepsilon}_i, x) K_h(X_i - x))^2}{1 + \hat{\eta}_n^t(x) g(\hat{\varepsilon}_i, x) K_h(X_i - x)} \\ &=: D_1 + D_2 + D_3 \end{aligned}$$

Wir zeigen weiter unten mittels der Hilfsaussagen 4.9 (i) - (iv), dass $D_3 = o_p\left(\frac{1}{\sqrt{nh}}\right)$. Dann folgt mit 4.9 (v):

$$0_{k \times 1} = \frac{1}{nh} \sum_{i=1}^n g(\hat{\varepsilon}_i, x) K_h(X_i - x) - \frac{1}{nh} \sum_{i=1}^n g(\hat{\varepsilon}_i, x) g^t(\hat{\varepsilon}_i, x) K_h^2(X_i - x) \hat{\eta}_n(x) + o_p\left(\frac{1}{\sqrt{nh}}\right)$$

Stellt man diese Gleichung um, so folgt:

$$\begin{aligned} \frac{1}{nh} \sum_{i=1}^n g(\hat{\varepsilon}_i, x) K_h(X_i - x) &= \frac{1}{nh} \sum_{i=1}^n g(\hat{\varepsilon}_i, x) g^t(\hat{\varepsilon}_i, x) K_h^2(X_i - x) \hat{\eta}_n(x) \\ &\stackrel{(1)}{=} \Sigma(x) \hat{\eta}_n(x) + o_p\left(\frac{1}{\sqrt{nh}}\right), \end{aligned}$$

dabei ergibt sich (1) mittels Proposition 4.9 (v) + (ii) und somit

$$\hat{\eta}_n(x) = \Sigma^{-1}(x) \frac{1}{nh} \sum_{i=1}^n g(\hat{\varepsilon}_i, x) K_h(X_i - x) + o_p\left(\frac{1}{\sqrt{nh}}\right)$$

Zum Abschluss des Beweises fehlt noch die Abschätzung des Terms D_3 :

$$\begin{aligned} \|D_3\| &= \left\| \frac{1}{nh} \sum_{i=1}^n \frac{g(\hat{\varepsilon}_i, x) K_h(X_i - x) (\hat{\eta}_n^t(x) g(\hat{\varepsilon}_i, x) K_h(X_i - x))^2}{1 + \hat{\eta}_n^t(x) g(\hat{\varepsilon}_i, x) K_h(X_i - x)} \right\| \\ &\leq \|\hat{\eta}_n^t(x)\|^2 \frac{1}{nh} \sum_{i=1}^n \|g(\hat{\varepsilon}_i, x)\|^2 K_h^2(X_i - x) \cdot O_p(1) \\ &\stackrel{4.9}{=} O_p\left(\frac{1}{nh}\right) \cdot O_p(1) \\ &= o_p\left(\frac{1}{\sqrt{nh}}\right) \end{aligned}$$

□

Die Entwicklung von $\hat{\eta}_n(x)$, welche Proposition 4.10 angibt, hat noch nicht die im Weiteren benötigte Form, denn es wird eine Darstellung als Summe unabhängiger Zufallsvariablen zzgl. eines Resttermes von geeigneter Rate benötigt, es ist also $\frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) g(\hat{\varepsilon}_i, x)$ noch weiter zu entwickeln. Dies liefert die folgende Proposition:

Proposition 4.11 *Es gelten die Voraussetzungen (M1), (M2), (K), (H1) und (H2), (A). Dann gilt*

(i) *für Funktionen g , die (G1), (G2) und (M3) erfüllen:*

$$\frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) g(\hat{\varepsilon}_i, x) = \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) g(\varepsilon_i, x)$$

$$\begin{aligned}
& -\frac{1}{n} \sum_{i=1}^n \varepsilon_i \int \frac{1}{h} K_h(z-x) g'(y, x) \frac{1}{b} \tilde{K}_b(X_i - z) f_{\varepsilon|X}(y|z) d(y, z) \\
& -\frac{1}{n} \sum_{i=1}^n \int \frac{1}{h} K_h(z-x) g'(y, x) \frac{1}{b} \tilde{K}_b(X_i - z) \left(m(X_i) - m(z) \right) f_{\varepsilon|X}(y|z) d(y, z) \\
& + o_p\left(\frac{1}{\sqrt{nh}}\right)
\end{aligned}$$

(ii) im Falle, dass g die Gestalt $g(\varepsilon, x) = I\{\varepsilon \leq a(x)\} - d(x)$ besitzt:

$$\begin{aligned}
\frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) g(\hat{\varepsilon}_i, x) &= \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) I\{\varepsilon_i \leq a(x)\} - d(x) \\
&+ \frac{1}{n} \sum_{i=1}^n \varepsilon_i \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(a(x)|z) \frac{1}{b} \tilde{K}_b(X_i - z) dz \\
&+ \frac{1}{n} \sum_{i=1}^n \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(a(x)|z) \frac{1}{b} \tilde{K}_b(X_i - z) (m(X_i) - m(z)) dz \\
&+ o_p\left(\frac{1}{\sqrt{nh}}\right)
\end{aligned}$$

Beweis:

Zum Vorgehen: Dadurch, dass $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n$ von einander abhängen, ist die Entwicklung etwas länglich, die Techniken wiederholen sich aber immer wieder. Soweit möglich entwickelt man in Taylorreihen, nutzt die Voraussetzungen (M1) – (M3) um Restterme abzuschätzen bzw. nutzt für die die Zusatzinformation darstellende Funktion g die Voraussetzungen (G1) und (G2) um abschätzen zu können.

Wir beginnen mit dem Beweis zu (i), dabei gehen wir in zwei Schritten vor: Zuerst beweisen wir mittels Beweistechniken zur schwachen Konvergenz von stochastischen Prozessen die Darstellung:

$$\begin{aligned}
\frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) g(\hat{\varepsilon}_i, x) &= \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) g(\varepsilon_i, x) \\
(4.13) \quad &+ \int \frac{1}{h} K_h(z-x) g'(y, x) (m(z) - \hat{m}(z)) f_{\varepsilon, X}(y, z) d(y, z) \\
&+ o_p\left(\frac{1}{\sqrt{nh}}\right),
\end{aligned}$$

dabei ist $g'(y, x) := \frac{\partial g(y, x)}{\partial y}$. Im zweiten Schritt wird diese Darstellung dann noch weiter umgeformt.

Allein mittels Taylorentwicklung von $g(\hat{\varepsilon}_i, x)$ um $g(\varepsilon_i, x)$ und Berechnung von Erwartungswert und Varianz kommen wir nicht zu Darstellung (4.13), denn die Restterme sind nicht von passender Rate abzuschätzen da g nicht als beschränkt vorausgesetzt ist. In diesem Beweis nutzen wir die Eigenschaften (G1) und (G2) von g aus.

Zum ersten Beweisschritt:

Wir zeigen die folgende Aussage mittels Beweistechniken zur schwachen Konvergenz stochastischer Prozesse, dabei wenden wir Proposition 2.45 an.

Wir zeigen zunächst für $j = 1, \dots, k$ die schwache Konvergenz des folgenden empirischen Prozesses

$$(4.14) \quad G_n(\tilde{g}_{j,n}) := \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\tilde{g}_{j,n}(\varepsilon_i, X_i) - \int \int \tilde{g}_{j,n}(y, z) f_{\varepsilon, X}(y, z) dz dy \right),$$

welcher mit deterministischen Funktionen $\tilde{g}_{j,n} \in \mathcal{G}_{j,n}$, $j = 1, \dots, k$ indiziert ist, wobei

$$\mathcal{G}_{j,n} := \left\{ \tilde{g}_{j,n} : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R}, \tilde{g}_{j,n}(\varepsilon, z) = \left[g_j(\varepsilon + \tilde{h}(z), x) - g_j(\varepsilon, x) \right] \frac{1}{\sqrt{h}} K\left(\frac{z-x}{h}\right) \mid \tilde{h} \in \mathcal{H} \right\}$$

dabei ist die Abhängigkeit von n gegeben durch die von n abhängige Bandbreite $h = h_n$ und es gilt $\mathcal{H} := C_\delta^{1+\alpha}[0, 1]$, siehe zu dieser Klasse Definition 2.34. Für $(\hat{m} - m)$ gilt nach Akritas und van Keilegom (2001), dass $P(\hat{m} - m \in \mathcal{H}) \rightarrow 1$ für $n \rightarrow \infty$ ein Element der Klasse \mathcal{H} ist. Zusätzlich gilt:

$$\frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) g(\hat{\varepsilon}_i, x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) g(\varepsilon_i + (m - \hat{m})(X_i), x).$$

Nachzurechnen sind jetzt die Voraussetzungen der Proposition 2.45 mit $T = \mathcal{H}$, d.h. es ist zuerst zu zeigen, dass $(\mathcal{G}_{j,n})_{n \in \mathbb{N}}$ Donsker ist. Dies ist zur Übersicht ausgelagert und in Lemma 6.7 (i) zu finden. Der dort geführte Beweis wird darüber geführt, dass gezeigt wird, dass $\mathcal{G}_{j,n}$ asymptotisch stoch. gleich-stetig bzgl. einer Semimetrik ρ . Zusätzlich gilt $\|m - \hat{m}\|_\infty = o(1)$. Betrachtet man die im Beweis zu 6.7 (i) genutzte Semimetrik $\rho(s, t) = \|s - t\|_\infty$, so ergibt sich $\rho(c_n, 0) \rightarrow 0$ für alle $c_n \in \mathcal{H}$ mit $\|c_n\|_\infty = o(1)$ direkt.

In den Prozess G_n aus (4.14) setzen wir

$$\hat{g}_{j,n}(\varepsilon, z) = \left[g_j(\varepsilon + (m - \hat{m})(z), x) - g_j(\varepsilon, x) \right] \frac{1}{\sqrt{h}} K\left(\frac{z-x}{h}\right)$$

ein. Die Voraussetzungen von Proposition 2.45 sind also erfüllt.

Somit folgt nun, dass

$$\begin{aligned} G_n(\hat{g}_{j,n}) &:= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\left[g_j(\varepsilon_i + (m - \hat{m})(X_i), x) - g_j(\varepsilon, x) \right] \frac{1}{\sqrt{h}} K\left(\frac{X_i - x}{h}\right) \right. \\ &\quad \left. - \int \int \left[g_j(y + (m - \hat{m})(z), x) - g_j(y, x) \right] \frac{1}{\sqrt{h}} K\left(\frac{z-x}{h}\right) f_{\varepsilon, X}(y, z) dz dy \right) = o_p(1), \end{aligned}$$

oder in der Form, in der wir die Darstellung benötigen:

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^n \left[g_j(\varepsilon_i + (m - \hat{m})(X_i), x) - g_j(\varepsilon, x) \right] \frac{1}{h} K\left(\frac{X_i - x}{h}\right) \\ &= \int \int \left[g_j(y + (m - \hat{m})(z), x) - g_j(y, x) \right] \frac{1}{h} K\left(\frac{z-x}{h}\right) f_{\varepsilon, X}(y, z) dz dy + o_p\left(\frac{1}{\sqrt{nh}}\right). \end{aligned}$$

Um die in Proposition 4.11 (i) angegebene Entwicklung zu erhalten ist nun noch weiter zu entwickeln. Wir zeigen:

$$\begin{aligned} &\int \int (g_j(y + (m - \hat{m})(z), x) - g_j(y, x)) \frac{1}{h} K\left(\frac{z-x}{h}\right) f_{\varepsilon, X}(y, z) dz dy \\ (4.15) \quad &= \int \int g'_j(y, x) (m - \hat{m})(z) \frac{1}{h} K\left(\frac{z-x}{h}\right) f_{\varepsilon, X}(y, z) dz dy + o_p\left(\frac{1}{\sqrt{nh}}\right) \end{aligned}$$

Setze dazu:

$$A_n := \sqrt{nh} \int \int (g_j(y + (m - \hat{m})(z), x) - g_j(y, x) - g'_j(y, x) (m - \hat{m})(z)) \frac{1}{h} K\left(\frac{z-x}{h}\right) f_{\varepsilon, X}(y, z) dz dy$$

und betrachte die Wahrscheinlichkeit, dass $A_n > \rho$:

$$P(A_n > \rho) = P(A_n > \rho, \|m - \hat{m}\|_\infty \leq \delta) + P(A_n > \rho, \|m - \hat{m}\|_\infty > \delta) =: a_1 + a_2$$

Nach Proposition 2.53 (i) und den Bandbreitenbedingungen (H1) gilt:

$$a_2 \leq P(\|m - \hat{m}\|_\infty > \delta) = o(1) \quad \text{für alle } \delta > 0$$

Zum Beweis der $o(1)$ -Rate von a_1 nutzen wir die Eigenschaft (G1), (H1), (H2), speziell das Verhältnis von h und b zueinander und die in 2.53 (i) gegebene *f.s.* geltende Rate $\|m - \hat{m}\|_\infty = O((n^{-1}b^{-1} \log(b^{-1}))^{1/2})$ aus:

$$a_1 \leq P(\sqrt{nh}C \int \left| \frac{1}{h} K\left(\frac{u-x}{h}\right) \right| \cdot |(m - \hat{m})(u)|^{1+\beta} f_X(u) du > \rho, \|m - \hat{m}\|_\infty \leq \delta) = o(1)$$

Dabei ist für den letzten Schritt nötig, dass $\sqrt{\frac{h}{b}} \cdot \frac{\log(b^{-1})^{(1+\beta)/2}}{(nb)^{\beta/2}}$, vgl. (H2). Insgesamt ergibt sich die Darstellung (4.13). Damit ist der erste Beweisschritt abgeschlossen.

Zweiter Beweisschritt:

Es ist jetzt noch $\int \frac{1}{h} K_h(z-x) g'(y, x) (m(z) - \hat{m}(z)) f_{\varepsilon, X}(y, z) d(y, z)$ zu entwickeln und es ergibt sich Proposition 4.11 (i).

Dazu ersetzen wir den Nadaraya-Watson-Schätzer $\hat{m}(z)$ durch die in Definition 2.49 gegebene Darstellung und $Y_i = \varepsilon_i + m(X_i)$:

$$\begin{aligned} & \int \frac{1}{h} K_h(z-x) g'(y, x) (m(z) - \hat{m}(z)) f_{\varepsilon, X}(y, z) d(y, z) \\ &= - \int \frac{1}{h} K_h(z-x) g'(y, x) \hat{f}_X^{-1}(z) \frac{1}{nb} \sum_{i=1}^n \tilde{K}_b(X_i - z) \left((\varepsilon_i + m(X_i)) - m(z) \right) f_{\varepsilon, X}(y, z) d(y, z) \\ &= - \int \frac{1}{h} K_h(z-x) g'(y, x) \hat{f}_X^{-1}(z) \frac{1}{nb} \sum_{i=1}^n \tilde{K}_b(X_i - z) \varepsilon_i f_{\varepsilon, X}(y, z) d(y, z) \\ &\quad - \int \frac{1}{h} K_h(z-x) g'(y, x) \hat{f}_X^{-1}(z) \frac{1}{nb} \sum_{i=1}^n \tilde{K}_b(X_i - z) \left(m(X_i) - m(z) \right) f_{\varepsilon, X}(y, z) d(y, z) \\ &= - \int \frac{1}{h} K_h(z-x) g'(y, x) \frac{1}{nb} \sum_{i=1}^n \tilde{K}_b(X_i - z) \varepsilon_i f_{\varepsilon|X}(y|z) d(y, z) \\ &\quad - \int \frac{1}{h} K_h(z-x) g'(y, x) \frac{1}{nb} \sum_{i=1}^n \tilde{K}_b(X_i - z) \left(m(X_i) - m(z) \right) f_{\varepsilon|X}(y|z) d(y, z) \\ &\quad - \int \frac{1}{h} K_h(z-x) g'(y, x) \frac{f_X(z) - \hat{f}_X(z)}{\hat{f}_X(z)} \frac{1}{nb} \sum_{i=1}^n \tilde{K}_b(X_i - z) \varepsilon_i f_{\varepsilon|X}(y|z) d(y, z) \end{aligned}$$

$$\begin{aligned}
& - \int \frac{1}{h} K_h(z-x) g'(y,x) \frac{f_X(z) - \hat{f}_X(z)}{\hat{f}_X(z)} \frac{1}{nb} \sum_{i=1}^n \tilde{K}_b(X_i - z) \left(m(X_i) - m(z) \right) f_{\varepsilon|X}(y|z) d(y, z) \\
& =: T_1 + T_2 - T_3 - T_4
\end{aligned}$$

Die ersten beiden Terme, T_1 und T_2 , übernehmen wir direkt in die Entwicklung. Man könnte hier noch das Integral mittels Substitution und Taylorentwicklung berechnen, für die später folgende Berechnung der Varianz hat es aber Vorteile an dieser Stelle darauf zu verzichten.

Für die Terme T_3 und T_4 betrachten wir die folgende Zerlegung:

$$\begin{aligned}
(4.16) \quad T_3 &= \frac{1}{nh} \sum_{i=1}^n \varepsilon_i \int K_h(z-x) g'(y,x) \frac{f_X(z) - \hat{f}_X(z)}{\hat{f}_X(z)} \frac{1}{b} \tilde{K}_b(X_i - z) f_{\varepsilon|X}(y|z) d(z, y) \\
&= \frac{1}{nh} \sum_{i=1}^n \varepsilon_i \int K_h(z-x) g'(y,x) \frac{f_X(z) - \hat{f}_X(z)}{f_X(z)} \frac{1}{b} \tilde{K}_b(X_i - z) f_{\varepsilon|X}(y|z) d(z, y) \\
&+ \frac{1}{nh} \sum_{i=1}^n \varepsilon_i \int K_h(z-x) g'(y,x) \frac{(f_X(z) - \hat{f}_X(z))^2}{f_X(z) \cdot \hat{f}_X(z)} \frac{1}{b} \tilde{K}_b(X_i - z) f_{\varepsilon|X}(y|z) d(z, y)
\end{aligned}$$

Zu zeigen ist jetzt, dass diese beiden Summanden die Rate $o_p(\frac{1}{\sqrt{nh}})$ besitzen.

Den zweiten Term schätzen wir mittels Substitution, Proposition 2.55 (a) und den Voraussetzungen (M2), (K) und (H1), (H2) wie folgt grob ab:

$$\begin{aligned}
& \left| \frac{1}{nh} \sum_{i=1}^n \varepsilon_i \int K_h(z-x) g'(y,x) \frac{(f_X(z) - \hat{f}_X(z))^2}{f_X(z) \cdot \hat{f}_X(z)} \frac{1}{b} \tilde{K}_b(X_i - z) f_{\varepsilon|X}(y|z) d(z, y) \right| \\
& \leq \sup_{\tilde{x} \in [0,1]} \left| \frac{(f_X(\tilde{x}) - \hat{f}_X(\tilde{x}))^2}{f_X(\tilde{x}) \cdot \hat{f}_X(\tilde{x})} \right| \cdot \sup_{\tilde{x} \in [0,1]} \frac{1}{nh} \sum_{i=1}^n |\varepsilon_i| \int |K_h(z-\tilde{x})| |g'(y,x) f_{\varepsilon|X}(y|z)| \left| \frac{1}{b} \tilde{K}_b(X_i - z) \right| d(z, y) \\
& \leq \sup_{\tilde{x} \in [0,1]} \left| \frac{(f_X(\tilde{x}) - \hat{f}_X(\tilde{x}))^2}{f_X(\tilde{x}) \cdot \hat{f}_X(\tilde{x})} \right| \cdot \|2K\|_\infty \cdot \frac{1}{h} \|\tilde{K}\|_\infty \cdot \frac{1}{n} \sum_{i=1}^n |\varepsilon_i| \sup_{z \in [0,1]} \int |g'(y,x) f_{\varepsilon|X}(y|z)| dy (1 + o_p(1)) \\
& \leq O\left(\frac{\log(b^{-1})}{nb}\right) \cdot O\left(\frac{1}{h}\right) \cdot O_p(1) \\
& \stackrel{(H2)}{\leq} o_p\left(\frac{1}{\sqrt{nh}}\right)
\end{aligned}$$

Zum ersten Summanden in (4.16) berechnen wir den Erwartungswert und die Varianz. Der Erwartungswert ist wegen $E[\varepsilon_1 | X_1] = 0$ identisch 0. Die Varianz zu berechnen ist etwas aufwändiger. Setze dazu im weiteren $E[\varepsilon_i^2 | X_i] = \sigma^2(X_i)$ (vgl. Voraussetzung (A), S. 41) Wir zerlegen

die Varianz der Summe zuerst in die Summe Varianz und Kovarianz-Anteil, nutzen dann für den Kovarianz-Anteil aus, dass für $i \neq j$ $E[\varepsilon_i \cdot \varepsilon_j | X_1, \dots, X_n] = 0$ gilt:

$$\begin{aligned}
& \text{Var}\left(\frac{1}{n} \sum_{i=1}^n \varepsilon_i \int \frac{1}{h} K_h(z-x) g'(y,x) \frac{f_X(z) - \hat{f}_X(z)}{f_X(z)} \frac{1}{b} \tilde{K}_b(X_i - z) f_{\varepsilon|X}(y|z) d(z,y)\right) \\
&= \frac{1}{n^2} \sum_{i=1}^n E[\varepsilon_i^2 \left(\int \frac{1}{h} K_h(z-x) g'(y,x) \frac{f_X(z) - \hat{f}_X(z)}{f_X(z)} \frac{1}{b} \tilde{K}_b(X_i - z) f_{\varepsilon|X}(y|z) d(z,y)\right)^2] \\
&= \frac{1}{n^2} \sum_{i=1}^n E[E[\varepsilon_i^2 | X_i] \left(\int \frac{1}{h} K_h(z-x) g'(y,x) \frac{f_X(z) - \hat{f}_X(z)}{f_X(z)} \frac{1}{b} \tilde{K}_b(X_i - z) f_{\varepsilon|X}(y|z) d(z,y)\right)^2] \\
&= \frac{1}{n^2} \sum_{i=1}^n E[\sigma^2(X_i) \left(\int \frac{1}{h} K_h(z-x) g'(y,x) \frac{f_X(z) - \hat{f}_X(z)}{f_X(z)} \frac{1}{b} \tilde{K}_b(X_i - z) f_{\varepsilon|X}(y|z) d(z,y)\right)^2] \\
&= \frac{1}{n^2} \sum_{i=1}^n E[\sigma^2(X_i) \left(\int \frac{1}{h} K_h(z-x) g'(y,x) \frac{f_X(z) - \frac{1}{nb} \sum_{j=1, j \neq i}^n \tilde{K}_b(X_j - z) - \frac{1}{nb} \tilde{K}_b(X_i - z)}{f_X(z)} \right. \\
&\quad \left. \cdot \frac{1}{b} \tilde{K}_b(X_i - z) f_{\varepsilon|X}(y|z) d(z,y)\right)^2] \\
&= \frac{1}{n} E[\sigma^2(X_1) \left(\int \frac{1}{h} K_h(z-x) g'(y,x) \frac{f_X(z) - \frac{1}{nb} \sum_{j=1, j \neq 1}^n \tilde{K}_b(X_j - z)}{f_X(z)} \cdot \frac{1}{b} \tilde{K}_b(X_1 - z) f_{\varepsilon|X}(y|z) d(z,y) \right. \\
&\quad \left. - \int \frac{1}{h} K_h(z-x) g'(y,x) \frac{\frac{1}{nb} \tilde{K}_b(X_1 - z)}{f_X(z)} \cdot \frac{1}{b} \tilde{K}_b(X_1 - z) f_{\varepsilon|X}(y|z) d(z,y)\right)^2] \\
&= \frac{1}{n} E[\sigma^2(X_1) \left(\int \frac{1}{h} K_h(z-x) g'(y,x) \frac{f_X(z) - \frac{1}{nb} \sum_{j=1, j \neq 1}^n \tilde{K}_b(X_j - z)}{f_X(z)} \right. \\
&\quad \left. \cdot \frac{1}{b} \tilde{K}_b(X_1 - z) f_{\varepsilon|X}(y|z) d(z,y)\right)^2] \\
&\quad - 2 \frac{1}{n} E[\sigma^2(X_1) \int \frac{1}{h} K_h(z-x) g'(y,x) \frac{f_X(z) - \frac{1}{nb} \sum_{j=1, j \neq 1}^n \tilde{K}_b(X_j - z)}{f_X(z)} \cdot \frac{1}{b} \tilde{K}_b(X_1 - z) f_{\varepsilon|X}(y|z) d(z,y) \\
&\quad \cdot \int \frac{1}{h} K_h(z-x) g'(y,x) \frac{\frac{1}{nb} \tilde{K}_b(X_1 - z)}{f_X(z)} \cdot \frac{1}{b} \tilde{K}_b(X_1 - z) f_{\varepsilon|X}(y|z) d(z,y)] \\
&\quad + \frac{1}{n} E[\sigma^2(X_1) \left(\int \frac{1}{h} K_h(z-x) g'(y,x) \frac{\frac{1}{nb} \tilde{K}_b(X_1 - z)}{f_X(z)} \cdot \frac{1}{b} \tilde{K}_b(X_1 - z) f_{\varepsilon|X}(y|z) d(z,y)\right)^2] \\
&= \frac{1}{n} E[\sigma^2(X_1) \left(\int \frac{1}{h} K_h(z-x) g'(y,x) \frac{f_X(z) - \frac{1}{nb} \sum_{j=2}^n \tilde{K}_b(X_j - z)}{f_X(z)} \cdot \frac{1}{b} \tilde{K}_b(X_1 - z) f_{\varepsilon|X}(y|z) d(z,y)\right)^2] \\
&\quad - 2 \frac{1}{n^2 b} E[\sigma^2(X_1) \int \frac{1}{h} K_h(z-x) g'(y,x) \frac{f_X(z) - \frac{1}{nb} \sum_{j=2}^n \tilde{K}_b(X_j - z)}{f_X(z)} \cdot \frac{1}{b} \tilde{K}_b(X_1 - z) f_{\varepsilon|X}(y|z) d(z,y)
\end{aligned}$$

$$\begin{aligned}
& \cdot \int \frac{1}{h} K_h(z-x) g'(y, x) f_X^{-1}(z) \cdot \frac{1}{b} \tilde{K}_b^2(X_1-z) f_{\varepsilon|X}(y|z) d(z, y) \Big] \\
& + \frac{1}{n} \frac{1}{n^2 b^2} E[\sigma^2(X_1) \left(\int \frac{1}{h} K_h(z-x) g'(y, x) f_X^{-1}(z) \cdot \frac{1}{b} \tilde{K}_b^2(X_1-z) f_{\varepsilon|X}(y|z) d(z, y) \right)^2] \\
& \leq \frac{1}{n} E[\sigma^2(X_1) \int \int (f_X^{-1}(z_1))^2 (f_X^{-1}(z_2))^2 \frac{1}{h} K_h(z_1-x) \frac{1}{h} K_h(z_2-x) g'(y_1, x) g'(y_2, x) \\
& \quad \cdot (f_X(z_1) - \frac{1}{nb} \sum_{j=2}^n \tilde{K}_b(X_j - z_1))(f_X(z_2) - \frac{1}{nb} \sum_{j=2}^n \tilde{K}_b(X_j - z_2)) \\
& \quad \cdot \frac{1}{b} \tilde{K}_b(X_1 - z_1) \cdot \frac{1}{b} \tilde{K}_b(X_1 - z_2) f(y_1|z_1) f(y_2|z_2) d(z_1, y_1) d(z_2, y_2)] \\
& \quad - O\left(\frac{1}{n^2 b^2}\right) \\
& \quad + O\left(\frac{1}{n^3 b^3}\right) \\
& = o\left(\frac{1}{nh}\right) - O\left(\frac{1}{n^2 b^2}\right) + O\left(\frac{1}{n^3 b^3}\right) \\
& = o\left(\frac{1}{nh}\right)
\end{aligned}$$

Ähnlich zum Vorgehen zur Abschätzung von T_3 ist das Vorgehen im Falle von T_4 :

$$\begin{aligned}
T_4 &= \frac{1}{nh} \sum_{i=1}^n \int K_h(z-x) g'(y, x) \frac{f_X(z) - \hat{f}_X(z)}{\hat{f}_X(z)} \frac{1}{b} \tilde{K}_b(X_i - z) \left(m(X_i) - m(z) \right) f_{\varepsilon|X}(y|z) d(z, y) \\
&= \frac{1}{nh} \sum_{i=1}^n \int K_h(z-x) g'(y, x) \frac{f_X(z) - \hat{f}_X(z)}{f_X(z)} \frac{1}{b} \tilde{K}_b(X_i - z) \left(m(X_i) - m(z) \right) f_{\varepsilon|X}(y|z) d(z, y) \\
&\quad + \frac{1}{nh} \sum_{i=1}^n \int K_h(z-x) g'(y, x) \frac{(f_X(z) - \hat{f}_X(z))^2}{f_X(z) \cdot \hat{f}_X(z)} \frac{1}{b} \tilde{K}_b(X_i - z) \left(m(X_i) - m(z) \right) f_{\varepsilon|X}(y|z) d(z, y)
\end{aligned}$$

Diese beiden Summanden sind wiederum durch $o_p\left(\frac{1}{\sqrt{nh}}\right)$ abschätzbar, dazu schätzen wir wie schon zuvor den zweiten Term grob ab und berechnen für den ersten Term den Erwartungswert (dieser ist gleich $o\left(\frac{1}{\sqrt{nh}}\right)$) und die Varianz (ist gleich $O\left(\frac{1}{n}\right)$).

Insgesamt erhält man Proposition 4.11 (i). Zu Proposition 4.11 (ii):

Für g von der Gestalt $g(\varepsilon, x) = I\{\varepsilon \leq a(x)\} - d(x)$ ergibt sich als Entwicklung:

$$\frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) I\{\hat{\varepsilon}_i \leq a(x)\} = \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) I\{\varepsilon_i \leq a(x)\} - d(x)$$

$$\begin{aligned}
& + \frac{1}{n} \sum_{i=1}^n \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(a(x)|z) \frac{1}{b} \tilde{K}_b(X_i-z) (m(X_i) - m(z)) dz \\
& + \frac{1}{n} \sum_{i=1}^n \varepsilon_i \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(a(x)|z) \frac{1}{b} \tilde{K}_b(X_i-z) dz \\
& + o_p\left(\frac{1}{\sqrt{nh}}\right)
\end{aligned}$$

Beweis:

Im weiteren Verlauf der Arbeit folgt ein nahezu identischer Beweis (Beweis zu Proposition 5.4). Darauf wollen wir hier verweisen. Dort wird der Beweis gleichmäßig im hier $a(x)$ genannten Parameter geführt, an dieser Stelle ist die gleichmäßige Abschätzung aber nicht nötig. Der Beweis zu 4.11 ist damit abgeschlossen. □

Zur Erinnerung:

Die Entwicklungen wurden durchgeführt um die bisher hergeleitete Darstellung für $\hat{\eta}_n^t(x)$ aus Proposition 4.10 im Weiteren ebenfalls entwickeln zu können. Diese weitere Entwicklung ist nötig, da in der Entwicklung aus Proposition 4.10 immer noch die Schätzung $\hat{\varepsilon}_i$ enthalten ist und sie ist im nächsten Korollar zusammengefasst:

Ziel der vorangegangenen Proposition 4.11 war die Darstellung von $\sum g(\hat{\varepsilon}_i, x) K_h(X_i - x)$ als Summe unabhängiger Zufallsvariablen. Proposition 4.10 besagt, dass:

$$\hat{\eta}_n(x) = \Sigma^{-1}(x) \frac{1}{nh} \sum_{i=1}^n g(\hat{\varepsilon}_i, x) K_h(X_i - x) + o_p\left(\frac{1}{\sqrt{nh}}\right).$$

Mithilfe von Proposition 4.11 kann nun leicht, passend zum jeweiligen Typ von g , eine Entwicklung als Summe unabhängiger Zufallsvariablen angeben werden:

Proposition 4.12 *Es gelten Modell (3.1) und die Voraussetzungen (M1), (M2), (K), (A), (H1) und (H2). $\hat{\eta}_n(x)$ sei wie in 4.10 definiert. Dann gilt:*

(E1) *für Funktionen g , welche (G1), (G2) und (M3) erfüllen:*

$$\hat{\eta}_n(x) = \Sigma^{-1}(x) \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) g(\varepsilon_i, x)$$

$$\begin{aligned}
& -\Sigma^{-1}(x) \frac{1}{n} \sum_{i=1}^n \varepsilon_i \int \frac{1}{h} K_h(z-x) g'(y,x) \frac{1}{b} \tilde{K}_b(X_i-z) f_{\varepsilon|X}(y|z) d(y,z) \\
& -\Sigma^{-1}(x) \frac{1}{n} \sum_{i=1}^n \int \frac{1}{h} K_h(z-x) g'(y,x) \frac{1}{b} \tilde{K}_b(X_i-z) \left(m(X_i) - m(z) \right) f_{\varepsilon|X}(y|z) d(y,z) \\
& + o_p\left(\frac{1}{\sqrt{nh}}\right)
\end{aligned}$$

(E2) für Funktionen vom Typ $g(\varepsilon, x) = I\{\varepsilon \leq a(x)\} - d(x)$:

$$\begin{aligned}
\hat{\eta}_n(x) &= \Sigma^{-1}(x) \frac{1}{nh} \sum_{i=1}^n K_h(X_i-x) g(\varepsilon_i, x) \\
& + \Sigma^{-1}(x) \frac{1}{n} \sum_{i=1}^n \varepsilon_i \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(a(x)|z) \frac{1}{b} \tilde{K}_b(X_i-z) dz \\
& + \Sigma^{-1}(x) \frac{1}{n} \sum_{i=1}^n \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(a(x)|z) \frac{1}{b} \tilde{K}_b(X_i-z) (m(X_i) - m(z)) dz \\
& + o_p\left(\frac{1}{\sqrt{nh}}\right)
\end{aligned}$$

Beide Aussagen folgen mittels Proposition 4.11 und Proposition 4.10.

□

In diesem Abschnitt haben wir uns hauptsächlich mit der Entwicklung von $\hat{\eta}_n^t(x)$ in Summen unabhängiger Zufallsvariablen zzgl. eines Resttermes von hinreichend schneller Rate, $o_p\left(\frac{1}{\sqrt{nh}}\right)$, beschäftigt. Diese Entwicklung werden wir im weiteren Verlauf nutzen um $\hat{F}_n^*(y|x) - F_{\varepsilon|X}(y|x)$ zu entwickeln.

Bemerkung 4.13 *In diesem Kapitel treten die Bandbreiten b und h nur einmal, im Beweis zu Gleichung (4.15) in Wechselwirkung. Dort ist zum Beweis einer Rate nötig, dass $\sqrt{\frac{h}{b} \frac{\log(b^{-1})^{(1+\beta)/2}}{(nb)^{\beta/2}}} = o(1)$. Es werden aber an vielen weiteren Stellen schon explizit die in (H1) und (H2) angegebenen Raten für die Bandbreiten genutzt, so speziell in einigen Abschätzungen von Resttermen die Voraussetzung an die Bandbreite b , so dass für die genutzten Kernschätzer \hat{m} und \hat{f}_X die Abstände $\|\hat{m} - m\|_\infty$ und $\|\hat{f}_X - f_X\|_\infty$ von der Größenordnung $O\left(\sqrt{\frac{\log b^{-1}}{nb}}\right)$ sind.*

5 Entwicklung des Empirical-Likelihood-Schätzers

5.1 Zerlegung in drei Summanden

Im Weiteren soll nun die Differenz $\hat{F}_n^*(y|x) - F_{\varepsilon|X}(y|x)$ so entwickelt werden, dass sich eine Summe unabhängiger Zufallsvariablen zuzüglich eines Resttermes von geeigneter, in $y \in \mathbb{R}$ gleichmäßiger Rate ergibt. Dazu zerlege wie folgt:

$$(5.1) \quad \begin{aligned} \hat{F}_n^*(y|x) - F_{\varepsilon|X}(y|x) \\ = (\hat{F}_n^*(y|x) - \hat{F}_n(y|x)) + (\hat{F}_n(y|x) - F_n(y|x)) + (F_n(y|x) - F_{\varepsilon|X}(y|x)), \end{aligned}$$

dabei siehe für die einzelnen Bauteile die Definitionen 4.1, 4.2 und 4.6. Die erste Differenz ist die zwischen dem Empirical-Likelihood-Schätzer und dem Kernschätzer der bedingten Verteilungsfunktion der Residuen gegeben $X = x$ auf Basis der nichtparametrisch mittels Kernschätzern geschätzten Fehler. Die zweite Differenz ist eine zwischen verschiedenen Kernschätzern der bedingten Verteilungsfunktion der Fehler gegeben $X = x$, einmal auf Basis der geschätzten, einmal mit den original Fehler. Dazu ist zu bemerken, dass der zuletzt genannte Schätzer in der Praxis nicht nutzbar ist, da die Fehler nicht direkt beobachtbar sind. Im Weiteren setze zum Zwecke der Übersicht $\mathcal{X}_n = (X_1, \dots, X_n)$.

In den Entwicklungen dieses Abschnittes tritt der Kerndichteschätzer $\hat{f}_X(x)$ in zwei Weisen auf, zu unterscheiden nur durch die jeweils andere genutzte Bandbreite h bzw. b , vgl. (H1) und (H2) und die genutzten Kerne K und \tilde{K} . Zur Kennzeichnung nutze für der Kerndichteschätzer in diesem Kapitel (und nur hier) die Bezeichnungen $\hat{f}_{h,X}(x)$ und $\hat{f}_{b,X}(x)$, in den vorangegangenen Kapiteln meint $\hat{f}_X(x)$ immer $\hat{f}_{b,X}(x)$.

5.2 Entwicklung der drei Summanden

Wir betrachten die drei Summanden der Darstellung (5.1) noch genauer und zeigen:

Proposition 5.1 *Unter den Voraussetzungen (M1) – (M3), (K), (A), (H1), (H2) und (G1), (G2), (M3) bzw. g von der Form $g(\varepsilon, x) = I\{\varepsilon \leq a(x)\} - d(x)$ gilt gleichmäßig in $y \in \mathbb{R}$:*

$$(i) \quad \hat{F}_n^*(y|x) - \hat{F}_n(y|x) = -\hat{\eta}_n^t(x) \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = x] + o_p\left(\frac{1}{\sqrt{nh}}\right)$$

$$\begin{aligned}
\text{(ii)} \quad \hat{F}_n(y|x) - F_n(y|x) &= f_X^{-1}(x) \frac{1}{n} \sum_{i=1}^n \int \frac{1}{h} K\left(\frac{z-x}{h}\right) f_{\varepsilon|X}(y|z) \frac{1}{b} \tilde{K}\left(\frac{X_i-z}{b}\right) (m(X_i) - m(z)) dz \\
&\quad + f_X^{-1}(x) \frac{1}{n} \sum_{i=1}^n \varepsilon_i \int \frac{1}{h} K\left(\frac{z-x}{h}\right) f_{\varepsilon|X}(y|z) \frac{1}{b} \tilde{K}\left(\frac{X_i-z}{b}\right) dz \\
&\quad + o_p\left(\frac{1}{\sqrt{nh}}\right)
\end{aligned}$$

$$\text{(iii)} \quad F_n(y|x) - F_{\varepsilon|X}(y|x) = f_X^{-1}(x) \frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i-x}{h}\right) (I\{\varepsilon_i \leq y\} - F_{\varepsilon|X}(y|x)) + o_p\left(\frac{1}{\sqrt{nh}}\right)$$

Wir beginnen mit dem Beweis zu Proposition 5.1 (iii):

Mithilfe von Proposition 2.55 ergibt sich:

$$\begin{aligned}
F_n(y|x) - F_{\varepsilon|X}(y|x) &= \frac{\frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) (I\{\varepsilon_i \leq y\} - F_{\varepsilon|X}(y|x))}{\hat{f}_{h,X}(x)} \\
&= \frac{\frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) (I\{\varepsilon_i \leq y\} - F_{\varepsilon|X}(y|x))}{f_X(x)} \\
&\quad - \frac{\frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) (I\{\varepsilon_i \leq y\} - F_{\varepsilon|X}(y|x)) (\hat{f}_{h,X}(x) - f_X(x))}{\hat{f}_{h,X}(x) f_X(x)} \\
&= \frac{\frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) (I\{\varepsilon_i \leq y\} - F_{\varepsilon|X}(y|x))}{f_X(x)} + o_p\left(\frac{1}{\sqrt{nh}}\right).
\end{aligned}$$

Den zweiten Summanden werden wir grob abschätzen. Der Nenner des Zweiten Summanden ist von der Größenordnung $O_p(1)$ und größer 0, da $f_X(x) > 0$ für alle x aus $[0, 1]$ nach Voraussetzung. Für die Abschätzung des Zählers des zweiten Summanden nutzen wir die f.s. geltende Rate $\|\hat{f}_{h,X} - f\|_\infty = O\left(\sqrt{\frac{\log(h^{-1})}{nh}}\right)$ aus 2.55. Nun ist noch der erste Teil des Zählers gleichmäßig in $y \in \mathbb{R}$ abzuschätzen. Wir multiplizieren den Ausdruck zur Vereinfachung mit h und zeigen:

$$\sup_{y \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) (I\{\varepsilon_i \leq y\} - F_{\varepsilon|X}(y|x)) \right| = o_p(h^2).$$

Zum Beweis nutze Proposition 2.39. Die zugrundeliegende Funktionenklasse hat die Form

$$\mathcal{F}_n = \{ \varphi_{n,y} : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R} \mid \varphi_{n,y}(\varepsilon, z) = K\left(\frac{z-x}{h}\right) (I\{\varepsilon \leq y\} - F_{\varepsilon|X}(y|x)), y \in \mathbb{R} \},$$

man wähle $\delta_n^2 = h \cdot \log(n)$ und $\alpha_n = h/\log(n)^{\sqrt{1+\gamma}}$, $\gamma > 0$ aus Voraussetzung (H2), dann folgt die Aussage wegen $\delta_n^2 \alpha_n = h^2/\log(n)^{\sqrt{1+\gamma}-1} = o(h^2)$ und $\frac{\log(n)}{n\delta_n^2 \alpha_n^2} = \frac{\log(n)^{1+\gamma}}{h^3 n} \rightarrow 0$, siehe ebenfalls (H2). Ein Beweis für eine kompliziertere Funktionenklassen mittels Proposition 2.39 folgt auf Seite 75, auf das genaue Vorrechnen der weiteren Bedingungen in diesem Fall verzichten wir hier. Damit ist (iii) bewiesen.

Als nächstes betrachte den ersten Summanden $\hat{F}_n^*(y|x) - \hat{F}_n(y|x)$ aus der Differenzen-Darstellung (5.1) vor Proposition 5.1 zu Beginn dieses Abschnittes und beweise die Darstellung und Rate aus Proposition 5.1 (i).

Ziel ist es auch hier wieder den Summanden darzustellen mittels einer Summe unabhängiger und identisch verteilter Zufallsvariablen zzgl. eines Resttermes von in $y \in \mathbb{R}$ gleichmäßiger Rate $o_p(\frac{1}{\sqrt{nh}})$:

$$\begin{aligned}
& \hat{F}_n^*(y|x) - \hat{F}_n(y|x) \\
&= \sum_{i=1}^n \left(\frac{\frac{1}{nh} K_h(X_i - x) \hat{p}_i(x)}{\frac{1}{nh} \sum_{j=1}^n K_h(X_j - x) \hat{p}_j(x)} - \frac{\frac{1}{nh} K_h(X_i - x)}{\frac{1}{nh} \sum_{j=1}^n K_h(X_j - x)} \right) I\{\hat{\varepsilon}_i \leq y\} \\
&= \sum_{i=1}^n \frac{\frac{1}{nh} K_h(X_i - x)}{\frac{1}{nh} \sum_{j=1}^n K_h(X_j - x) \hat{p}_j(x)} \frac{1}{\frac{1}{nh} \sum_{j=1}^n K_h(X_j - x)} \left(\frac{1}{nh} \sum_{j=1}^n K_h(X_j - x) (\hat{p}_i(x) - \hat{p}_j(x)) \right) I\{\hat{\varepsilon}_i \leq y\} \\
&= \sum_{i=1}^n C(X_i, x, h, \mathcal{X}_n, \dots) \cdot \left(\frac{1}{nh} \sum_{j=1}^n K_h(X_j - x) (\hat{p}_i(x) - \hat{p}_j(x)) \right) I\{\hat{\varepsilon}_i \leq y\} \\
&= \sum_{i=1}^n C(X_i, x, h, \mathcal{X}_n, \dots) \cdot \left(\frac{1}{n^2 h} \sum_{j=1}^n K_h(X_j - x) \right) \cdot \\
&\quad \left(\frac{1}{1 + \hat{\eta}_n^t(x) g(\hat{\varepsilon}_i, x) K_h(X_i - x)} - \frac{1}{1 + \hat{\eta}_n^t(x) g(\hat{\varepsilon}_j, x) K_h(X_j - x)} \right) I\{\hat{\varepsilon}_i \leq y\} \\
&= \sum_{i=1}^n C(X_i, x, h, \mathcal{X}_n, \dots) \cdot \frac{1}{n^2 h} \sum_{j=1}^n K_h(X_j - x) \left(-\hat{\eta}_n^t(x) g(\hat{\varepsilon}_i, x) K_h(X_i - x) + \hat{\eta}_n^t(x) g(\hat{\varepsilon}_j, x) K_h(X_j - x) \right. \\
&\quad \left. + \frac{(\hat{\eta}_n^t(x) g(\hat{\varepsilon}_i, x) K_h(X_i - x))^2}{1 + \hat{\eta}_n^t(x) g(\hat{\varepsilon}_i, x) K_h(X_i - x)} - \frac{(\hat{\eta}_n^t(x) g(\hat{\varepsilon}_j, x) K_h(X_j - x))^2}{1 + \hat{\eta}_n^t(x) g(\hat{\varepsilon}_j, x) K_h(X_j - x)} \right) I\{\hat{\varepsilon}_i \leq y\} \\
&=: \hat{A}_1(n, x, y) + \hat{A}_2(n, x, y) + \hat{A}_3(n, x, y) + \hat{A}_4(n, x, y)
\end{aligned}$$

Dabei benutze die in Gleichung (4.8) angegebene Darstellung für $\hat{p}_i(x)$; die vorletzte Gleichheit nutzt $\frac{1}{1+x} = 1 - x + \frac{x^2}{1+x}$.

Berechne nun im Weiteren die Raten der Terme $\hat{A}_1(n, x, y), \dots, \hat{A}_4(n, x, y)$ getrennt. Dabei zeigt sich, dass nur der erste Summand einen Beitrag liefert, die anderen Summanden sind von kleinerer (in y gleichmäßiger) Rate:

$$\begin{aligned}\hat{F}_n^*(y|x) - \hat{F}_n(y|x) &= \hat{A}_1(n, x, y) + \hat{A}_2(n, x, y) + \hat{A}_3(n, x, y) + \hat{A}_4(n, x, y) \\ &= \hat{A}_1(n, x, y) + o_p\left(\frac{1}{\sqrt{nh}}\right) \\ &= -\hat{\eta}_n^t(x) \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = x] + o_p\left(\frac{1}{\sqrt{nh}}\right)\end{aligned}$$

Zum Beweis der Raten der Summanden $\hat{A}_2(n, x, y), \dots, \hat{A}_4(n, x, y)$ sowie der Darstellung von $\hat{A}_1(n, x, y)$ siehe die folgende Proposition:

Proposition 5.2 *Unter den Annahmen (M1) – (M3), (K), (A), (H1), (H2), (G1) und (G2) gilt gleichmäßig in $y \in \mathbb{R}$:*

$$\begin{aligned}\text{(i)} \quad \hat{A}_1(n, x, y) &= -\hat{\eta}_n^t(x) \sum_{i=1}^n \frac{\frac{1}{nh} K_h(X_i - x) \frac{1}{n^2 h} \sum_{j=1}^n K_h(X_j - x) g(\hat{\varepsilon}_i, x) K_h(X_i - x)}{\frac{1}{nh} \sum_{j=1}^n K_h(X_j - x) \hat{p}_j(x) \frac{1}{nh} \sum_{j=1}^n K_h(X_j - x)} I\{\hat{\varepsilon}_i \leq y\} \\ &= -\hat{\eta}_n^t(x) \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = x] + o_p\left(\frac{1}{\sqrt{nh}}\right) \\ \text{(ii)} \quad \hat{A}_2(n, x, y) &= \hat{\eta}_n^t(x) \sum_{i=1}^n \frac{\frac{1}{nh} K_h(X_i - x) \frac{1}{n^2 h} \sum_{j=1}^n K_h(X_j - x) g(\hat{\varepsilon}_j, x) K_h(X_j - x)}{\frac{1}{nh} \sum_{j=1}^n K_h(X_j - x) \hat{p}_j(x) \frac{1}{nh} \sum_{j=1}^n K_h(X_j - x)} I\{\hat{\varepsilon}_i \leq y\} \\ &= o_p\left(\frac{1}{\sqrt{nh}}\right) \\ \text{(iii)} \quad \hat{A}_3(n, x, y) &= \sum_{i=1}^n \frac{\frac{1}{nh} K_h(X_i - x) \frac{1}{n^2 h} \sum_{j=1}^n K_h(X_j - x) \frac{(\hat{\eta}_n^t(x) g(\hat{\varepsilon}_i, x) K_h(X_i - x))^2}{1 + \hat{\eta}_n^t(x) g(\hat{\varepsilon}_j, x) K_h(X_j - x)}}{\frac{1}{nh} \sum_{j=1}^n K_h(X_j - x) \hat{p}_j(x) \frac{1}{nh} \sum_{j=1}^n K_h(X_j - x)} I\{\hat{\varepsilon}_i \leq y\} \\ &= o_p\left(\frac{1}{\sqrt{nh}}\right) \\ \text{(iv)} \quad \hat{A}_4(n, x, y) &= \sum_{i=1}^n \frac{\frac{1}{nh} K_h(X_i - x) \frac{1}{n^2 h} \sum_{j=1}^n K_h(X_j - x) \frac{(\hat{\eta}_n^t(x) g(\hat{\varepsilon}_j, x) K_h(X_j - x))^2}{1 + \hat{\eta}_n^t(x) g(\hat{\varepsilon}_j, x) K_h(X_j - x)}}{\frac{1}{nh} \sum_{j=1}^n K_h(X_j - x) \hat{p}_j(x) \frac{1}{nh} \sum_{j=1}^n K_h(X_j - x)} I\{\hat{\varepsilon}_i \leq y\} \\ &= o_p\left(\frac{1}{\sqrt{nh}}\right)\end{aligned}$$

Bemerkung 5.3 *Im Weiteren muss auch noch das in der Darstellung (i) auftretende $\hat{\eta}_n(x)$ als eine Summe unabhängiger Zufallsvariablen plus einem Anteil geeigneter Rate dargestellt werden. Die Entwicklung liefert Proposition 4.12, die Ersetzung erfolgt an dieser Stelle aber noch nicht,*

zu einem späteren Zeitpunkt kommen wir darauf zurück.

Beweis zu Proposition 5.2 (i):

$$\begin{aligned}
\hat{A}_1(n, x, y) &= -\hat{\eta}_n^t(x) \sum_{i=1}^n \frac{\frac{1}{n^2 h} K_h^2(X_i - x) g(\hat{\varepsilon}_i, x)}{\frac{1}{n h} \sum_{j=1}^n K_h(X_j - x) \hat{p}_j(x)} I\{\hat{\varepsilon}_i \leq y\} \\
&= -\hat{\eta}_n^t(x) \cdot \left[\sum_{i=1}^n \frac{\frac{1}{n h} K_h^2(X_i - x) g(\varepsilon_i, x)}{\frac{1}{h} \sum_{j=1}^n K_h(X_j - x) \hat{p}_j(x)} I\{\varepsilon_i \leq y\} \right. \\
&\quad + \sum_{i=1}^n \frac{\frac{1}{n h} K_h^2(X_i - x) (g(\hat{\varepsilon}_i, x) - g(\varepsilon_i, x))}{\frac{1}{h} \sum_{j=1}^n K_h(X_j - x) \hat{p}_j(x)} I\{\hat{\varepsilon}_i \leq y\} \\
&\quad \left. - \sum_{i=1}^n \frac{\frac{1}{n h} K_h^2(X_i - x) g(\varepsilon_i, x)}{\frac{1}{h} \sum_{j=1}^n K_h(X_j - x) \hat{p}_j(x)} (I\{\varepsilon_i \leq y\} - I\{\hat{\varepsilon}_i \leq y\}) \right] \\
&=: -\hat{\eta}_n^t(x) (\hat{A}_{11} + \hat{A}_{12} - \hat{A}_{13})(n, x, y)
\end{aligned}$$

Wir zeigen $\sup_{y \in \mathbb{R}} |\hat{A}_{12}(n, x, y)| = o_p(1)$ und $\sup_{y \in \mathbb{R}} |\hat{A}_{13}(n, x, y)| = o_p(1)$. Hier sei nur der Beweis zu $\hat{A}_{12}(n, x, y)$ angegeben, $\hat{A}_{13}(n, x, y)$ folgt mit ähnlichen Methoden.

Im weiteren Verlauf der Arbeit wird für den Nenner des Terms $\hat{A}_{12}(n, x, y)$ die Rate $c + o_p(1)$ mit $c \neq 0$ (siehe Term $\hat{A}_{112}(n, x, y)$ auf der folgenden Seite) bewiesen. Betrachte hier also nur den Zähler. Für diesen gilt nach der Hölder-Ungleichung und den Hilfsaussagen 4.9:

$$\begin{aligned}
&\left\| \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K_h^2(X_i - x) (g(\hat{\varepsilon}_i, x) - g(\varepsilon_i, x)) I\{\hat{\varepsilon}_i \leq y\} \right\| \\
&\leq \frac{1}{n} \sum_{i=1}^n \left\| \frac{1}{\sqrt{h}} K_h(X_i - x) (g(\hat{\varepsilon}_i, x) - g(\varepsilon_i, x)) \cdot \frac{1}{\sqrt{h}} K_h(X_i - x) I\{\hat{\varepsilon}_i \leq y\} \right\| \\
&\leq \left(\frac{1}{n} \sum_{i=1}^n \frac{1}{h} K_h^2(X_i - x) I\{\hat{\varepsilon}_i \leq y\} \right)^{1/2} \cdot \left(\frac{1}{n} \sum_{i=1}^n \frac{1}{h} K_h^2(X_i - x) \|g(\hat{\varepsilon}_i, x) - g(\varepsilon_i, x)\|^2 \right)^{1/2} \\
&\leq \left(\frac{1}{n} \sum_{i=1}^n \frac{1}{h} K_h^2(X_i - x) \right)^{1/2} \cdot \left(\frac{1}{n} \sum_{i=1}^n \frac{1}{h} K_h^2(X_i - x) \|g(\hat{\varepsilon}_i, x) - g(\varepsilon_i, x)\|^2 \right)^{1/2} \\
&= o_p(1),
\end{aligned}$$

dazu benutze Proposition 4.9 (iv). Der einen Beitrag von relevanter Rate liefernde Term ist $\hat{A}_{11}(n, x, y)$. Dieser hat die Gestalt:

$$\begin{aligned}
\hat{A}_{11}(n, x, y) &= \sum_{i=1}^n \frac{\frac{1}{nh} K_h^2(X_i - x) g(\varepsilon_i, x)}{\frac{1}{h} \sum_{j=1}^n K_h(X_j - x) \hat{p}_j(x)} I\{\varepsilon_i \leq y\} \\
(5.2) \qquad &= \frac{\frac{1}{nh} \sum_{i=1}^n K_h^2(X_i - x) g(\varepsilon_i, x) I\{\varepsilon_i \leq y\}}{\frac{1}{nh} \sum_{j=1}^n K_h(X_j - x) \frac{1}{1 + \hat{\eta}_n^t(x) g(\hat{\varepsilon}_j, x) K_h(X_j - x)}} \\
&=: \frac{\hat{A}_{111}(n, x, y)}{\hat{A}_{112}(n, x, y)}.
\end{aligned}$$

Wir betrachten nun den Zähler und Nenner aus (5.2) getrennt. Es ergibt sich für den Nenner mittels Proposition 4.9 und Taylorentwicklungen bzw. Substitutionen wie sie schon häufiger auftraten:

$$\begin{aligned}
\hat{A}_{112}(n, x, y) &= \frac{1}{nh} \sum_{j=1}^n K_h(X_j - x) \frac{1}{1 + \hat{\eta}_n^t(x) g(\hat{\varepsilon}_j, x) K_h(X_j - x)} \\
&= \frac{1}{nh} \sum_{j=1}^n K_h(X_j - x) (1 - \hat{\eta}_n^t(x) g(\hat{\varepsilon}_j, x) K_h(X_j - x) + \frac{(\hat{\eta}_n^t(x) g(\hat{\varepsilon}_j, x) K_h(X_j - x))^2}{1 + \hat{\eta}_n^t(x) g(\hat{\varepsilon}_j, x) K_h(X_j - x)}) \\
&= \left(\frac{1}{n} \sum_{j=1}^n \frac{1}{h} K_h(X_j - x) - f_X(x) \right) + f_X(x) + o_p(1) \\
&= (\hat{f}_X(x) - f_X(x)) + f_X(x) + o_p(1) \\
&= f_X(x) + o_p(1)
\end{aligned}$$

Den Zähler wollen wir durch dessen Erwartungswert zzgl. eines Terms von geeigneter Rate ersetzen. Da der Zähler von y abhängt, muss diese Darstellung gleichmäßig in y bewiesen werden. Bevor wir mit der Betrachtung des Zählers beginnen, verweisen wir auf Theorem 2.39 und die dazugehörige Bemerkung 2.40, welche es zusammen ermöglichen eine gleichmäßige Rate anzugeben.

Theorem 2.39 und Bemerkung 2.40 wenden wir nun auf den Zähler von (5.2) an, dazu sind verschiedenste Bedingungen zu prüfen. Für den Zähler von (5.2) würden wir gerne gleichmäßig in $y \in \mathbb{R}$ zeigen, dass

$$\frac{1}{nh} \sum_{i=1}^n K_h^2(X_i - x) g(\varepsilon_i, x) I\{\varepsilon_i \leq y\} = \mu_0^{K^2} f_X(x) E[g(\varepsilon, x) I\{\varepsilon \leq y\} | X = x] + o_p(1).$$

Wir multiplizieren diese Gleichung noch mit h , was die folgenden Berechnungen übersichtlicher macht. Wir zeigen also

$$\sup_{y \in \mathbb{R}} \left\| \frac{1}{n} \sum_{i=1}^n K_h^2(X_i - x) g(\varepsilon_i, x) I\{\varepsilon_i \leq y\} - h \cdot \mu_0^{K^2} f_X(x) E[g(\varepsilon, x) I\{\varepsilon \leq y\} | X = x] \right\| = o_p(h).$$

Da g ggf. mehrdimensional ist, gehen im Weiteren komponentenweise vor. Sei $j \in \{1, \dots, k\}$, zur Anwendung von Theorem 2.39 und Bemerkung 2.40 definieren wir die Funktionenklasse \mathcal{F}_n wie folgt, x sei dabei fest:

$$\mathcal{F}_n = \mathcal{F}_n^x := \{ \varphi_{n,y} : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R} \mid \varphi_{n,y}(\varepsilon, z) = K^2\left(\frac{z-x}{h}\right) g_j(\varepsilon, x) I\{\varepsilon \leq y\}, y \in \mathbb{R} \}.$$

Auf diese Klasse wollen wir nun Theorem 2.39 anwenden. Da die Funktionenklasse keine konstante Einhüllende unabhängig von n besitzt sind zusätzlich die verallgemeinerten Bedingungen aus Bemerkung 2.40 nachzurechnen. Zuerst berechnen wir den Erwartungswert und das zweite Moment von $\varphi_{y,n}(\varepsilon_1, X_1)$ und bestimmen darüber δ_n und α_n . Darauf folgend rechnen wir die Bedingungen aus Bemerkung 2.40 nach, zeigen insbesondere, dass die Überdeckungszahl der Klasse \mathcal{F}_n für alle n der dort angegebenen Bedingung genügt. Es gilt gleichmäßig in $y \in \mathbb{R}$:

$$\begin{aligned} E[\varphi_{n,y}(\varepsilon_1, X_1)] &= E[K_h^2(X_1 - x) g_j(\varepsilon_1, x) I\{\varepsilon_1 \leq y\}] \\ &= h \cdot \int K^2(u) g_j(\varepsilon, x) I\{\varepsilon \leq y\} f_{\varepsilon, X}(\varepsilon, x + uh) du d\varepsilon \\ &= h \cdot \mu_0^{K^2} f_X(x) \int g_j(\varepsilon, x) I\{\varepsilon \leq y\} f_{\varepsilon|X}(\varepsilon|x) d\varepsilon + O(h^2) \\ &= h \cdot \mu_0^{K^2} f_X(x) E[g_j(\varepsilon, x) I\{\varepsilon \leq y\} | X = x] + O(h^2) \\ &= O(h) \end{aligned}$$

$$\begin{aligned} E[\varphi_{n,y}^2(\varepsilon_1, X_1)] &= E[K_h^4(X_1 - x) g_j^2(\varepsilon_1, x) I\{\varepsilon_1 \leq y\}] \\ &= h \cdot \mu_0^{K^4} \cdot f_X(x) E[g_j^2(\varepsilon, x) I\{\varepsilon \leq y\} | X = x] \\ &= O(h) \leq C \cdot h^{3/4}. \end{aligned}$$

Für jedes $\varphi_{y,n} \in \mathcal{F}_n$ existiert also der Erwartungswert von $\varphi_{y,n}(\varepsilon_1, X_1)$ und das zweite Moment und die Raten dazu ergeben sich wie oben angegeben gleichmäßig in $y \in \mathbb{R}$. Damit ist im

Hinblick auf Proposition 2.39 $\delta_n^2 = h^{3/4}$ und $\alpha_n = h^{1/4}/\log(n)^{1/2}$ und somit $o(\delta_n^2 \alpha_n) = o(h)$. Ebenso gilt unter der Voraussetzung (H1) die für Proposition 2.39 notwendige Voraussetzung $\frac{\log(n)}{n \delta_n^2 \alpha_n^2} = \frac{\log(n)^2}{n h^{5/4}} \rightarrow 0$ für $n \rightarrow \infty$, siehe (H2).

Im folgenden rechne nun die Bedingungen aus Bemerkung 2.40 nach, beginnend mit (ii) und (iii), dabei ist $\Phi(\varepsilon, z) = K^2(\frac{z-x}{h}) |g_j(\varepsilon, x)|$ die Einhüllende.

Zu (iii), nach Voraussetzung gilt $E[\Phi^2(\varepsilon, X)] < \infty$ und damit nach dem starken Gesetz der großen Zahlen, dass $\sup_{n \in \mathbb{N}} \frac{1}{n} \sum_{i=1}^n \Phi^2(\varepsilon, X) < \infty$ f.s. Voraussetzung (ii) ist gültig, denn es folgt mit der Markov-Ungleichung:

$$\begin{aligned} P\left(\sup_{\Phi \in \mathcal{F}_n} \frac{1}{n} \sum_{i=1}^n \varphi^2(\varepsilon_i, X_i) > h^{3/4}\right) &\leq P\left(\frac{1}{n} \sum_{i=1}^n K_h^4(X_i - x) g_j^2(\varepsilon_i, x) > h^{3/4}\right) \\ &\leq \frac{E(K_h^4(X_i - x) g_j^2(\varepsilon_i, x))}{h^{3/4}} \\ &= O\left(\frac{h}{h^{3/4}}\right) \\ &= o(1). \end{aligned}$$

Zur Voraussetzung (i):

Für die Überdeckungszahl der Klasse \mathcal{F}_n (s.o.) gilt die in Bemerkung 2.40 (i) geforderte Rate, siehe dazu Lemma 6.6 (i).

Nun kann man Proposition 2.39 bzw. Bemerkung 2.40 anwenden und erhält gleichmäßig in $y \in \mathbb{R}$:

$$\frac{1}{n} \sum_{i=1}^n K_h^2(X_i - x) g(\varepsilon_i, x) I\{\varepsilon_i \leq y\} = h \cdot \mu_0^{K^2} f_X(x) E[g(\varepsilon, x) I\{\varepsilon \leq y\} | X = x] + o_p(h)$$

Insgesamt ergibt sich gleichmäßig in $y \in \mathbb{R}$:

$$\hat{A}_{11}(n, x, y) = \frac{\mu_0^{K^2} f_X(x) E[g(\varepsilon, x) I\{\varepsilon \leq y\} | X = x]}{f_X(x)} + o_p(1)$$

und somit

$$\hat{A}_1(n, x, y) = -\hat{\eta}_n^t(x) \mu_0^{K^2} E[g(\varepsilon, x) I\{\varepsilon \leq y\} | X = x] + o_p\left(\frac{1}{\sqrt{nh}}\right)$$

$\hat{A}_2(n, x, y)$ zerlegt man ähnlich zu $\hat{A}_1(n, x, y)$, ersetzt $\hat{\varepsilon}_i$ durch ε_i auf die gleiche Weise wie dort und zeigt gleichmäßig in $y \in \mathbb{R}$:

$$\begin{aligned}\hat{A}_2(n, x, y) &= \hat{\eta}_n^t(x) \sum_{i=1}^n \frac{\frac{1}{nh} K_h(X_i - x) \frac{1}{nh} \sum_{j=1}^n K_h(X_j - x) g(\hat{\varepsilon}_j, x) K_h(X_j - x)}{\frac{1}{h} \sum_{j=1}^n K_h(X_j - x) \hat{p}_j(x) \frac{1}{nh} \sum_{j=1}^n K_h(X_j - x)} I\{\hat{\varepsilon}_i \leq y\} \\ &= \hat{\eta}_n^t(x) \frac{\sum_{i=1}^n \frac{1}{nh} K_h(X_i - x) I\{\varepsilon_i \leq y\} \frac{1}{nh} \sum_{j=1}^n K_h^2(X_j - x) g(\varepsilon_j, x)}{f_X^2(x)} + o_p\left(\frac{1}{\sqrt{nh}}\right) \\ &= o_p\left(\frac{1}{\sqrt{nh}}\right),\end{aligned}$$

dabei benutzt man maßgeblich für die letzte Gleichheit, dass für den Erwartungswert und die Varianz des zweiten Teil des Zählers die Rate $o(1)$ folgt.

Es fehlen noch die Terme $\hat{A}_3(n, x, y)$ und $\hat{A}_4(n, x, y)$. Betrachte Term $\hat{A}_3(n, x, y)$ genauer, $\hat{A}_4(n, x, y)$ schätze dann analog ab. Der Nenner von

$$\begin{aligned}\hat{A}_3(n, x, y) &= \sum_{i=1}^n \frac{\frac{1}{nh} K_h(X_i - x) \frac{1}{n^2 h} \sum_{j=1}^n K_h(X_j - x) \left(\frac{(\hat{\eta}_n^t(x) g(\hat{\varepsilon}_i, x) K_h(X_i - x))^2}{1 + \hat{\eta}_n^t(x) g(\hat{\varepsilon}_i, x) K_h(X_i - x)} \right)}{\frac{1}{nh} \sum_{j=1}^n K_h(X_j - x) \hat{p}_j(x) \frac{1}{nh} \sum_{j=1}^n K_h(X_j - x)} I\{\hat{\varepsilon}_i \leq y\} \\ &= \sum_{i=1}^n \frac{\frac{1}{nh} K_h(X_i - x) \left(\frac{(\hat{\eta}_n^t(x) g(\hat{\varepsilon}_i, x) K_h(X_i - x))^2}{1 + \hat{\eta}_n^t(x) g(\hat{\varepsilon}_i, x) K_h(X_i - x)} \right)}{\frac{1}{h} \sum_{j=1}^n K_h(X_j - x) \hat{p}_j(x)} I\{\hat{\varepsilon}_i \leq y\}\end{aligned}$$

ist von der Ordnung $O_p(1)$, siehe $\hat{A}_{112}(n, x, y)$. Betrachte noch den Zähler, nutze wiederum Hilfsatz 4.9:

$$\begin{aligned}& \left| \sum_{i=1}^n \frac{1}{nh} K_h^3(X_i - x) \frac{(\hat{\eta}_n^t(x) g(\hat{\varepsilon}_i, x))^2}{1 + \hat{\eta}_n^t(x) g(\hat{\varepsilon}_i, x) K_h(X_i - x)} I\{\hat{\varepsilon}_i \leq y\} \right| \\ & \leq \|\hat{\eta}_n(x)\|^2 \max_{1 \leq i \leq n} \left| \frac{1}{1 + \hat{\eta}_n^t(x) g(\hat{\varepsilon}_i, x) K_h(X_i - x)} \right| \cdot \frac{1}{nh} \sum_{i=1}^n |K_h^2(X_i - x) g(\hat{\varepsilon}_i, x) g^t(\hat{\varepsilon}_i, x)| \|K\|_\infty \\ & = O_p\left(\frac{1}{nh}\right) \cdot O_p(1) \cdot O_p(1) \\ & = o_p\left(\frac{1}{\sqrt{nh}}\right)\end{aligned}$$

Damit ist Proposition 5.2 und somit auch 5.1 (i) bewiesen.

□

Beweise als letztes die in 5.1 (ii) gegebene Differenz und ihre Darstellung. Es gilt:

$$\begin{aligned}
\hat{F}_n(y|x) - F_n(y|x) &= \frac{1}{nh} \sum_{i=1}^n \frac{K_h(X_i - x)}{\frac{1}{nh} \sum_{j=1}^n K_h(X_j - x)} \cdot (I\{\hat{\varepsilon}_i \leq y\} - I\{\varepsilon_i \leq y\}) \\
&= \frac{1}{nh} \sum_{i=1}^n \frac{K_h(X_i - x)}{f_X(x)} \cdot (I\{\hat{\varepsilon}_i \leq y\} - I\{\varepsilon_i \leq y\}) \\
&\quad - \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) \cdot (I\{\hat{\varepsilon}_i \leq y\} - I\{\varepsilon_i \leq y\}) \cdot \frac{\hat{f}_{h,X}(x) - f_X(x)}{f_X(x) \cdot \hat{f}_{h,X}(x)} \\
&=: \hat{B}_1(n, x, y) + \hat{B}_2(n, x, y)
\end{aligned}$$

Für den zweiten Term erhält man mittels der Rechnungen zum Term $\hat{B}_1(n, x, y)$ und $\|f_X - \hat{f}_{h,X}\|_\infty = o(1)$ gleichmäßig in $y \in \mathbb{R}$ die Rate $o_p(\frac{1}{\sqrt{nh}})$. Um die Darstellung Proposition 5.1 (ii) zu erhalten muss der Term $\hat{B}_1(n, x, y)$ noch weiter entwickelt werden. Für den Beweis der Entwicklung nutze Proposition 2.45 :

Proposition 5.4 *Unter den Annahmen (M1) – (M3), (K), (H1), (H2) gilt:*

$$\begin{aligned}
\sup_{y \in \mathbb{R}} \left| \frac{1}{nh} \sum_{i=1}^n \frac{K_h(X_i - x) \left(I\{\hat{\varepsilon}_i \leq y\} - I\{\varepsilon_i \leq y\} \right)}{f_X(x)} \right. \\
\left. - f_X^{-1}(x) \int \frac{1}{h} K_h(z - x) \left(I\{\varepsilon \leq y - m(z) + \hat{m}(z)\} - I\{\varepsilon \leq y\} \right) f_{\varepsilon, X}(\varepsilon, z) d(\varepsilon, z) \right| = o_p\left(\frac{1}{\sqrt{nh}}\right)
\end{aligned}$$

Beweis:

Zum Beweis der Proposition werden wir Proposition 2.45 anwenden. Wir zeigen, dass

$$\sup_{\varphi \in \mathcal{F}_n} \left| \frac{1}{n} \sum_{i=1}^n \varphi(\varepsilon_i, X_i) - E[\varphi(\varepsilon_1, X_1)] \right| = o\left(\frac{1}{\sqrt{n}}\right)$$

dabei sei φ aus der Funktionenklasse

$$\mathcal{F}_n := \left\{ \varphi : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R} \mid \varphi(\varepsilon, z) = \frac{1}{\sqrt{h}} K\left(\frac{z - x}{h}\right) \left(I\{\varepsilon \leq y - \tilde{h}(z)\} - I\{\varepsilon \leq y\} \right) \tilde{h} \in \mathcal{H}, y \in \mathbb{R} \right\}$$

mit $\mathcal{H} = C_\delta^{1+\alpha}[0, 1]$, siehe Definition 2.34. Die Abhängigkeit der Funktionenklasse von n ist dabei gegeben durch die von n abhängige Bandbreite $h = h_n$. Die Funktionenfolge $(\mathcal{F}_n)_{n \in \mathbb{N}}$ ist Donsker. Dies gilt nach Beispiel 6.7 (vi). Der dort geführte Beweis wird über den Nachweis der asympt. stochastischen gleich-Stetigkeit des Prozesses bzgl. der (im Beweis angegebenen) Semimetrik ρ geführt. Um die obige Rate und Darstellung nachzuweisen, sind nun die weiteren Voraussetzungen von Proposition 2.45 nachzurechnen. Betrachte die zufällige Funktion

$$\hat{\varphi}_n(\varepsilon, z) = \frac{1}{\sqrt{h}} K\left(\frac{z-x}{h}\right) \left(I\{\varepsilon \leq y + \hat{m}(z) - m(z)\} - I\{\varepsilon \leq y\} \right)$$

Für $(\hat{m} - m)$ gilt nach Akritas und van Keilegom (2001), dass $P(\hat{m} - m \in \mathcal{H}) \rightarrow 1$ für $n \rightarrow \infty$ und somit φ_n asymptotisch ein Element der Klasse \mathcal{F}_n ist. Ebenso gilt $\|\hat{m} - m\|_\infty = o(1)$ in Wahrscheinlichkeit. Betrachtet man die im Beweis zu 6.7 (vi) genutzte Semimetrik ρ , so ergibt sich $\|\rho((y, c_n), (y, 0))\| = o(1)$ für alle $c_n \in \mathcal{H}$ mit $\|c_n\|_\infty = o(1)$. Damit sind die Voraussetzungen von Proposition 2.45 erfüllt und es gilt:

$$\sup_{\varphi_n \in \mathcal{F}_n} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\varphi}_n(\varepsilon_i, X_i) - \int \hat{\varphi}_n dP \right| = o_p(1),$$

wobei P die Verteilung von (ε, X) bezeichnet. Es folgt Proposition 5.4

□

Wir benutzen diese Proposition um den Term $B_1(n, x, y)$ umzuformen und führen im nächsten Schritt die Entwicklung noch weiter mittels Substitution und Taylorentwicklung:

$$\begin{aligned} \hat{B}_1(n, x, y) &= \frac{1}{nh} \sum_{i=1}^n \frac{K_h(X_i - x)}{f_X(x)} \cdot (I\{\hat{\varepsilon}_i \leq y\} - I\{\varepsilon_i \leq y\}) \\ &\stackrel{5.4}{=} f_X^{-1}(x) \int \frac{1}{h} K_h(z - x) \left(I\{\varepsilon \leq y - m(z) + \hat{m}(z)\} - I\{\varepsilon \leq y\} \right) f_{\varepsilon, X}(\varepsilon, z) d(\varepsilon, z) + o_p\left(\frac{1}{\sqrt{nh}}\right) \\ &= f_X^{-1}(x) \int \frac{1}{h} K_h(z - x) \int \left(I\{\varepsilon \leq y - m(z) + \hat{m}(z)\} - I\{\varepsilon \leq y\} \right) f_{\varepsilon|X}(\varepsilon|z) d\varepsilon f_X(z) dz + o_p\left(\frac{1}{\sqrt{nh}}\right) \\ &= f_X^{-1}(x) \int \frac{1}{h} K_h(z - x) \left(F_{\varepsilon|X}(y - m(z) + \hat{m}(z)|z) - F_{\varepsilon|X}(y|z) \right) f_X(z) dz + o_p\left(\frac{1}{\sqrt{nh}}\right) \\ &= f_X^{-1}(x) \int \frac{1}{h} K_h(z - x) \left(f_{\varepsilon|X}(y|z)(\hat{m}(z) - m(z)) + \frac{1}{2} \frac{\partial f_{\varepsilon|X}(\varepsilon|u)}{\partial \varepsilon} \Big|_{(\varepsilon|u)=(\xi|z)} (\hat{m}(z) - m(z))^2 \right) f_X(z) dz \end{aligned}$$

$$\begin{aligned}
& + o_p\left(\frac{1}{\sqrt{nh}}\right) \\
\stackrel{(1)}{=} & f_X^{-1}(x) \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(y|z) (\hat{m}(z) - m(z)) f_X(z) dz + o_p\left(\frac{1}{\sqrt{nh}}\right) \\
\stackrel{Def.}{=} & f_X^{-1}(x) \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(y|z) \left(\frac{1}{n} \sum_{i=1}^n \frac{1}{b} \tilde{K}_b(X_i - z) Y_i \frac{1}{\hat{f}_{b,X}(z)} - m(z) \right) f_X(z) dz + o_p\left(\frac{1}{\sqrt{nh}}\right) \\
= & f_X^{-1}(x) \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(y|z) \frac{1}{\hat{f}_{b,X}(z)} \frac{1}{n} \sum_{i=1}^n \frac{1}{b} \tilde{K}_b(X_i - z) (m(X_i) - m(z)) f_X(z) dz \\
& + f_X^{-1}(x) \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(y|z) \frac{1}{\hat{f}_{b,X}(z)} \frac{1}{n} \sum_{i=1}^n \frac{1}{b} \tilde{K}_b(X_i - z) \varepsilon_i f_X(z) dz + o_p\left(\frac{1}{\sqrt{nh}}\right) \\
\stackrel{(2)}{=} & f_X^{-1}(x) \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(y|z) \frac{1}{n} \sum_{i=1}^n \frac{1}{b} \tilde{K}_b(X_i - z) (m(X_i) - m(z)) dz \\
& + f_X^{-1}(x) \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(y|z) \frac{1}{n} \sum_{i=1}^n \frac{1}{b} \tilde{K}_b(X_i - z) \varepsilon_i dz \\
& + f_X^{-1}(x) \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(y|z) \frac{f_X(z) - \hat{f}_{b,X}(z)}{\hat{f}_{b,X}(z)} \frac{1}{n} \sum_{i=1}^n \frac{1}{b} \tilde{K}_b(X_i - z) (m(X_i) - m(z)) dz \\
& + f_X^{-1}(x) \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(y|z) \frac{f_X(z) - \hat{f}_{b,X}(z)}{\hat{f}_{b,X}(z)} \frac{1}{n} \sum_{i=1}^n \frac{1}{b} \tilde{K}_b(X_i - z) \varepsilon_i dz + o_p\left(\frac{1}{\sqrt{nh}}\right) \\
=: & \hat{B}_{11}(n, x, y) + \hat{B}_{12}(n, x, y) + \hat{B}_{13}(n, x, y) + \hat{B}_{14}(n, x, y) + o_p\left(\frac{1}{\sqrt{nh}}\right)
\end{aligned}$$

Zu Zeile (1): hier hat man den zweiten Term der Taylorentwicklung mithilfe der folgenden Rechnung abgeschätzt:

$$\begin{aligned}
& \left| \int \frac{1}{h} K_h(z-x) \frac{\partial f_{\varepsilon|X}(\varepsilon|t)}{\partial \varepsilon} \Big|_{(\varepsilon|t)=(\xi|z)} (\hat{m}(z) - m(z))^2 f_X(z) dz \right| \\
& \leq \|(\hat{m} - m)^2\|_{\infty} \int \left| \frac{1}{h} K_h(z-x) \frac{\partial f_{\varepsilon|X}(\varepsilon|t)}{\partial \varepsilon} \Big|_{(\varepsilon|t)=(\xi|z)} f_X(z) \right| dz \\
& = \|(\hat{m} - m)^2\|_{\infty} \int \left| K(u) \frac{\partial f_{\varepsilon|X}(\varepsilon|t)}{\partial \varepsilon} \Big|_{(\varepsilon|t)=(\xi|x+uh)} f_X(x+uh) \right| du \\
& \leq 2 \|K\|_{\infty} \cdot \sup_{y \in \mathcal{R}, u \in [0,1]} \left| \frac{\partial f_{\varepsilon|X}(\varepsilon|u)}{\partial \varepsilon} \Big|_{\varepsilon=y} \cdot \|f_X\|_{\infty} \cdot \|(\hat{m} - m)^2\|_{\infty} \\
& \leq O(1) O_p\left(\frac{\log(b^{-1})}{nb}\right) = o_p\left(\frac{1}{\sqrt{nh}}\right),
\end{aligned}$$

wobei die letzte Rate mit den Voraussetzungen (H2) folgt.

Für den ersten und einzigen noch übrig bleibenden Summanden der Taylorentwicklung setzt man in der darauf folgenden Zeile die Definition des Schätzers der Modellfunktion, \hat{m} , ein und schiebt zur weiteren Vereinfachung noch Terme so dazwischen (siehe (2)), dass $\frac{1}{f_X(z)}$ in den führenden Termen nicht mehr auftritt.

$\hat{B}_{11}(n, x, y)$ und $\hat{B}_{12}(n, x, y)$ übernehmen wir in die Enddarstellung. Betrachte im folgenden noch die Summanden $\hat{B}_{13}(n, x, y)$, $\hat{B}_{14}(n, x, y)$ und zeige für beide gleichmäßig in $y \in \mathbb{R}$ die Rate $o_p(\frac{1}{\sqrt{nh}})$.

Betrachte Term $\hat{B}_{14}(n, x, y)$ genauer:

$$\hat{B}_{14}(n, x, y) = \int \frac{f_X(z) - \hat{f}_{b,X}(z)}{\hat{f}_{b,X}(z)} \frac{1}{h} K_h(z - x) f_{\varepsilon|X}(y|z) \frac{1}{n} \sum_{i=1}^n \frac{1}{b} \tilde{K}_b(X_i - z) \varepsilon_i dz$$

Problematisch ist, dass $\hat{f}_{b,X}(x)$ im Nenner auftritt. Schiebe Terme so dazwischen, dass im Nenner $f_X(x)$ anstelle des Schätzers steht:

$$(5.3) \quad \begin{aligned} \hat{B}_{14}(n, x, y) &= \frac{1}{n} \sum_{i=1}^n \int \frac{f_X(z) - \hat{f}_{b,X}(z)}{f_X(z)} \frac{1}{h} K_h(z - x) f_{\varepsilon|X}(y|z) \frac{1}{b} \tilde{K}_b(X_i - z) \varepsilon_i dz \\ &+ \int \frac{(f_X(z) - \hat{f}_{b,X}(z))^2}{\hat{f}_{b,X}(z) f_X(z)} \frac{1}{h} K_h(z - x) f_{\varepsilon|X}(y|z) \frac{1}{n} \sum_{i=1}^n \frac{1}{b} \tilde{K}_b(X_i - z) \varepsilon_i dz \end{aligned}$$

Zum Beweis der $o_p(\frac{1}{\sqrt{nh}})$ Rate des zweiten Summanden schätze den Term mittels Supremum grob ab und zeige für den Anteil $\frac{1}{b} \sup_{z \in [0,1]} \frac{1}{n} \sum_{i=1}^n |\tilde{K}_b(X_i - z) \varepsilon_i|$ die Rate $O_p(1)$ mittels Proposition 2.39 und Bemerkung 2.40, dabei ist $\alpha_n = \frac{1}{\log(n)}$ und $\delta_n = \sqrt{b \log(n)}$. Zusätzlich nutze $\sup_{z \in [0,1]} |(f_X(z) - \hat{f}_{b,X}(z))^2| = O(n^{-1} b^{-1} \log(b^{-1}))$ aus 2.55 (a^*).

Den ersten Summanden der Darstellung (5.3) betrachte gesondert. Hier ist wieder eine Rate gleichmäßig in $y \in \mathbb{R}$ nötig, so dass es nicht ausreicht Erwartungswert und Varianz zu berechnen. Wir wenden Proposition 2.45 an. Betrachte dazu den folgenden Prozess:

$$(5.4) \quad G_n(\varphi_y) := \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\varphi_y(\varepsilon_i, X_i) - \int \int \varphi_y(\nu, z) f_{\varepsilon, X}(\nu, z) dz d\nu \right),$$

mit der von n abhängenden Funktionenklasse:

$$\mathcal{F}_n := \{ \varphi_{y,n} : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R} \mid \varphi_{y,n}(\varepsilon, z) = \varepsilon \cdot \int \tilde{h}(u) \frac{1}{\sqrt{h}} K\left(\frac{u-x}{h}\right) f_{\varepsilon|X}(y|u) \frac{1}{b} \tilde{K}\left(\frac{z-u}{b}\right) du, y \in \mathbb{R}, \tilde{h} \in \mathcal{H} \}$$

$\mathcal{H} := C_{\delta}^{1+\alpha}[0, 1]$, siehe zu dieser Klasse Definition 2.34. Im folgenden sind nun die Voraussetzungen von Proposition 2.45 nachzurechnen. $(\mathcal{F}_n)_{n \in \mathbb{N}}$ ist nach Lemma 6.7 (vii) Donsker und der Prozess ist asympt. stoch. gleich-stetig bzgl. der im Beweis angegebenen Semimetrik ρ , für $\hat{f}_{b,X} - f_X$ gilt $\|\hat{f}_{b,X} - f_X\|_{\infty} = o(1)$. Betrachtet man die im Beweis zu 6.7 (vii) genutzte Semimetrik ρ , so ergibt sich $\|\rho(y, c_n) - \rho(y, 0)\| = o(1)$ für alle $c_n \in \mathcal{H}$ mit $\|c_n\|_{\infty} = o(1)$.

In den Prozess G_n aus (5.4) setzen wir

$$\hat{\varphi}_{y,n}(\varepsilon, z) = \varepsilon \cdot \int (\hat{f}_{b,X}(u) - f_X(u)) \frac{1}{\sqrt{h}} K\left(\frac{u-x}{h}\right) f_{\varepsilon|X}(y|u) \frac{1}{b} \tilde{K}\left(\frac{z-u}{b}\right) du$$

Für $(\hat{f}_{b,X} - f_X)$ gilt nach Akritas und van Keilegom (2001), dass $P(\hat{f}_{b,X} - f_X \in \mathcal{H}) \rightarrow 1$ für $n \rightarrow \infty$ und somit $\varphi_{y,n}$ asymptotisch ein Element der Klasse \mathcal{F}_n ist.

Damit sind alle Voraussetzungen von Proposition 2.45 erfüllt und wir erhalten gleichmäßig in y die Darstellung

$$\begin{aligned} \hat{B}_{14}(n, x, y) &= \int \int \varepsilon \cdot \int (\hat{f}_{b,X}(u) - f_X(u)) \frac{1}{h} K\left(\frac{u-x}{h}\right) f_{\varepsilon|X}(y|u) \frac{1}{b} \tilde{K}\left(\frac{z-u}{b}\right) du f_{\varepsilon,X}(\varepsilon, z) dz d\varepsilon + o_p\left(\frac{1}{\sqrt{nh}}\right) \\ &= o_p\left(\frac{1}{\sqrt{nh}}\right) \end{aligned}$$

Es ergibt sich also gleichmäßig in y : $\hat{B}_{14}(n, x, y) = o_p\left(\frac{1}{\sqrt{nh}}\right)$

Setzt man nun alle Teile zusammen, so ergibt sich die Darstellung 5.1 (ii):

$$\begin{aligned} \hat{F}_n(y|x) - F_n(y|x) &= \hat{B}(n, x, y) \\ &= \hat{B}_1(n, x, y) + \hat{B}_2(n, x, y) \\ &= \hat{B}_1(n, x, y) + o_p\left(\frac{1}{\sqrt{nh}}\right) \\ &= \hat{B}_{11}(n, x, y) + \hat{B}_{12}(n, x, y) + \hat{B}_{13}(n, x, y) + \hat{B}_{14}(n, x, y) + o_p\left(\frac{1}{\sqrt{nh}}\right) \end{aligned}$$

$$\begin{aligned}
(5.5) \quad &= \hat{B}_{11}(n, x, y) + \hat{B}_{12}(n, x, y) + o_p\left(\frac{1}{\sqrt{nh}}\right) \\
&= f_X^{-1}(x) \frac{1}{n} \sum_{i=1}^n \varepsilon_i \int \frac{1}{h} K\left(\frac{z-x}{h}\right) f_{\varepsilon|X}(y|z) \frac{1}{b} \tilde{K}\left(\frac{X_i-z}{b}\right) dz \\
&\quad + f_X^{-1}(x) \frac{1}{n} \sum_{i=1}^n \int \frac{1}{h} K\left(\frac{z-x}{h}\right) f_{\varepsilon|X}(y|z) \frac{1}{b} \tilde{K}\left(\frac{X_i-z}{b}\right) (m(X_i) - m(z)) dz \\
&\quad + o_p\left(\frac{1}{\sqrt{nh}}\right)
\end{aligned}$$

Der Beweis zu Proposition 5.1 (Entwicklung der Differenz $\hat{F}_n^*(y|x) - F_{\varepsilon|X}(y|x)$ als Summe unabhängiger Zufallsvariablen zzgl. eines Resttermes der (in y gleichmäßigen) Rate $o_p(\frac{1}{\sqrt{nh}})$) ist damit abgeschlossen.

□

Bemerkung 5.5 *In diesem Kapitel treten die Bandbreiten b und h nicht in Wechselwirkung, es werden aber explizit die in den Voraussetzungen (H1) und (H2) zusammengefassten Raten der Bandbreiten genutzt, speziell die daraus folgenden Raten für $\|\hat{m} - m\|_\infty$, $\|\hat{f}_{h,X} - f_X\|_\infty$ etc. und die zum Beweis der Donskereigenschaft auftretender Funktionenklassen nötigen Bandbreitenannahmen. Da die Beweise zur Donskereigenschaft von Funktionenklassen in das nächste Kapitel, Lemma 6.7 ausgelagert sind, wird darauf nochmals eingegangen werden.*

Mit diesen Entwicklungen wollen wir nun zur Berechnung des Erwartungswertes übergehen.

6 Erwartungswert, Kovarianz und schwache Konvergenz

Bemerkung 6.1 *Wir betrachten der Übersicht halber im Weiteren nur eindimensionale die Nebeninformation beschreibende Funktionen g vom Typ $g(\varepsilon, x) = I\{\varepsilon \leq a(x)\} - d(x)$. Analoge Aussagen lassen sich aber auf die gleiche Weise auch für diesen Zusatzinformationstyp mehrdimensional angeben. Für Schätzer auf Basis die Zusatzinformation beschreibender Funktionen g vom Typ (G1) und (G2) geben wir mehrdimensionale Resultate an; ebenso können auch für Mischungen der beiden Typen g Resultate angegeben werden, diese ergeben sich direkt mit den Rechnungen zu beiden Typen und werden in dieser Arbeit nicht gesondert angegeben.*

Im vorherigen Kapitel haben wir für die Differenz $\hat{F}_n^*(y|x) - F_{\varepsilon|X}(y|x)$ eine Darstellung als Summe unabhängiger Zufallsvariablen zzgl. eines Resttermes der (in $y \in \mathbb{R}$ gleichmäßigen) Rate $o_p(\frac{1}{\sqrt{nh}})$ hergeleitet. In diesem Kapitel werden wir nun Erwartungswert und Varianz der Darstellung ohne den Restterm berechnen. Setze im Weiteren

$$(6.1) \quad A(y, x) := \mu_0^{K^2} \cdot E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = x]$$

$$\begin{aligned} B_n(y, x) &:= f_X^{-1}(x) \frac{1}{n} \sum_{i=1}^n \int \frac{1}{h} K\left(\frac{z-x}{h}\right) f_{\varepsilon|X}(y|z) \frac{1}{b} \tilde{K}\left(\frac{X_i-z}{b}\right) (m(X_i) - m(z)) dz \\ &\quad + f_X^{-1}(x) \frac{1}{n} \sum_{i=1}^n \varepsilon_i \int \frac{1}{h} K\left(\frac{z-x}{h}\right) f_{\varepsilon|X}(y|z) \frac{1}{b} \tilde{K}\left(\frac{X_i-z}{b}\right) dz \end{aligned}$$

$$C_n(y, x) := \frac{\frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i-x}{h}\right) (I\{\varepsilon_i \leq y\} - F_{\varepsilon|X}(y|x))}{f_X(x)}$$

Es gilt nach den vorangegangenen Rechnungen (siehe Gleichung (5.1) und Proposition 5.1) gleichmäßig in $y \in \mathbb{R}$ die Entwicklung:

$$\hat{F}_n^*(y|x) - F_{\varepsilon|X}(y|x) = -\hat{\eta}_n^t(x) \cdot A(y, x) + B_n(y, x) + C_n(y, x) + o_p\left(\frac{1}{\sqrt{nh}}\right)$$

Mithilfe der Entwicklung von $\hat{\eta}_n^t(x)$ (siehe Proposition 4.12) erhält man:

$$(6.2) \quad \hat{F}_n^*(y|x) - F_{\varepsilon|X}(y|x) = R_n(x, y) + o_p\left(\frac{1}{\sqrt{nh}}\right)$$

wobei

$$(6.3) \quad R_n(x, y) := -(\Sigma^{-1}(x) \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) g(\varepsilon_i, x))^t \cdot A(y, x) + B_n(y, x) + C_n(y, x) \\ - r_n^t(x) \cdot A(y, x)$$

$r_n(x)$ ist je nach der Gestalt von g definiert. Wir wollen im folgenden den Erwartungswert von $R_n(x, y)$ berechnen. Für eine die Zusatzinformation beschreibende Funktion g mit der Entwicklung aus Proposition 4.12 (E1) ist

$$r_n(x) =: r_{n,1}(x) = -\Sigma^{-1}(x) \frac{1}{n} \sum_{i=1}^n \varepsilon_i \int \frac{1}{h} K_h(z - x) g'(y, x) \frac{1}{b} \tilde{K}_b(X_i - z) f_{\varepsilon|X}(y|z) d(y, z) \\ - \Sigma^{-1}(x) \frac{1}{n} \sum_{i=1}^n \int \frac{1}{h} K_h(z - x) g'(y, x) \frac{1}{b} \tilde{K}_b(X_i - z) (m(X_i) - m(z)) f_{\varepsilon|X}(y|z) d(y, z) \\ + o_p\left(\frac{1}{\sqrt{nh}}\right)$$

Für Funktionen g mit der Entwicklung aus Proposition 4.12 (E2) ist

$$r_n(x) =: r_{n,2}(x) = \Sigma^{-1}(x) \frac{1}{n} \sum_{i=1}^n \varepsilon_i \int \frac{1}{h} K_h(z - x) f_{\varepsilon|X}(a(x)|z) \frac{1}{b} \tilde{K}_b(X_i - z) dz \\ + \Sigma^{-1}(x) \frac{1}{n} \sum_{i=1}^n \int \frac{1}{h} K_h(z - x) f_{\varepsilon|X}(a(x)|z) \frac{1}{b} \tilde{K}_b(X_i - z) (m(X_i) - m(z)) dz \\ + o_p\left(\frac{1}{\sqrt{nh}}\right),$$

dabei gilt die Rate $o_p(\frac{1}{\sqrt{nh}})$ der abgeschätzten Restterme jeweils gleichmäßig in $y \in \mathbb{R}$.

6.1 Erwartungswert

Im folgenden berechnen wir den Erwartungswert von $R_n(x, y)$. Wie im letzten Kapitel betrachte zur Berechnung die Summanden getrennt; das Vorgehen ist jetzt, da die Ausdrücke durch die Darstellung in Summen unabhängiger Zufallsvariablen vereinfacht sind, nicht mehr sonderlich problematisch.

Im Anschluss folgt dann noch eine Proposition zur Kovarianzfunktion des Prozesses $G(\cdot, x) := \sqrt{nh} (\hat{F}_n^*(\cdot | x) - F_{\varepsilon|X}(\cdot | x) - b_n^*(\cdot, x))$, dabei wird der abzuziehende Korrekturterm $b_n^*(\cdot, x)$ mittels der obigen Proposition zum Erwartungswert und Varianzberechnungen festgelegt, d.h. er wird gleich denjenigen Termen der Entwicklung gesetzt, die im Erwartungswert die Rate $O(\frac{1}{\sqrt{nh}})$ und für die Varianz die Rate $o(\frac{1}{nh})$ liefern. Die Terme, welche nicht Erwartungswert gleich 0 und eine Varianz der Ordnung $O(\frac{1}{nh})$ haben werden durch Abziehen des Erwartungswertes erwartungstreu gemacht. Korrigierend fließt der Erwartungswert dann noch in den Korrekturterm ein. Die genaue Aufteilung folgt an späterer Stelle.

Proposition 6.2 *Für die obige Entwicklung (6.2) und (6.3) von $\hat{F}_n^*(y|x) - F_{\varepsilon|X}(y|x)$ gilt unter den Voraussetzungen (M1) - (M3), (K), (H1), (H2), mit g erfüllend (G1), (G2) oder g Indikatorfunktion der Gestalt $g(\varepsilon, x) = I\{\varepsilon \leq a(x)\} - d(x)$:*

$$\begin{aligned}
E[R_n(x, y)] &= E\left[-(\Sigma^{-1}(x) \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) g(\varepsilon_i, x))^t \cdot A(y, x)\right] \\
&\quad + E[B_n(y, x)] \\
&\quad + E[C_n(y, x)] \\
&\quad - E[r_n^t(x) \cdot A(y, x)] \\
&= -\frac{h^2}{2} \mu_2^K \frac{\partial^2 E[g^t(\varepsilon, x) | X = t] \cdot f_X(t)}{\partial t^2} \Big|_{t=x} \cdot \Sigma^{-1}(x)^t \cdot A(y, x) + o(h^2) \\
&\quad + \frac{b^2}{2} \mu_2^{\tilde{K}} f_X^{-1}(x) f_{\varepsilon|X}(y|x) B(x) + o(b^2) \\
&\quad + \frac{h^2}{2} \cdot \mu_2^K \cdot (F_{\varepsilon|X}^{(2)}(y|x) + \frac{2f'_X(x)}{f_X(x)} F_{\varepsilon|X}^{(1)}(y|x)) + o(h^2) \\
&\quad - E[r_n^t(x) \cdot A(y, x)]
\end{aligned}$$

und mit $r_n(x)$ jeweils passend zu der die Nebeninformation beschreibenden Funktion g (vgl. (6.3) ff.):

$$E[r_{n,1}(x)] = -\Sigma^{-1}(x) \frac{b^2}{2} \mu_2^{\tilde{K}} E[g'(\varepsilon, X) | X = x] B(x) + o(b^2)$$

$$E[r_{n,2}(x)] = \Sigma^{-1}(x) \frac{b^2}{2} \mu_2^{\tilde{K}} f_{\varepsilon|X}(a(x)|x) B(x) + o(b^2)$$

Dabei ist $B(x) := (m \cdot f_X)''(x) - m(x) f_X''(x)$ und $F_{\varepsilon|X}^{(j)}(y|x) := \frac{\partial^j F_{\varepsilon|X}(y|x)}{\partial x^j}$, $j = 1, 2$.

Beweis von Proposition 6.2:

Zur Berechnung des Erwartungswertes von Term $C_n(y, x)$ (siehe (6.3)) nutze Proposition 2.54, welche Eigenschaften des Kerndichteschätzers zusammen trägt. Damit ergibt sich:

$$\begin{aligned} E[C_n(y, x)] &= f_X^{-1}(x) \cdot E\left[\frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) (I\{\varepsilon_i \leq y\} - F_{\varepsilon|X}(y|x))\right] \\ (6.4) \quad &\stackrel{2.54}{=} f_X^{-1}(x) \cdot \left(\frac{h^2}{2!} \cdot \mu_2^K \cdot \left(\int_{(-\infty, y]} \ddot{f}_{\varepsilon, X}(\varepsilon, x) d\varepsilon - F_{\varepsilon|X}(y|x) f_X''(x)\right)\right) + o(h^2) \\ &= \frac{h^2}{2!} \cdot \mu_2^K \cdot \left(F_{\varepsilon|X}^{(2)}(y|x) + \frac{2 f_X'(x)}{f_X(x)} F_{\varepsilon|X}^{(1)}(y|x)\right) + o(h^2) \end{aligned}$$

Die letzte Umformung ergibt sich mit $F_{\varepsilon|X}^{(j)}(y|x) := \frac{\partial^j F_{\varepsilon|X}(y|x)}{\partial x^j}$, $j = 1, 2$, $\ddot{f}_{\varepsilon, X}(\varepsilon, x) := \frac{\partial^2 f(\varepsilon, x)}{\partial x^2}$ und einigen Zeilen von Umformungen, vgl. dazu auch Hall, Wolff & Yao (1999). In den Beispielen in Kapitel 7 greifen wir auf die besser berechenbare erste Darstellung zurück.

Für den ersten Summanden, $D_n(y, x) := -(\Sigma^{-1}(x) \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) g(\varepsilon_i, x))^t \cdot A(y, x)$ aus (6.3), ergibt sich:

$$\begin{aligned} E[D_n(y, x)] &= E\left[-(\Sigma^{-1}(x) \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) g(\varepsilon_i, x))^t \cdot A(y, x)\right] \\ &= -E\left[\frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) g^t(\varepsilon_i, x)\right] \cdot \Sigma^{-1}(x)^t \cdot A(y, x) \\ &= -\int K(u) g^t(\varepsilon, x) f_{\varepsilon, X}(\varepsilon, x + uh) d(\varepsilon, u) \cdot \Sigma^{-1}(x)^t \cdot A(y, x) \\ &= -\int K(u) g^t(\varepsilon, x) f_{\varepsilon|X}(\varepsilon|x + uh) f_X(x + uh) d(\varepsilon, u) \cdot \Sigma^{-1}(x)^t \cdot A(y, x) \\ &= -\int K(u) E[g^t(\varepsilon, x) | X = x + uh] \cdot f_X(x + uh) du \cdot \Sigma^{-1}(x)^t \cdot A(y, x) \\ &= -\left[f_X(x) \cdot \int \sum_{i=1}^n K(u) g^t(\varepsilon, x) f_{\varepsilon|X}(\varepsilon|x) d(\varepsilon, u)\right] \end{aligned}$$

$$\begin{aligned}
& + \frac{h^2}{2} \mu_2^K \frac{\partial^2 E[g^t(\varepsilon, x) | X = t] \cdot f_X(t)}{\partial t^2} \Big|_{t=x} \Big] \cdot \Sigma^{-1}(x)^t \cdot A(y, x) + o(h^2) \\
& = - \frac{h^2}{2} \mu_2^K \frac{\partial^2 E[g^t(\varepsilon, x) | X = t] \cdot f_X(t)}{\partial t^2} \Big|_{t=x} \cdot \Sigma^{-1}(x)^t \cdot A(y, x) + o(h^2)
\end{aligned}$$

Nun fehlt noch der Erwartungswert der unter $B_n(y, x)$ zusammengefassten Summanden aus der Entwicklung (6.3). $B_n(y, x)$ besteht aus zwei Teilen. Der Erwartungswert des ersten Terms ist leicht zu berechnen, der des zweiten etwas umständlicher:

$$\begin{aligned}
E[f_X^{-1}(x) \frac{1}{n} \sum_{i=1}^n \varepsilon_i \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(y|z) \frac{1}{b} \tilde{K}_b(X_i - z) dz] \\
& = E[f_X^{-1}(x) E[\varepsilon_i | X_i] \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(y|z) \frac{1}{b} \tilde{K}_b(X_i - z) dz] \\
& = 0 \\
E[f_X^{-1}(x) \frac{1}{n} \sum_{i=1}^n \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(y|z) \frac{1}{b} \tilde{K}_b(X_i - z) (m(X_i) - m(z)) dz] \\
& = f_X^{-1}(x) \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(y|z) \int \frac{1}{b} \tilde{K}\left(\frac{u-z}{b}\right) (m(u) - m(z)) f_X(u) du dz
\end{aligned}$$

Betrachte hier zuerst das innere Integral. Entwickele in einer Taylorreihe bis zur zweiten Ableitung, $\xi_1 := \xi_1(z, t, b)$ und $\xi_2 := \xi_2(z, t, b)$ bezeichnen dabei die dazugehörigen Zwischenstellen:

$$\begin{aligned}
& \int \frac{1}{b} \tilde{K}\left(\frac{u-z}{b}\right) (m(u) - m(z)) f_X(u) du \\
& = \int \tilde{K}(t) (m(z+tb) - m(z)) f_X(z+tb) dt \\
& = \int \tilde{K}(t) m(z+tb) f_X(z+tb) dt - \int \tilde{K}(t) m(z) f_X(z+tb) dt \\
& = \int \tilde{K}(t) ((m \cdot f_X)(z) + tb(m \cdot f_X)'(z)) dt \\
& \quad + \int t^2 \frac{b^2}{2} ((m \cdot f_X)''(z) + t^2 \frac{b^2}{2} ((m \cdot f_X)''(\xi_1) - (m \cdot f_X)''(z))) dt \\
& \quad - \int \tilde{K}(t) m(z) (f_X(z) + tb f_X'(z) + t^2 \frac{b^2}{2} f_X''(z) + t^2 \frac{b^2}{2} (f_X''(\xi_2) - f_X''(z))) dt \\
& = \int \frac{b^2}{2} t^2 \tilde{K}(t) ((m \cdot f_X)''(z) + R_1(\xi_1, t, b, z)) dt - \int \frac{b^2}{2} t^2 \tilde{K}(t) m(z) (f_X''(z) + R_2(\xi_2, t, b, z)) dt \\
& = \frac{b^2}{2} \mu_2^K \left((m \cdot f_X)''(z) - m(z) f_X''(z) \right) + \frac{b^2}{2} \int \tilde{K}(t) t^2 (R_1(\xi_1, t, b, z) - m(z) R_2(\xi_2, t, b, z)) dt
\end{aligned}$$

$$=: \frac{b^2}{2} \mu_2^{\tilde{K}} B(z) + \frac{b^2}{2} \int \tilde{K}(t) t^2 (R_1(\xi_1, t, b, z) - m(z) R_2(\xi_2, t, b, z)) dt$$

Dabei ist:

$$(6.5) \quad B(z) := (m \cdot f_X)''(z) - m(z) f_X''(z)$$

Die Zwischenstellenterme haben die Rate $o(b^2)$, was im Weiteren, wenn auch noch das äußere Integral ausgewertet ist, ersichtlich wird, dabei nutzt man direkt die Voraussetzungen (M2) und (K), d. h. die gleichmäßige Beschränktheit der Bauteile dieser Restterme. Betrachte jetzt also das äußere (Erwartungswert-)Integral:

$$\begin{aligned} E[f_X^{-1}(x) \frac{1}{n} \sum_{i=1}^n \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(y|z) \frac{1}{b} \tilde{K}_b(X_i - z) (m(X_i) - m(z)) dz] \\ &= f_X^{-1}(x) \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(y|z) \int \frac{1}{b} \tilde{K}\left(\frac{u-z}{b}\right) (m(u) - m(z)) f_X(u) du dz \\ &\stackrel{(1)}{=} \frac{b^2}{2} f_X^{-1}(x) \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(y|z) \mu_2^{\tilde{K}} B(z) dz \\ &\quad + \frac{b^2}{2} f_X^{-1}(x) \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(y|z) \int \tilde{K}(t) t^2 (R_1(\xi_1, t, b, z) - m(z) R_2(\xi_2, t, b, z)) dt dz \\ &\stackrel{(M2)}{=} \frac{b^2}{2} f_X^{-1}(x) \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(y|z) \mu_2^{\tilde{K}} B(z) dz + o(b^2) \\ &= \frac{b^2}{2} \mu_2^{\tilde{K}} f_X^{-1}(x) \int K(u) f_{\varepsilon|X}(y|x+uh) B(x+uh) du + o(b^2) \\ &\stackrel{(M2)}{=} \frac{b^2}{2} \mu_2^{\tilde{K}} f_X^{-1}(x) \int K(u) f_{\varepsilon|X}(y|x+uh) B(x) du + o(b^2) \\ &= \frac{b^2}{2} \mu_2^{\tilde{K}} f_X^{-1}(x) \int K(u) B(x) \left(f_{\varepsilon|X}(y|x) + uh \frac{\partial (f_{\varepsilon|X}(y|x))}{\partial x} + uh R_3(\xi_3(h, u, x)) \right) du \\ &\quad + o(b^2) \\ &= \frac{b^2}{2} \mu_2^{\tilde{K}} f_X^{-1}(x) f_{\varepsilon|X}(y|x) B(x) + o(b^2) \end{aligned}$$

Zu (1), hier nutze die vorangegangene Umformung des inneren Integrales. Zur Abschätzung der Restterme haben wir die Voraussetzungen (M2) genutzt, speziell auch die dort vorausgesetzte Lipschitzstetigkeit von $B(x)$.

Zum Ende der Erwartungswertbetrachtungen berechne noch den Erwartungswert derjenigen Terme, welche über die Entwicklung von $\hat{\eta}_n^t(x)$ dazu gekommen sind (vgl. (6.3) ff.) und von der Gestalt $E[r_n^t(x)] \cdot A(y, x)$ sind. Nutze dabei die Voraussetzung $E[\varepsilon | X = x] = 0$ (siehe (M1), (A)) aus.

Betrachte zuerst $E[r_{n,1}(x)]$, wobei man vorherige Rechnungen (siehe die Rechnung um Gleichung (6.5)) ausnutzt; $r_{n,2}(x)$ folgt später analog:

$$\begin{aligned}
E[r_{n,1}(x)] &= -\Sigma^{-1}(x) \left(\frac{1}{n} \sum_{i=1}^n E[\varepsilon_i \int \frac{1}{h} K_h(z-x) g'(y, x) \frac{1}{b} \tilde{K}_b(X_i - z) f_{\varepsilon|X}(y|z) d(y, z)] \right. \\
&\quad \left. + \frac{1}{n} \sum_{i=1}^n E\left[\int \frac{1}{h} K_h(z-x) g'(y, x) \frac{1}{b} \tilde{K}_b(X_i - z) \left(m(X_i) - m(z) \right) f_{\varepsilon|X}(y|z) d(y, z) \right] \right) \\
&= 0 \\
&\quad -\Sigma^{-1}(x) \frac{1}{n} \sum_{i=1}^n E\left[\int \frac{1}{h} K_h(z-x) g'(y, x) \frac{1}{b} \tilde{K}_b(X_i - z) \left(m(X_i) - m(z) \right) f_{\varepsilon|X}(y|z) d(y, z) \right] \\
&= -\Sigma^{-1}(x) \int \frac{1}{h} K_h(z-x) g'(y, x) f_{\varepsilon|X}(y|z) \int \frac{1}{b} \tilde{K}_b(u-z) (m(u) - m(z)) f_X(u) du d(y, z) \\
&= -\Sigma^{-1}(x) \int \frac{1}{h} K_h(z-x) g'(y, x) f_{\varepsilon|X}(y|z) \\
&\quad \cdot \left(\frac{b^2}{2} \mu_2^{\tilde{K}} B(z) + \frac{b^2}{2} \int \tilde{K}(t) t^2 (R_1(\xi_1, t, b, z) - m(z) R_2(\xi_2, t, b, z)) dt \right) d(y, z) \\
&= -\Sigma^{-1}(x) \frac{b^2}{2} \mu_2^{\tilde{K}} \int \frac{1}{h} K_h(z-x) g'(y, x) f_{\varepsilon|X}(y|z) B(z) d(y, z) + o(b^2)
\end{aligned}$$

Die Zwischenstellenterme haben die Rate $o(b^2)$, dabei nutzt man direkt die Voraussetzungen (M2) und (K) zur gleichmäßigen Beschränktheit der Bauteile dieser Restterme. Jetzt substituieren ein weiteres Mal, nutze die Lipschitzstetigkeit von $B(x)$ und entwickle $f_{\varepsilon|X}(\varepsilon | x + uh)$ um x bis zu einer Zwischenstelle der ersten Ableitung als Taylorreihe. Es ergibt sich:

$$E[r_{n,1}(x)] = -\Sigma^{-1}(x) \frac{b^2}{2} \mu_2^{\tilde{K}} E[g'(\varepsilon, X) | X = x] B(x) + o(b^2)$$

Für den Term $E[r_{n,2}(x)]$ ergibt sich auf die gleiche Weise:

$$\begin{aligned}
E[r_{n,2}(x)] &= E\left[\Sigma^{-1}(x) \frac{1}{n} \sum_{i=1}^n \varepsilon_i \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(a(x)|z) \frac{1}{b} \tilde{K}_b(X_i - z) dz \right] \\
&\quad + E\left[\Sigma^{-1}(x) \frac{1}{n} \sum_{i=1}^n \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(a(x)|z) \frac{1}{b} \tilde{K}_b(X_i - z) (m(X_i) - m(z)) dz \right]
\end{aligned}$$

$$= \Sigma^{-1}(x) \frac{b^2}{2} \mu_2^{\bar{K}} f_{\varepsilon|X}(a(x)|x) B(x) + o(b^2)$$

Damit ist Proposition 6.2 bewiesen.

□

6.2 Angabe des Prozesses

Jetzt ist mittels der Rechnungen zum Erwartungswert und weiterer Varianzrechnungen festzulegen, welche Summanden in den Prozess eingehen und welche in den Bias. Diejenigen Summanden, die nicht zentriert sind und eine Varianz der Rate $O(\frac{1}{nh})$ aufweisen, werden durch Abzug des Erwartungswertes der Terme zentriert, welcher dann folglich auch im abzuziehenden Bias enthalten sein muss.

Diejenigen Terme, welche eine Varianz der Rate $o(\frac{1}{nh})$ aufweisen, werden ebenfalls mittels Abzug des Erwartungswertes zentriert, so dass der Erwartungswert ebenfalls im Bias wieder hinzuge-rechnet werden muss.

Zu betrachten ist nun, von welcher Rate die Varianz der Summanden der Entwicklung sind. Die Bezeichnungen entnehmen wir der Proposition 6.2 zum Erwartungswert. Die Varianzrechnungen der meisten Terme werden zur Berechnung der Kovarianz noch genau angegeben, hier werden die Rechnungen deshalb knapp gehalten oder entfallen ganz. Es gilt:

$$(6.6) \quad \text{Var}(C_n(y, x)) = O\left(\frac{1}{nh}\right)$$

Der Term $C_n(y, x)$ hat nach obiger Proposition Erwartung ungleich 0, auch keine Varianz der Ordnung $o(\frac{1}{nh})$, multipliziert mit \sqrt{nh} würde er einen Beitrag zum Erwartungswert und zur Varianz liefern. Um Erwartungswert gleich Null zu erhalten werden wir im weiteren Verlauf deshalb $C_n(y, x) - E[C_n(y, x)]$ zu dem uns interessierenden Prozess hinzu nehmen, denn dieser Term hat Erwartung 0 und Varianz $O(\frac{1}{nh})$. Zum Ausgleich wird wieder $E[C_n(y, x)]$ zum Bias-Term des uns interessierenden Prozesses (6.8) hinzugerechnet.

$$(6.7) \quad \text{Var}(D_n(y, x)) = \text{Var}\left(-(\Sigma^{-1}(x) \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) g(\varepsilon_i, x))^t \cdot A(y, x)\right) = O\left(\frac{1}{nh}\right)$$

Dieser Term ist ebenfalls nicht zentriert (vgl. Erwartungswert-Berechnungen) und hat auch keine Varianz der Ordnung $o(\frac{1}{nh})$, wir gehen deshalb ebenso wie zu Term $C_n(y, x)$ vor.

$$\begin{aligned} B_n(y, x) &= f_X^{-1}(x) \frac{1}{n} \sum_{i=1}^n \varepsilon_i \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(y|z) \frac{1}{b} \tilde{K}_b(X_i - z) dz \\ &\quad + f_X^{-1}(x) \frac{1}{n} \sum_{i=1}^n \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(y|z) \frac{1}{b} \tilde{K}_b(X_i - z) (m(X_i) - m(z)) dz \end{aligned}$$

Der erste Summand ist zentriert und hat eine Varianz der Rate $O(\frac{1}{nh})$, Der zweite Summand ist zwar nicht zentriert, hat aber eine deutlich kleinere Varianz:

$$\text{Var}(f_X^{-1}(x) \frac{1}{n} \sum_{i=1}^n \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(y|z) \frac{1}{b} \tilde{K}_b(X_i - z) (m(X_i) - m(z)) dz) = o(\frac{1}{nh}).$$

Ebendies gilt auch für die zentrierten Anteile der durch den jeweils vorliegenden Typ von Nebeninformationen g hinzukommenden Summanden, diese haben eine Varianz der Rate $O(\frac{1}{nh})$.

Für die nicht zentrierten Anteile gilt:

Für die Zusatzinformation beschreibende Funktionen g , welche (G1) und (G2) erfüllen:

$$\begin{aligned} \text{Var}(-\Sigma^{-1}(x) \frac{1}{n} \sum_{i=1}^n \int \frac{1}{h} K_h(z-x) g'(y, x) \frac{1}{b} \tilde{K}_b(X_i - z) (m(X_i) - m(z)) f_{\varepsilon|X}(y|z) d(y, z)) \\ = o(\frac{1}{nh}) \end{aligned}$$

Für die Zusatzinformation beschreibende Funktionen g der Gestalt $g(\varepsilon, x) = I\{\varepsilon \leq a(x)\} - d(x)$:

$$\begin{aligned} \text{Var}(\Sigma^{-1}(x) \frac{1}{n} \sum_{i=1}^n \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(a(x)|z) \frac{1}{b} \tilde{K}_b(X_i - z) (m(X_i) - m(z)) dz) \\ = o(\frac{1}{nh}) \end{aligned}$$

Ähnliche Varianzrechnungen wie oben nötig folgen an späterer Stelle noch häufiger, hier verzichten wir auf die Angabe. Nach den Berechnungen zum Erwartungswert der in $y \in \mathbb{R}$ gleichmäßigen Entwicklung von $\hat{F}_n^*(y|x) - F_{\varepsilon|X}(y|x)$ (siehe (6.3)) und den Bemerkungen bezüglich der Varianzen wollen wir nun den Prozess angeben und die Kovarianz betrachten, dazu betrachten wir zunächst nur die Zusatzinformation beschreibende Funktionen g , welche (G1) und (G2)

erfüllen; für Funktionen g von Indikatorgestalt folgen Aussagen im Anschluss.

Mittels der Entwicklung von $\hat{F}_n^*(y|x) - F_{\varepsilon|X}(y|x)$ (Proposition 6.2) und den vorausgegangenen Rechnungen legen wir den Korrekturterm für den folgenden Prozess fest:

$$(6.8) \quad G_n(\cdot, x) := \sqrt{nh} (\hat{F}_n^*(\cdot|x) - F_{\varepsilon|X}(\cdot|x) - b_n^*(\cdot, x))$$

Wie schon zuvor ausgeführt setzten wir $b_n^*(\cdot, x)$ gleich dem Erwartungswert der beiden Terme, welche im Erwartungswert eine Rate ungleich 0 und eine Varianz der Ordnung $o(\frac{1}{nh})$ liefern, dem Erwartungswert des Terms $C_n(y, x) = f_X^{-1}(x) \cdot \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) (I\{\varepsilon_i \leq y\} - F_{\varepsilon|X}(y|x))$, im Weiteren mit $E[C_n(y, x)]$ bezeichnet und dem Erwartungswert von $D_n(y, x) := -(\Sigma^{-1}(x) \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) g(\varepsilon_i, x))^t \cdot A(y, x)$, vgl. dazu auch (6.6) bzw. (6.7) und (6.1):

$$\begin{aligned} b_n^*(\cdot, x) &:= E[D_n(\cdot, x)] \\ &+ (\Sigma^{-1}(x) \frac{b^2}{2} \mu_2^{\tilde{K}} E[g'(\varepsilon, X) | X = x] B(x))^t \cdot \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq \cdot\} | X = x] \\ &+ \frac{b^2}{2} \mu_2^{\tilde{K}} f_X^{-1}(x) f_{\varepsilon|X}(\cdot|x) B(x) \\ &+ E[C_n(\cdot, x)] \\ &= -\frac{h^2}{2} \mu_2^K \frac{\partial^2 E[g^t(\varepsilon, x) | X = t] \cdot f_X(t)}{\partial t^2} \Big|_{t=x} \cdot \Sigma^{-1}(x)^t \cdot \mu_0^{K^2} \cdot E[g(\varepsilon, X) I\{\varepsilon \leq \cdot\} | X = x] \\ &+ (\Sigma^{-1}(x) \frac{b^2}{2} \mu_2^{\tilde{K}} E[g'(\varepsilon, X) | X = x] B(x))^t \cdot \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq \cdot\} | X = x] \\ &+ \frac{b^2}{2} \mu_2^{\tilde{K}} f_X^{-1}(x) f_{\varepsilon|X}(\cdot|x) B(x) \\ &+ \frac{h^2}{2} \cdot \mu_2^K \cdot (F_{\varepsilon|X}^{(2)}(\cdot|x) + \frac{2 f_X'(x)}{f_X(x)} F_{\varepsilon|X}^{(1)}(\cdot|x)) \\ &+ o(h^2) \end{aligned}$$

Zu den einzelnen Bauteilen siehe Proposition 6.2, bzw. (6.5) zu $B(x)$. Es gilt $o(h^2) = o(\frac{1}{\sqrt{nh}})$, siehe (H2).

Damit erhält man zusammengefasst:

$$(6.9) \quad \hat{F}_n^*(\cdot|x) - F_{\varepsilon|X}(\cdot|x) = H_n(\cdot, x) + b_n^*(\cdot, x) + o_p\left(\frac{1}{\sqrt{nh}}\right)$$

wobei

$$(6.10) \quad H_n(\cdot, x) =: H_{n,1}(\cdot, x) + H_{n,2}(\cdot, x) + H_{n,3}(\cdot, x) + H_{n,4}(\cdot, x)$$

und

$$H_{n,1}(\cdot, x) = -(\Sigma^{-1}(x) \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) g(\varepsilon_i, x))^t \cdot \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq \cdot\} | X = x] - E[D_n(\cdot, x)]$$

$$H_{n,2}(\cdot, x) = \frac{1}{n} \sum_{i=1}^n \varepsilon_i \int \frac{1}{h} K_h(z - x) (g'(y, x))^t \frac{1}{b} \tilde{K}_b(X_i - z) f_{\varepsilon|X}(y|z) d(y, z) \\ \cdot \Sigma^{-1}(x)^t \cdot \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq \cdot\} | X = x]$$

$$H_{n,3}(\cdot, x) = f_X^{-1}(x) \frac{1}{n} \sum_{i=1}^n \varepsilon_i \int \frac{1}{h} K_h(z - x) f_{\varepsilon|X}(\cdot|z) \frac{1}{b} \tilde{K}_b(X_i - z) dz$$

$$H_{n,4}(\cdot, x) = \left(f_X^{-1}(x) \cdot \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) (I\{\varepsilon_i \leq \cdot\} - F_{\varepsilon|X}(\cdot|x)) - E[C_n(\cdot, x)] \right)$$

Bemerkung 6.3 (Bandbreiten) Der Korrekturterm $b_n^*(\cdot, x)$ enthält nur Summanden welche jeweils quadratisch von den Bandbreiten h und b abhängen, hier sind keine Mischungen zu beachten. Es wird sich nach Berechnung der Kovarianz herausstellen, dass sich, bei geeigneter Wahl des Verhältnisses von h und b , die Kovarianzen der Summanden $H_{n,1}(y, x) - H_{n,4}(y, x)$ etwas vereinfachen können. Darauf gehen wir nach Berechnung der Kovarianzstruktur bzw. des Beweises der schwachen Konvergenz und der dort benötigten Bandbreitenannahmen an h und b nochmals ein.

6.3 Kovarianzfunktion

Im Weiteren werden wir die Kovarianzfunktion des Prozesses $H_n(\cdot, x)$ berechnen. Der Beweis zur schwachen Konvergenz des Prozesses folgt im Anschluss. Es sei $y_1 \leq y_2$ und damit

$$\text{Cov}(H_n(y_1, x), H_n(y_2, x)) = \text{Cov}(H_{n,1}(y_1, x) + H_{n,2}(y_1, x) + H_{n,3}(y_1, x) + H_{n,4}(y_1, x), \\ H_{n,1}(y_2, x) + H_{n,2}(y_2, x) + H_{n,3}(y_2, x) + H_{n,4}(y_2, x))$$

$$\begin{aligned}
(6.11) \quad &= \text{Cov}(H_{n,1}(y_1, x), H_{n,1}(y_2, x)) \\
&+ \text{Cov}(H_{n,2}(y_1, x), H_{n,2}(y_2, x)) \\
&+ \text{Cov}(H_{n,3}(y_1, x), H_{n,3}(y_2, x)) \\
&+ \text{Cov}(H_{n,1}(y_1, x), H_{n,3}(y_2, x)) + \text{Cov}(H_{n,3}(y_1, x), H_{n,1}(y_2, x)) \\
&+ \text{Cov}(H_{n,2}(y_1, x), H_{n,3}(y_2, x)) + \text{Cov}(H_{n,3}(y_1, x), H_{n,2}(y_2, x)) \\
&+ \text{Cov}(H_{n,1}(y_1, x), H_{n,2}(y_2, x)) + \text{Cov}(H_{n,2}(y_1, x), H_{n,1}(y_2, x)) \\
&+ \text{Cov}(H_{n,4}(y_1, x), H_{n,4}(y_2, x)) \\
&+ \text{Cov}(H_{n,4}(y_1, x), H_{n,1}(y_2, x)) + \text{Cov}(H_{n,1}(y_1, x), H_{n,4}(y_2, x)) \\
&+ \text{Cov}(H_{n,4}(y_1, x), H_{n,2}(y_2, x)) + \text{Cov}(H_{n,2}(y_1, x), H_{n,4}(y_2, x)) \\
&+ \text{Cov}(H_{n,4}(y_1, x), H_{n,3}(y_2, x)) + \text{Cov}(H_{n,3}(y_1, x), H_{n,4}(y_2, x))
\end{aligned}$$

Die einzelnen Kovarianzen sind nun zu berechnen. Es gilt die folgende Proposition:

Proposition 6.4 *Betrachte die Darstellung (6.9) und (6.10) mit $y_1 \leq y_2$. Es gilt unter den Voraussetzungen (M1) – (M3), (H1), (H2), (K), (A) und g erfüllend (G1), (G2):*

$$(i) \quad \text{Cov}(H_{n,1}(y_1, x), H_{n,1}(y_2, x))$$

$$\begin{aligned}
&= \frac{1}{nh} (\mu_0^{K^2})^3 f_X(x) E \left[g^t(\varepsilon, U) \cdot (\Sigma^{-1}(U))^t \cdot E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = U] \right. \\
&\quad \left. \cdot g^t(\varepsilon, U) \cdot (\Sigma^{-1}(U))^t \cdot E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = U] \Big| U = x \right] + o\left(\frac{1}{nh}\right)
\end{aligned}$$

$$(ii) \quad \text{Cov}(H_{n,2}(y_1, x), H_{n,2}(y_2, x))$$

$$\begin{aligned}
&= \frac{1}{nh} \cdot \sigma^2(x) \cdot (\mu_0^{K^2})^2 f_X(x) \cdot E[(g'(\varepsilon, X))^t | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x] \\
&\quad \cdot E[(g'(\varepsilon, X))^t | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = x] \\
&\quad \cdot \int \int \int K(v + \frac{b}{h}(u-w)) K(v) \tilde{K}(u) \tilde{K}(w) du dv dw + o\left(\frac{1}{nh}\right)
\end{aligned}$$

$$(iii) \quad \text{Cov}(H_{n,3}(y_1, x), H_{n,3}(y_2, x))$$

$$\begin{aligned}
&= f_X^{-1}(x) \frac{1}{nh} E[\varepsilon^2 | X = x] f_{\varepsilon|X}(y_1|x) f_{\varepsilon|X}(y_2|x) \int K(v + \frac{b}{h}(u-w)) \tilde{K}(w) K(v) \tilde{K}(u) d(u, v, w) \\
&\quad + o\left(\frac{1}{nh}\right)
\end{aligned}$$

$$(iv) \text{Cov}(H_{n,1}(y_1, x), H_{n,3}(y_2, x))$$

$$= -\frac{1}{nh} \mu_0^{K^2} f_{\varepsilon|X}(y_2|x) E[(g(\varepsilon, X))^t \cdot \varepsilon | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x] \\ \cdot \int \int K(u + \frac{b}{h}v) K(u) \tilde{K}(v) dudv + o(\frac{1}{nh})$$

$$(v) \text{Cov}(H_{n,3}(y_1, x), H_{n,1}(y_2, x))$$

$$= -\frac{1}{nh} \mu_0^{K^2} f_{\varepsilon|X}(y_1|x) E[(g(\varepsilon, X))^t \cdot \varepsilon | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = x] \\ \cdot \int \int K(u + \frac{b}{h}v) K(u) \tilde{K}(v) dudv + o(\frac{1}{nh})$$

$$(vi) \text{Cov}(H_{n,1}(y_1, x), H_{n,2}(y_2, x))$$

$$= -\frac{1}{nh} f_X(x) (\mu_0^{K^2})^2 E[(g(\varepsilon, X))^t \varepsilon | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x] \\ \cdot E[(g'(\varepsilon, X))^t | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = x] \\ \cdot \int \int K(u + \frac{b}{h}v) K(u) \tilde{K}(v) dudv + o(\frac{1}{nh})$$

$$(vii) \text{Cov}(H_{n,2}(y_1, x), H_{n,1}(y_2, x))$$

$$= -\frac{1}{nh} f_X(x) (\mu_0^{K^2})^2 E[(g'(\varepsilon, X))^t | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = x] \\ \cdot E[(g(\varepsilon, X))^t \varepsilon | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x] \\ \cdot \int \int K(u + \frac{b}{h}v) K(u) \tilde{K}(v) dudv + o(\frac{1}{nh})$$

$$(viii) \text{Cov}(H_{n,3}(y_1, x), H_{n,2}(y_2, x))$$

$$= \frac{1}{nh} E[\varepsilon^2 | X = x] E[(g'(\varepsilon, X))^t | X = x] (\Sigma^{-1}(x))^t \cdot \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = x] \\ \cdot f_{\varepsilon|X}(y_1|x) \int \int \int K(v + \frac{b}{h}(u-w)) K(v) \tilde{K}(u) \tilde{K}(w) du dv dw + o(\frac{1}{nh})$$

$$(ix) \text{Cov}(H_{n,2}(y_1, x), H_{n,3}(y_2, x))$$

$$= \frac{1}{nh} E[\varepsilon^2 | X = x] E[(g'(\varepsilon, X))^t | X = x] (\Sigma^{-1}(x))^t \cdot \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x] \\ \cdot f_{\varepsilon|X}(y_2|x) \int \int \int K(v + \frac{b}{h}(u-w)) K(v) \tilde{K}(u) \tilde{K}(w) du dv dw + o(\frac{1}{nh})$$

$$(x) \text{Cov}(H_{n,4}(y_1, x), H_{n,4}(y_2, x))$$

$$= +\frac{1}{nh} \mu_0^{K^2} f_X^{-1}(x) F_{\varepsilon|X}(\min(y_1, y_2) | x) - \frac{1}{nh} \mu_0^{K^2} f_X^{-1}(x) F_{\varepsilon|X}(y_2 | x) F_{\varepsilon|X}(y_1 | x) + o\left(\frac{1}{nh}\right)$$

$$\text{(xi)} \quad \text{Cov}(H_{n,1}(y_1, x), H_{n,4}(y_2, x))$$

$$= -\frac{1}{nh} (\mu_0^{K^2})^2 E[(g(\varepsilon, X))^t I\{\varepsilon \leq y_2\} | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x] + o\left(\frac{1}{nh}\right)$$

$$\text{(xii)} \quad \text{Cov}(H_{n,4}(y_1, x), H_{n,1}(y_2, x))$$

$$= -\frac{1}{nh} (\mu_0^{K^2})^2 E[(g(\varepsilon, X))^t I\{\varepsilon \leq y_1\} | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = x] + o\left(\frac{1}{nh}\right)$$

$$\text{(xiii)} \quad \text{Cov}(H_{n,2}(y_1, x), H_{n,4}(y_2, x))$$

$$= \frac{1}{nh} E[\varepsilon I\{\varepsilon \leq y_2\} | X = x] E[(g'(\varepsilon, x))^t | X = x] (\Sigma^{-1}(x))^t \cdot E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x] \cdot \mu_0^{K^2} \int \int K(u - v \frac{b}{h}) \tilde{K}(u) K(v) du dv + o\left(\frac{1}{nh}\right)$$

$$\text{(xiv)} \quad \text{Cov}(H_{n,4}(y_1, x), H_{n,2}(y_2, x))$$

$$= \frac{1}{nh} E[\varepsilon I\{\varepsilon \leq y_1\} | X = x] E[(g'(\varepsilon, x))^t | X = x] (\Sigma^{-1}(x))^t \cdot E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = x] \cdot \mu_0^{K^2} \int \int K(u - \frac{b}{h} v) \tilde{K}(u) K(v) du dv + o\left(\frac{1}{nh}\right)$$

$$\text{(xv)} \quad \text{Cov}(H_{n,3}(y_1, x), H_{n,4}(y_2, x))$$

$$= \frac{1}{nh} f_X^{-1}(x) f_{\varepsilon|X}(y_1 | x) E[\varepsilon I\{\varepsilon \leq y_2\} | X = x] \cdot \int \int K(u - \frac{b}{h} v) \tilde{K}(u) K(v) du dv + o\left(\frac{1}{nh}\right)$$

$$\text{(xvi)} \quad \text{Cov}(H_{n,4}(y_1, x), H_{n,3}(y_2, x))$$

$$= \frac{1}{nh} f_X^{-1}(x) f_{\varepsilon|X}(y_2 | x) E[\varepsilon I\{\varepsilon \leq y_1\} | X = x] \cdot \int \int K(u - \frac{b}{h} v) \tilde{K}(u) K(v) du dv + o\left(\frac{1}{nh}\right)$$

Beweis:

(i), es sei $A(y_1, x) := \mu_0^{K^2} (\Sigma^{-1}(x))^t \cdot E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x]$, $A(y_2, x)$ analog.

$$\begin{aligned}
& \text{Cov}(H_{n,1}(y_1, x), H_{n,1}(y_2, x)) \\
&= \text{Cov}\left([\Sigma^{-1}(x) \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) g(\varepsilon_i, x)]^t \cdot \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x] - E[D_n(y_1, x)], \right. \\
&\quad \left. [\Sigma^{-1}(x) \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) g(\varepsilon_i, x)]^t \cdot \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = x] - E[D_n(y_2, x)]\right) \\
&= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \text{Cov}\left(\frac{1}{h} K_h(X_i - x) g^t(\varepsilon_i, x) \cdot A(y_1, x), \frac{1}{h} K_h(X_j - x) g^t(\varepsilon_j, x) \cdot A(y_2, x)\right) \\
&= \frac{1}{n^2} \sum_{i=1}^n \text{Cov}\left(\frac{1}{h} K_h(X_i - x) g^t(\varepsilon_i, x) \cdot A(y_1, x), \frac{1}{h} K_h(X_i - x) g^t(\varepsilon_i, x) \cdot A(y_2, x)\right) \\
&\quad + \frac{1}{n^2} \sum_{i \neq j} \text{Cov}\left(\frac{1}{h} K_h(X_i - x) g^t(\varepsilon_i, x) \cdot A(y_1, x), \frac{1}{h} K_h(X_j - x) g^t(\varepsilon_j, x) \cdot A(y_2, x)\right) \\
&= \frac{1}{n} \text{Cov}\left(\frac{1}{h} K_h(X_1 - x) g^t(\varepsilon_1, x) \cdot A(y_1, x), \frac{1}{h} K_h(X_1 - x) g^t(\varepsilon_1, x) \cdot A(y_2, x)\right) \\
&= \frac{1}{nh} E\left[\frac{1}{h} K_h^2(X_1 - x) g^t(\varepsilon_1, x) \cdot A(y_1, x) \cdot g^t(\varepsilon_1, x) \cdot A(y_2, x)\right] \\
&\quad - \frac{1}{n} E\left[\frac{1}{h} K_h(X_1 - x) g^t(\varepsilon_1, x) \cdot A(y_1, x)\right] \cdot E\left[\frac{1}{h} K_h(X_1 - x) g^t(\varepsilon_1, x) \cdot A(y_2, x)\right] \\
&= \frac{1}{nh} E\left[\frac{1}{h} K_h^2(X_1 - x) g^t(\varepsilon_1, x) \cdot A(y_1, x) \cdot g^t(\varepsilon_1, x) \cdot A(y_2, x)\right] + O\left(\frac{h^4}{n}\right) \\
&= \frac{1}{nh} E\left[\frac{1}{h} K_h^2(X_1 - x) g^t(\varepsilon_1, x) \cdot A(y_1, x) \cdot g^t(\varepsilon_1, x) \cdot A(y_2, x)\right] + o\left(\frac{1}{nh}\right) \\
&= \frac{1}{nh} (\mu_0^{K^2})^2 E\left[\frac{1}{h} K_h^2(X_1 - x) g^t(\varepsilon_1, x) \cdot (\Sigma^{-1}(x))^t \cdot E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x] \right. \\
&\quad \left. \cdot g^t(\varepsilon_1, x) \cdot (\Sigma^{-1}(x))^t \cdot E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = x]\right] + o\left(\frac{1}{nh}\right) \\
&= \frac{1}{nh} (\mu_0^{K^2})^3 f_X(x) E\left[g^t(\varepsilon, U) \cdot (\Sigma^{-1}(U))^t \cdot E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = U] \right. \\
&\quad \left. \cdot g^t(\varepsilon, U) \cdot (\Sigma^{-1}(U))^t \cdot E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = U] \Big| U = x\right] + o\left(\frac{1}{nh}\right)
\end{aligned}$$

Die letzte Zeile folgt dabei mittels Substitution und Taylorentwicklung. Der Restterm der Taylorentwicklung wurde mittels der Voraussetzungen (M1) – (M3) abgeschätzt.

(iii):

$$\text{Cov}(H_{n,3}(y_1, x), H_{n,3}(y_2, x)) = \text{Cov}\left(f_X^{-1}(x) \frac{1}{n} \sum_{i=1}^n \varepsilon_i \int \frac{1}{h} K_h(z - x) f_{\varepsilon|X}(y_1|z) \frac{1}{b} \tilde{K}_b(X_i - z) dz, \right.$$

$$\begin{aligned}
& f_X^{-1}(x) \frac{1}{n} \sum_{i=1}^n \varepsilon_i \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(y_2|z) \frac{1}{b} \tilde{K}_b(X_i - z) dz \\
= & (f_X^{-1}(x))^2 \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \text{Cov} \left(\varepsilon_i \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(y_1|z) \frac{1}{b} \tilde{K}_b(X_i - z) dz, \right. \\
& \left. \varepsilon_j \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(y_2|z) \frac{1}{b} \tilde{K}_b(X_j - z) dz \right) \\
(6.12) \quad = & (f_X^{-1}(x))^2 \frac{1}{n^2} \sum_{i=1}^n E \left[\varepsilon_i^2 \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(y_1|z) \frac{1}{b} \tilde{K}_b(X_i - z) dz \right. \\
& \left. \cdot \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(y_2|z) \frac{1}{b} \tilde{K}_b(X_i - z) dz \right] \\
= & (f_X^{-1}(x))^2 \frac{1}{n^2} \sum_{i=1}^n E \left[\varepsilon_i^2 \int \int \frac{1}{h} K_h(z_1-x) f_{\varepsilon|X}(y_1|z_1) \frac{1}{b} \tilde{K}_b(X_i - z_1) \right. \\
& \left. \cdot \frac{1}{h} K_h(z_2-x) f_{\varepsilon|X}(y_2|z_2) \frac{1}{b} \tilde{K}_b(X_i - z_2) dz_1 dz_2 \right] \\
= & (f_X^{-1}(x))^2 \frac{1}{n} E \left[\varepsilon_i^2 \int \int \frac{1}{h} K_h(z_1-x) f_{\varepsilon|X}(y_1|z_1) \frac{1}{b} \tilde{K}_b(X_i - z_1) \right. \\
& \left. \cdot \frac{1}{h} K_h(z_2-x) f_{\varepsilon|X}(y_2|z_2) \frac{1}{b} \tilde{K}_b(X_i - z_2) dz_1 dz_2 \right]
\end{aligned}$$

Setze $E[\varepsilon_i^2 | X_i] =: \sigma^2(X_i)$. damit gilt weiter:

$$\begin{aligned}
& = (f_X^{-1}(x))^2 \frac{1}{n} E \left[\sigma^2(X_i) \int \int \frac{1}{h} K_h(z_1-x) f_{\varepsilon|X}(y_1|z_1) \frac{1}{b} \tilde{K}_b(X_i - z_1) \right. \\
& \quad \left. \cdot \frac{1}{h} K_h(z_2-x) f_{\varepsilon|X}(y_2|z_2) \frac{1}{b} \tilde{K}_b(X_i - z_2) dz_1 dz_2 \right] \\
& = (f_X^{-1}(x))^2 \frac{1}{n} \int \sigma^2(t) \int \int \frac{1}{h} K_h(z_1-x) f_{\varepsilon|X}(y_1|z_1) \frac{1}{b} \tilde{K}_b(t - z_1) \\
& \quad \cdot \frac{1}{h} K_h(z_2-x) f_{\varepsilon|X}(y_2|z_2) \frac{1}{b} \tilde{K}_b(t - z_2) f_X(t) dz_1 dz_2 dt
\end{aligned}$$

Um dieses Integral zu lösen substituieren wir an drei Stellen wie folgt:

$$v = \frac{z_2-x}{h}, \quad w = \frac{t-z_1}{b}, \quad u = \frac{t-z_2}{b} \quad \text{und damit} \quad z_2 = x + hv, \quad t = z_2 + bu = x + hv + bu \quad \text{und} \\
z_1 = t - wb = x + hv + bu - bw \Rightarrow \frac{z_1-x}{h} = \frac{x+hv+bu-bw-x}{h} = v + \frac{b}{h}(u-w).$$

$$\begin{aligned}
& = (f_X^{-1}(x))^2 \frac{1}{n} \int \int \int \sigma^2(x + hv + bu) \frac{1}{h} K(v + \frac{b}{h}(u-w)) f_{\varepsilon|X}(y_1|x + hv + bu - bw) \tilde{K}(w) \\
& \quad \cdot K(v) f_{\varepsilon|X}(y_2|x + hv) \tilde{K}(u) f_X(x + hv + bu) du dw dv
\end{aligned}$$

Nun entwickelt man die drei vorkommenden Dichten, $f_{\varepsilon|X}(y_2|\cdot)$, $f_{\varepsilon|X}(y_1|\cdot)$, $f_X(\cdot)$ und $\sigma^2(\cdot) := E[\varepsilon^2 | X = \cdot]$ um x in Taylorreihen bis zur einer Zwischenstelle der ersten Ableitung, die Funktion $K(v + \frac{b}{h}(u - w))$ entwickle man nicht. Die dabei entstehenden Restterme sind aufgrund der Beschränktheit aller Bauteile von der Rate $o(\frac{1}{nh})$ (Vgl. (M1) – (M3), (K) etc.). Der erste Term der Entwicklung hat die Ordnung $O(\frac{1}{nh})$:

$$= f_X^{-1}(x) \frac{1}{nh} E[\varepsilon^2 | X = x] f_{\varepsilon|X}(y_1|x) f_{\varepsilon|X}(y_2|x) \int K(v + \frac{b}{h}(u - w)) \tilde{K}(w) K(v) \tilde{K}(u) d(u, w, v) + o(\frac{1}{nh})$$

Zu (ii):

$$\begin{aligned} & \text{Cov}(H_{n,2}(y_1, x), H_{n,2}(y_2, x)) \\ &= (\mu_0^{K^2})^2 \cdot \text{Cov} \left(\int \frac{1}{h} K_h(z - x) \frac{1}{nb} \sum_{i=1}^n \tilde{K}_b(X_i - z) \varepsilon_i (g'(y, x))^t f_{\varepsilon|X}(y|z) d(y, z) \right. \\ & \quad \left. \cdot (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x], \right. \\ & \quad \left. \int \frac{1}{h} K_h(z - x) \frac{1}{nb} \sum_{i=1}^n \tilde{K}_b(X_i - z) \varepsilon_i (g'(y, x))^t f_{\varepsilon|X}(y|z) d(y, z) \right. \\ & \quad \left. \cdot (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = x] \right) \\ &= \frac{1}{n} (\mu_0^{K^2})^2 \cdot \text{Cov} \left(\int \frac{1}{h} K_h(z - x) \frac{1}{b} \tilde{K}_b(X_i - z) \varepsilon_i (g'(y, x))^t f_{\varepsilon|X}(y|z) d(y, z) \right. \\ & \quad \left. \cdot (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x], \right. \\ & \quad \left. \int \frac{1}{h} K_h(z - x) \frac{1}{b} \tilde{K}_b(X_i - z) \varepsilon_i (g'(y, x))^t f_{\varepsilon|X}(y|z) d(y, z) \cdot (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = x] \right) \\ &= \frac{1}{n} (\mu_0^{K^2})^2 \cdot E \left[\varepsilon_i^2 \int \int \frac{1}{h} K_h(z_a - x) \frac{1}{b} \tilde{K}_b(X_i - z_a) (g'(y_a, x))^t f_{\varepsilon|X}(y_a|z_a) d(y_a, z_a) \right. \\ & \quad \left. \cdot (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x] \right. \\ & \quad \left. \cdot \frac{1}{h} K_h(z_b - x) \frac{1}{n} \tilde{K}_b(X_i - z_b) (g'(y_b, x))^t f_{\varepsilon|X}(y_b|z_b) d(y_b, z_b) \cdot (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = x] \right] \\ &= \frac{1}{nh} \cdot f_X(x) \cdot (\mu_0^{K^2})^2 E[\varepsilon^2 | X = x] E[(g'(\varepsilon, X))^t | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x] \\ & \quad \cdot E[(g'(\varepsilon, X))^t | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = x] \\ & \quad \cdot \int \int \int K(v + \frac{b}{h}(u - w)) K(v) \tilde{K}(u) \tilde{K}(w) du dv dw + o(\frac{1}{nh}) \\ &= \frac{1}{nh} \cdot \sigma^2(x) \cdot (\mu_0^{K^2})^2 f_X(x) \cdot E[(g'(\varepsilon, X))^t | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x] \\ & \quad \cdot E[(g'(\varepsilon, X))^t | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = x] \\ & \quad \cdot \int \int \int K(v + \frac{b}{h}(u - w)) K(v) \tilde{K}(u) \tilde{K}(w) du dv dw + o(\frac{1}{nh}) \end{aligned}$$

Die Substitution und Taylorentwicklung von der drittletzten auf die vorletzte Gleichheit ist dabei wie im Beweis zu (iii).

Der Beweis zu (iv) ist ähnlich zu (i) und (iii). Man zeigt mit einfachen Umformungen und Resttermabschätzungen unter Ausnutzung der Voraussetzungen (M1) – (M3) etc.:

$$\begin{aligned}
& \text{Cov}(H_{n,1}(y_1, x), H_{n,3}(y_2, x)) \\
&= \text{Cov}\left(-[\Sigma^{-1}(x) \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) g(\varepsilon_i, x)]^t \cdot \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x], \right. \\
&\quad \left. f_X^{-1}(x) \frac{1}{n} \sum_{i=1}^n \varepsilon_i \int \frac{1}{h} K_h(z - x) f_{\varepsilon|X}(y_2|z) \frac{1}{b} \tilde{K}_b(X_i - z) dz\right) \\
&= -\frac{1}{n^2 h} \sum_{i=1}^n \text{Cov}\left(K_h(X_i - x) (g(\varepsilon_i, x))^t, \varepsilon_i \int \frac{1}{h} K_h(z - x) f_{\varepsilon|X}(y_2|z) \frac{1}{b} \tilde{K}_b(X_i - z) dz\right) \\
&\quad \cdot (\Sigma^{-1}(x))^t \frac{\mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x]}{f_X(x)} \\
&= -\frac{1}{nh} \text{Cov}\left(K_h(X_1 - x) (g(\varepsilon_1, x))^t, \varepsilon_1 \int \frac{1}{h} K_h(z - x) f_{\varepsilon|X}(y_2|z) \frac{1}{b} \tilde{K}_b(X_1 - z) dz\right) \\
&\quad \cdot (\Sigma^{-1}(x))^t \frac{\mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x]}{f_X(x)} \\
&= -\frac{1}{nh} E\left[K_h(X_1 - x) (g(\varepsilon_1, x))^t \cdot \varepsilon_1 \int \frac{1}{h} K_h(z - x) f_{\varepsilon|X}(y_2|z) \frac{1}{b} \tilde{K}_b(X_1 - z) dz\right] \\
&\quad \cdot (\Sigma^{-1}(x))^t \frac{\mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x]}{f_X(x)} \\
&= -\frac{1}{nh} \int \int K_h(t - x) (g(\varepsilon, x))^t \cdot \varepsilon \cdot \frac{1}{h} K_h(z - x) f_{\varepsilon|X}(y_2|z) \frac{1}{b} \tilde{K}_b(t - z) f_{\varepsilon, X}(\varepsilon, t) dz d(\varepsilon, t) \\
&\quad \cdot (\Sigma^{-1}(x))^t \frac{\mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x]}{f_X(x)}
\end{aligned}$$

Jetzt ist an zwei Stellen (2. und 3. Kern) zu substituieren, in Taylorreihen zu entwickeln und man erhält:

$$\begin{aligned}
&= -\frac{1}{nh} \mu_0^{K^2} \cdot f_{\varepsilon|X}(y_2|x) E[(g(\varepsilon, X))^t \cdot \varepsilon | X = x] \cdot (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x] \\
&\quad \cdot \int \int K(u + \frac{b}{h}v) K(u) \tilde{K}(v) dudv + o(\frac{1}{nh})
\end{aligned}$$

(v) analog.

(vi):

$$\begin{aligned}
& \text{Cov}(H_{n,1}(y_1, x), H_{n,2}(y_2, x)) \\
&= \text{Cov}\left(-\frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) (g(\varepsilon_i, x))^t (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x], \right. \\
&\quad \left. \int \frac{1}{h} K_h(z - x) \frac{1}{nb} \sum_{i=1}^n \tilde{K}_b(X_i - z) \varepsilon_i (g'(y, x))^t f_{\varepsilon|X}(y|z) d(y, z) \right) \\
&\quad \cdot (\mu_0^{K^2})^2 (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = x] \\
&= -\frac{1}{n} \text{Cov}\left(\frac{1}{h} K_h(X_i - x) (g(\varepsilon_i, x))^t (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x], \right. \\
&\quad \left. \int \frac{1}{h} K_h(z - x) \frac{1}{b} \tilde{K}_b(X_i - z) \varepsilon_i (g'(y, x))^t f_{\varepsilon|X}(y|z) d(y, z) \right) \\
&\quad \cdot (\mu_0^{K^2})^2 (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = x] \\
&= -\frac{1}{n} E\left[\frac{1}{h} K_h(X_i - x) (g(\varepsilon_i, x))^t \varepsilon_i (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x] \right. \\
&\quad \left. \cdot \int \frac{1}{h} K_h(z - x) \frac{1}{b} \tilde{K}_b(X_i - z) (g'(y, x))^t f_{\varepsilon|X}(y|z) d(y, z) \right] \\
&\quad \cdot (\mu_0^{K^2})^2 (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = x] \\
&= -\frac{1}{n} \int \int \frac{1}{h} K_h(u - x) \int (g(\varepsilon, x))^t \varepsilon f_{\varepsilon|X}(\varepsilon|u) d\varepsilon (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x] \\
&\quad \cdot \frac{1}{h} K_h(z - x) \frac{1}{b} \tilde{K}_b(u - z) (g'(y, x))^t f_{\varepsilon|X}(y|z) f_X(u) d(y, z) du \\
&\quad \cdot (\mu_0^{K^2})^2 (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = x] \\
&= -\frac{1}{nh} f_X(x) (\mu_0^{K^2})^2 E[(g(\varepsilon, X))^t \varepsilon | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x] \\
&\quad \cdot E[(g'(\varepsilon, X))^t | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = x] \\
&\quad \cdot \int \int K(u + \frac{b}{h}v) K(u) \tilde{K}(v) dudv + o\left(\frac{1}{nh}\right)
\end{aligned}$$

(vii) analog.

(viii):

$$\begin{aligned}
& \text{Cov}(H_{n,3}(y_1, x), H_{n,2}(y_2, x)) \\
&= \text{Cov}\left(\frac{1}{n} \sum_{i=1}^n \varepsilon_i \int \frac{1}{h} K_h(z - x) f_{\varepsilon|X}(y_1|z) \frac{1}{b} \tilde{K}_b(X_i - z) dz, \right.
\end{aligned}$$

$$\begin{aligned}
& \int \frac{1}{h} K_h(z-x) \frac{1}{nb} \sum_{i=1}^n \tilde{K}_b(X_i - z) \varepsilon_i (g'(y, x))^t f_{\varepsilon|X}(y|z) d(y, z) \\
& \quad \cdot (\Sigma^{-1}(x))^t \cdot \frac{\mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = x]}{f_X(x)} \\
&= \frac{1}{n} \text{Cov} \left(\varepsilon_i \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(y_1|z) \frac{1}{b} \tilde{K}_b(X_i - z) dz, \right. \\
& \quad \left. \varepsilon_i \int \frac{1}{h} K_h(z-x) \frac{1}{b} \tilde{K}_b(X_i - z) (g'(y, x))^t f_{\varepsilon|X}(y|z) d(y, z) \right) \\
& \quad \cdot (\Sigma^{-1}(x))^t \cdot \frac{\mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = x]}{f_X(x)} \\
&= \frac{1}{n} E \left[\varepsilon_i^2 \int \int \frac{1}{h} K_h(z_1 - x) f_{\varepsilon|X}(y_1|z_1) \frac{1}{b} \tilde{K}_b(X_i - z_1) \frac{1}{h} K_h(z_2 - x) \right. \\
& \quad \left. \cdot \frac{1}{b} \tilde{K}_b(X_i - z_2) (g'(y, x))^t f_{\varepsilon|X}(y|z_2) dz_1 d(y, z_2) \right] \\
& \quad \cdot (\Sigma^{-1}(x))^t \cdot \frac{\mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = x]}{f_X(x)} \\
&= \frac{1}{nh} E[\varepsilon^2 | X = x] E[(g'(\varepsilon, X))^t | X = x] (\Sigma^{-1}(x))^t \cdot \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = x] \\
& \quad \cdot f_{\varepsilon|X}(y_1|x) \int \int \int K(v + \frac{b}{h}(u-w)) K(v) \tilde{K}(u) \tilde{K}(w) du dv dw
\end{aligned}$$

(ix) analog.

(x):

Nach Voraussetzung ist $y_1 \leq y_2$.

$$\begin{aligned}
& \text{Cov} (H_{n,4}(y_1, x), H_{n,4}(y_2, x)) \\
&= \text{Cov} \left(f_X^{-1}(x) \cdot \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) (I\{\varepsilon_i \leq y_1\} - F_{\varepsilon|X}(y_1|x)) - E[C_n(y_1, x)], \right. \\
& \quad \left. f_X^{-1}(x) \cdot \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) (I\{\varepsilon_i \leq y_2\} - F_{\varepsilon|X}(y_2|x)) - E[C_n(y_2, x)] \right) \\
&= \frac{1}{n} E \left[(f_X^{-1}(x))^2 \frac{1}{h} K_h(X_i - x) (I\{\varepsilon_i \leq y_1\} - F_{\varepsilon|X}(y_1|x)) \frac{1}{h} K_h(X_i - x) (I\{\varepsilon_i \leq y_2\} - F_{\varepsilon|X}(y_2|x)) \right] \\
& \quad - \frac{1}{n} E[C_n(y_1, x)] \cdot E[C_n(y_2, x)] \\
&= + \frac{1}{nh} (f_X^{-1}(x))^2 E \left[\frac{1}{h} K_h^2(X_1 - x) (I\{\varepsilon_1 \leq y_1\} - F_{\varepsilon|X}(y_1|x)) (I\{\varepsilon_1 \leq y_2\} - F_{\varepsilon|X}(y_2|x)) \right] \\
& \quad - O\left(\frac{h^4}{n}\right) \\
&= + \frac{1}{nh} (f_X^{-1}(x))^2 E \left[\frac{1}{h} K_h^2(X_1 - x) (I\{\varepsilon_1 \leq y_1\} - F_{\varepsilon|X}(y_1|x)) (I\{\varepsilon_1 \leq y_2\} - F_{\varepsilon|X}(y_2|x)) \right]
\end{aligned}$$

$$\begin{aligned}
& +o\left(\frac{1}{nh}\right) \\
& = +\frac{1}{nh} (f_X^{-1}(x))^2 E\left[\frac{1}{h} K_h^2(X_1 - x) I\{\varepsilon_1 \leq y_1\} I\{\varepsilon_1 \leq y_2\}\right] - \frac{1}{nh} \mu_0^{K^2} f_X^{-1}(x) F_{\varepsilon|X}(y_2|x) \cdot F_{\varepsilon|X}(y_1|x) \\
& +o\left(\frac{1}{nh}\right) \\
& = +\frac{1}{nh} \mu_0^{K^2} f_X^{-1}(x) F_{\varepsilon|X}(\min(y_1, y_2)|x) - \frac{1}{nh} \mu_0^{K^2} f_X^{-1}(x) F_{\varepsilon|X}(y_2|x) \cdot F_{\varepsilon|X}(y_1|x) + o\left(\frac{1}{nh}\right)
\end{aligned}$$

(xi):

$$\begin{aligned}
& \text{Cov}(H_{n,1}(y_1, x), H_{n,4}(y_2, x)) \\
& = \text{Cov}\left(-(\Sigma^{-1}(x)) \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) g(\varepsilon_i, x))^t \cdot \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x] - E[D_n(y_1, x)], \right. \\
& \quad \left. f_X^{-1}(x) \cdot \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) (I\{\varepsilon_i \leq y_2\} - F_{\varepsilon|X}(y_2|x)) - E[C_n(y_2, x)]\right) \\
& = -E\left[(\Sigma^{-1}(x)) \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) g(\varepsilon_i, x))^t \cdot \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x] \right. \\
& \quad \left. \cdot f_X^{-1}(x) \cdot \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) (I\{\varepsilon_i \leq y_2\} - F_{\varepsilon|X}(y_2|x))\right] \\
& \quad + E\left[(\Sigma^{-1}(x)) \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) g(\varepsilon_i, x))^t \cdot \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x]\right] \\
& \quad \cdot E\left[f_X^{-1}(x) \cdot \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) (I\{\varepsilon_i \leq y_2\} - F_{\varepsilon|X}(y_2|x))\right] \\
& = -\frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n E\left[(\Sigma^{-1}(x)) \frac{1}{h} K_h(X_i - x) g(\varepsilon_i, x))^t \cdot \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x] \right. \\
& \quad \left. \cdot f_X^{-1}(x) \cdot \frac{1}{h} K_h(X_j - x) (I\{\varepsilon_j \leq y_2\} - F_{\varepsilon|X}(y_2|x))\right] \\
& \quad + E\left[(\Sigma^{-1}(x)) \frac{1}{h} K_h(X_i - x) g(\varepsilon_i, x))^t \cdot \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x]\right] \\
& \quad \cdot E\left[f_X^{-1}(x) \cdot \frac{1}{h} K_h(X_i - x) (I\{\varepsilon_i \leq y_2\} - F_{\varepsilon|X}(y_2|x))\right] \\
& = -\frac{1}{n} E\left[(\Sigma^{-1}(x)) \frac{1}{h} K_h(X_i - x) g(\varepsilon_i, x))^t \cdot \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x] \right. \\
& \quad \left. \cdot f_X^{-1}(x) \cdot \frac{1}{h} K_h(X_i - x) (I\{\varepsilon_i \leq y_2\} - F_{\varepsilon|X}(y_2|x))\right] + O\left(\frac{h^4}{n}\right) \\
& = -\frac{1}{nh} \cdot f_X^{-1}(x) \mu_0^{K^2} \cdot E\left[\frac{1}{h} K_h^2(X_1 - x) (g(\varepsilon_1, x))^t (I\{\varepsilon_1 \leq y_2\} - F_{\varepsilon|X}(y_2|x))\right] \\
& \quad \cdot (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x] + o\left(\frac{1}{nh}\right)
\end{aligned}$$

$$= -\frac{1}{nh} (\mu_0^{K^2})^2 E[(g(\varepsilon, X))^t I\{\varepsilon \leq y_2\} | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x] + o\left(\frac{1}{nh}\right)$$

(xii) analog.

(xiii):

$$\begin{aligned} & \text{Cov}(H_{n,2}(y_1, x), H_{n,4}(y_2, x)) \\ &= \text{Cov}\left(\frac{1}{n} \sum_{i=1}^n \varepsilon_i \int \frac{1}{h} K_h(z-x) (g'(y, x))^t \frac{1}{b} \tilde{K}_b(X_i - z) f_{\varepsilon|X}(y|z) d(y, z) \right. \\ & \quad \cdot (\Sigma^{-1}(x))^t \cdot \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x], \\ & \quad \left. f_X^{-1}(x) \cdot \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) (I\{\varepsilon_i \leq y_2\} - F_{\varepsilon|X}(y_2|x)) - E[C_n(y_2, x)]\right) \\ &= E\left[\frac{1}{n} \sum_{i=1}^n \varepsilon_i \int \frac{1}{h} K_h(z-x) (g'(y, x))^t \frac{1}{b} \tilde{K}_b(X_i - z) f_{\varepsilon|X}(y|z) d(y, z) \right. \\ & \quad \cdot (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x] \mu_0^{K^2} f_X^{-1}(x) \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) (I\{\varepsilon_i \leq y_2\} - F_{\varepsilon|X}(y_2|x)) \left. \right] \\ &= \frac{1}{n} E\left[\varepsilon_1 \int \frac{1}{h} K_h(z-x) (g'(y, x))^t \frac{1}{b} \tilde{K}_b(X_1 - z) f_{\varepsilon|X}(y|z) d(y, z) \cdot \frac{1}{h} K_h(X_1 - x) I\{\varepsilon_1 \leq y_2\} \right] \\ & \quad \cdot (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x] \mu_0^{K^2} f_X^{-1}(x) \\ &= \frac{1}{nh} E[\varepsilon I\{\varepsilon \leq y_2\} | X = x] E[(g'(\varepsilon, x))^t | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x] \cdot \mu_0^{K^2} \\ & \quad \cdot \int \int K(u - \frac{b}{h} v) \tilde{K}(u) K(v) du dv + o\left(\frac{1}{nh}\right) \end{aligned}$$

(xiv) analog.

(xv):

$$\begin{aligned} & \text{Cov}(H_{n,3}(y_1, x), H_{n,4}(y_2, x)) \\ &= \text{Cov}\left(f_X^{-1}(x) \frac{1}{n} \sum_{i=1}^n \varepsilon_i \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(y_1|z) \frac{1}{b} \tilde{K}_b(X_i - z) dz, \right. \\ & \quad \left. f_X^{-1}(x) \cdot \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) (I\{\varepsilon_i \leq y_2\} - F_{\varepsilon|X}(y_2|x)) - E[C_n(y_2, x)]\right) \\ &= \text{Cov}\left(f_X^{-1}(x) \frac{1}{n} \sum_{i=1}^n \varepsilon_i \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(y_1|z) \frac{1}{b} \tilde{K}_b(X_i - z) dz, f_X^{-1}(x) \cdot \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) I\{\varepsilon_i \leq y_2\}\right) \end{aligned}$$

$$\begin{aligned}
&= E \left[(f_X^{-1}(x))^2 \frac{1}{n} \sum_{i=1}^n \varepsilon_i \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(y_1|z) \frac{1}{b} \tilde{K}_b(X_i - z) dz \cdot \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) I\{\varepsilon_i \leq y_2\} \right] \\
&= \frac{1}{n} (f_X^{-1}(x))^2 E \left[\varepsilon_1 \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(y_1|z) \frac{1}{b} \tilde{K}_b(X_1 - z) \frac{1}{h} K_h(X_1 - x) I\{\varepsilon_1 \leq y_2\} dz \right] \\
&= \frac{1}{nh} f_X^{-1}(x) f_{\varepsilon|X}(y_1|x) E[\varepsilon I\{\varepsilon \leq y_2\} | X = x] \cdot \int \int K(u - v \frac{b}{h}) \tilde{K}(u) K(v) du dv + o(\frac{1}{nh})
\end{aligned}$$

(xvi) analog.

Damit ist der Beweis zu Proposition 6.4 abgeschlossen.

□

Bemerkung 6.5 (Bandbreiten) In Proposition 6.4 treten die Bandbreiten b und h in Wechselwirkung. Die in den einzelnen Rechnungen durchgeführten Substitutionen sind so durchgeführt worden, dass die Varianzen jeweils von der Rate $O(\frac{1}{nh})$ und nicht von der Rate $O(\frac{1}{nb})$ sind, denn damit wäre die Rate zu schlecht abgeschätzt. Viele der Kovarianzterme enthalten noch Integrale der Art $\int K(u - v \frac{b}{h}) K(u) K(v) d(u, v)$. Mit $\lim_{n \rightarrow \infty} \frac{b}{h} = 0$, also b schneller gegen Null laufend als h (siehe (H1) und (H2)), vereinfachen sich diese Integrale asymptotisch noch zu $\int \int K^2(u) K(v) du dv = \mu_0^{K^2}$. Zu Beginn des nächsten Kapitels wird betrachtet werden, ob es eine zulässige Wahl des Verhältnisses von h und b gibt, welche ganz allgemein zu einer Verbesserung des Bias und der Varianz des dieser Arbeit betrachteten Empirical-Likelihood Schätzers im Vergleich zum ebenfalls betrachteten herkömmlichen Kernschätzer führt. Es zeigt sich aber, dass bei einem anderen Verhältnis die Kernfunktion und ihre Eigenschaften, speziell Ableitungen, eine große Rolle spielen. Es existiert aufgrund von Wechselwirkungen des Kernes und der (unbekannten) bedingten Dichte von ε gegeben X kein gleichmäßig bester Kern für alle möglichen bedingten Verteilungen, noch nicht einmal eingeschränkt auf die Klasse der Normalverteilungen.

6.4 Beweis der schwachen Konvergenz der Prozesse

Wir beginnen mit dem Beweis der Euklidizität einiger Funktionenklassen:

Lemma 6.6 Es gelten die Voraussetzungen (A), (K), (H2). Die folgenden Funktionenklassen sind euklidisch und die in der Abschätzung der Überdeckungsanzahl auftretenden Konstanten sind unabhängig von n wählbar:

$$(i) \mathcal{G}_n := \{ \varphi_{n,y} : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R} \mid \varphi_{n,y}(\varepsilon, z) = K^2\left(\frac{z-x}{h}\right) I\{\varepsilon \leq y\}, y \in \mathbb{R} \}$$

$$(ii) \mathcal{F}_n := \{ \varphi_{n,y} : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R} \mid \varphi_{n,y}(\varepsilon, z) = K^2\left(\frac{z-x}{h}\right) g_j(\varepsilon, x) I\{\varepsilon \leq y\}, y \in \mathbb{R} \}$$

wobei die n -Abhängigkeit durch $h = h_n$ gegeben ist.

Beweis zu (i):

Sei $c > 0$. Für n genügend groß ist $h < c$ erfüllt, so dass wir im Weiteren die Klassen $\mathcal{G}_n, n \in \mathbb{N}$ vereinigen und uns auf

$$\mathcal{G} := \{ \varphi_{h,y} : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R} \mid \varphi_{h,y}(\varepsilon, z) = K^2\left(\frac{z-x}{h}\right) I\{\varepsilon \leq y\}, h \in (0, c), y \in \mathbb{R} \}$$

beschränken können. Von dieser Funktionenklasse ist nun die Euklidizität zu beweisen. Die Klassen $\mathcal{G}_1 := \{ \varphi_h : [0, 1] \rightarrow \mathbb{R} \mid \varphi_h(z) = K^2\left(\frac{z-x}{h}\right), h \in (0, c) \}$ und $\mathcal{G}_2 := \{ \varphi_y : \mathbb{R} \rightarrow \mathbb{R} \mid \varphi_y(\varepsilon) = I\{\varepsilon \leq y\}, y \in \mathbb{R} \}$ sind unter der Annahme, dass der Kern K von beschränkter Variation euklidisch, siehe dazu Neumeyer (2006), S. 152, Lemma A.13 bzw. Beispiel 2.28 und Proposition 2.33 dieser Arbeit. Mit Proposition 2.20 übertragen sich die Überdeckungszahlen, so dass die Klasse \mathcal{G} ebenfalls euklidisch ist, aber hier die Überdeckungszahl dann abschätzbar durch das Produkt der (in ϵ polynomiellen) Schranken der Klassen \mathcal{G}_1 und \mathcal{G}_2 ist.

Die Aussage (ii) folgt direkt mit (i) und Proposition 2.31.

□

Als letzte Vorarbeit vor dem Beweis der schwachen Konvergenz der in dieser Arbeit im vorherigen Kapitel herausgearbeiteten stochastischen Prozesse betrachten wir einige Funktionenklassen und zeigen deren Donskereigenschaft, dabei nutzen wir auch die vorangegangene Proposition:

Lemma 6.7 *Es gelten die Voraussetzungen (M1) - (M3), (A), (K), (H1), (H2) und es sei $\mathcal{H} := C_\delta^{1+\alpha}[0, 1]$ die in 2.34 definierte Funktionenklasse. Die folgenden von n abhängigen Funktionenklassen sind Donskerklassen:*

$$(i) \mathcal{F}_{1,n} = \{ \varphi_{n,\tilde{h}} : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R} \mid \varphi_{n,\tilde{h}}(\varepsilon, z) = \frac{1}{\sqrt{h}} K\left(\frac{z-x}{h}\right) (g_j(\varepsilon + \tilde{h}(z), x) - g_j(\varepsilon, x)), \tilde{h} \in \mathcal{H} \}$$

$$(ii) \mathcal{F}_{2,n} = \{ \varphi_{n,\tilde{h}} : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R} \mid \varphi_{n,\tilde{h}}(\varepsilon, z) = \frac{1}{\sqrt{h}} K\left(\frac{z-x}{h}\right) g_j(\varepsilon + \tilde{h}(z), x), \tilde{h} \in \mathcal{H} \}$$

$$(iii) \mathcal{F}_{3,n} := \{ \varphi_{n,y} : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R} \mid \varphi_{n,y}(\tilde{\varepsilon}, z) = \frac{1}{\sqrt{h}} K_h(z-x) g_{j_1}(\tilde{\varepsilon}, x) \\ \cdot E[g_{j_2}(\varepsilon, X) I\{\varepsilon \leq y\} \mid X = x], y \in \mathbb{R} \}$$

$$(iv) \mathcal{F}_{5,n} := \{ \varphi_{n,y} : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R} \mid \varphi_{n,y}(\varepsilon, z) = \varepsilon \int \frac{1}{\sqrt{h}} K_h(u-x) f_{\varepsilon|X}(y|u) \frac{1}{b} \tilde{K}_b(z-u) du, y \in \mathbb{R} \}$$

$$(v) \mathcal{F}_{6,n} := \{ \varphi_{n,y} : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R} \mid \varphi_{n,y}(\varepsilon, z) = \frac{1}{\sqrt{h}} K_h(z-x) (I\{\varepsilon \leq y\} - F_{\varepsilon|X}(y|x)), y \in \mathbb{R} \}$$

$$(vi) \mathcal{F}_{7,n} := \{ \varphi_{n,t} : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R}, \mid \varphi_{n,t}(\varepsilon, z) = \frac{1}{\sqrt{h}} K\left(\frac{z-x}{h}\right) (I\{\varepsilon \leq y - \tilde{h}(z)\} - I\{\varepsilon \leq y\}), \\ t = (y, \tilde{h}), \tilde{h} \in \mathcal{H}, y \in \mathbb{R} \}$$

$$(vii) \mathcal{F}_{8,n} := \{ \varphi_{n,t} : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R} \mid \varphi_{n,t}(\varepsilon, z) = \varepsilon \cdot \int \tilde{h}(u) \frac{1}{\sqrt{h}} K\left(\frac{u-x}{h}\right) f_{\varepsilon|X}(y|u) \frac{1}{b} \tilde{K}\left(\frac{z-u}{b}\right) du, \\ t = (y, \tilde{h}), \tilde{h} \in \mathcal{H}, y \in \mathbb{R} \}$$

Beweis:

Da die obigen Funktionenklassen von n abhängen und keine von n unabhängige beschränkte Einhüllende Φ besitzen wird Proposition 2.41 und Bemerkung 2.42 angewendet. Der Erwartungswert und die Varianz wurde für Prozesse auf Basis dieser Klassen schon in diesem Kapitel in Proposition 6.2 bzw. Proposition 6.4 bzw. an der Stelle des Auftretens der jeweiligen Klasse in vorangegangenen Beweisen berechnet.

Wir beginnen mit dem Beweis zu (i) und zeigen die restlichen für Proposition 2.41 und die Verallgemeinerung 2.42 nötigen Voraussetzungen (2.5), (2.6), (2.7) und (2.9), dabei werden wir hier die Voraussetzungen an g , d.h. (G1) und (G2) bzw. die Folgerung (3.3) teilweise direkt nutzen.

Für die erste Klasse, $\mathcal{F}_{1,n}$, ist

$$\Phi_n(\varepsilon, z) := \frac{1}{\sqrt{h}} \left| K\left(\frac{z-x}{h}\right) \right| \sup_{y \in \mathbb{R}: |y| \leq \delta} |g_j(\varepsilon + y, x) - g_j(\varepsilon, x)|$$

eine Einhüllende, vergleiche dazu die Definition 2.34 der Klasse $\mathcal{H} := C_\delta^{1+\alpha}[0, 1]$. Es gilt:

$$E[\Phi_n^2(\varepsilon, X)] = E\left[\frac{1}{h} K^2\left(\frac{X-x}{h}\right) E\left[\sup_{y \in \mathbb{R}: |y| \leq \delta} (g_j(\varepsilon + y, x) - g_j(\varepsilon, x))^2 \mid X\right]\right] = O(1),$$

denn mittels Substitution entfällt das $\frac{1}{h}$, für den inneren Erwartungswert nutze direkt die Folgerung (3.3) und damit folgt (2.6). Zum Beweis der Gültigkeit der Gleichung (2.7) betrachte folgende Rechnung, dabei ist $m > 0$:

$$\begin{aligned}
& E[\Phi_n^2(\varepsilon, X) I\{\Phi_n(\varepsilon, X) > \beta\sqrt{n}\}] \\
&= E[\Phi_n^2(\varepsilon, X) I\{\frac{\Phi_n(\varepsilon, X)}{\beta\sqrt{n}} > 1\}] \\
&\leq E[\frac{\Phi_n^{m+2}(\varepsilon, X)}{\beta^m(n)^{m/2}}] \\
&= \frac{1}{\beta^m(nh)^{m/2}} E[\frac{1}{h} K^{2+m}(\frac{X-x}{h}) E[\sup_{y \in \mathbb{R}: |y| \leq \delta} (g_j(\varepsilon + y, x) - g_j(\varepsilon, x))^{2+m} | X]] \\
&= \frac{1}{\beta^m(nh)^{m/2}} \int \frac{1}{h} K^{2+m}(\frac{z-x}{h}) E[\sup_{y \in \mathbb{R}: |y| \leq \delta} (g_j(\varepsilon + y, x) - g_j(\varepsilon, x))^{2+m} | X = z] f_X(z) dz \\
&= \frac{1}{\beta^m(nh)^{m/2}} \int K^{2+m}(u) E[\sup_{y \in \mathbb{R}: |y| \leq \delta} (g_j(\varepsilon + y, x) - g_j(\varepsilon, x))^{2+m} | X = x + uh] f_X(x + uh) du \\
&= o(1),
\end{aligned}$$

dabei nutze Voraussetzung (G2).

Zum Beweis der Gültigkeit der Gleichung (2.5). Es seien $s, t \in C_\delta^{1+\alpha}[0, 1]$, die nötige Semimetrik sei gegeben durch

$$\rho(s, t) := \|s - t\|_\infty.$$

Zu zeigen ist, dass $(C_\delta^{1+\alpha}[0, 1], \rho)$ totalbeschränkter semimetrischer Raum ist. Dies folgt mit einfachen Rechnungen, denn ρ ist Semimetrik, was sich direkt aus der Definition ergibt. Zur Totalbeschränktheit: man nutze die Eigenschaft $\|s - t\|_\infty \leq 2\delta$ für alle $s, t \in C_\delta^{1+\alpha}[0, 1]$, die Totalbeschränktheit liefert Kapitel 2.7 aus van der Vaart und Wellner (2000). Es folgt weiter:

$$\begin{aligned}
& \sup_{\rho(s,t) < \delta_n} E[(\varphi_{n,s}(\varepsilon, X) - \varphi_{n,t}(\varepsilon, X))^2] \\
&= \sup_{\rho(s,t) < \delta_n} \frac{1}{h} E[K^2(\frac{X-x}{h}) (g_j(\varepsilon + s(X), x) - g_j(\varepsilon + t(X), x))^2] \\
&= \sup_{\rho(s,t) < \delta_n} \frac{1}{h} E[K^2(\frac{X-x}{h}) E[(g_j(\varepsilon + s(X), x) - g_j(\varepsilon + t(X), x))^2 | X]]
\end{aligned}$$

$$\begin{aligned}
&= \sup_{\rho(s,t) < \delta_n} \frac{1}{h} \int K^2\left(\frac{z-x}{h}\right) E[(g_j(\varepsilon + s(X), x) - g_j(\varepsilon + t(X), x))^2 | X = z] f_X(z) dz \\
&= \sup_{\rho(s,t) < \delta_n} \int K^2(u) E[(g_j(\varepsilon + s(X), x) - g_j(\varepsilon + t(X), x))^2 | X = x + uh] f_X(x + uh) du \\
&\leq \int K^2(u) E\left[\sup_{y_1, y_2 \in \mathbb{R}: |y_1| < \delta, |y_2| < \delta, |y_1 - y_2| < \delta_n} (g_j(\varepsilon + y_1, x) - g_j(\varepsilon + y_2, x))^2 | X = x + uh\right] f_X(x + uh) du \\
&\stackrel{(G2)}{=} o(1)
\end{aligned}$$

Zuletzt folgt Bedingung (2.9). Es ist zu zeigen, dass für $(\varepsilon, X) \sim \mathcal{P}$ gilt:

$$\int_0^{\delta_n} \sqrt{\log \mathcal{N}_{[]}(\varepsilon \|\Phi_n\|_{2, \mathcal{P}}, \mathcal{F}_{1,n}, L_2(\mathcal{P}))} d\varepsilon \rightarrow 0 \quad \text{für jedes } \delta_n \rightarrow 0$$

Beweis:

Setze $\tilde{\varepsilon} := (\frac{\varepsilon}{4C^2 \|K^2\|_\infty \|f_X\|_\infty})^{\kappa/2}$ mit C und κ aus (G2). Dann folgt mit Proposition 2.35, dass

$$(6.13) \quad \log \mathcal{N}(\tilde{\varepsilon}, C_\delta^{1+\alpha}[0, 1], \|\cdot\|_\infty) \leq K_1 \left(\frac{1}{\tilde{\varepsilon}}\right)^{1/(1+\alpha)} = K_1 \left(\frac{4C\mu_0^{K^2}}{\varepsilon}\right)^{\kappa/(1+\alpha)}$$

Betrachtet man $\tilde{\varepsilon} \geq \delta$, so ist nach der Definition von $C_\delta^{1+\alpha}[0, 1]$ die Überdeckungszahl $\mathcal{N}(\tilde{\varepsilon}, C_\delta^{1+\alpha}[0, 1], \|\cdot\|_\infty) = 1$. Von Interesse ist also nur der Fall $\tilde{\varepsilon} < \delta \Rightarrow \varepsilon < \delta^{1/\kappa} 4C\mu_0^{K^2}$. Sei $\lambda = \mathcal{N}(\tilde{\varepsilon}, C_\delta^{1+\alpha}[0, 1], \|\cdot\|_\infty)$ und die Zentren der $\tilde{\varepsilon}$ -Bälle der Überdeckung von $C_\delta^{1+\alpha}[0, 1]$ gegeben durch c_1, \dots, c_λ . Nun wollen wir die Funktionenklasse $\mathcal{F}_{1,n}$ mittels Klammern überdecken, deren Grundlage die Zentren bilden. Betrachte die folgende Klammer:

$$\left[\frac{1}{\sqrt{h}} K\left(\frac{z-x}{h}\right) \left((g_j(\varepsilon + c_i(z), x) - g_j(\varepsilon, x)) - \sup_{\substack{z, \tilde{z} \in \mathbb{R}: |z| \leq \delta, \\ |\tilde{z}| \leq \delta, |z - \tilde{z}| \leq \tilde{\varepsilon}}} (g_j(\varepsilon + z, x) - g_j(\varepsilon + \tilde{z}, x)) \right), \right. \\
\left. \frac{1}{\sqrt{h}} K\left(\frac{z-x}{h}\right) \left((g_j(\varepsilon + c_i(z), x) - g_j(\varepsilon, x)) + \sup_{\substack{z, \tilde{z} \in \mathbb{R}: |z| \leq \delta, \\ |\tilde{z}| \leq \delta, |z - \tilde{z}| \leq \tilde{\varepsilon}}} (g_j(\varepsilon + z, x) - g_j(\varepsilon + \tilde{z}, x)) \right) \right]$$

Diese λ Klammern überdecken $\mathcal{F}_{1,n}$, denn die λ $\tilde{\varepsilon}$ -Bälle überdecken $C_\delta^{1+\alpha}[0, 1]$ und damit lässt sich nach Konstruktion der Klasse $\mathcal{F}_{1,n}$ auch diese mit λ Klammern überdecken.

Ebenfalls gilt, dass diese Klammer höchstens $L_2(\mathcal{P})$ -Länge ε hat, denn

$$\begin{aligned}
& 2E\left[\frac{1}{h}K^2\left(\frac{X-x}{h}\right)\left(\sup_{\substack{z, \tilde{z} \in \mathbb{R}: |z| \leq \delta, \\ |\tilde{z}| \leq \delta, |z-\tilde{z}| \leq \tilde{\epsilon}}} (g_j(\varepsilon+z, x) - g_j(\varepsilon+\tilde{z}, x))^2\right)\right] \\
&= 2E\left[\frac{1}{h}K^2\left(\frac{X-x}{h}\right)E\left[\left(\sup_{\substack{z, \tilde{z} \in \mathbb{R}: |z| \leq \delta, \\ |\tilde{z}| \leq \delta, |z-\tilde{z}| \leq \tilde{\epsilon}}} (g_j(\varepsilon+z, x) - g_j(\varepsilon+\tilde{z}, x))^2 \mid X\right)\right]\right] \\
&\stackrel{(G2)}{\leq} 2 \int \frac{1}{h}K^2\left(\frac{u-x}{h}\right) f_X(u) du \cdot C^2 \tilde{\epsilon}^{2/\kappa} \\
&\leq 4 \|K^2\|_\infty \|f_X\|_\infty \cdot C^2 \tilde{\epsilon}^{2/\kappa} \\
&\leq \epsilon.
\end{aligned}$$

Die Überdeckungsanzahl der Klasse $C_\delta^{1+\alpha}[0, 1]$ überträgt sich also auf die Klammeranzahl der Klasse $\mathcal{F}_{n,1}$. Es gilt nach Gleichung (6.13) dass für $\kappa < 2(1 + \alpha)$ (siehe (G2))

$$\int_0^{\delta_n} \sqrt{\log \mathcal{N}_{[]}(\epsilon \|\Phi_n\|_{2, \mathcal{P}}, \mathcal{F}_{1,n}, L_2(\mathcal{P}))} d\epsilon \leq K_3 \int_0^{\delta_n} \epsilon^{-\frac{\kappa}{2(1+\alpha)}} d\epsilon \xrightarrow{n \rightarrow \infty} 0.$$

Damit folgt Proposition 6.7 (i). Der Beweis zu (ii) folgt mit identischen Rechnungen:

Als nächstes betrachten wir die in (iii) gegebene Funktionenklasse $\mathcal{F}_{n,3}$ und gehen in gleicher Reihenfolge wie in (i) vor.

Es ist mit der Voraussetzung $\sup_{y \in \mathbb{R}} |E[g_{j_2}(\varepsilon, X) I\{\varepsilon \leq y\} \mid X = x]| \leq C$, siehe (M3),

$$\Phi_n(\tilde{\varepsilon}, z) := \frac{1}{\sqrt{h}} |K_h(z-x) g_{j_1}(\tilde{\varepsilon}, x)| \cdot C$$

eine Einhüllende. Es folgt:

$$\begin{aligned}
E[\Phi_n^2(\varepsilon, X)] &= C^2 \cdot E\left[\frac{1}{h} (K_h(X-x) g_{j_1}(\varepsilon, x))^2\right] \\
&= O(1),
\end{aligned}$$

denn mittels Substitution entfällt das $\frac{1}{h}$, entwickle die gemeinsame Dichte in einer Taylorreihe und nutze die Voraussetzungen (A) und (M1). Sei m so gewählt, dass $r_1 \geq m > 0$ (vgl. (A)).

Dann gilt weiter:

$$\begin{aligned}
E[\Phi_n^2(\varepsilon, X) I\{\Phi_n(\varepsilon, X) > \beta\sqrt{n}\}] &= E[\Phi_n^2(\varepsilon, X) I\{\frac{\Phi_n(\varepsilon, X)}{\beta\sqrt{n}} > 1\}] \\
&\leq E[\frac{\Phi_n^{m+2}(\varepsilon, X)}{\beta^m(n)^{m/2}}] \\
&= \frac{C^{m+2}}{\beta^m(nh)^{m/2}} \frac{1}{h} E[(K_h(X-x)g_{j_1}(\varepsilon, x))^{m+2}] \\
&= O(\frac{1}{(nh)^{m/2}}) \\
&= o(1),
\end{aligned}$$

dabei folgt die vorletzte Gleichheit aus (M3) (viii). Zum Beweis der Gültigkeit der Gleichung (2.5). Es seien $s, t \in \mathbb{R}$, die nötige Semimetrik sei gegeben durch

$$\begin{aligned}
\rho(s, t) &:= E[|g_{j_2}(\varepsilon, X)| | I\{\varepsilon \leq s\} - I\{\varepsilon \leq t\} | | X = x] \\
&= \int |g_{j_2}(\varepsilon, x)| | I\{\varepsilon \leq s\} - I\{\varepsilon \leq t\} | f_{\varepsilon|X}(\varepsilon|x) d\varepsilon
\end{aligned}$$

Zu zeigen ist, dass (\mathbb{R}, ρ) totalbeschränkter semimetrischer Raum ist. Dies folgt mit einfachen Rechnungen: ρ ist Semimetrik, was sich direkt aus der Definition ergibt. Es fehlt noch die Totalbeschränktheit von (\mathbb{R}, ρ) . Sei $\epsilon > 0$ beliebig. Es gilt nach Voraussetzung, dass

$$2 E[|g_{j_2}(\varepsilon, X)| | X = x] = 2 \int_{\mathbb{R}} |g_{j_2}(\varepsilon, x)| f_{\varepsilon|X}(\varepsilon|x) d\varepsilon = 2C < \infty$$

Zerlege nun \mathbb{R} so in $n_0 = \lceil C/\epsilon \rceil$ disjunkte Intervalle A_1, \dots, A_{n_0} , so dass für jedes $A_i, i = 1, \dots, n_0$ gilt:

$$\int_{A_i} 2 |g_{j_2}(\varepsilon, x)| f_{\varepsilon|X}(\varepsilon|x) d\varepsilon \leq 2\epsilon$$

Seien nun a_1, \dots, a_{n_0} die Zentren der Intervalle A_i , so gewählt, dass für alle $s \in A_i$ gilt: $\rho(s, a_i) \leq \epsilon$. Zu jedem $\epsilon > 0$ haben wir also endlich viele, genauer, n_0 ϵ -Bälle gefunden, welche ganz \mathbb{R} überdecken und damit ist (\mathbb{R}, ρ) totalbeschränkter semimetrischer Raum.

Es folgt weiter:

$$\begin{aligned}
& \sup_{\rho(s,t) < \delta_n} E[(\varphi_{n,s}(\varepsilon, X) - \varphi_{n,t}(\varepsilon, X))^2] \\
&= \sup_{\rho(s,t) < \delta_n} \frac{1}{h} E[(K_h(X-x)g_{j_1}(\varepsilon, x))^2] \cdot (E[g_{j_2}(\varepsilon, X)I\{\varepsilon \leq s\} - g_{j_2}(\varepsilon, X)I\{\varepsilon \leq t\} | X=x])^2 \\
&= O(1) \cdot \sup_{\rho(s,t) < \delta_n} (E[g_{j_2}(\varepsilon, X)(I\{\varepsilon \leq s\} - I\{\varepsilon \leq t\}) | X=x])^2 \\
&\leq O(\delta_n^2) \\
&= o(1)
\end{aligned}$$

Bedingung (2.9) wird in Lemma 6.8 bewiesen und somit Aussage 6.7 (iii).

Als nächstes betrachten wir die in (iv) gegebene Funktionenklasse $\mathcal{F}_{5,n}$.

Die Funktionenklasse hat die Gestalt

$$\mathcal{F}_{5,n} := \{ \varphi_{n,y} : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R} \mid \varphi_{n,y}(\varepsilon, z) = \varepsilon \int \frac{1}{\sqrt{h}} K_h(u-x) f_{\varepsilon|X}(y|u) \frac{1}{b} \tilde{K}_b(z-u) du, y \in \mathbb{R} \}$$

Wegen $\sup_{y \in \mathbb{R}, x \in [0,1]} |f_{\varepsilon|X}(y|x)| < c$, vgl. (M2), ist

$$\Phi_n(\varepsilon, z) := c \cdot |\varepsilon| \int \frac{1}{\sqrt{h}} K_h(u-x) \frac{1}{b} \tilde{K}_b(z-u) du$$

eine Einhüllende. Es folgt direkt mit den Rechnungen zu Proposition 6.4 (iii):

$$E[\Phi_n^2(\varepsilon, X)] = c^2 E[\varepsilon^2 (\int \frac{1}{\sqrt{h}} K_h(u-x) \frac{1}{b} \tilde{K}_b(X-u) du)^2] = O(1).$$

Sei $r_0 \geq m > 0$. Dann gilt:

$$\begin{aligned}
E[\Phi_n^2(\varepsilon, X) I\{\Phi_n(\varepsilon, X) > \beta\sqrt{n}\}] &= E[\Phi_n^2(\varepsilon, X) I\{\frac{\Phi_n(\varepsilon, X)}{\beta\sqrt{n}} > 1\}] \\
&\leq E[\frac{\Phi_n^{m+2}(\varepsilon, X)}{\beta^m(n)^{m/2}}]
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\beta^m (nh)^{m/2} h} E\left[(|\varepsilon| \int |K_h(u-x) \frac{1}{b} \tilde{K}_b(X-u)| du)^{m+2} \right] \\
&\stackrel{(M3)}{=} O\left(\frac{1}{(nh)^{m/2}}\right) \\
&= o(1)
\end{aligned}$$

Zum Beweis der Gültigkeit der Gleichung (2.5). Es seien $s, t \in \mathbb{R}$, die nötige Semimetrik sei gegeben durch

$$\rho(s, t) := |f_{\varepsilon|X}(s|x) - f_{\varepsilon|X}(t|x)|$$

Zu zeigen ist, dass (\mathbb{R}, ρ) totalbeschränkter semimetrischer Raum ist. Dies folgt mit einfachen Rechnungen aus der Beschränktheit und dem Randverhalten der Dichte, vergleiche dazu auch den Beweis zu (ii). Es folgt weiter:

$$\begin{aligned}
&\sup_{\rho(s,t) < \delta_n} E[(\varphi_{n,s}(\varepsilon, X) - \varphi_{n,t}(\varepsilon, X))^2] \\
&= \sup_{\rho(s,t) < \delta_n} \frac{1}{h} E\left[\left(\varepsilon \int K_h(u-x) f_{\varepsilon|X}(s|u) \frac{1}{b} \tilde{K}_b(X-u) du - \varepsilon \int K_h(u-x) f_{\varepsilon|X}(t|u) \frac{1}{b} \tilde{K}_b(X-u) du \right)^2 \right] \\
&= \sup_{\rho(s,t) < \delta_n} \frac{1}{h} E\left[\left(\varepsilon \int K_h(u-x) (f_{\varepsilon|X}(s|u) - f_{\varepsilon|X}(t|u)) \frac{1}{b} \tilde{K}_b(X-u) du \right)^2 \right] \\
&\leq O(\delta_n^2) + o(1) \\
&= o(1)
\end{aligned}$$

Zuletzt folgt Bedingung (2.9) aus Lemma 6.8 und somit Aussage 6.7 (iv)

Als nächstes betrachten wir die in (v) gegebene Funktionenklasse $\mathcal{F}_{6,n}$.

Es ist

$$\Phi_n(\varepsilon, z) := \frac{1}{\sqrt{h}} K_h(z-x)$$

eine Einhüllende und mittels Substitution folgt direkt:

$$E[\Phi_n^2(\varepsilon, X)] = E\left[\left(\frac{1}{\sqrt{h}} K_h(X-x)\right)^2\right] = O(1).$$

Sei wie oben wieder $r_0 \geq m > 0$. Dann gilt:

$$\begin{aligned}
E[\Phi_n^2(\varepsilon, X)I\{\Phi_n(\varepsilon, X) > \beta\sqrt{n}\}] &= E[\Phi_n^2(\varepsilon, X)I\{\frac{\Phi_n(\varepsilon, X)}{\beta\sqrt{n}} > 1\}] \\
&\leq E[\frac{\Phi_n^{m+2}(\varepsilon, X)}{\beta^m(n)^{m/2}}] \\
&= \frac{1}{\beta^m(nh)^{m/2}} \frac{1}{h} E[(K_h(X-x))^{m+2}] \\
&\stackrel{(K)}{=} O(\frac{1}{(nh)^{m/2}}) \\
&= o(1)
\end{aligned}$$

Zum Beweis der Gültigkeit der Gleichung (2.5). Es seien $s, t \in \mathbb{R}$, die nötige Semimetrik sei gegeben durch

$$\rho(s, t) := |F_{\varepsilon|X}(s|x) - F_{\varepsilon|X}(t|x)|$$

Zu zeigen ist, dass (\mathbb{R}, ρ) totalbeschränkter semimetrischer Raum ist. Dies folgt mit einfachen Rechnungen, vergleiche dazu auch den Beweis zu (ii). Es folgt weiter:

$$\begin{aligned}
&\sup_{\rho(s,t) < \delta_n} E[(\varphi_{n,s}(\varepsilon, X) - \varphi_{n,t}(\varepsilon, X))^2] \\
&= \sup_{\rho(s,t) < \delta_n} \frac{1}{h} E\left[\left(K_h(X-x)(I\{\varepsilon \leq s\} - F_{\varepsilon|X}(s|x)) - K_h(X-x)(I\{\varepsilon \leq t\} - F_{\varepsilon|X}(t|x))\right)^2\right] \\
&= \sup_{\rho(s,t) < \delta_n} \frac{1}{h} E\left[\left(K_h(X-x)(I\{\varepsilon \leq s\} - I\{\varepsilon \leq t\} - F_{\varepsilon|X}(s|x) + F_{\varepsilon|X}(t|x))\right)^2\right] \\
&= \sup_{\rho(s,t) < \delta_n} \frac{1}{h} E\left[\left(K_h(X-x)(I\{\varepsilon \leq s\} - I\{\varepsilon \leq t\}) + K_h(X-x) \cdot (F_{\varepsilon|X}(t|x) - F_{\varepsilon|X}(s|x))\right)^2\right] \\
&\leq \sup_{\rho(s,t) < \delta_n} \frac{2}{h} \left(E[K_h^2(X-x)(I\{\varepsilon \leq s\} - I\{\varepsilon \leq t\})^2] + E[K_h^2(X-x)] \cdot (F_{\varepsilon|X}(t|x) - F_{\varepsilon|X}(s|x))^2\right) \\
&\stackrel{(1)}{\leq} O(\delta_n) + o(1) \\
&= o(1)
\end{aligned}$$

Erklärung zu (1). Das voran stehende $\frac{1}{h}$ entfällt nach Substitution. Beide Summanden kann man mit einigen Rechnungen auf die Semimetrik zurückführen, dabei nutzt man maßgeblich

Eigenschaften des Kernes K und die in (M2) gegebenen Voraussetzungen an $F_{\varepsilon|X}(y|x)$, welche die Entwicklung in Taylorreihen um x ermöglichen, aus, so dass gilt

$$\begin{aligned} E\left[\frac{1}{h} K_h^2(X-x) (I\{\varepsilon \leq s\} - I\{\varepsilon \leq t\})\right] &= E\left[\frac{1}{h} K_h^2(X-x) E[I\{\varepsilon \leq s\} - I\{\varepsilon \leq t\} | X]\right] \\ &= E\left[\frac{1}{h} K_h^2(X-x) (F_{\varepsilon|X}(s|X) - F_{\varepsilon|X}(t|X))\right] \\ &\leq C \cdot |F_{\varepsilon|X}(s|x) - F_{\varepsilon|X}(t|x)| + O(h) \end{aligned}$$

dabei gilt die letzte Zeile für n hinreichend groß. Der Restterm der Taylorentwicklung kann dabei unabhängig von s und t durch $O(h)$ abgeschätzt werden, da $\|F_{\varepsilon|X}\|_{\infty} = 1$ und $\|f_{\varepsilon|X}\|_{\infty} < \infty$ nach (M2).

Eine ähnliche Rechnung ist auch für den quadratischen Term durchführbar, so dass sich insgesamt (1) ergibt.

Zuletzt folgt Bedingung (2.9) aus Lemma 6.8 und somit Aussage 6.7 (v).

Als nächstes betrachten wir die in (vi) gegebene Funktionenklasse $\mathcal{F}_{7,n}$.

Für diese Klasse sind die Rechnungen zu (i) und (v) meist übertragbar. Es ist

$$\Phi_n(\varepsilon, z) := \frac{1}{\sqrt{h}} K\left(\frac{z-x}{h}\right)$$

eine Einhüllende. Die Voraussetzungen (2.6) und (2.7) ergeben sich direkt. Zu (2.5), hier ist die Semimetrik gegeben durch

$$\begin{aligned} \rho(s, t) = \rho((s_1, s_2), (t_1, t_2)) &= |F_{\varepsilon|X}(s_2 - s_1(x)|x) - F_{\varepsilon|X}(t_2 - t_1(x)|x)| \\ &\quad + |F_{\varepsilon|X}(s_2|x) - F_{\varepsilon|X}(t_2|x)| \end{aligned}$$

Zu zeigen ist, dass $(\mathcal{H} \times \mathbb{R}, \rho)$ totalbeschränkter, semimetrischer Raum ist. Diesen Beweis werden wir wieder genauer betrachten. Der Beweis, dass $(\mathcal{H} \times \mathbb{R}, \rho)$ semimetrischer Raum ist folgt direkt aus den Eigenschaften der Verteilungsfunktion.

Für die Totalbeschränktheit betrachtet man beide Terme separat und zeigt die Totalbeschränktheit wie im Beweis zu (ii), hierbei nutzt man aus, dass $F_{\varepsilon|X}$ monoton wachsend und beschränkt und konstruiert die Bälle durch Zerlegung von $[0, 1]$. Insgesamt ist $(\mathcal{H} \times \mathbb{R}, \rho)$ totalbeschränkter,

semimetrischer Raum.

Es folgt weiter:

$$\begin{aligned}
& \sup_{\rho(s,t) < \delta_n} E[(\varphi_{n,s}(\varepsilon, X) - \varphi_{n,t}(\varepsilon, X))^2] \\
&= \sup_{\rho(s,t) < \delta_n} \frac{1}{h} E[K^2\left(\frac{X-x}{h}\right) \left(I\{\varepsilon \leq s_2 - s_1(X)\} - I\{\varepsilon \leq s_2\} - I\{\varepsilon \leq t_2 - t_1(X)\} + I\{\varepsilon \leq t_2\} \right)^2] \\
&\stackrel{(1)}{\leq} \sup_{\rho(s,t) < \delta_n} \frac{2}{h} E[K^2\left(\frac{X-x}{h}\right) \left((I\{\varepsilon \leq s_2 - s_1(X)\} - I\{\varepsilon \leq t_2 - t_1(X)\})^2 + (I\{\varepsilon \leq s_2\} - I\{\varepsilon \leq t_2\})^2 \right)] \\
&= \sup_{\rho(s,t) < \delta_n} \left(2 E\left[\frac{1}{h} K^2\left(\frac{X-x}{h}\right) (I\{\varepsilon \leq s_2 - s_1(X)\} - I\{\varepsilon \leq t_2 - t_1(X)\})^2\right] \right. \\
&\quad \left. + 2 E\left[\frac{1}{h} K^2\left(\frac{X-x}{h}\right) (I\{\varepsilon \leq s_2\} - I\{\varepsilon \leq t_2\})^2\right] \right) \\
&= \sup_{\rho(s,t) < \delta_n} \left(2 E\left[\frac{1}{h} K^2\left(\frac{X-x}{h}\right) E[(I\{\varepsilon \leq s_2 - s_1(X)\} - I\{\varepsilon \leq t_2 - t_1(X)\})^2 | X]\right] \right. \\
&\quad \left. + 2 E\left[\frac{1}{h} K^2\left(\frac{X-x}{h}\right) (I\{\varepsilon \leq s_2\} - I\{\varepsilon \leq t_2\})^2\right] \right) \\
&= \sup_{\rho(s,t) < \delta_n} \left(2 E\left[\frac{1}{h} K^2\left(\frac{X-x}{h}\right) |F_{\varepsilon|X}(s_2 - s_1(X)|X) - F_{\varepsilon|X}(t_2 - t_1(X)|X)|\right] \right. \\
&\quad \left. + 2 E\left[\frac{1}{h} K^2\left(\frac{X-x}{h}\right) (I\{\varepsilon \leq s_2\} - I\{\varepsilon \leq t_2\})^2\right] \right) \\
&= \sup_{\rho(s,t) < \delta_n} \left(2 E\left[\frac{1}{h} K^2\left(\frac{X-x}{h}\right) |F_{\varepsilon|X}(s_2 - s_1(X)|X) - F_{\varepsilon|X}(t_2 - t_1(X)|X)|\right] \right. \\
&\quad \left. + 2 E\left[\frac{1}{h} K^2\left(\frac{X-x}{h}\right) |F_{\varepsilon|X}(s_2|X) - F_{\varepsilon|X}(t_2|X)|\right] \right) \\
&\leq C_1 \cdot |F_{\varepsilon|X}(s_2 - s_1(x)|x) - F_{\varepsilon|X}(t_2 - t_1(x)|x)| + C_2 \cdot |F_{\varepsilon|X}(s_2|x) - F_{\varepsilon|X}(t_2|x)| + o(1) \\
&\leq C_3 \cdot (|F_{\varepsilon|X}(s_2 - s_1(x)|x) - F_{\varepsilon|X}(t_2 - t_1(x)|x)| + |F_{\varepsilon|X}(s_2|x) - F_{\varepsilon|X}(t_2|x)|) + o(1) \\
&= O(\delta_n) + o(1) \\
&= o(1)
\end{aligned}$$

(1) folgt dabei aus $(a + b)^2 \leq 2a^2 + 2b^2$.

Zuletzt folgt Bedingung (2.9) aus Lemma 6.8 und somit Aussage 6.7 (vi).

Als letztes betrachten wir die in (vii) gegebene Funktionenklasse $\mathcal{F}_{8,n}$, für diese Klasse sind die Rechnungen aus dem Beweis zu (vi) teils übertragbar.

Es gilt für jedes $\tilde{h} \in \mathcal{H}$ nach Voraussetzung, dass $|\tilde{h}| < \delta$ und damit ergibt sich als Einhüllende:

$$\Phi_n(\varepsilon, z) := \delta \cdot |\varepsilon| \cdot \int \frac{1}{\sqrt{h}} K\left(\frac{u-x}{h}\right) f_{\varepsilon|X}(y|u) \frac{1}{b} \tilde{K}\left(\frac{z-u}{b}\right) du.$$

Mit den Rechnungen zu Proposition 6.4 (iii) folgt

$$E[\Phi_n^2(\varepsilon, X)] = \delta^2 \cdot E[\varepsilon^2 \cdot (\int \frac{1}{\sqrt{h}} K\left(\frac{u-x}{h}\right) f_{\varepsilon|X}(y|u) \frac{1}{b} \tilde{K}\left(\frac{X-u}{b}\right) du)^2] = O(1).$$

Sei $r_0 \geq m > 0$. Dann folgt mit den Eigenschaften der Bandbreite h aus (H2), Substitution und den Eigenschaften der Kerne K und \tilde{K} :

$$\begin{aligned} E[\Phi_n^2(\varepsilon, X) I\{\Phi_n(\varepsilon, X) > \beta\sqrt{n}\}] &= E[\Phi_n^2(\varepsilon, X) I\{\frac{\Phi_n(\varepsilon, X)}{\beta\sqrt{n}} > 1\}] \\ &\leq E[\frac{\Phi_n^{m+2}(\varepsilon, X)}{\beta^m (nh)^{m/2}}] \\ &= \frac{1}{\beta^m (nh)^{m/2}} \frac{1}{h} E[(\delta \cdot |\varepsilon| \cdot \int K\left(\frac{u-x}{h}\right) f_{\varepsilon|X}(y|u) \frac{1}{b} \tilde{K}\left(\frac{X-u}{b}\right) du)^{m+2}] \\ &= O\left(\frac{1}{(nh)^{m/2}}\right) = o(1). \end{aligned}$$

Zum Beweis der Gültigkeit der Gleichung (2.5). Es seien $s, t \in \mathcal{H} \times \mathbb{R}$, die nötige Semimetrik sei gegeben durch

$$\rho(s, t) = \rho((s_1, s_2), (t_1, t_2)) := |s_1(x) f_{\varepsilon|X}(s_2|x) - t_1(x) f_{\varepsilon|X}(t_2|x)|$$

Zu zeigen ist, dass $(\mathcal{H} \times \mathbb{R}, \rho)$ totalbeschränkter semimetrischer Raum ist, was mit ähnlichen Argumenten und Rechnungen wie zu (vi) folgt. Es folgt weiter:

$$\begin{aligned} &\sup_{\rho(s,t) < \delta_n} E[(\varphi_{n,s}(\varepsilon, X) - \varphi_{n,t}(\varepsilon, X))^2] \\ &= \sup_{\rho(s,t) < \delta_n} E[\varepsilon^2 \cdot (\int (s_1(u) f_{\varepsilon|X}(s_2|u) - t_1(u) f_{\varepsilon|X}(t_2|u)) \frac{1}{\sqrt{h}} K\left(\frac{u-x}{h}\right) \frac{1}{b} \tilde{K}\left(\frac{X-u}{b}\right) du)^2] \\ &= O(\delta_n) + o(1) \\ &= o(1) \end{aligned}$$

Zuletzt folgt Bedingung (2.9) aus Lemma 6.8 und somit Aussage 6.7 (vii).

Der Beweis von Lemma 6.7 ist damit abgeschlossen. □

Für den Beweis der Donskereigenschaft in dieser Arbeit auftretenden Funktionenklassen sind noch einige technische Voraussetzungen zu prüfen. Der Beweis dieser ausgelagerten Voraussetzungen folgt im Weiteren:

Lemma 6.8 *Es sei $(\varepsilon, X) \sim \mathcal{P}$ und es gelten die Voraussetzungen (M1) - (M3), (A), (K), (H1), (H2) und es sei $\mathcal{H} := C_\delta^{1+\alpha}[0, 1]$ die in 2.34 definierte Funktionenklasse. Es gilt:*

- (i) *Für die von n abhängigen Funktionenklassen $\mathcal{F}_n \in \{\mathcal{F}_{5,n}, \mathcal{F}_{7,n}, \mathcal{F}_{8,n}\}$ aus 6.7 (iii) - (vii) mit der dort im Beweis angegebenen Einhüllenden Φ_n :*

$$(6.14) \quad \int_0^{\delta_n} \sqrt{\log \mathcal{N}_{[]}(\epsilon \|\Phi_n\|_{2,P}, \mathcal{F}_n, L_2(P))} d\epsilon \rightarrow 0 \quad \text{für jedes } \delta_n \rightarrow 0, n \rightarrow \infty$$

- (ii) *Für die Funktionenklassen $\mathcal{F}_n \in \{\mathcal{F}_{3,n}, \mathcal{F}_{6,n}\}$ aus 6.7 (iii) - (vii) mit der dort im Beweis angegebenen Einhüllenden Φ_n :*

$$(6.15) \quad \sup_{Q \in \mathcal{Q}} \int_0^{\delta_n} \sqrt{\log \mathcal{N}_{[]}(\epsilon \|\Phi_n\|_{2,Q}, \mathcal{F}_n, L_2(Q))} d\epsilon \rightarrow 0 \quad \text{für jedes } \delta_n \rightarrow 0.$$

Zum Beweis werden wir die (logarithmierten) Überdeckungszahlen bzw. (logarithmierten) Klammerungszahlen der obigen Funktionenklassen polynomiell abschätzen, meist durch Angabe von Klammern und/oder Zurückführung auf VC-Klassen, für welche Abschätzungen bekannt sind. Wir beginnen mit dem Beweis zur Funktionenklasse $\mathcal{F}_{3,n}$ und schreiben die Funktionenklasse

$$\mathcal{F}_{3,n} = \{ \varphi_{n,y} : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R} \mid \varphi_{n,y}(\tilde{\varepsilon}, z) = \frac{1}{\sqrt{h}} K_h(z - x) g_{j_1}(\tilde{\varepsilon}, x) \cdot E[g_{j_2}(\varepsilon, X) I\{\varepsilon \leq y\} \mid X = x], y \in \mathbb{R} \}$$

als Produkt zweier Funktionenklassen

$$\mathcal{F}_n := \{ f_n : [0, 1] \rightarrow \mathbb{R} \mid f_n(z) := \frac{1}{\sqrt{h}} K_h(z - x) \}$$

(diese enthält nur eine Funktion) und

$$\mathcal{G} := \{ \varphi_y : \mathbb{R} \rightarrow \mathbb{R} \mid \varphi_y(\tilde{\varepsilon}) = g_{j_1}(\tilde{\varepsilon}, x) E[g_{j_2}(\varepsilon, X) I\{\varepsilon \leq y\} \mid X = x], y \in \mathbb{R} \}.$$

Die dazugehörigen Einhüllenden haben die Gestalt:

$$F_n(z) = \frac{1}{\sqrt{h}} K_h(z - x) \text{ bzw. } G(\tilde{\varepsilon}) = |g_{j_1}(\tilde{\varepsilon}, x)| \cdot C,$$

wobei $C \geq E[|g_{j_2}(\varepsilon, X)| \mid X = x]$ vgl. Voraussetzung (A).

Betrachte zuerst die folgende Funktionenklasse in $\tilde{\varepsilon}$ konstanter Funktionen:

$$\mathcal{G}_1 := \{ \varphi_y : \mathbb{R} \rightarrow \mathbb{R} \mid \varphi_y(\tilde{\varepsilon}) = E[g_{j_2}(\varepsilon, X) I\{\varepsilon \leq y\} \mid X = x], y \in \mathbb{R} \}.$$

\mathcal{G}_1 ist VC-Klasse, da die Funktionen in dieser Klasse konstant in $\tilde{\varepsilon}$ sind. Mittels Proposition 2.31 folgt damit, dass \mathcal{G} euklidisch ist.

Jetzt ist noch zu begründen, dass die Überdeckungszahl der Klasse $\mathcal{F}_{3,n} := \{ f_n \cdot g : f_n \in \mathcal{F}_n, g \in \mathcal{G} \}$ mit Einhüllender $\Phi_n := F_n \cdot G$ abschätzbar ist durch die Überdeckungszahl der Klasse \mathcal{G} , also insbesondere unabhängig von n abschätzbar ist. Die Einhüllende ergibt sich direkt, vgl. Proposition 6.7 (ii). Für den weiteren Beweis nutze, dass \mathcal{F}_n nur ein Element enthält und wende Proposition 1.29 aus Nagel (2006) bzw. den Beweis des Theorems 9.15 aus Kosorok (2008) an, welche besagt, dass für jedes feste $\epsilon > 0$ und diskretes Wahrscheinlichkeitsmaß \mathcal{P} mit $0 < \|\Phi_n\|_{2,\mathcal{P}} < \infty$ gilt, dass

$$\begin{aligned} \mathcal{N}(\epsilon \|\Phi_n\|_{2,\mathcal{P}}, \mathcal{F}_{3,n}, L_2(\mathcal{P})) &\leq \mathcal{N}(\epsilon \|G\|_{2,\mathcal{P}^*}, \mathcal{G}, L_2(\mathcal{P}^*)) \\ &\leq \sup_{\tilde{\mathcal{P}}} \mathcal{N}(\epsilon \cdot \|G\|_{2,\tilde{\mathcal{P}}}, \mathcal{G}, L_2(\tilde{\mathcal{P}})) \end{aligned}$$

dabei ist $d\mathcal{P}^* := F_n^2 d\mathcal{P} / \|F_n\|_{2,\mathcal{P}}$ und das Supremum gebildet über alle diskreten Wahrscheinlichkeitsmaße $\tilde{\mathcal{P}}$ mit $0 < \|G\|_{\tilde{\mathcal{P}}} < \infty$. Siehe dazu auch Proposition 11.22, Kosorok (2008), S. 226.

Die Überdeckungsanzahl der Funktionenklasse $\mathcal{F}_{3,n}$ ist somit polynomiell abschätzbar und es gilt nach Proposition 2.43 die Gleichung (6.15).

Damit ist die Aussage des Lemmas für die Funktionenklasse $\mathcal{F}_{3,n}$ bewiesen.

Als nächsten wollen wir uns dem Beweis zur Funktionenklasse $\mathcal{F}_{5,n}$ zuwenden, dazu direkt Klammern angeben und die Klammeranzahl abschätzen:

Die Funktionenklasse hat die Form

$$\mathcal{F}_{5,n} := \left\{ \varphi_{n,y} : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R} \mid \varphi_{n,y}(\varepsilon, z) = \varepsilon \int \frac{1}{\sqrt{h}} K_h(u-x) f_{\varepsilon|X}(y|u) \frac{1}{b} \tilde{K}_b(z-u) du, y \in \mathbb{R} \right\}$$

mit Einhüllender Φ_n , vgl. Proposition 6.7. Im weiteren geben wir zuerst eine kurze Skizze des Beweises und führen ihn dann genauer aus:

Wir konstruieren zu Beginn Klammern für die Klasse:

$$\tilde{\mathcal{F}}_{5,n} := \left\{ \varphi_{n,y} : \mathbb{R} \times [0, 1] \times [0, 1] \rightarrow \mathbb{R} \mid \varphi_{n,y}(\varepsilon, u, z) = \varepsilon \frac{1}{\sqrt{h}} K_h(u-x) f_{\varepsilon|X}(y|u) \tilde{K}_b(z-u), y \in \mathbb{R} \right\}.$$

Wegen der Monotonie des Integrales übertragen sich die Klammern für diese Klasse dann auf die Klasse $\mathcal{F}_{5,n}$, dabei fügen wir an dieser Stelle auch das noch fehlende $\frac{1}{b}$ hinzu. Die Klasse $\tilde{\mathcal{F}}_{5,n}$ lässt sich noch weiter zerlegen in einen Anteil, welcher identisch ist für jede Funktion aus $\tilde{\mathcal{F}}_{5,n}$, dies ist $\psi(\varepsilon, z, u) := \varepsilon \frac{1}{\sqrt{h}} K_h(u-x) \tilde{K}_b(z-u)$ und den Anteil $f_{\varepsilon|X}(y|u)$.

Klammern für die Klasse

$$\mathcal{G}_1 := \left\{ \varphi : [0, 1] \rightarrow \mathbb{R} \mid \varphi(u) = f_{\varepsilon|X}(y|u), y \in \mathbb{R} \right\}$$

können durch multiplizieren mit $\psi(\varepsilon, z, u)$ als Grundlage für Klammern der Klasse $\tilde{\mathcal{F}}_{5,n}$ genutzt werden. Aus (M2) ergibt sich c als Einhüllende für \mathcal{G}_1 , zusätzlich gilt nach Voraussetzung, dass $\sup_{x \in [0,1]} \left| \frac{\partial^2 f_{\varepsilon|X}(y|x)}{\partial x^2} \right| < c < \infty$ (ebenso für die erste Ableitung) und damit $\mathcal{G}_1 \subset C_c^2[0, 1]$. Im Weiteren

werden wir die Anzahl der ϵ -Klammern für \mathcal{G}_1 abschätzen und damit die Anzahl der ϵ -Klammern für $\mathcal{F}_{5,n}$ abschätzen. Das Vorgehen und die Klammern betrachten wir jetzt genauer:

Es gilt $\mathcal{G}_1 \subset C_c^2[0, 1]$ und damit liefert Theorem 2.7.1 aus van der Vaart und Wellner (2000), Seite 155 eine Abschätzung für die Überdeckungszahl von \mathcal{G}_1 :

$$\log(\mathcal{N}(\epsilon/2, \mathcal{G}_1, \|\cdot\|_\infty)) \leq K\epsilon^{-1/2}.$$

Die Zentren f_i der Bälle liefern die Grundlage für die ϵ -Klammern $[f_i(u) - \epsilon/2, f_i(u) + \epsilon/2]$, so dass ebenfalls

$$\log(\mathcal{N}_{[]}(\epsilon, \mathcal{G}_1, \|\cdot\|_\infty)) \leq \tilde{K}\epsilon^{-1/2}.$$

Aus diesen Klammern lassen sich Klammern für die Klasse $\tilde{\mathcal{F}}_{5,n}$ konstruieren. Dazu betrachte $\psi(\epsilon, z, u)$ genauer. $\psi(\epsilon, z, u)$ ist negativ für $\epsilon < 0$, sonst positiv und es gilt $\|\psi\|_{2,Q} < \infty$, dabei ist $Q = P_1 \times P$ mit P_1 die Verteilung von (ϵ, X) und P die Verteilung von U , gewählt als die Gleichverteilung auf $[0, 1]$, bezeichnet mit f_U . Es bezeichne $I_A(\epsilon)$ im weiteren die Indikatorfunktion. Durch multiplizieren der ϵ -Klammern von \mathcal{G}_1 mit ψ konstruiere die ϵ -Klammern für $\tilde{\mathcal{F}}_{5,n}$ wie folgt:

Betrachtet man eine ϵ -Klammer $[l_1, l_2]$ der Klasse \mathcal{G}_1 bzgl. $\|\cdot\|_\infty$, so ist

$$[\psi(\epsilon, z, u) \cdot (l_1(u)I_{\mathbb{R}^+}(\epsilon) + l_2(u)(1 - I_{\mathbb{R}^+}(\epsilon))), \psi(\epsilon, z, u) \cdot (l_2(u)I_{\mathbb{R}^+}(\epsilon) + l_1(u)(1 - I_{\mathbb{R}^+}(\epsilon)))]$$

eine Klammer für die Klasse $\tilde{\mathcal{F}}_{5,n}$. Die Länge der Klammer ist hier noch nicht von Interesse. Von Interesse ist aber die Länge der daraus für $\mathcal{F}_{5,n}$ resultierenden Klammer. Aufgrund der Monotonie des Integrales überträgt sich die Anzahl der Klammern ebenfalls auf $\mathcal{F}_{5,n}$. Diese sind ϵ -Klammern, denn:

$$\begin{aligned} & \left\| \int \frac{1}{b} ((\psi(\epsilon, z, u) \cdot l_2(u)) - (\psi(\epsilon, z, u) \cdot l_1(u))) f_U(u) du \right\|_{2, P_1} \\ &= \left\| \int \frac{1}{b} \psi(\epsilon, z, u) (l_2 - l_1)(u) f_U(u) du \right\|_{2, P_1} \\ &\leq \sqrt{\int \left(\int \frac{1}{b} |\psi(\epsilon, z, u)| f_U(u) du \right)^2 f_{\epsilon|X}(\epsilon|z) f_X(z) d(\epsilon, z)} \|(l_2 - l_1)\|_\infty \\ &\leq \sqrt{2 \|\sigma^2\|_\infty \|f_X\|_\infty \|K\|_\infty^2 \|\tilde{K}\|_\infty^2} \cdot \|(l_2 - l_1)\|_\infty \\ &< \sqrt{2 \|\sigma^2\|_\infty \|f_X\|_\infty \|K\|_\infty^2 \|\tilde{K}\|_\infty^2} \cdot \epsilon \end{aligned}$$

Es kann also eine Konstante $C := \frac{1}{\sqrt{\|\sigma^2\|_\infty \|f_X\|_\infty \|K\|_\infty^2 \|\tilde{K}\|_\infty^2}}$ festgelegt werden, dass aus $C \cdot \epsilon$ -Klammern für \mathcal{G}_1 ϵ -Klammern für $\mathcal{F}_{5,n}$ resultieren. Mit obiger Abschätzung der logarithmierten Klammerungszahl für die Klasse \mathcal{G}_1 folgt dann übertragen auf die Klasse $\mathcal{F}_{5,n}$ Gleichung (6.14).

Beweis zur Funktionenklasse $\mathcal{F}_{6,n}$:

Die Funktionenklasse hat die Form

$$\mathcal{F}_{6,n} := \{ \varphi_{n,y} : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R} \mid \varphi_{n,y}(\varepsilon, z) = \frac{1}{\sqrt{h}} K_h(z - x) (I\{\varepsilon \leq y\} - F_{\varepsilon|X}(y|x)), y \in \mathbb{R} \}.$$

Wie schon im Beweis der Gültigkeit von Gleichung (6.15) für die Funktionenklasse $\mathcal{F}_{3,n}$ reicht es hier aus zu zeigen, dass die Funktionenklasse

$$\mathcal{G} := \{ g_y : \mathbb{R} \rightarrow \mathbb{R} \mid g_y(\varepsilon) = I\{\varepsilon \leq y\} - F_{\varepsilon|X}(y|x), y \in \mathbb{R} \}$$

euklidisch ist. Dies gilt, da $\tilde{g}_y(\varepsilon) = F_{\varepsilon|X}(y|x)$ konstant ist, also die Klasse $\{ F_{\varepsilon|X}(y|x), y \in \mathbb{R} \}$ eine VC -Klasse bildet und die Klasse der Indikatorfunktionen eine VC -Klasse bildet. Beide Klassen sind damit euklidisch. Nach Neumeyer (2006), S. 150 folgt, dass damit die Differenz der Klassen, also \mathcal{G} euklidisch ist. Mit Proposition 2.43 folgt die Gleichung (6.15).

Zur Funktionenklasse $\mathcal{F}_{7,n}$ mit Einhüllender Φ_n gehen wir wie folgt vor. Es gilt $\mathcal{F}_{7,n} = \mathcal{G}_{7,n} - \mathcal{H}_{7,n}$ mit

$$\begin{aligned} \mathcal{G}_{7,n} &:= \{ \varphi_{n,t} : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R}, \mid \varphi_{n,t}(\varepsilon, z) = \frac{1}{\sqrt{h}} K\left(\frac{z-x}{h}\right) I\{\varepsilon \leq y - \tilde{h}(z)\}, t = (y, \tilde{h}), \tilde{h} \in \mathcal{H}, y \in \mathbb{R} \} \\ \mathcal{H}_{7,n} &:= \{ \varphi_{n,t} : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R}, \mid \varphi_{n,y}(\varepsilon, z) = \frac{1}{\sqrt{h}} K\left(\frac{z-x}{h}\right) I\{\varepsilon \leq y\}, y \in \mathbb{R} \} \end{aligned}$$

Für beide Klassen schätzen wir die Klammerungszahl ab. Mit Proposition 2.20 lässt sich dann die Klammerungszahl von $\mathcal{F}_{7,n}$ durch das Produkt der Klammerungszahlen von $\mathcal{G}_{7,n}$ und $\mathcal{H}_{7,n}$ abschätzen. Wir beginnen mit der Abschätzung der Klammerungszahl der Klasse $\mathcal{H}_{7,n}$. Da diese

in $\mathcal{G}_{7,n}$ enthalten ist können wir die Klammerungszahl direkt durch diejenige von $\mathcal{G}_{7,n}$ abschätzen. Betrachte nun die Funktionenklasse $\mathcal{G}_{7,n}$.

Akritas und van Keilegom (2001) betrachten in Beweis Lemma 1, Gleichung (20) ff., die Funktionenklasse

$$\mathcal{G}_1 := \{ \varphi_t : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R}, | \varphi_t(\varepsilon, z) = I\{\varepsilon \leq y - \tilde{h}(z)\}, t = (y, \tilde{h}), \tilde{h} \in \mathcal{H}, y \in \mathbb{R} \}$$

und konstruieren $L_2(P)$ ϵ -Klammern. Für die Funktionenklasse \mathcal{H} nutzen Akritas und van Keilegom (2001) dabei ϵ^2 -Klammern bzgl. $L_2(P)$. Mit van der Vaart und Wellner (2000), Theorem 2.7.1, S. 155 lassen sich wie im Beweis zur Klasse $\mathcal{F}_{5,n}$ auch ϵ^2 -Klammern für \mathcal{H} bzgl. $\| \cdot \|_\infty$ konstruieren, mit einer bis auf Konstante gleichen Abschätzung der logarithmierten Klammerungszahl. Mit dem Beweis von Akritas und van Keilegom (2001), Lemma 1, Gleichung (20) ff., ergeben sich Klammern $[l, u]$ für \mathcal{G}_1 , so dass:

$$\| u(\varepsilon, z) - l(\varepsilon, z) \|_{L_2(P), \infty} := \sup_{z \in [0, 1]} \sqrt{\int (u(\varepsilon, z) - l(\varepsilon, z))^2 f_{\varepsilon|X}(\varepsilon|z) d\varepsilon} < \epsilon,$$

die Konstruktion der Klammern wird im weiteren noch genau ausgeführt. Damit erhält man

$$\begin{aligned} \left\| \frac{1}{\sqrt{h}} K\left(\frac{z-x}{h}\right) (u(\varepsilon, z) - l(\varepsilon, z)) \right\|_{2,P}^2 &= \int \frac{1}{h} K^2\left(\frac{z-x}{h}\right) (u(\varepsilon, z) - l(\varepsilon, z))^2 f_{\varepsilon, X}(\varepsilon, z) d(\varepsilon, z) \\ (6.16) \qquad \qquad \qquad &\leq \int \frac{1}{h} K^2\left(\frac{z-x}{h}\right) f_X(z) dz \sup_{\tilde{z} \in [0, 1]} \int (u(\varepsilon, \tilde{z}) - l(\varepsilon, \tilde{z}))^2 f_{\varepsilon|X}(\varepsilon|\tilde{z}) d\varepsilon \\ &= \|K^2\|_\infty \|f_X\|_\infty \|u(\varepsilon, z) - l(\varepsilon, z)\|_{L_2(P), \infty}^2 \\ &< \|K^2\|_\infty \|f_X\|_\infty \cdot \epsilon^2 \end{aligned}$$

Die Konstruktion der Klammern für \mathcal{G}_1 werden wir jetzt genauer ausführen:

Nach oben genanntem Theorem von van der Vaart und Wellner (2000) gilt $m = \mathcal{N}_{[]}(\tilde{\epsilon}^2, \mathcal{H}, \| \cdot \|_\infty) \leq \exp(K\tilde{\epsilon}^{-2/(1+\alpha)})$. Es seien $d_1^L \leq d_1^U, \dots, d_m^L \leq d_m^U$ die dazugehörigen $\tilde{\epsilon}^2$ -Klammern. Damit existiert für jedes y fest und $\tilde{h} \in \mathcal{H}$ ein $i \in \{1, \dots, m\}$, so dass

$$I\{\varepsilon \leq y - d_i^U(X)\} \leq I\{\varepsilon \leq y - \tilde{h}(X)\} \leq I\{\varepsilon \leq y - d_i^L(X)\}$$

Setze $F_i^L(y|X) := P(\varepsilon \leq y + d_i^L(X)|X)$ und zerlege \mathbb{R} so in Teile $y_{i,k}^L(X)$, $k = 1, \dots, O(\tilde{\epsilon}^{-2})$,

so dass die einzelnen Teilstücke unter F_i^L eine Wahrscheinlichkeit von weniger oder gleich $\tilde{\varepsilon}^2$ besitzen. Analog definiere $F_i^U(y|X)$ und eine Zerlegung $y_{i,k}^U(X)$, $k = 1, \dots, O(\tilde{\varepsilon}^{-2})$. Für y und gegebenes X definiere daraus die folgende Klammer:

$$y_{i,k_1}^L(X) \leq y \leq y_{i,k_2}^U(X),$$

wobei jeweils k_1 größtmöglich, so dass $y_{i,k_1}^L(X) \leq y$ und k_2 kleinstmöglich, so dass $y \leq y_{i,k_2}^U(X)$. Nun ist zu zeigen, dass dies zu den folgenden Klammern für \mathcal{G}_1 führt:

$$I\{\varepsilon \leq y_{i,k_1}^L(X) - d_i^U(X)\} \leq I\{\varepsilon \leq y - \tilde{h}(X)\} \leq I\{\varepsilon \leq y_{i,k_2}^U(X) - d_i^L(X)\}$$

Nun wollen wir die Größe der Klammern bestimmen:

$$\begin{aligned} & \| I\{\varepsilon \leq y_{i,k_2}^U(X) - d_i^U(X)\} - I\{\varepsilon \leq y_{i,k_1}^L(X) - d_i^L(X)\} \|_{L_2(P),\infty}^2 \\ &= \sup_{z \in [0,1]} (F_i^U(y_{i,k_2}^U(z)|z) - F_i^L(y_{i,k_1}^L(z)|z)) \\ &\leq \sup_{z \in [0,1]} (F_i^U(y(z)|z) - F_i^L(y(z)|z)) + C_1 \tilde{\varepsilon}^2, \end{aligned}$$

wobei $y(z) \in [y_{i,k_1}^L(z), y_{i,k_2}^U(z)]$. Es folgt die letzte Gleichheit aus der Konstruktion, denn durch den Übergang von $y_{i,k_2}^U(z)$ zu $y(z)$ verringert sich $F_i^U(\cdot|X=z)$ um höchstens $\tilde{\varepsilon}^2$, ebenso wächst $F_i^L(\cdot|X=z)$ um höchstens $\tilde{\varepsilon}^2$ an und es gibt eine Konstante $C_1 \leq 2$, so dass die letzte Gleichheit gilt. Die Differenz $F_i^U(y(z)|z) - F_i^L(y(z)|z)$ kann mittels Taylorentwicklung abgeschätzt werden, so dass gilt:

$$\begin{aligned} \sup_{z \in [0,1]} (F_i^U(y(z)|z) - F_i^L(y(z)|z)) &= \sup_{z \in [0,1]} (P(\varepsilon \leq y(z) + d_i^U(z)|z) - P(\varepsilon \leq y(z) + d_i^L(z)|z)) \\ &= \sup_{z \in [0,1]} (F_{\varepsilon|X}(y(z) + d_i^U(z)|z) - F_{\varepsilon|X}(y(z) + d_i^L(z)|z)) \\ &= \sup_{z \in [0,1]} (f_{\varepsilon|X}(\xi_1|z)(d_i^U(z) - d_i^L(z))) \\ &\leq \sup_{z \in [0,1], y \in \mathbb{R}} f_{\varepsilon|X}(y|z) \sup_{z \in [0,1]} ((d_i^U(z) - d_i^L(z))) \\ &< C_2 \cdot \tilde{\varepsilon}^2, \end{aligned}$$

also ergibt sich insgesamt:

$$\| I\{\varepsilon \leq y_{i,k_2}^U(X) - d_i^L(X)\} - I\{\varepsilon \leq y_{i,k_1}^L(X) - d_i^U(X)\} \|_{L_2(P),\infty}^2 < (C_1 + C_2) \tilde{\varepsilon}^2,$$

damit gilt also $\mathcal{N}_{[\cdot]}(\tilde{\varepsilon}, \mathcal{G}_1, \|\cdot\|_{L_2(P), \infty}) \leq \exp(K\tilde{\varepsilon}^{-2/(1+\delta)})$. Mit der Rechnung (6.16) ergibt sich nun die $L_2(P)$ -Klammerungszahl von $\mathcal{G}_{7,n}$ und zusammengesetzt mit $\mathcal{H}_{7,n}$ erhält man insgesamt die Gleichung (6.15).

Zur Funktionenklasse $\mathcal{F}_{8,n}$:

Die Funktionenklasse hat die Form

$$\mathcal{F}_{8,n} := \left\{ \varphi_{n,t} : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R} \mid \varphi_{n,t}(\varepsilon, z) = \varepsilon \cdot \int \tilde{h}(u) \frac{1}{\sqrt{h}} K\left(\frac{u-x}{h}\right) f_{\varepsilon|X}(y|u) \frac{1}{b} \tilde{K}\left(\frac{z-u}{b}\right) du, \right. \\ \left. t = (y, \tilde{h}), \tilde{h} \in \mathcal{H}, y \in \mathbb{R} \right\}.$$

Das Vorgehen ist identisch zu dem Vorgehen zu Klasse $\mathcal{F}_{5,n}$. Nutze hier aus, dass für die Klasse $\mathcal{H} \subset C_{\delta}^{1+\alpha}[0, 1]$, damit also Abschätzungen der (logarithmierten) Klammerungszahl, versehen mit der Supremumsnorm gegeben sind (vgl. Theorem 2.7.1 aus van der Vaart und Wellner (2000), Seite 155) und, dass $f_{\varepsilon|X}(y|u) \cdot \tilde{h}(u) \in C_{\delta \cdot c}^{1+\alpha}[0, 1]$.

Damit ist Proposition 6.8 und somit auch die noch ausstehenden Teile von Proposition 6.7 bewiesen. □

Jetzt kann mithilfe der Propositionen zum Erwartungswert und zur Kovarianz des Prozesses (6.8) eine Aussage zur schwachen Konvergenz aufgestellt werden:

Proposition 6.9 *Es gilt unter den Voraussetzungen (M1) – (M3), (H1), (H2), (K) und g erfüllend (G1) und (G2), dass*

$$\hat{F}_n^*(\cdot | x) - F_{\varepsilon|X}(\cdot | x) = H_n(\cdot, x) + b_n^*(\cdot, x) + o_p\left(\frac{1}{\sqrt{nh}}\right)$$

wobei

$$b_n^*(\cdot, x) := -\frac{h^2}{2} \mu_2^K \frac{\partial^2 E[g^t(\varepsilon, x) | X = t] \cdot f_X(t)}{\partial t^2} \Big|_{t=x} \cdot \Sigma^{-1}(x)^t \cdot \mu_0^{K^2} \cdot E[g(\varepsilon, X) I\{\varepsilon \leq \cdot\} | X = x]$$

$$\begin{aligned}
& + (\Sigma^{-1}(x) \frac{b^2}{2} \mu_2^{\tilde{K}} E[g'(\varepsilon, X) | X = x] B(x))^t \cdot \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq \cdot\} | X = x] \\
& + \frac{b^2}{2} \mu_2^{\tilde{K}} f_X^{-1}(x) f_{\varepsilon|X}(\cdot | x) B(x) \\
& + \frac{h^2}{2} \mu_2^K \cdot (F_{\varepsilon|X}^{(2)}(\cdot | x) + \frac{2 f'_X(x)}{f_X(x)} F_{\varepsilon|X}^{(1)}(\cdot | x))
\end{aligned}$$

und

$$\begin{aligned}
H_n(\cdot, x) := & - \left((\Sigma^{-1}(x) \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) g(\varepsilon_i, x))^t \cdot \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq \cdot\} | X = x] \right. \\
& \left. - E[(\Sigma^{-1}(x) \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) g(\varepsilon_i, x))^t \cdot \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq \cdot\} | X = x]] \right) \\
& + \frac{1}{n} \sum_{i=1}^n \varepsilon_i \int \frac{1}{h} K_h(z - x) (g'(y, x))^t \frac{1}{b} \tilde{K}_b(X_i - z) f_{\varepsilon|X}(y|z) d(y, z) \\
& \quad \cdot \Sigma^{-1}(x)^t \cdot \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq \cdot\} | X = x] \\
& + f_X^{-1}(x) \frac{1}{n} \sum_{i=1}^n \varepsilon_i \int \frac{1}{h} K_h(z - x) f_{\varepsilon|X}(\cdot | z) \frac{1}{b} \tilde{K}_b(X_i - z) dz \\
& + \left(f_X^{-1}(x) \cdot \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) (I\{\varepsilon_i \leq \cdot\} - F_{\varepsilon|X}(\cdot | x)) \right. \\
& \quad \left. - E[f_X^{-1}(x) \cdot \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) (I\{\varepsilon_i \leq \cdot\} - F_{\varepsilon|X}(\cdot | x))] \right)
\end{aligned}$$

Dann konvergiert $\sqrt{nh} (\hat{F}_n^*(\cdot | x) - F_{\varepsilon|X}(\cdot | x) - b_n^*(\cdot, x))$ schwach gegen einen zentrierten Gaußprozess $G(\cdot, x)$ mit Kovarianzfunktion

$$\begin{aligned}
& Cov(G(y_1, x), G(y_2, x)) \\
& = (\mu_0^{K^2})^3 f_X(x) E \left[g^t(\varepsilon, U) \cdot (\Sigma^{-1}(U))^t \cdot E[g(\varepsilon, x) I\{\varepsilon \leq y_1\} | X = U] \right. \\
& \quad \left. \cdot g^t(\varepsilon, U) \cdot (\Sigma^{-1}(U))^t \cdot E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = x] \Big| U = x \right] \\
& + \sigma^2(x) (\mu_0^{K^2})^2 f_X(x) E[(g'(\varepsilon, X))^t | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x] \\
& \quad \cdot E[(g'(\varepsilon, X))^t | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = x] \\
& \quad \cdot \int \int \int K(v + \lambda(u - w)) K(v) \tilde{K}(u) \tilde{K}(w) du dv dw \\
& + f_X^{-1}(x) E[\varepsilon^2 | X = x] f_{\varepsilon|X}(y_1 | x) f_{\varepsilon|X}(y_2 | x) \int K(v + \lambda(u - w)) \tilde{K}(w) K(v) \tilde{K}(u) d(u, w, v) \\
& - \mu_0^{K^2} \cdot f_{\varepsilon|X}(y_2 | x) E[(g(\varepsilon, X))^t \cdot \varepsilon | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x]
\end{aligned}$$

$$\begin{aligned}
& \cdot \int \int K(u + \lambda v) K(u) \tilde{K}(v) \, dudv \\
- \mu_0^{K^2} \cdot f_{\varepsilon|X}(y_1|x) E[(g(\varepsilon, X))^t \cdot \varepsilon | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = x] \\
& \cdot \int \int K(u + \lambda v) K(u) \tilde{K}(v) \, dudv \\
- f_X(x) (\mu_0^{K^2})^2 E[(g(\varepsilon, X))^t \varepsilon | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x] \\
& \cdot E[(g'(\varepsilon, X))^t | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = x] \\
& \cdot \int \int K(u + \lambda v) K(u) \tilde{K}(v) \, dudv \\
- f_X(x) (\mu_0^{K^2})^2 E[(g(\varepsilon, X))^t \varepsilon | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = x] \\
& \cdot E[(g'(\varepsilon, X))^t | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x] \\
& \cdot \int \int K(u + \lambda v) K(u) \tilde{K}(v) \, dudv \\
+ E[\varepsilon^2 | X = x] E[(g'(\varepsilon, X))^t | X = x] (\Sigma^{-1}(x))^t \cdot \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = x] \\
& \cdot f_{\varepsilon|X}(y_1|x) \int \int \int K(v + \lambda(u - w)) K(v) \tilde{K}(u) \tilde{K}(w) \, du \, dv \, dw \\
+ E[\varepsilon^2 | X = x] E[(g'(\varepsilon, X))^t | X = x] (\Sigma^{-1}(x))^t \cdot \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x] \\
& \cdot f_{\varepsilon|X}(y_2|x) \int \int \int K(v + \lambda(u - w)) K(v) \tilde{K}(u) \tilde{K}(w) \, du \, dv \, dw \\
+ \mu_0^{K^2} f_X^{-1}(x) F_{\varepsilon|X}(\min(y_1, y_2)|x) - \mu_0^{K^2} f_X^{-1}(x) F_{\varepsilon|X}(y_2|x) \cdot F_{\varepsilon|X}(y_1|x) \\
- (\mu_0^{K^2})^2 E[(g(\varepsilon, X))^t I\{\varepsilon \leq y_2\} | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x] \\
- (\mu_0^{K^2})^2 E[(g(\varepsilon, X))^t I\{\varepsilon \leq y_1\} | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = x] \\
+ E[\varepsilon I\{\varepsilon \leq y_2\} | X = x] E[(g'(\varepsilon, x))^t | X = x] (\Sigma^{-1}(x))^t \\
& \cdot E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x] \cdot \mu_0^{K^2} \int \int K(u - \lambda v) \tilde{K}(u) K(v) \, du \, dv \\
+ E[\varepsilon I\{\varepsilon \leq y_1\} | X = x] E[(g'(\varepsilon, x))^t | X = x] (\Sigma^{-1}(x))^t \\
& \cdot E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = x] \cdot \mu_0^{K^2} \int \int K(u - \lambda v) \tilde{K}(u) K(v) \, du \, dv \\
+ f_X^{-1}(x) f_{\varepsilon|X}(y_1|x) E[\varepsilon I\{\varepsilon \leq y_2\} | X = x] \cdot \int \int K(u - \lambda v) \tilde{K}(u) K(v) \, du \, dv \\
+ f_X^{-1}(x) f_{\varepsilon|X}(y_2|x) E[\varepsilon I\{\varepsilon \leq y_1\} | X = x] \cdot \int \int K(u - \lambda v) \tilde{K}(u) K(v) \, du \, dv,
\end{aligned}$$

dabei beschreibt λ das asymptotische Verhältnis von b und h , d.h. $\lambda = \lim_{n \rightarrow \infty} \frac{b}{h}$.

Beweis:

Die Entwicklung wurde sukzessive im vorherigen Kapitel hergeleitet. Mit den Rechnungen zu Proposition 6.2 und 6.4 folgt der Erwartungswert und obige Kovarianz. Das Vorgehen für den weiteren Beweis ist das Folgende: Wir nutzen die in Beispiel 6.7 bewiesene Donskereigenschaft

der vier Summanden des Prozesses. Zu klären ist dann noch, warum daraus die Donskereigenschaft des Gesamtprozesses folgt. Dazu betrachte Theorem 2.1 aus Kosorok (2008), S. 15. Dieses Theorem liefert notwendige und hinreichende Bedingungen für die schwache Konvergenz von Prozessen. Vorausgesetzt sei die schwache Konvergenz zweier Prozesse $G_n(t)$ und $F_n(t)$. Die Aussagen des Theorems 2.1 sind also für beide Prozesse gültig. Mit einfachen Rechnungen lässt sich Bedingung (ii) (asympt. stoch. equicontinuity) auf die Summe übertragen. Bedingung (i) gilt ebenfalls, denn sie folgt mit dem ZGWS aus der Konvergenz der endlich-dimensionalen Randverteilungen beider Prozesse, der Anwendung von Cramér-Wold und den vorangegangenen Rechnungen zur Kovarianz.

Nun wollen wir die Donskereigenschaft für die einzelnen Summanden nachweisen. Wir beginnen mit der Definition der zugrundeliegenden Funktionenklassen:

Beginne mit der Umformung des ersten Summanden, vgl. zur Notation die Zerlegung (6.10):

$$\begin{aligned}
H_{n,1}(\cdot, x) &= - \left((\Sigma^{-1}(x) \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) g(\varepsilon_i, x))^t \cdot \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq \cdot\} | X = x] \right. \\
&\quad \left. - E[(\Sigma^{-1}(x) \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) g(\varepsilon_i, x))^t \cdot \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq \cdot\} | X = x]] \right) \\
&= -\mu_0^{K^2} \frac{1}{n} \sum_{i=1}^n \frac{1}{h} (K_h(X_i - x) g^t(\varepsilon_i, x) - E[K_h(X_i - x) g^t(\varepsilon_i, x)]) \\
&\quad \cdot ((\Sigma^{-1}(x))^t \cdot E[g(\varepsilon, X) I\{\varepsilon \leq \cdot\} | X = x]) \\
&= -\mu_0^{K^2} \sum_{j_1=1}^k \sum_{j_2=1}^k ((\Sigma^{-1}(x))_{j_1, j_2}^t \frac{1}{n} \sum_{i=1}^n \frac{1}{h} (K_h(X_i - x) g_{j_1}(\varepsilon_i, x) - E[K_h(X_i - x) g_{j_1}(\varepsilon_i, x)])) \\
&\quad \cdot E[g_{j_2}(\varepsilon, X) I\{\varepsilon \leq \cdot\} | X = x],
\end{aligned}$$

dabei meint $((\Sigma^{-1}(x))_{s,t}^t)$ den (s, t) -ten Eintrag der Matrix $((\Sigma^{-1}(x))^t)$. Betrachte hierzu nun die folgende Funktionenklasse, dabei ist $x \in [0, 1]$ fest:

$$\mathcal{F}_{1,n} := \{ \varphi_y : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R} \mid \varphi_y(\varepsilon, z) = \frac{1}{\sqrt{h}} K_h(z - x) g_{j_1}(\varepsilon, x) \cdot E[g_{j_2}(\varepsilon, z) I\{\varepsilon \leq y\} | z = x], y \in \mathbb{R} \}$$

Diese Klasse ist nach Beispiel 6.7 (iii) Donskerklasse.

Zum zweiten Summanden

$$\begin{aligned}
H_{n,2}(\cdot, x) &= \frac{1}{n} \sum_{i=1}^n \varepsilon_i \int \frac{1}{h} K_h(z-x) (g'(y, x))^t \frac{1}{b} \tilde{K}_b(X_i - z) f_{\varepsilon|X}(y|z) d(y, z) \\
&\quad \cdot \Sigma^{-1}(x)^t \cdot \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq \cdot\} | X = x] \\
&= \mu_0^{K^2} \sum_{j_1=1}^k \sum_{j_2=1}^k ((\Sigma^{-1}(x))_{j_1, j_2}^t) \frac{1}{n} \sum_{i=1}^n \varepsilon_i \int \frac{1}{h} K_h(z-x) g'_{j_1}(y, x) \frac{1}{b} \tilde{K}_b(X_i - z) f_{\varepsilon|X}(y|z) d(y, z) \\
&\quad \cdot E[g_{j_2}(\varepsilon, X) I\{\varepsilon \leq \cdot\} | X = x].
\end{aligned}$$

Betrachte für diesen Prozess die Darstellung $H_{n,2}(\cdot, x) = \tilde{H}_{n,2}(x) \cdot \varphi(\cdot)$. Dann gilt nach dem zentralen Grenzwertsatz und den in den vorherigen Kapiteln ausgeführten Rechnungen zum Erwartungswert und zur Varianz, dass $\sqrt{nh} \tilde{H}_{n,2}(x)$ asymptotisch gegen eine normalverteilte Zufallsvariable Z mit Erwartungswert Null und hier nicht näher angegebener Varianz τ^2 konvergiert, also damit $H_{n,2}(\cdot, x)$ asymptotisch normalverteilt mit der gleichen Verteilung wie $Z \cdot \varphi(\cdot)$ ist.

Zum dritten Summanden

$$H_{n,3}(\cdot, x) = f_X^{-1}(x) \frac{1}{n} \sum_{i=1}^n \varepsilon_i \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(\cdot|z) \frac{1}{b} \tilde{K}_b(X_i - z) dz$$

Auch dieser Prozess ist wieder als Prozess indiziert mit Funktionen darzustellen, betrachte dazu die Funktionenklasse

$$\mathcal{F}_{3,n} := \{ \varphi_y : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R} \mid \varphi_y(\varepsilon, z) = \varepsilon \int \frac{1}{\sqrt{h}} K_h(u-x) f_{\varepsilon|X}(y|u) \frac{1}{b} \tilde{K}_b(z-u) du, y \in \mathbb{R} \}$$

Diese Klasse ist nach Beispiel 6.7 (iv) Donskerklasse.

Zum vierten und letzten Summanden

$$H_{n,4}(\cdot, x) = f_X^{-1}(x) \cdot \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) (I\{\varepsilon_i \leq \cdot\} - F_{\varepsilon|X}(\cdot|x))$$

$$\begin{aligned}
& -E[f_X^{-1}(x) \cdot \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) (I\{\varepsilon_i \leq \cdot\} - F_{\varepsilon|X}(\cdot|x))] \\
= & f_X^{-1}(x) \cdot \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K_h(X_i - x) (I\{\varepsilon_i \leq \cdot\} - F_{\varepsilon|X}(\cdot|x)) - E[\frac{1}{h} K_h(X_i - x) (I\{\varepsilon_i \leq \cdot\} - F_{\varepsilon|X}(\cdot|x))]
\end{aligned}$$

Auch dieser Prozess ist wieder als Prozess indiziert mit Funktionen darzustellen, betrachte dazu die Funktionenklasse

$$\mathcal{F}_{4,n} := \{ \varphi_y : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R} \mid \varphi_y(\varepsilon, z) = \frac{1}{\sqrt{h}} K_h(z - x) (I\{\varepsilon \leq y\} - F_{\varepsilon|X}(y|x)), y \in \mathbb{R} \}$$

Diese Klasse ist nach Beispiel 6.7 (v) ebenfalls Donskerklasse und somit folgt die Aussage des Satzes 6.9.

□

Eine ganz ähnliche Aussage lässt sich für den Schätzer $\hat{F}_n(\cdot|x)$, definiert in 4.1, zeigen, dabei kann die folgende Proposition direkt aus den Rechnungen die zu Proposition 6.9 nötig waren gefolgert werden, denn dort haben wir zur Herleitung die folgende Darstellung genutzt:

$$\hat{F}_n^*(y|x) - F_{\varepsilon|X}(y|x) = (\hat{F}_n^*(y|x) - \hat{F}_n(y|x)) + (\hat{F}_n(y|x) - F_n(y|x)) + (F_n(y|x) - F_{\varepsilon|X}(y|x))$$

Für die Entwicklung von $\hat{F}_n(\cdot|x)$ beziehen wir damit nur die Terme dieser Entwicklung ein, welche aus dem zweiten und dritten Summanden resultieren und erhalten:

Proposition 6.10 *Es gilt unter den Voraussetzungen (M1) – (M3), (H1), (H2), (K) dass*

$$\begin{aligned}
\hat{F}_n(\cdot|x) - F_{\varepsilon|X}(\cdot|x) &= H_{n,3}(\cdot, x) + H_{n,4}(\cdot, x) + b_n^{**}(\cdot, x) + o_p\left(\frac{1}{\sqrt{nh}}\right) \\
&=: H_n^*(\cdot, x) + b_n^{**}(\cdot, x) + o_p\left(\frac{1}{\sqrt{nh}}\right)
\end{aligned}$$

wobei

$$b_n^{**}(\cdot, x) := +\frac{b^2}{2} \mu_2^{\tilde{K}} f_X^{-1}(x) f_{\varepsilon|X}(\cdot|x) B(x) \\ +\frac{h^2}{2} \cdot \mu_2^K \cdot (F_{\varepsilon|X}^{(2)}(\cdot|x) + \frac{2 f_X'(x)}{f_X(x)} F_{\varepsilon|X}^{(1)}(\cdot|x))$$

und

$$H_n^*(\cdot, x) := f_X^{-1}(x) \frac{1}{n} \sum_{i=1}^n \varepsilon_i \int \frac{1}{h} K_h(z-x) f_{\varepsilon|X}(\cdot|z) \frac{1}{b} \tilde{K}_b(X_i - z) dz \\ + \left(f_X^{-1}(x) \cdot \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) (I\{\varepsilon_i \leq \cdot\} - F_{\varepsilon|X}(\cdot|x)) \right. \\ \left. - E[f_X^{-1}(x) \cdot \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) (I\{\varepsilon_i \leq \cdot\} - F_{\varepsilon|X}(\cdot|x))] \right).$$

Dann konvergiert $\sqrt{nh} (\hat{F}_n(\cdot|x) - F_{\varepsilon|X}(\cdot|x) - b_n^{**}(\cdot, x))$ schwach gegen einen zentrierten Gaußprozess $\tilde{G}(\cdot, x)$ mit Kovarianzfunktion

$$\text{Cov}(\tilde{G}(y_1, x), \tilde{G}(y_2, x)) = f_X^{-1}(x) E[\varepsilon^2 | X = x] f_{\varepsilon|X}(y_1|x) f_{\varepsilon|X}(y_2|x) \\ \cdot \int \int \int K(v + \lambda \cdot (u - w)) \tilde{K}(w) K(v) \tilde{K}(u) du dv dv \\ + \mu_0^{K^2} f_X^{-1}(x) F_{\varepsilon|X}(\min(y_1, y_2)|x) - \mu_0^{K^2} f_X^{-1}(x) F_{\varepsilon|X}(y_2|x) \cdot F_{\varepsilon|X}(y_1|x) \\ + f_X^{-1}(x) f_{\varepsilon|X}(y_1|x) E[\varepsilon I\{\varepsilon \leq y_2\} | X = x] \cdot \int \int K(u - \lambda v) \tilde{K}(u) K(v) du dv \\ + f_X^{-1}(x) f_{\varepsilon|X}(y_2|x) E[\varepsilon I\{\varepsilon \leq y_1\} | X = x] \cdot \int \int K(u - \lambda v) \tilde{K}(u) K(v) du dv$$

dabei beschreibt λ das asymptotische Verhältnis von b und h , d.h. $\lambda = \lim_{n \rightarrow \infty} \frac{b}{h}$.

Beweis:

Die Entwicklung wurde im vorherigen Kapitel hergeleitet. Die schwache Konvergenz folgt nach dem Beweis zu Proposition 6.9, denn die hier auftretenden Funktionenklassen treten auch dort auf und mit den Rechnungen zu Proposition 6.2 und 6.4 folgt der Erwartungswert und obige Kovarianz.

Nun fehlt noch eine Proposition zur schwachen Konvergenz im Falle g eindimensional und Indikatorfunktion der Gestalt $g(\varepsilon, x) = I\{\varepsilon \leq a(x)\} - d(x)$ (G3) erfüllend. Hier ist der Beweis und alle dazugehörigen Rechnungen ähnlich zum Beweis von Proposition 6.9, die auftretenden Funktionenklassen haben eine ähnliche Gestalt und die anzuwendenden Propositionen sind die gleichen.

Proposition 6.11 *Es gilt unter den Voraussetzungen (M1), (M2), (H1), (H2), (K) und g eindimensional von der Gestalt $g(\varepsilon, x) = I\{\varepsilon \leq a(x)\} - d(x)$:*

$$\hat{F}_n^*(\cdot | x) - F_{\varepsilon|X}(\cdot | x) = H_n(\cdot, x) + b_n^*(\cdot, x) + o_p\left(\frac{1}{\sqrt{nh}}\right)$$

wobei

$$\begin{aligned} b_n^*(\cdot, x) := & -f_X(x) \cdot \frac{h^2}{2} \mu_2^K \frac{\partial^2 E[g^t(\varepsilon, x) | X = t] \cdot f_X(t)}{\partial t^2} \Big|_{t=x} \cdot (\Sigma^{-1}(x))^t \cdot \mu_0^{K^2} \cdot E[g(\varepsilon, X) I\{\varepsilon \leq \cdot\} | X = x] \\ & - \Sigma^{-1}(x) \frac{b^2}{2} \mu_2^{\tilde{K}} f_{\varepsilon|X}(a(x) | x) B(x) \cdot \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq \cdot\} | X = x] \\ & + \frac{b^2}{2} \mu_2^{\tilde{K}} f_X^{-1}(x) f_{\varepsilon|X}(\cdot | x) B(x) \\ & + \frac{h^2}{2} \mu_2^K \cdot (F_{\varepsilon|X}^{(2)}(\cdot | x) + \frac{2f_X'(x)}{f_X(x)} F_{\varepsilon|X}^{(1)}(\cdot | x)) \end{aligned}$$

und

$$\begin{aligned} H_n(\cdot, x) := & - \left((\Sigma^{-1}(x) \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) g(\varepsilon_i, x))^t \cdot \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq \cdot\} | X = x] \right. \\ & \left. - E[(\Sigma^{-1}(x) \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) g(\varepsilon_i, x))^t \cdot \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq \cdot\} | X = x]] \right) \\ & - \frac{1}{n} \sum_{i=1}^n \varepsilon_i \int \frac{1}{h} K_h(z - x) f_{\varepsilon|X}(a(x) | z) \frac{1}{b} \tilde{K}_b(X_i - z) f_{\varepsilon|X}(y | z) d(y, z) \\ & \quad \cdot \Sigma^{-1}(x) \cdot \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq \cdot\} | X = x] \\ & + f_X^{-1}(x) \frac{1}{n} \sum_{i=1}^n \varepsilon_i \int \frac{1}{h} K_h(z - x) f_{\varepsilon|X}(\cdot | z) \frac{1}{b} \tilde{K}_b(X_i - z) dz \\ & + \left(f_X^{-1}(x) \cdot \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) (I\{\varepsilon_i \leq \cdot\} - F_{\varepsilon|X}(\cdot | x)) \right. \\ & \quad \left. - E[f_X^{-1}(x) \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) (I\{\varepsilon_i \leq \cdot\} - F_{\varepsilon|X}(\cdot | x))] \right). \end{aligned}$$

Dann konvergiert $\sqrt{nh} (\hat{F}_n^*(\cdot | x) - F_{\varepsilon|X}(\cdot | x) - b_n^*(\cdot, x))$ schwach gegen einen zentrierten Gaußprozess $G(\cdot, x)$ mit Kovarianzfunktion

$$\begin{aligned}
& \text{Cov}(G(y_1, x), G(y_2, x)) \\
&= (\mu_0^{K^2})^3 f_X(x) E \left[g^t(\varepsilon, U) \cdot (\Sigma^{-1}(U))^t \cdot E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = U] \right. \\
&\quad \left. \cdot g^t(\varepsilon, U) \cdot (\Sigma^{-1}(U))^t \cdot E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = U] \Big| U = x \right] \\
&+ \sigma^2(x) (\mu_0^{K^2})^2 f_X(x) f_{\varepsilon|X}(a(x) | x) (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x] \\
&\quad \cdot f_{\varepsilon|X}(a(x) | x) (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = x] \\
&\quad \cdot \int \int \int K(v + \lambda(u - w)) K(v) \tilde{K}(u) \tilde{K}(w) du dv dw \\
&+ f_X^{-1}(x) E[\varepsilon^2 | X = x] f_{\varepsilon|X}(y_1 | x) f_{\varepsilon|X}(y_2 | x) \int K(v + \lambda(u - w)) \tilde{K}(w) K(v) \tilde{K}(u) du dv dw \\
&- \mu_0^{K^2} \cdot f_{\varepsilon|X}(y_2 | x) E[(g(\varepsilon, X))^t \cdot \varepsilon | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x] \\
&\quad \cdot \int \int K(u + \lambda v) K(u) \tilde{K}(v) dudv \\
&- \mu_0^{K^2} \cdot f_{\varepsilon|X}(y_1 | x) E[(g(\varepsilon, X))^t \cdot \varepsilon | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = x] \\
&\quad \cdot \int \int K(u + \lambda v) K(u) \tilde{K}(v) dudv \\
&- f_X(x) (\mu_0^{K^2})^2 E[(g(\varepsilon, X))^t \varepsilon | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x] \\
&\quad \cdot f_{\varepsilon|X}(a(x) | x) (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = x] \\
&\quad \cdot \int \int K(u + \lambda v) K(u) \tilde{K}(v) dudv \\
&- f_X(x) (\mu_0^{K^2})^2 E[(g(\varepsilon, X))^t \varepsilon | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = x] \\
&\quad \cdot f_{\varepsilon|X}(a(x) | x) (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x] \\
&\quad \cdot \int \int K(u + \lambda v) K(u) \tilde{K}(v) dudv \\
&+ E[\varepsilon^2 | X = x] f_{\varepsilon|X}(a(x) | x) (\Sigma^{-1}(x))^t \cdot \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = x] \\
&\quad \cdot f_{\varepsilon|X}(y_1 | x) \int \int \int K(v + \lambda(u - w)) K(v) \tilde{K}(u) \tilde{K}(w) du dv dw \\
&+ E[\varepsilon^2 | X = x] f_{\varepsilon|X}(a(x) | x) (\Sigma^{-1}(x))^t \cdot \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x] \\
&\quad \cdot f_{\varepsilon|X}(y_2 | x) \int \int \int K(v + \lambda(u - w)) K(v) \tilde{K}(u) \tilde{K}(w) du dv dw \\
&+ \mu_0^{K^2} f_X^{-1}(x) F_{\varepsilon|X}(\min(y_1, y_2) | x) - \mu_0^{K^2} f_X^{-1}(x) F_{\varepsilon|X}(y_2 | x) \cdot F_{\varepsilon|X}(y_1 | x) \\
&- (\mu_0^{K^2})^2 E[(g(\varepsilon, X))^t I\{\varepsilon \leq y_2\} | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x] \\
&- (\mu_0^{K^2})^2 E[(g(\varepsilon, X))^t I\{\varepsilon \leq y_1\} | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = x] \\
&+ E[\varepsilon I\{\varepsilon \leq y_2\} | X = x] f_{\varepsilon|X}(a(x) | x) (\Sigma^{-1}(x))^t
\end{aligned}$$

$$\begin{aligned}
& \cdot E[g(\varepsilon, X) I\{\varepsilon \leq y_1\} | X = x] \cdot \mu_0^{K^2} \int \int K(u - \lambda v) \tilde{K}(u) K(v) du dv \\
& + E[\varepsilon I\{\varepsilon \leq y_1\} | X = x] f_{\varepsilon|X}(a(x)|x) (\Sigma^{-1}(x))^t \\
& \quad \cdot E[g(\varepsilon, X) I\{\varepsilon \leq y_2\} | X = x] \cdot \mu_0^{K^2} \int \int K(u - \lambda v) \tilde{K}(u) K(v) du dv \\
& + f_X^{-1}(x) f_{\varepsilon|X}(y_1|x) E[\varepsilon I\{\varepsilon \leq y_2\} | X = x] \cdot \int \int K(u - \lambda v) \tilde{K}(u) K(v) du dv \\
& + f_X^{-1}(x) f_{\varepsilon|X}(y_2|x) E[\varepsilon I\{\varepsilon \leq y_1\} | X = x] \cdot \int \int K(u - \lambda v) \tilde{K}(u) K(v) du dv,
\end{aligned}$$

dabei beschreibt λ das asymptotische Verhältnis von b und h , d.h. $\lambda = \lim_{n \rightarrow \infty} \frac{b}{h}$.

Der Beweis der schwachen Konvergenz folgt ähnlich dem der Proposition 6.9, ebenso der Beweis zur Kovarianz wie Proposition 6.4, dabei können die meisten Beweisideen und Rechnungen übernommen werden.

Bemerkung 6.12 (Bandbreiten) *In obigen Propositionen treten die Bandbreiten b und h in Wechselwirkung. Ist das Verhältnis λ der Bandbreiten > 0 , so müssten die Kerne taylorentwickelt werden und das genaue Aussehen der Kerne im Hinblick auf die Terme höherer Ordnung müsste bekannt sein. Da z.B. zusätzlich noch die Dichte $f_{\varepsilon|X}(y|x)$ in den Termen auftritt wird klar, dass die Wahl eines Kernes auch noch dadurch beeinflusst werden würde. Diese Dichte ist aber unbekannt und somit kann die Wahl des Kernes nicht darauf abgestimmt werden. Sinnvoll ist also die Wahl $\lambda = 0$, damit brauchen zunächst einmal keine weiteren Annahmen an die Kerne getroffen zu werden. Im nächsten Kapitel betrachten wir $\text{Var}(\hat{F}_n^*(y|x)) - \text{Var}(\hat{F}_n(y|x))$ beispielhaft für verschiedene bedingte Verteilungen und betrachten wie sich die Wahl des Kernes auswirkt.*

Mit diesen Aussagen zur schwachen Konvergenz wollen wir nun zur Beispielen übergehen.

7 Beispiele

In diesem Abschnitt wollen wir den in den vorherigen Kapiteln analysierten Schätzer, $\hat{F}_n^*(\cdot|x)$ mit dem Schätzer $\hat{F}_n(\cdot|x)$ vergleichen. Siehe dazu die Definitionen 4.1 und 4.6.

In den beiden vorherigen Kapiteln haben wir für beide Schätzer eine Entwicklung hergeleitet und schwache Konvergenz gegen einen Gaußprozess gezeigt, siehe Proposition 6.9 und 6.10. Die Varianz und den Bias der Asymptotik der beiden Schätzer wollen wir nun anhand von Beispielerteilungen vergleichen. Die im Weiteren genutzten Darstellungen finden sich in Proposition 6.9 und 6.10. Betrachte das folgende Modell:

- (B1)** Es gelte das Modell (3.1). $f_{\varepsilon|X}(\varepsilon|x)$ sei die Dichte der $\mathcal{N}(0, \sigma^2(x))$ Verteilung und $X \sim U[0, 1]$ verteilt. Ebenso seien Bandbreiten h und b und Kerne K, \tilde{K} so gewählt, dass (H1), (H2) und (K) erfüllt sind und $\lambda := \lim_{n \rightarrow \infty} \frac{b}{h} = 0$. Zur Vereinfachung sei noch $K = \tilde{K}$ angenommen.

Bemerkung 7.1 Die in (B1) genannten Dichten und Verteilungen erfüllen die Voraussetzungen (M1) und (M2), was einfach nachzurechnen ist.

7.1 Beispiel 1

Beispiel 7.2 Es gelte Modell (B1). Wir wollen in diesem Beispiel betrachten welche Auswirkungen die im Modell angenommene, aber nur im Schätzer $\hat{F}_n^*(y|x)$ explizit nutzbare Voraussetzung, dass $E[\varepsilon|X] = 0$, also $g(\varepsilon, x) = \varepsilon$ ist, hat. Dazu vergleichen wir diesen mit dem Schätzer $\hat{F}_n(y|x)$ aus Proposition 6.10, dabei bezeichne G den zu $\hat{F}_n^*(y|x)$, \tilde{G} den zu $\hat{F}_n(y|x)$ gehörigen Grenzprozess. Mit (B1) und dieser Wahl von g ist ebenfalls (M3) erfüllt. Wir berechnen jetzt zum Vergleich die Differenz der asymptotischen Varianz beider Schätzer, welche nach Proposition 6.9 und 6.10 gegeben sind und die Differenz der quadrierten Bias-Terme $(b_n^*(y, x))^2$ bzw. $(b_n^{**}(y, x))^2$.

Wir beginnen mit der Differenz der asymptotischen Varianzen. Benutze dazu, dass nach Proposition 4.9 (v) gilt:

$$\Sigma(x) = f_X(x) E[\varepsilon^2 | X = x] \mu_0^{K^2} = f_X(x) \sigma^2(x) \mu_0^{K^2}$$

$$\begin{aligned}
& \text{Var}(G(y, x)) - \text{Var}(\tilde{G}(y, x)) \\
&= (\mu_0^{K^2})^3 f_X(x) E \left[g^t(\varepsilon, U) \cdot (\Sigma^{-1}(U))^t \cdot E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = U] \right. \\
&\quad \left. \cdot g^t(\varepsilon, U) \cdot (\Sigma^{-1}(U))^t \cdot E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = U] \Big| U = x \right] \\
&+ \sigma^2(x) (\mu_0^{K^2})^2 f_X(x) E[(g'(\varepsilon, X))^t | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = x] \\
&\quad \cdot E[(g'(\varepsilon, X))^t | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = x] \\
&\quad \cdot \int \int \int K(v + \lambda(u - w)) K(v) \tilde{K}(u) \tilde{K}(w) du dv dw \\
&- 2 \mu_0^{K^2} \cdot f_{\varepsilon|X}(y|x) E[(g(\varepsilon, X))^t \cdot \varepsilon | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = x] \\
&\quad \cdot \int \int K(u - \lambda v) K(u) \tilde{K}(v) dudv \\
&- 2 f_X(x) (\mu_0^{K^2})^2 E[(g(\varepsilon, X))^t \varepsilon | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = x] \\
&\quad \cdot E[(g'(\varepsilon, X))^t | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = x] \\
&\quad \cdot \int \int K(u - \lambda v) K(u) \tilde{K}(v) dudv \\
&+ 2 E[\varepsilon^2 | X = x] E[(g'(\varepsilon, X))^t | X = x] (\Sigma^{-1}(x))^t \cdot \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = x] \\
&\quad \cdot f_{\varepsilon|X}(y|x) \int \int \int K(v + \lambda(u - w)) K(v) \tilde{K}(u) \tilde{K}(w) du dv dw \\
&- 2 (\mu_0^{K^2})^2 E[(g(\varepsilon, X))^t I\{\varepsilon \leq y\} | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = x] \\
&+ 2 E[\varepsilon I\{\varepsilon \leq y\} | X = x] E[(g'(\varepsilon, x))^t | X = x] (\Sigma^{-1}(x))^t \\
&\quad \cdot E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = x] \cdot \mu_0^{K^2} \int \int K(u - \lambda v) \tilde{K}(u) K(v) du dv \\
&= (\mu_0^{K^2})^2 \Sigma^{-1}(x) \cdot E^2[\varepsilon I\{\varepsilon \leq y\} | X = x] \\
&+ (\mu_0^{K^2})^2 f_X(x) \sigma^2(x) (\Sigma^{-1}(x))^2 \cdot E^2[\varepsilon I\{\varepsilon \leq y\} | X = x] \\
&\quad \cdot \int \int \int K(v + \lambda(u - w)) K(v) \tilde{K}(u) \tilde{K}(w) du dv dw \\
&- 2 f_X^{-1}(x) f_{\varepsilon|X}(y|x) E[\varepsilon I\{\varepsilon \leq y\} | X = x] \\
&\quad \cdot \int \int K(u - \lambda v) K(u) \tilde{K}(v) dudv \\
&- 2 \cdot \Sigma^{-1}(x) \mu_0^{K^2} E^2[\varepsilon I\{\varepsilon \leq y\} | X = x] \\
&\quad \cdot \int \int K(u - \lambda v) K(u) \tilde{K}(v) dudv \\
&+ 2 f_X^{-1}(x) f_{\varepsilon|X}(y|x) E[\varepsilon I\{\varepsilon \leq y\} | X = x] \\
&\quad \cdot \int \int \int K(v + \lambda(u - w)) K(v) \tilde{K}(u) \tilde{K}(w) du dv dw \\
&- 2 \Sigma^{-1}(x) \cdot (\mu_0^{K^2})^2 E^2[\varepsilon I\{\varepsilon \leq y\} | X = x] \\
&+ 2 E^2[\varepsilon I\{\varepsilon \leq y\} | X = x] \Sigma^{-1}(x) \mu_0^{K^2} \int \int K(u - \lambda v) \tilde{K}(u) K(v) du dv \\
&= \mu_0^{K^2} f_X(x) \Sigma^{-1}(x) \cdot E^2[\varepsilon I\{\varepsilon \leq y\} | X = x]
\end{aligned}$$

$$\begin{aligned}
& \cdot \int \int \int K(v + \lambda(u - w)) K(v) \tilde{K}(u) \tilde{K}(w) du dv dw \\
-2 f_X^{-1}(x) f_{\varepsilon|X}(y|x) E[\varepsilon I\{\varepsilon \leq y\} | X = x] \\
& \cdot \int \int K(u - \lambda v) K(u) \tilde{K}(v) dudv \\
+2 f_X^{-1}(x) f_{\varepsilon|X}(y|x) E[\varepsilon I\{\varepsilon \leq y\} | X = x] \\
& \cdot \int \int \int K(v + \lambda(u - w)) K(v) \tilde{K}(u) \tilde{K}(w) du dv dw \\
- \Sigma^{-1}(x) \cdot (\mu_0^{K^2})^2 E^2[\varepsilon I\{\varepsilon \leq y\} | X = x]
\end{aligned}$$

Bisher haben wir noch nicht die genaue Gestalt der Dichten, wie sie in (B1) vorausgesetzt ist, genutzt. Setzt man nun wie nach Voraussetzung $\lambda = 0$ und $\tilde{K} = K$, so zeigt sich, dass die Differenz der Varianzen unabhängig von der Wahl der Verteilung und des Kernes K Null ergibt. Unabhängig vom (die Voraussetzungen aus Kapitel 3 erfüllenden) Modell ergibt sich also im Falle $g(\varepsilon, x) = \varepsilon$ und $\frac{b}{h} = o(1)$ für beide Schätzer dieselbe Varianz. dabei ist zu beachten, dass kein spezielles $\sigma^2(x)$ gewählt wurde.

Zu den Bias-Ausdrücken:

Die Biasausdrücke haben ein ähnliches Aussehen, genauer $b_n^{**}(y, x)$ ist in $b_n^*(y, x)$ enthalten. Es gilt speziell in diesem Beispiel, dass

$$\left. \frac{\partial^2 E[\varepsilon | X = t] \cdot f_X(t)}{\partial t^2} \right|_{t=x} = 0, \quad F_{\varepsilon|X}^{(2)}(y|x) = 0, \quad \frac{2 f'_X(x)}{f_X(x)} F_{\varepsilon|X}^{(1)}(y|x) = 0 \quad \text{und} \quad \Sigma(x) = \sigma^2(x) \mu_0^{K^2}.$$

Für die Differenz der quadrierten Bias-Ausdrücke ergibt sich damit:

$$\begin{aligned}
& (b_n^*(y, x))^2 - (b_n^{**}(y, x))^2 \\
& = \left(-\frac{h^2}{2} \mu_K^2 \left. \frac{\partial^2 E[g^t(\varepsilon, x) | X = t] \cdot f_X(t)}{\partial t^2} \right|_{t=x} \cdot (\Sigma^{-1}(x))^t \cdot \mu_0^{K^2} \cdot E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = x] \right. \\
& \quad \left. + (\Sigma^{-1}(x) \frac{b^2}{2} \mu_{\tilde{K}}^2 E[g'(\varepsilon, X) | X = x] B(x))^t \cdot \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = x] \right)^2 \\
& + 2 \left(-\frac{h^2}{2} \mu_K^2 \left. \frac{\partial^2 E[g^t(\varepsilon, x) | X = t] \cdot f_X(t)}{\partial t^2} \right|_{t=x} \right. \\
& \quad \left. \cdot (\Sigma^{-1}(x))^t \cdot \mu_0^{K^2} \cdot E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = x] \right)
\end{aligned}$$

$$\begin{aligned}
& + (\Sigma^{-1}(x) \frac{b^2}{2} \mu_2^{\tilde{K}} E[g'(\varepsilon, X) | X = x] B(x))^t \cdot \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = x] \\
& \cdot \left(\frac{b^2}{2} \mu_2^{\tilde{K}} f_X^{-1}(x) f_{\varepsilon|X}(y|x) B(x) + \frac{h^2}{2} \cdot \mu_2^K \cdot (F_{\varepsilon|X}^{(2)}(y|x) + \frac{2f'_X(x)}{f_X(x)} F_{\varepsilon|X}^{(1)}(y|x)) \right) \\
= & \left((\Sigma^{-1}(x) \frac{b^2}{2} \mu_2^{\tilde{K}} E[g'(\varepsilon, X) | X = x] B(x))^t \cdot \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = x] \right)^2 \\
& + 2 \left((\Sigma^{-1}(x) \frac{b^2}{2} \mu_2^{\tilde{K}} E[g'(\varepsilon, X) | X = x] B(x))^t \cdot \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = x] \right) \\
& \cdot \left(\frac{b^2}{2} \mu_2^{\tilde{K}} f_X^{-1}(x) f_{\varepsilon|X}(y|x) B(x) + \frac{h^2}{2} \cdot \mu_2^K \cdot (F_{\varepsilon|X}^{(2)}(y|x) + \frac{2f'_X(x)}{f_X(x)} F_{\varepsilon|X}^{(1)}(y|x)) \right) \\
\stackrel{b=o(h)}{=} & o(h^4)
\end{aligned}$$

In diesem Beispiel gibt es also weder in der Varianz noch im Bias einen Unterschied.

7.2 Beispiel 2

Beispiel 7.3 Es gelte Modell (B1). Zusätzlich sei noch gegeben, dass $E[\varepsilon^2 | X = x] = \sigma^2(x)$ bekannt ist, also $g(\varepsilon, x) = \varepsilon^2 - \sigma^2(x) = \varepsilon^2 - \sigma^2(x)$. Mit (B1) und dieser Wahl von g ist ebenfalls (M3) erfüllt. Wir berechnen jetzt zum Vergleich für beide Schätzer die nach Proposition 6.9 und 6.10 gegebene asymptotische Varianz, also $\text{Cov}(G(y, x), G(y, x))$ bzw. $\text{Cov}(\tilde{G}(y, x), \tilde{G}(y, x))$ und den Bias $b_n^*(y, x)$ bzw. $b_n^{**}(y, x)$ und betrachten wie im Beispiel zuvor die Differenz bzw. die Differenz der Quadrate.

Wir beginnen mit der Varianz. Benutze dazu, dass nach Proposition 4.9 (v) bzw. den Voraussetzungen gilt:

$$E[g'(\varepsilon, X) | X = x] = 2E[\varepsilon | X = x] = 0$$

$$\begin{aligned}
E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = x] &= \int_{-\infty}^y (\varepsilon^2 - \sigma^2(x)) f_{\varepsilon|X}(\varepsilon|x) d\varepsilon \\
&= -\sigma^2(x) \cdot (y f_{\varepsilon|X}(y|x) - F_{\varepsilon|X}(y|x)) - \sigma^2(x) \cdot F_{\varepsilon|X}(y|x) \\
&= -\sigma^2(x) \cdot y f_{\varepsilon|X}(y|x)
\end{aligned}$$

$$\begin{aligned}
E[g^2(\varepsilon, X) | X = x] &= \int_{\mathbb{R}} (\varepsilon^2 - \sigma^2(x))^2 f_{\varepsilon|X}(\varepsilon|x) d\varepsilon \\
&= \int_{\mathbb{R}} \varepsilon^4 f_{\varepsilon|X}(\varepsilon|x) d\varepsilon - (\sigma^2(x))^2
\end{aligned}$$

$$\begin{aligned}
&= 3(\sigma^2(x))^2 - (\sigma^2(x))^2 \\
&= 2(\sigma^2(x))^2
\end{aligned}$$

$$\Sigma(x) = f_X(x) E[g^2(\varepsilon, X) | X = x] \mu_0^{K^2} = \mu_0^{K^2} f_X(x) \cdot 2(\sigma^2(x))^2$$

$$E[\varepsilon g(\varepsilon, X) | X = x] = \int_{\mathbb{R}} \varepsilon(\varepsilon^2 - \sigma^2(x)) f_{\varepsilon|X}(\varepsilon|x) d\varepsilon = 0$$

und damit:

$$\begin{aligned}
&Var(G(y, x)) - Var(\tilde{G}(y, x)) \\
&= (\mu_0^{K^2})^3 f_X(x) E \left[g^t(\varepsilon, U) \cdot (\Sigma^{-1}(U))^t \cdot E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = U] \right. \\
&\quad \left. \cdot g^t(\varepsilon, U) \cdot (\Sigma^{-1}(U))^t \cdot E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = U] \Big| U = x \right] \\
&+ \sigma^2(x) (\mu_0^{K^2})^3 f_X(x) E[(g'(\varepsilon, X))^t | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = x] \\
&\quad \cdot E[(g'(\varepsilon, X))^t | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = x] \\
&- 2(\mu_0^{K^2})^2 \cdot f_{\varepsilon|X}(y|x) E[(g(\varepsilon, X))^t \cdot \varepsilon | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = x] \\
&- 2f_X(x) (\mu_0^{K^2})^3 E[(g(\varepsilon, X))^t \varepsilon | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = x] \\
&\quad \cdot E[(g'(\varepsilon, X))^t | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = x] \\
&+ 2E[\varepsilon^2 | X = x] E[(g'(\varepsilon, X))^t | X = x] (\Sigma^{-1}(x))^t \cdot (\mu_0^{K^2})^2 E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = x] \cdot f_{\varepsilon|X}(y|x) \\
&- 2(\mu_0^{K^2})^2 E[(g(\varepsilon, X))^t I\{\varepsilon \leq y\} | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = x] \\
&+ 2E[\varepsilon I\{\varepsilon \leq y\} | X = x] E[(g'(\varepsilon, X))^t | X = x] (\Sigma^{-1}(x))^t \cdot E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = x] \cdot (\mu_0^{K^2})^2 \\
&= (\mu_0^{K^2})^3 f_X(x) E \left[g^t(\varepsilon, U) \cdot (\Sigma^{-1}(U))^t \cdot E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = U] \right. \\
&\quad \left. \cdot g^t(\varepsilon, U) \cdot (\Sigma^{-1}(U))^t \cdot E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = U] \Big| U = x \right] \\
&- 2(\mu_0^{K^2})^2 E[(g(\varepsilon, X))^t I\{\varepsilon \leq y\} | X = x] (\Sigma^{-1}(x))^t E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = x] \\
&= - \frac{\mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = x]^2}{f_X(x) E[g^2(\varepsilon, X) | X = x]} \\
&= - \frac{\mu_0^{K^2} (-\sigma^2(x))^2 y^2 f_{\varepsilon|X}^2(y|x)}{f_X(x) 2(\sigma^2(x))^2} \\
&= - \frac{\mu_0^{K^2}}{2f_X(x)} y^2 f_{\varepsilon|X}^2(y|x)
\end{aligned}$$

Also ist die Varianz des ELE unabhängig von der Gestalt von $\sigma^2(x)$ kleiner als die des KE.

Im Folgenden berechnen wir die Bias-Differenzen und geben darauf die Differenz des MSE der beiden Schätzer an.

Zu den Bias-Ausdrücken. Es gilt mit $b^2 = o(h^2)$:

$$\begin{aligned}
b_n^*(y, x) &= b_n^{**}(y, x) \\
&\quad + \frac{h^2}{2} \cdot \mu_2^K \frac{\partial^2 E[g(\varepsilon, x) | X = t] \cdot f_X(t)}{\partial t^2} \Big|_{t=x} \cdot (\Sigma^{-1}(x))^t \cdot \mu_0^{K^2} \cdot E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = x] \\
&\quad + o(h^2) \\
&= b_n^{**}(y, x) + \frac{h^2}{2} \frac{\mu_2^K}{f_X(x)} \frac{\partial^2 E[g(\varepsilon, x) | X = t] \cdot f_X(t)}{\partial t^2} \Big|_{t=x} \cdot \frac{E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = x]}{E[g^2(\varepsilon, X) | X = x]} + o(h^2) \\
&= b_n^{**}(y, x) + \frac{h^2}{4f_X(x)} \cdot \mu_2^K \frac{\partial^2 E[g(\varepsilon, x) | X = t] \cdot f_X(t)}{\partial t^2} \Big|_{t=x} \cdot \frac{y \cdot f_{\varepsilon|X}(y|x)}{\sigma^2(x)} + o(h^2),
\end{aligned}$$

dabei nutzt die letzte Gleichheit die Normalverteilungsannahme. Ist $\text{Var}(\varepsilon) = \vartheta x$ und $g(\varepsilon, x) = \varepsilon^2 - \vartheta x$, so ist $b_n^{**}(y, x) = b_n^*(y, x) + o(h^2)$. Für $\text{Var}(\varepsilon) = \exp(\vartheta x)$ und $g(\varepsilon, x) = \varepsilon^2 - \exp(\vartheta x)$ ergibt sich unter der Annahme X gleichverteilt auf $[0, 1]$:

$$b_n^*(y, x) = b_n^{**}(y, x) - \vartheta^2 \frac{h^2}{4} \cdot \mu_2^K \cdot y f_{\varepsilon|X}(y|x) + o(h^2).$$

Für die Differenz der quadrierten Bias-Ausdrücke erhält man:

$$\begin{aligned}
&(b_n^*(y, x))^2 - (b_n^{**}(y, x))^2 \\
&= +2 b_n^{**}(y, x) \cdot \left(\frac{h^2}{4} \cdot \mu_2^K \frac{\partial^2 E[g(\varepsilon, x) | X = t] \cdot f_X(t)}{\partial t^2} \Big|_{t=x} \cdot \frac{y \cdot f_{\varepsilon|X}(y|x)}{\sigma^2(x)} \right) \\
&\quad + \left(\frac{h^2}{4} \cdot \mu_2^K \frac{\partial^2 E[g(\varepsilon, x) | X = t] \cdot f_X(t)}{\partial t^2} \Big|_{t=x} \cdot \frac{y \cdot f_{\varepsilon|X}(y|x)}{\sigma^2(x)} \right)^2 \\
&\quad + o(h^4).
\end{aligned}$$

speziell für $g(\varepsilon, x) = \varepsilon^2 - \exp(\vartheta x)$ ergibt sich:

$$b_n^{**}(y, x) = \frac{h^2}{2} \mu_2^K f_X^{-1}(x) \cdot \left(\int_{(-\infty, y]} \ddot{f}_{\varepsilon, X}(\varepsilon, x) d\varepsilon - F_{\varepsilon|X}(y|x) f_X''(x) \right) + o(h^2)$$

$$\begin{aligned}
&= \frac{h^2}{2} \mu_2^K \int_{(-\infty, y]} \ddot{f}_{\varepsilon, X}(\varepsilon, x) d\varepsilon + o(h^2) \\
&\stackrel{(1)}{=} \frac{h^2}{2} \mu_2^K \frac{\vartheta^2}{4} y \cdot f_{\varepsilon|X}(y|x) - \frac{h^2}{2} \mu_2^K \frac{\vartheta^2}{4 \cdot \exp(\vartheta x)} y^3 \cdot f_{\varepsilon|X}(y|x) + o(h^2),
\end{aligned}$$

dabei erfordert (1) eine längere Rechnung. Es gilt $\lim_{n \rightarrow \infty} \sqrt{nh} h^2 = c$. Setze zur Vereinfachung im Weiteren $c = 1$. Damit gilt:

$$\begin{aligned}
(b^{**}(y, x)) &= \lim_{n \rightarrow \infty} \sqrt{nh} b_n^{**}(y, x) \\
&= \mu_2^K \frac{\vartheta^2}{8} y \cdot f_{\varepsilon|X}(y|x) - \mu_2^K \frac{\vartheta^2}{8 \cdot \exp(\vartheta x)} y^3 \cdot f_{\varepsilon|X}(y|x)
\end{aligned}$$

und somit asymptotisch:

$$\begin{aligned}
(b^*(y, x))^2 - (b^{**}(y, x))^2 &:= \lim_{n \rightarrow \infty} nh (b_n^*(y, x))^2 - (b_n^{**}(y, x))^2 \\
&= -2 b^{**}(y, x) \cdot \vartheta^2 \frac{1}{4} \cdot \mu_2^K \cdot y f_{\varepsilon|X}(y|x) \\
&\quad + \vartheta^4 \frac{1}{4^2} \cdot (\mu_2^K)^2 \cdot y^2 f_{\varepsilon|X}^2(y|x) \\
&= +\vartheta^4 \frac{1}{4^2 \exp(\vartheta x)} \cdot (\mu_2^K)^2 \cdot y^4 f_{\varepsilon|X}^2(y|x)
\end{aligned}$$

und damit

$$\begin{aligned}
&\lim_{n \rightarrow \infty} MSE(\hat{F}_n^*(y|x)) - MSE(\hat{F}_n^{**}(y|x)) \\
&= \text{Var}(G(y, x)) + (b^*(y, x))^2 - \text{Var}(\tilde{G}(y, x)) - (b^{**}(y, x))^2 \\
&= -\frac{\mu_0^{K^2}}{2} y^2 f_{\varepsilon|X}^2(y|x) \\
&\quad + \vartheta^4 \frac{1}{4^2 \exp(\vartheta x)} \cdot (\mu_2^K)^2 \cdot y^4 f_{\varepsilon|X}^2(y|x)
\end{aligned}$$

Im folgenden Kapitel werden MSE und MISE für beide Beispiele und verschiedene ϑ simuliert.

Bemerkung 7.4 (weitere Beispiele) Auf weitere Beispiele wollen wir verzichten, da obige Rechnungen meist sehr unübersichtlich werden und nicht sehr aussagekräftig sind, da man das Modell

sehr speziell zu wählen hat. Nebenbedingungen in Quantilsform und Mischungen werden wir aber in den nun folgenden Simulationen betrachten, ebenso mehrdimensionale Nebeninformationen.

7.3 Zusammenfassung der Ergebnisse

Die Berechnung des Bias und der Varianz war für die Schätzer $\hat{F}_n^*(y|x)$ und $\hat{F}_n(y|x)$ in speziellen Modelle recht aufwändig und unübersichtlich. Zwar konnten wir für einige spezielle Modelle teils eine Verkleinerung der Varianz nachweisen, dies war aber aufgrund der auf vielfältige Weise in die Varianz- und Biasausdrücke eingehenden Modelleigenheiten nicht für große Klassen allgemein beweisbar. Im folgenden Kapitel werden $\hat{F}_n^*(y|x)$ und $\hat{F}_n(y|x)$ für deutlich mehr Beispielveilungen in Simulationen verglichen und dort wird im Gegensatz zu obigen unübersichtlichen Termen der Vorteil des Schätzers $\hat{F}_n^*(y|x)$ im Vergleich zu $\hat{F}_n(y|x)$ in einigen Fällen deutlich.

8 Simulationen

8.1 Einleitung

In diesem Kapitel werden verschiedenste Modelle betrachtet, die Voraussetzungen überprüft und für die beiden Schätzer, $\hat{F}_n^*(y|x)$ und $\hat{F}_n(y|x)$ (siehe Definition 4.2 und 4.6), jeweils $MSE(y, x) = E[(\hat{F}(y|x) - F(y|x))^2]$ und $MISE(x) = E[(\int_{\mathbb{R}} \hat{F}(y|x) - F(y|x))^2 dy]$ approximiert. Dabei approximieren wir $MSE(y, x)$ in dem wir D mal unabhängig Daten der gegebenen Modellverteilung erzeugen und mit diesen Beobachtungen

$$((X_1, Y_1)_k, \dots, (X_n, Y_n)_k)_{k=1, \dots, D} = ((x_1, y_1)_k, \dots, (x_n, y_n)_k)_{k=1, \dots, D}$$

für jedes k die Schätzer $\hat{F}_{n_k}^*(y|x)$ und $\hat{F}_{n_k}(y|x)$ bestimmen. Durch Mittlung ergibt sich für den MSE von $\hat{F}_n(y|x)$ die Approximation

$$\widehat{MSE}(y, x) = \frac{1}{D} \sum_{k=1}^m (\hat{F}_{n_k}(y|x) - F(y|x))^2$$

analog ergibt sich die Approximation für $\hat{F}_{n_k}^*(y|x)$. Diese Approximation führen wir an jedem Datensatz für verschiedene $x \in [0, 1]$ durch, im allgemeinen setzen wir hier ein 1/10-Schritt Gitter.

Ebenso approximieren wir an eben diesen Datensätzen $MISE(x)$ für verschiedene $x \in [0, 1]$ durch

$$\widehat{MISE}(x) = \frac{1}{D} \sum_{k=1}^D \sum_{l=1}^{L-1} (\hat{F}_{n_k}(\xi_l|x) - F(\xi_l|x))^2 (\xi_{l+1} - \xi_l),$$

dabei ist die Zerlegung $(\xi_l)_{l=1, \dots, L}$ gleichmäßig und für die Normalverteilungsbeispiele auf $[-4, 4]$ und mit 1001 Gitterpunkten gewählt worden. Andere Beispiele weichen vom Intervall $[-4, 4]$ ab, die Anzahl der Gitterpunkte bleibt aber gleich. Ebenso wie zum $MSE(y, x)$ wurde der $MISE(x)$ auch wieder für verschiedene $x \in [0, 1]$ parallel approximiert, d.h. es wurden dieselben Daten genutzt. Insbesondere sind die jeweils angegebenen Grafiken zum MSE und $MISE$ auf Basis derselben Stichproben erstellt worden.

8.2 Berechnung der Schätzer

Zur Erinnerung: Der Modellansatz dieser Arbeit hat die Form $Y = m(X) + \varepsilon$. Für jedes Beispiel wurde die unbekannte Modellfunktion m gewählt als $m(x) = 5 \cdot x$, es wurden zu den dort gegebenen Verteilungen unabhängige Datensätze $(X_1, Y_1), \dots, (X_n, Y_n)$ erzeugt und daraus mittels des Nadaraya-Watson-Schätzers (vgl. Definition 2.2) $(X_1, \hat{\varepsilon}_1), \dots, (X_n, \hat{\varepsilon}_n)$ gebildet, die Grundlage beider Schätzer $\hat{F}_{\varepsilon|X}(y|x)$ der Verteilungsfunktion $F_{\varepsilon|X}(y|x)$.

Für die Berechnung der Schätzer wurden i.A. die Bandbreiten $h = n^{-1/5}$ und $b = n^{-1/4}$ genutzt, vgl. (H1) und (H2). Der genutzte Kern ist im Nadaraya-Watson-Schätzer ein Gaußkern, an allen anderen Stellen nutzen wir den Epanechnikov-Kern (siehe Definition 2.46 ff.). Simulationen mit Gaußkernen an allen Stellen ergaben keine nennenswerten Abweichungen zu den Ergebnissen dieser Simulationen.

Um die Gewichte $p_i(x)$ des Schätzers $\hat{F}_n^*(y|x)$ zu schätzen ist $\eta(x)$ zu schätzen, vgl. Definition 4.6 und Gleichung (4.8). Für $\eta(x)$ nutzen wir die definierende Gleichung (4.5) und lösen diese numerisch. In einigen Fällen konnte für eine gegebene Stichprobe diese Gleichung nur unzureichend gelöst werden, in diesen Fällen wurden automatisch die Gewichte als $\frac{1}{n}$ gesetzt, also identisch zu denen des Kernschätzeransatzes.

Im Weiteren geben wir nun verschiedenste Beispiele, beginnend mit dem Normalverteilungsmodell aus dem vorherigen Kapitel. Es folgen neben Beispielen mit mehrdimensionalen Nebeninformationen auch noch Beispiele, welche die in dieser Arbeit getroffenen Voraussetzungen nicht erfüllen und Beispiele mit geschätzten Nebeninformationen.

8.3 Datenerzeugung

Die den Simulationen zugrundeliegenden Stichproben $(X_1, Y_1), \dots, (X_n, Y_n)$ wurden weitestgehend mittels in der freien Statistik-Software R vorhandener Funktionen erzeugt. Die Schätzer sowie einige Funktionen zur Erzeugung von Zufallsvariablen wurden ebenfalls in R programmiert.

8.4 Simulationsbeispiel 1 (Normalverteilungsmodell)

Es sei X gleichverteilt auf $[0, 1]$ und ε gegeben $X = x$ normalverteilt mit Erwartungswert 0 und Varianz $\sigma^2(x) = 1 + \vartheta \cdot x$. Die Verteilung der Residuen hängt also (für $\vartheta \neq 0$) von X ab.

Mit obigen Verteilungen wurde ein Datensatz $(X_1, Y_1), \dots, (X_n, Y_n)$ erzeugt und mittels des Nadaraya-Watson-Schätzers (vgl. Definition 2.2) $(X_1, \hat{\varepsilon}_1), \dots, (X_n, \hat{\varepsilon}_n)$ gebildet. Ein Beispielda-

tensatz $(X_1, Y_1), \dots, (X_n, Y_n)$ bzw. $(X_1, \hat{\varepsilon}_1), \dots, (X_n, \hat{\varepsilon}_n)$ mit $\vartheta = 1$ und $n = 100$ hat die folgende Gestalt:

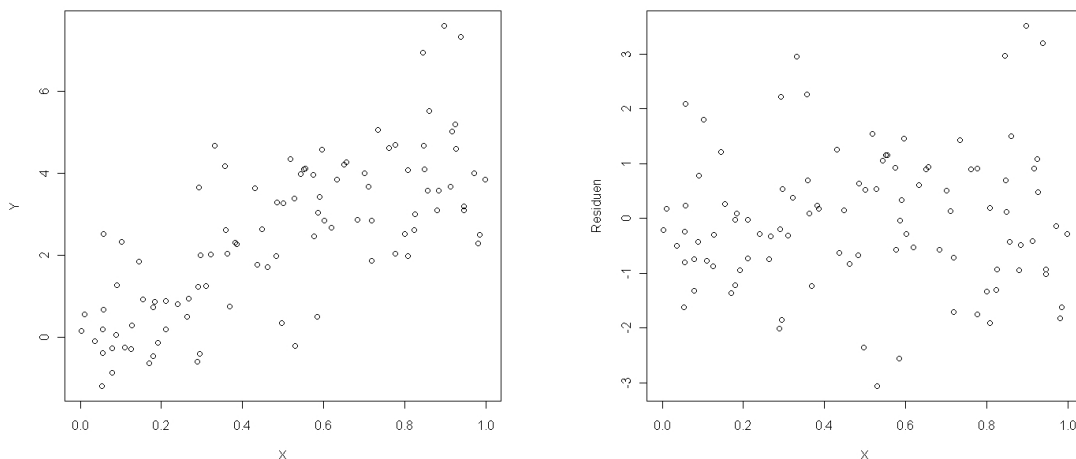


Abb.1: Beispieldatensatz $(X_1, Y_1), \dots, (X_n, Y_n)$ bzw. $(X_1, \hat{\varepsilon}_1), \dots, (X_n, \hat{\varepsilon}_n)$

Die X -Abhängigkeit der Varianz der Zufallsvariablen ε ist hier nicht deutlich zu erkennen. Die Klasse der Normalverteilungen ist eine wichtige Beispielklasse, da sie häufig angenommen wird. In diesem grundlegenden Beispiel wird neben dem Einbinden verschiedenster gegebener Nebeninformationen untersucht wie sich eine veränderte Wahl der Bandbreite b auf die Simulationsergebnisse auswirkt. Es werden dazu in Beispiel (a) drei Spezifikationen betrachtet, dabei ist $n = 100$ und die Anzahl der wiederholten Datenerzeugungen $D = 250$.

- (a) (Beispielvariation 1) Es gelte obiges Modell mit $\vartheta = 1$ und $n = 100$. Die Nebeninformation sei gegeben durch $g(\varepsilon, x) = \varepsilon$. Diese Zusatzinformation der Zentriertheit der Fehler ist schon in den Modellannahmen für beide Schätzer enthalten (vgl. die Modellvoraussetzungen (M1)). Grund ist der Nadaraya-Watson Regressionsschätzer, welcher den nicht zentrierten Anteil der Fehler zur Schätzung \hat{m} hinzunimmt. Es ist also interessant zu sehen wie sich in diesem Beispiel die Schätzer zueinander verhalten.

$MSE(x, y)$ und $MISE(x)$ werden für $h = n^{-1/5}$ und drei verschiedene Wahlen der Bandbreite b ($n^{-1/4}$, $n^{-1/5}$, Cross-Validation-Methode) approximiert. Cross-Validation-Methoden sind Bandbreitenwahl-Methoden, welche zu jeder Stichprobe eine spezielle Bandbreite b , Lösung einer Minimierungsaufgabe, wählen, genauer werden wir darauf aber nicht eingehen.

Die Schätzkurven haben, beispielhaft ausgewertet an einem Datensatz, $x = 0.8$ und der Wahl $b = n^{-1/4}$, die folgende Gestalt:

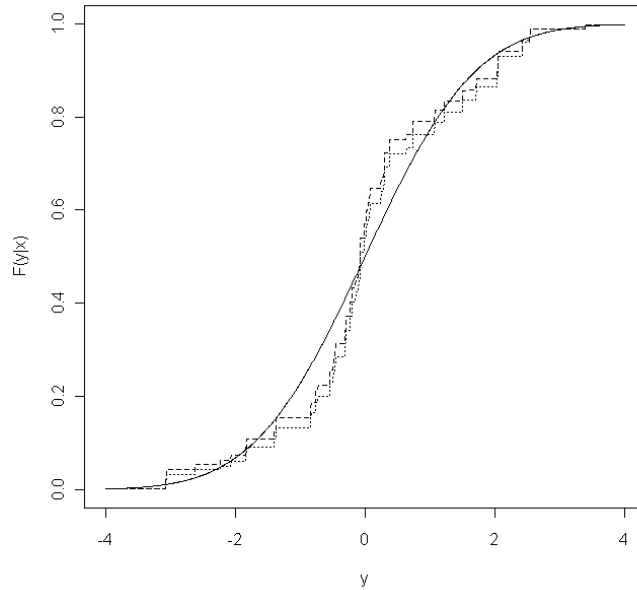


Abb.2: Schätzkurven für $x = 0.8$

Dabei ist $\hat{F}_n^*(y|x)$ die gestrichelte, $\hat{F}_n(y|x)$ die gepunktete und $F(y|x)$ die durchgezogene Kurve. Es ergeben sich die folgenden Grafiken für die approximierte $MSE(x,y)$ -Differenz $\Delta \widehat{MSE} = \widehat{MSE}(\hat{F}_n(y|x)) - \widehat{MSE}(\hat{F}_n^*(y|x))$ und den $MISE(x)$ mit den verschiedenen Bandbreitenwahlen b . Es ist die Kurve des Empirical-Likelihood-Schätzers der gestrichelte, die des Kernschätzers der durchgezogene Graph:

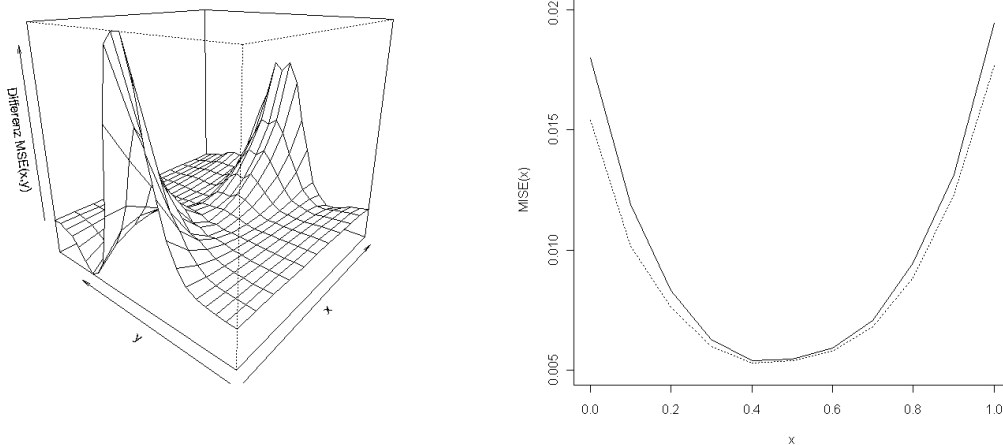


Abb.3: $\Delta \widehat{MSE}$ und \widehat{MISE} für die Bandbreitenwahl $b = n^{-1/4}$

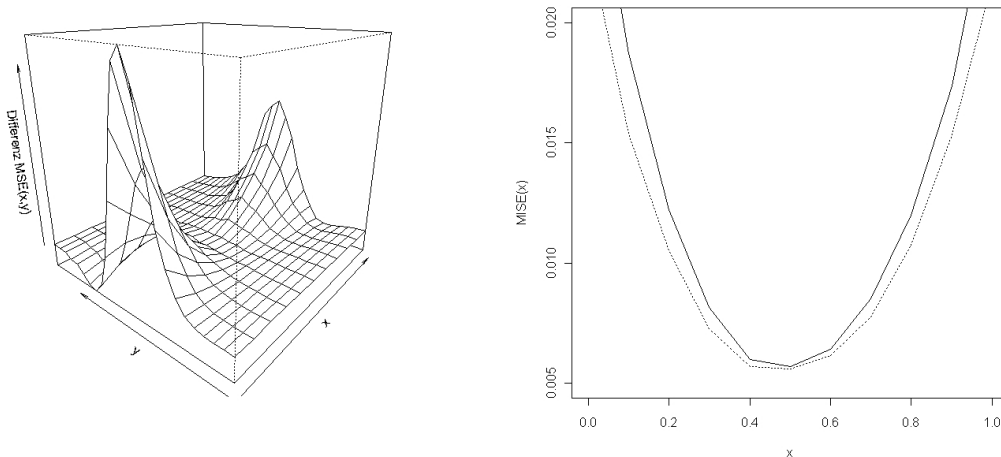


Abb.4: $\widehat{\Delta MSE}$ und \widehat{MISE} für die Bandbreitenwahl $b = n^{-1/5}$

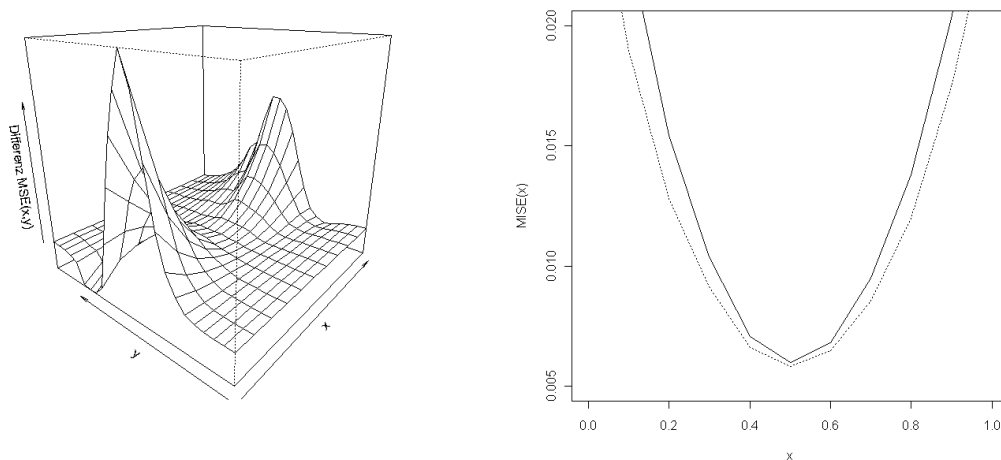


Abb.5: $\widehat{\Delta MSE}$ und \widehat{MISE} für die Bandbreitenwahl mittels cross-validation-Methode

Zu erkennen ist zum einen, dass sich die unterschiedlichen Bandbreitenwahlen nicht (oder nur sehr schwach) auf die obigen Schätzkurven des MSE und $MISE$ auswirken und zum anderen, dass der Schätzer $\hat{F}_n^*(y|x)$, welcher die vorausgesetzte Nebeninformation der Zentriertheit der Fehler explizit nutzt, einen gleichmäßig besseren $MISE(x)$ aufweist. Die Grafik zur $MSE(x, y)$ -Differenz zeigt speziell im Bereich sehr kleiner und sehr großer x -Werte Verbesserungen auf. Dazu betrachte man auch noch die folgenden drei Schnitte des $\Delta MSE(x, y)$ für $x = 0, 0.5, 1$ und die Bandbreite $b = n^{-1/4}$, man beachte dabei die unterschiedlichen Maßstäbe:

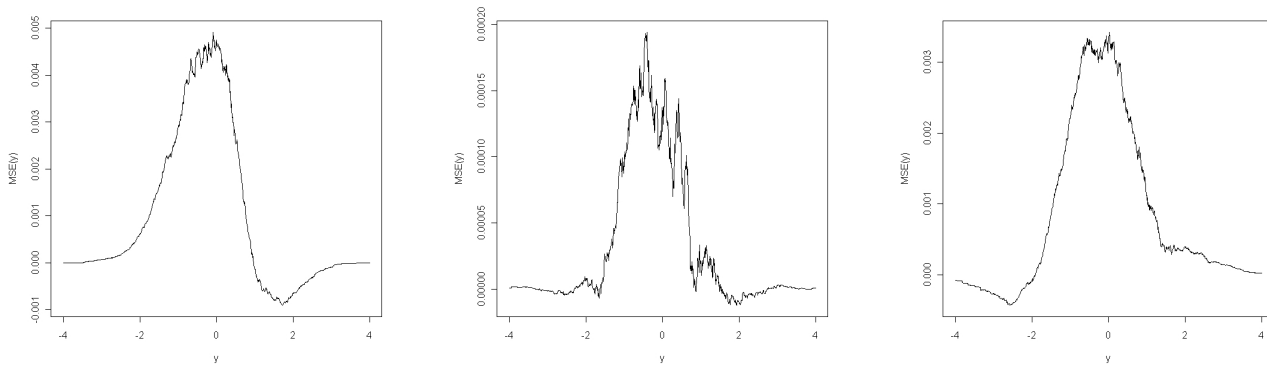


Abb.6: $\widehat{\Delta MSE}(0, y)$ bzw. $\widehat{\Delta MSE}(0.5, y)$ bzw. $\widehat{\Delta MSE}(1, y)$

Die verschiedenen Bandbreitenwahlen beeinflussen die MSE -Kurven nicht oder nur schwach, der $MISE$ ist besonders klein für die Wahl $b = n^{-1/4}$. Nachdem wir die recht einfache und in den Modellannahmen schon enthaltene Nebeninformation $g(\varepsilon, x) = \varepsilon$ betrachtet haben gehen wir zu einem weiteren Beispiel über und betrachten im Weiteren nur noch die Bandbreite $b = n^{-1/4}$.

- (b) (Beispielvariation 2) Es gelte obiges Normalverteilungsmodell mit Parameter $\vartheta = 3$ und es sei $n = 100$. Die Nebeninformation sei gegeben durch $g(\varepsilon, x) = \varepsilon^2 - (1 + 3x)$. $MSE(x, y)$ und $MISE(x)$ werden für $h = n^{-1/5}$ und $b = n^{-1/4}$ approximiert. Es ergeben sich die folgenden Graphen:

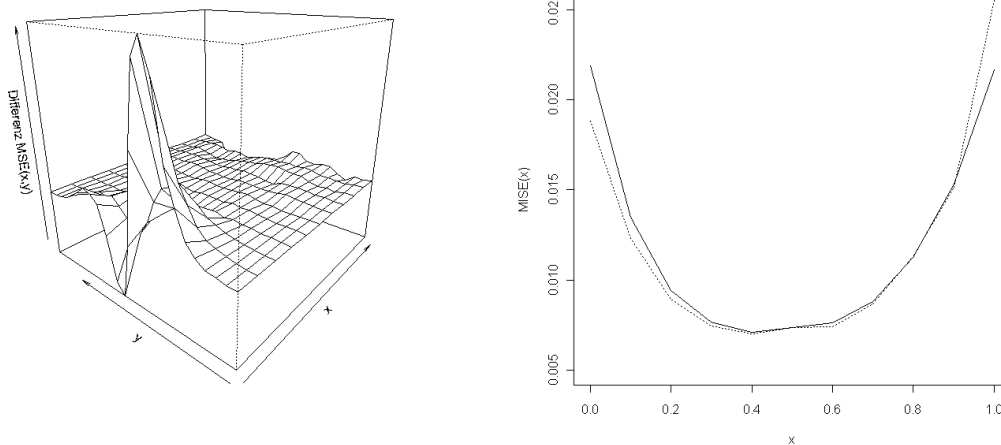


Abb.7: $\widehat{\Delta MSE}(x, y)$ bzw. $\widehat{MISE}(x)$

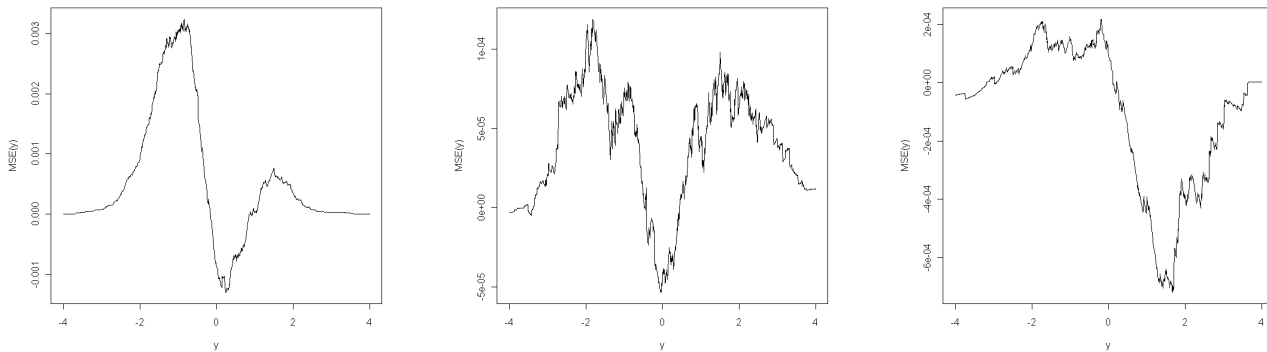


Abb.8: $\widehat{\Delta MSE}(0, y)$ bzw. $\widehat{\Delta MSE}(0.5, y)$ bzw. $\widehat{\Delta MSE}(1, y)$

wählt man nicht $n = 100$ sondern $n = 500$, so ergibt sich folgendes Bild:

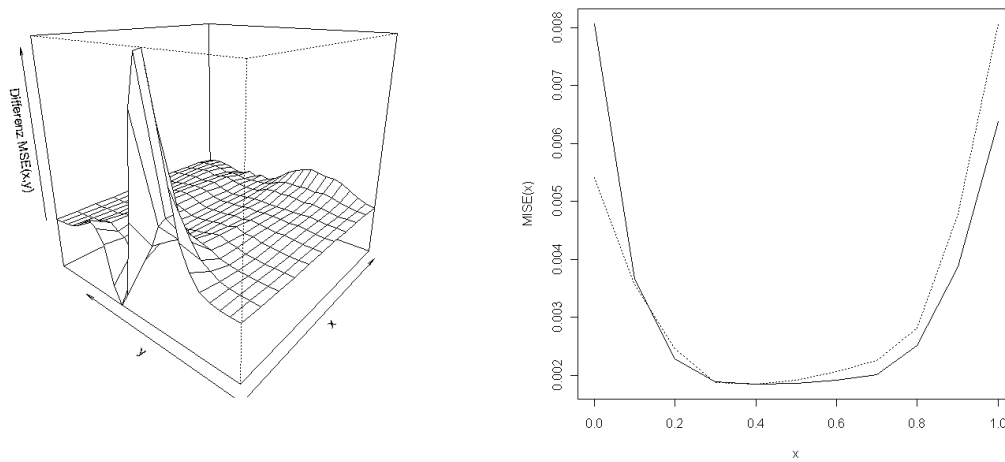


Abb.9: $\widehat{\Delta MSE}(x, y)$ bzw. $\widehat{MISE}(x)$

wählt man $n = 50$, so ergibt sich folgendes Bild:

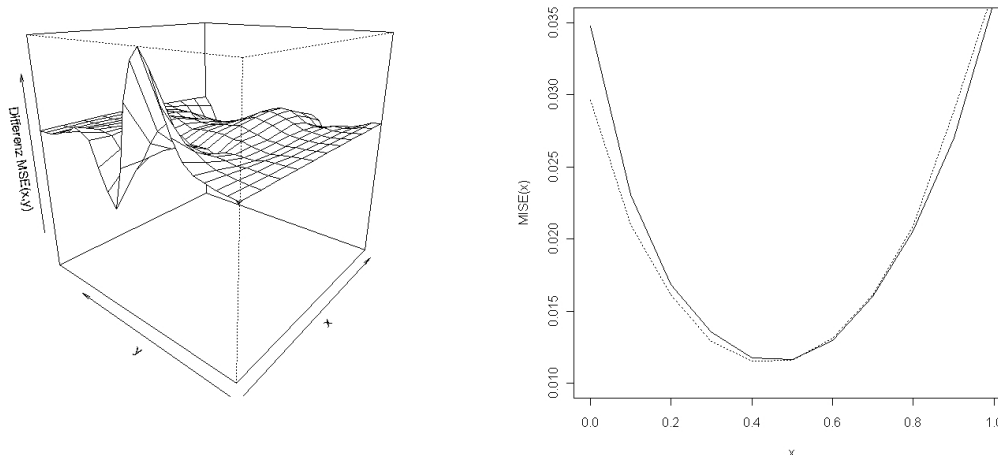


Abb.10: $\widehat{\Delta MSE}(x, y)$ bzw. $\widehat{MISE}(x)$

Betrachtet man die ΔMSE -Kurve zu zeigen sich Verbesserungen auch in diesem Beispiel, aber nicht gleichmäßig für alle x . Insbesondere für kleine Werte von x und y in $[-2, 0]$ gibt es eine Verbesserung, für x nahe 1 eine Verschlechterung. Des Weiteren ist zu bemerken, dass die Verbesserung für wachsenden Stichprobenumfang nachlässt. Der $MISE$ zeigt ebenfalls eine Verschlechterung für große Werte von x , eine Verbesserung für kleine Werte von x auf.

- (c) (Beispielvariation 3) Es gelte obiges Normalverteilungsmodell mit Parameter $\vartheta = -0.5$ und es sei $n = 100$. Die Nebeninformation sei gegeben durch $g(\varepsilon, x) = \varepsilon^2 - (1 - 0.5x)$, also einer Varianz der Art $\sigma^2(x) = 1 - 0.5x$, wie sie in Beispiel 2 des letzten Abschnittes diskutiert wurde. $MSE(x, y)$ und $MISE(x)$ werden für $h = n^{-1/5}$ und $b = n^{-1/4}$ approximiert. Es ergeben sich die folgenden Graphen für $\widehat{\Delta MSE}(x, y)$ und $\widehat{MISE}(x)$:

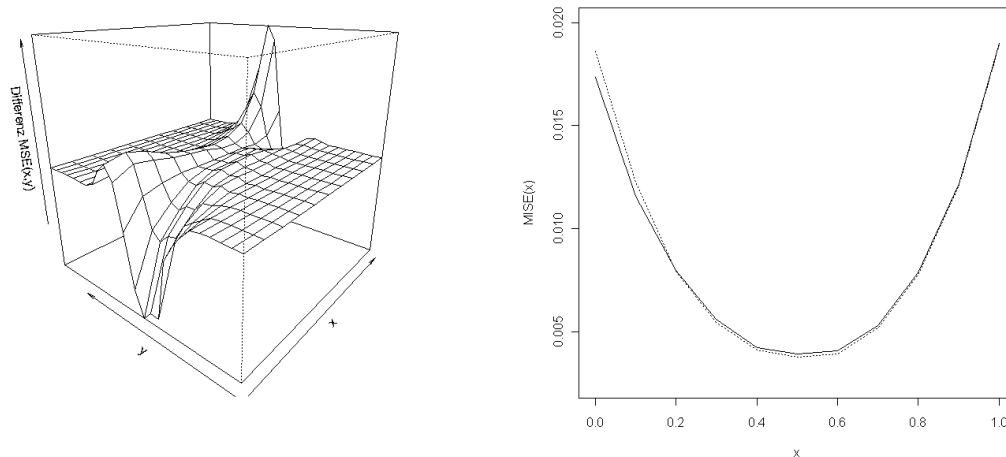


Abb.11: $\widehat{\Delta MSE}(x, y)$ bzw. $\widehat{MISE}(x)$

Vergleiche auch hier wieder die simulierte MSE -Differenz Abb. 11 mit der theoretischen MSE -Differenz aus dem letzten Kapitel: Weder für die $MISE$ - noch für die $\Delta - MSE$ -Kurven zeigen sich in diesem Beispiel gleichmäßige Verbesserungen.

Als vierte Variation betrachte nochmals eine Nebeninformation g hergeleitet aus bekannter Varianz:

- (d) (Beispielvariation 4) Für die Fehler ε gelte obiges Normalverteilungsmodell mit $\text{Var}(\varepsilon) = \exp(\vartheta x)$. Die Nebeninformation sei gegeben durch $g(\varepsilon, x) = \varepsilon^2 - (\exp(\vartheta x))$. $MSE(x, y)$ und $MISE(x)$ werden für $\vartheta = 1$, $h = n^{-1/5}$ und $b = n^{-1/4}$ approximiert. Es ergeben sich die folgenden Graphen, dabei beachte zu Abb. 13 die unterschiedlichen Maßstäbe:

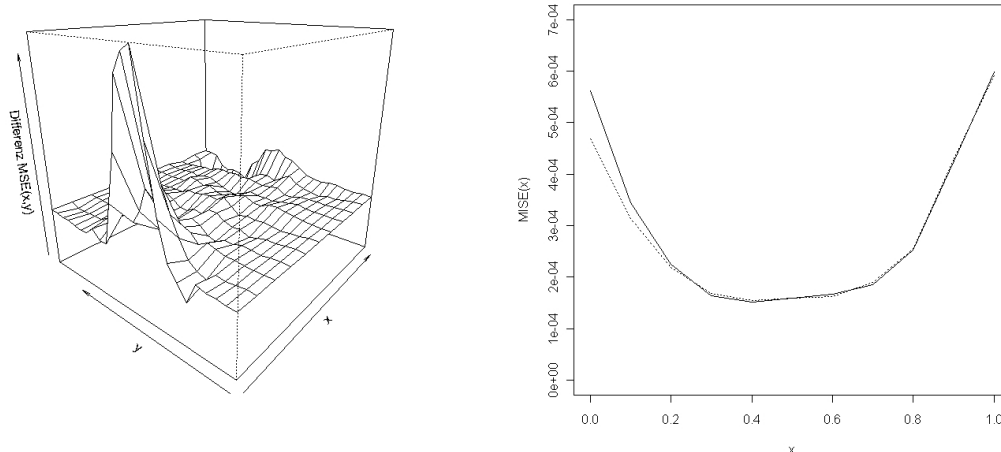


Abb.12: $\Delta \widehat{MSE}(x, y)$ bzw. $\widehat{MISE}(x)$

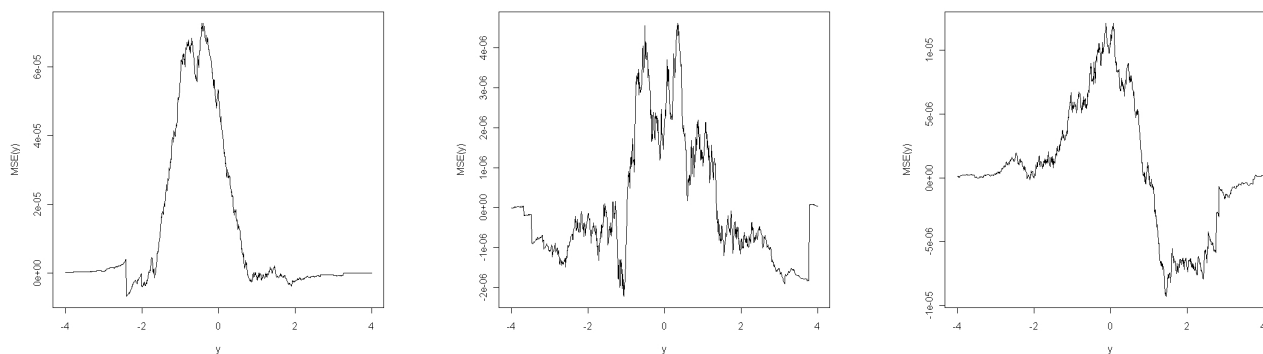


Abb.13: $\Delta \widehat{MSE}(0, y)$ bzw. $\Delta \widehat{MSE}(0.5, y)$ bzw. $\Delta \widehat{MSE}(1, y)$

Hier ergibt sich für den ELE im Vergleich zum KE im Hinblick auf den MSE eine Verschlechterung speziell für kleine Werte von x und y , für kleine Werte von x und große Werte von y ist eine (im Vergleich deutlich schwächere) Verbesserung zu beobachten. Für große Werte von x , also kleinen Varianzen ist ebenfalls eine (schwache) Verbesserung zu beobachten.

- (e) (Beispielvariation 5) Als weitere Variation des ersten Beispiels betrachte das Modell aus (a) und die zweidimensionale Nebeninformation $g(\varepsilon, x) = (\varepsilon, \varepsilon^2 - (1 + 3x))^T$. $MSE(x, y)$ und $MISE(x)$ werden für $h = n^{-1/5}$ und $b = n^{-1/4}$ approximiert. Es ergeben sich die folgenden Graphen:

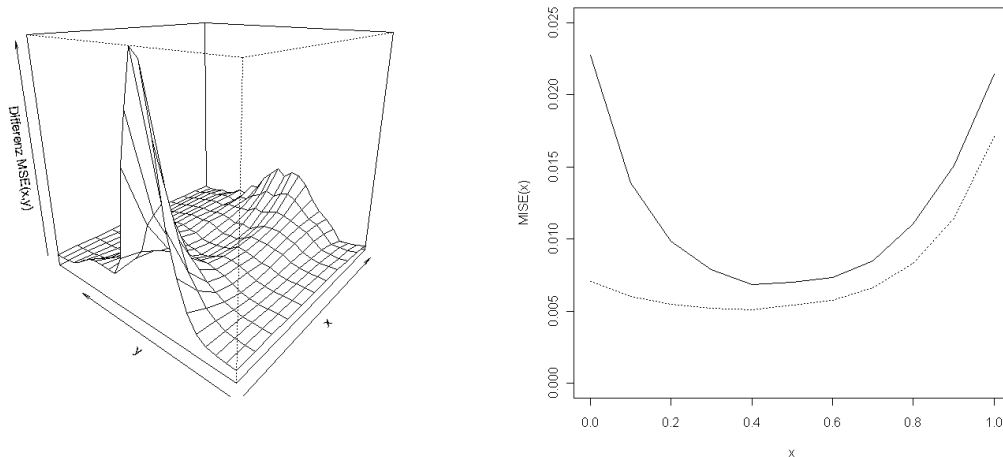


Abb.14: $\widehat{\Delta MSE}(x, y)$ bzw. $\widehat{MISE}(x)$

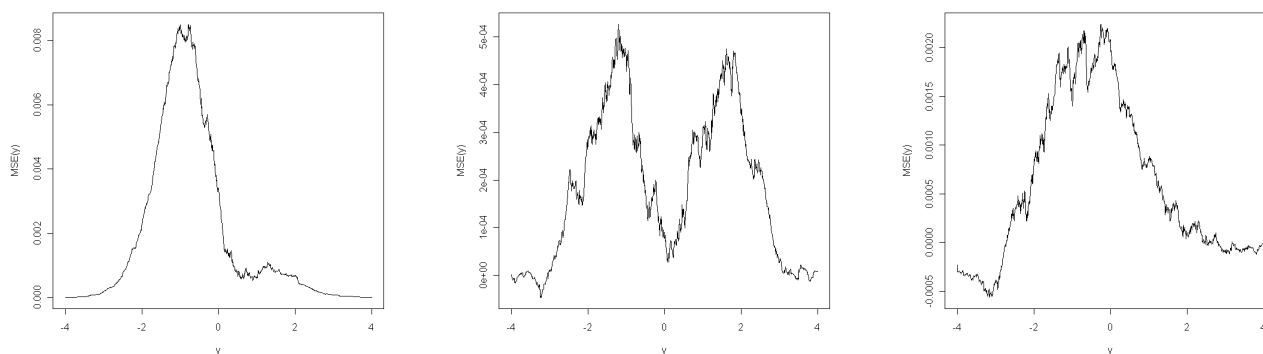


Abb.15: $\widehat{\Delta MSE}(0, y)$ bzw. $\widehat{\Delta MSE}(0.5, y)$ bzw. $\widehat{\Delta MSE}(1, y)$

Die in Beispiel (a) und (c) beobachteten Verbesserungen fügen sich hier zusammen.

- (f) (Beispielvariation 6) Als vorletzte Variation des ersten Beispielles betrachte das Modell aus (a) und die zweidimensionale Nebeninformation $g(\varepsilon, x) = (\varepsilon, I\{\varepsilon \leq 0\} - 1/2)^T$. Die Nebeninformationen sind sich sehr ähnlich, in diesem Modell sind Median und Erwartungswert identisch. $MSE(x, y)$ und $MISE(x)$ werden für $h = n^{-1/5}$ und $b = n^{-1/4}$ approximiert. Es ergeben sich die folgenden Graphen:

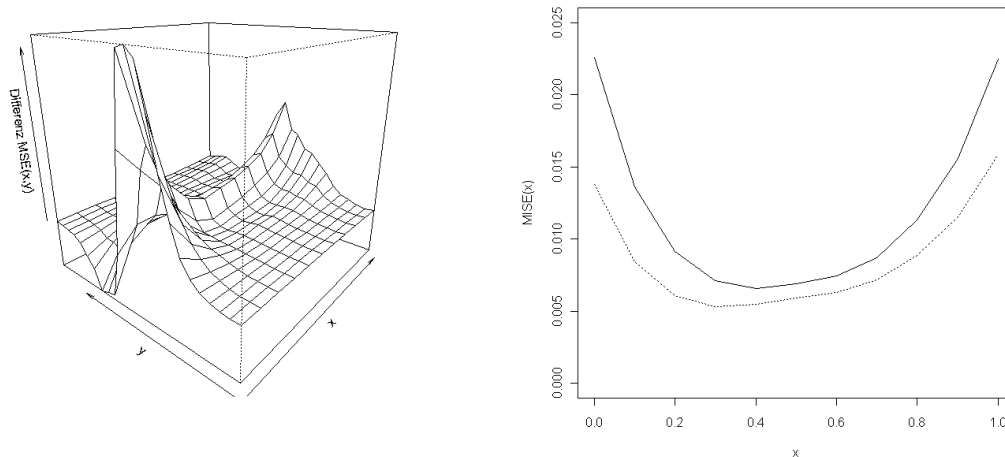


Abb.16: $\Delta \widehat{MSE}(x, y)$ bzw. $\widehat{MISE}(x)$

Zu erkennen ist wiederum eine Verbesserung, dabei wirkt sich der hinzu genommene Indikator gleichmäßig über alle x aus, die Momentebedingung größtenteils am Rand.

- (f) (Beispielvariation 7) Als letzte Variation des ersten Beispiels betrachte das Modell aus (a) und die dreidimensionale Nebeninformation

$$g(\varepsilon, x) = (\varepsilon, I\{\varepsilon \leq 0\} - 1/2, I\{\varepsilon \leq \sqrt{1 + 3x}\} - 0.84)^T.$$

Die ersten beiden Teile der Nebeninformation sind sich wiederum sehr ähnlich. $MSE(x, y)$ und $MISE(x)$ werden für $h = n^{-1/5}$ und $b = n^{-1/4}$ approximiert. Es ergeben sich die folgenden Graphen:

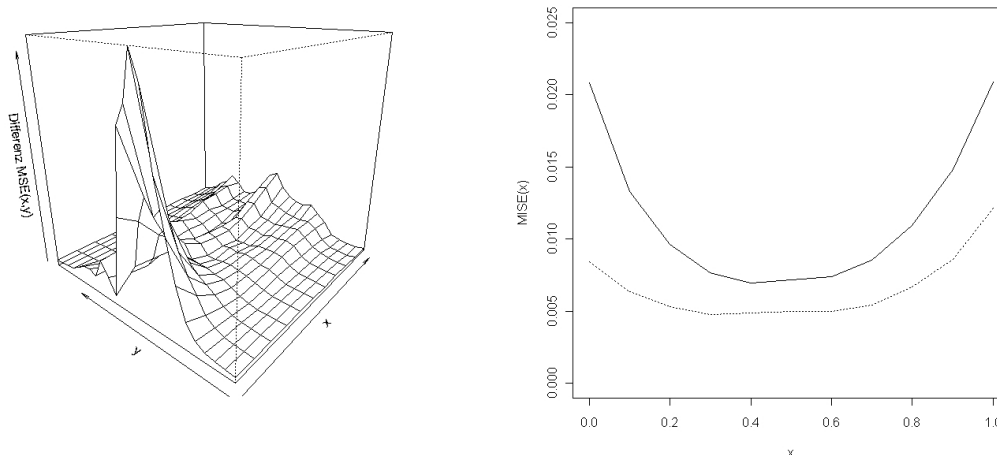


Abb.17: $\widehat{\Delta MSE}(x, y)$ bzw. $\widehat{MISE}(x)$

Bemerkung 8.1 (Zusammenfassung Beispiel 1) Es zeigt sich, dass für alle sieben Beispielvariationen für bestimmte $x \in [0, 1]$ eine Verbesserung des Empirical-Likelihood-Schätzers (ELE) im Vergleich zum Kernschätzer (KE) im Hinblick auf den MSE und MISE zu beobachten ist, speziell für x für welche $\text{Var}(\varepsilon|X = x) = \sigma^2(x)$ tendenziell nahe bei 1 liegt. Auch die Nebeninformation $E[\varepsilon] = 0$ verbessert den ELE im Vergleich zum KE.

In Beispiel (d) kommt es für große Varianzen zu einer Verschlechterung des ELE im Vergleich zum KE im Bezug auf den MSE, was der MISE aber nicht wiedergibt. Beispiel (b), Abb. 7 (a) – (d) zeigen deutlich, dass hier besonders für kleine Stichprobenumfänge der ELE einen Vorteil im Vergleich zum KE hat.

Beispiel (b) und (c) zeigen die Probleme der Schätzung auf. Obwohl die Variation von (b) nach (c) eher minimal ist, nur eine Variation in der Struktur der Nebeninformation darstellt, ergeben sich für den MSE deutliche Auswirkungen, was gegen den Nutzen des ELE spricht.

Nebeninformationen in Indikatorgestalt bringen in diesem Modell deutliche Verbesserungen. Der Vergleich der Beispiele (a) und (f) zeigt, dass die Nebeninformation $g(\varepsilon, x) = I\{\varepsilon \leq 0\} - 1/2$ noch Verbesserungen des ELE im Vergleich zum KE im Hinblick auf den MSE bringt, diese sind, anders als bei den Momentebedingungen, nicht nur für besonders kleine bzw. große Werte von x zu beobachten, sondern gleichmäßig über alle x Werte gegeben. Die MISE-Kurven verändern sich nicht so deutlich. Ebenso zeigen die letzten Beispiele (e) und (f) die Verbesserungsmöglichkeiten auf, welche aus Verknüpfung von Momenten- und Quantilsnebenbedingungen auch in Modellen entstehen können, welche Quantilsbedingungen aus den Momenten folgern lassen. Dabei ist noch einmal zu bemerken, dass speziell die Quantil- y_0 -Nebenbedingungen für eine

Verbesserung des MSE an den Stellen $MSE(y_0, x)$, $x \in [0, 1]$ erreichen, die Momentebedingungen zu Verbesserungen von $MSE(y_0, x)$ hauptsächlich an den Rändern, d.h. für x nahe 1 bzw. nahe 0 führen.

8.5 Simulationsbeispiel 2 (schiefe Verteilung)

X sei gleichverteilt auf $[0, 1]$ und ε gegeben $X = x$ doppel exponentialverteilt mit Erwartungswert 0 und Varianz $\sigma^2(x)$, die Abhängigkeit wird später noch genauer angegeben. Die Dichte einer doppel exponentialverteilten Zufallsgrößen U hat die Form

$$f_U(u) = \frac{1}{p} \exp\left(-\frac{u-q}{p}\right) \cdot \exp\left(-\exp\left(-\frac{u-q}{p}\right)\right), \quad u \in \mathbb{R}$$

mit $E[U] = q + p \cdot 0.577216$.

Um wie vorausgesetzt Erwartungswert gleich Null zu erhalten muss gelten: $q = -p \cdot 0.577216$. Mit einigen Rechnungen kann man zeigen, dass $\text{Var}(U) = E[U^2] = \frac{p^2 \pi^2}{6}$, so dass über p die X -Abhängigkeit festgelegt werden kann. Die Dichte von ε gegeben $X = x$ sei in diesem speziellen Modell damit gegeben durch:

$$f_{\varepsilon|X}(u|x) = \frac{1}{p(x)} \exp\left(-\frac{u + p(x) \cdot 0.577216}{p(x)}\right) \cdot \exp\left(-\exp\left(-\frac{u + p(x) \cdot 0.577216}{p(x)}\right)\right).$$

Für die Erstellung der Schätzer $\hat{F}_n^*(y|x)$ und $\hat{F}_n(y|x)$ wurde mit obigen Verteilungen ein Datensatz $(X_1, Y_1), \dots, (X_n, Y_n)$ erzeugt und mittels des Nadaraya-Watson-Schätzers (vgl. Definition 2.2) und der Bandbreite $b = n^{-1/4}$ $(X_1, \hat{\varepsilon}_1), \dots, (X_n, \hat{\varepsilon}_n)$ gebildet.

Ein Beispieldatensatz $(X_1, Y_1), \dots, (X_n, Y_n)$ bzw. $(X_1, \hat{\varepsilon}_1), \dots, (X_n, \hat{\varepsilon}_n)$ mit $p(x) = 1 + 3x$ hat die folgende Gestalt:

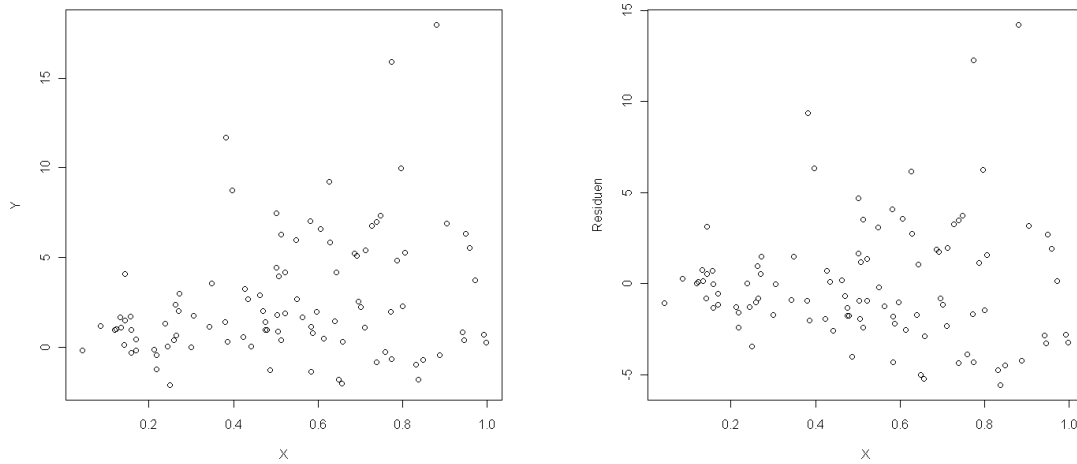


Abb.18: Beispieldatensatz

Die mit wachsendem x zunehmende Streuung ist deutlich zu erkennen.

Es werden nun wieder die Schätzer $\widehat{MSE}(x, y)$ und $\widehat{MISE}(x)$ für $MSE(x, y)$ und $MISE(x)$ verglichen, dabei betrachten wir verschiedene Nebeninformationen $g(\varepsilon, x) = \varepsilon^2 - \frac{p^2(x)\pi^2}{6}$, mit $n = 100$, die Anzahl der wiederholten Datenerzeugungen D gleich 250, $b = n^{-1/4}$ und $h = n^{-1/5}$:

- (a) Mit $p(x) = 1 + 3x$, also $g(\varepsilon, x) = \varepsilon^2 - \frac{(1+3x)^2\pi^2}{6}$ und $x = 0.8$ ergeben sich (beispielhaft) die Schätzkurven:

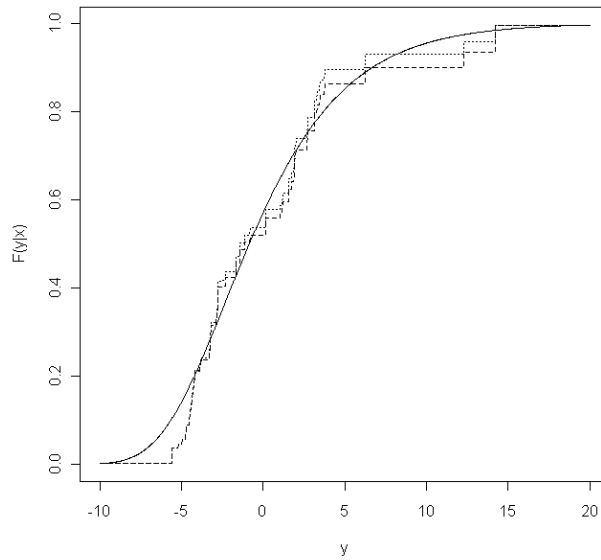


Abb.19: Schätzkurven für $x = 0.8$

Dabei ist $\hat{F}_n^*(y|x)$ die gestrichelte, $\hat{F}_n(y|x)$ die gepunktete und $F(y|x)$ die durchgezogene Kurve. Es ergeben sich die folgenden Graphen für $\Delta \widehat{MSE}(x, y)$ und $\widehat{MISE}(x)$:

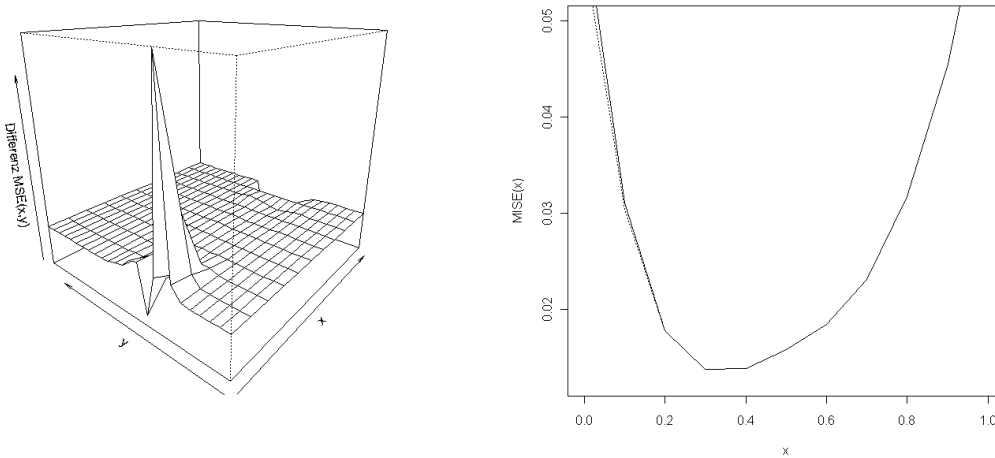


Abb.20: $\Delta \widehat{MSE}(x, y)$ bzw. $\widehat{MISE}(x)$

(b) Mit $p(x) = 1 + 3 \exp(-x)$, also $g(\varepsilon, x) = \varepsilon^2 - \frac{(1+3 \exp(-x))^2 \pi^2}{6}$ ergeben sich die folgenden Graphen für $\Delta MSE(x, y)$ und $MISE(x)$:

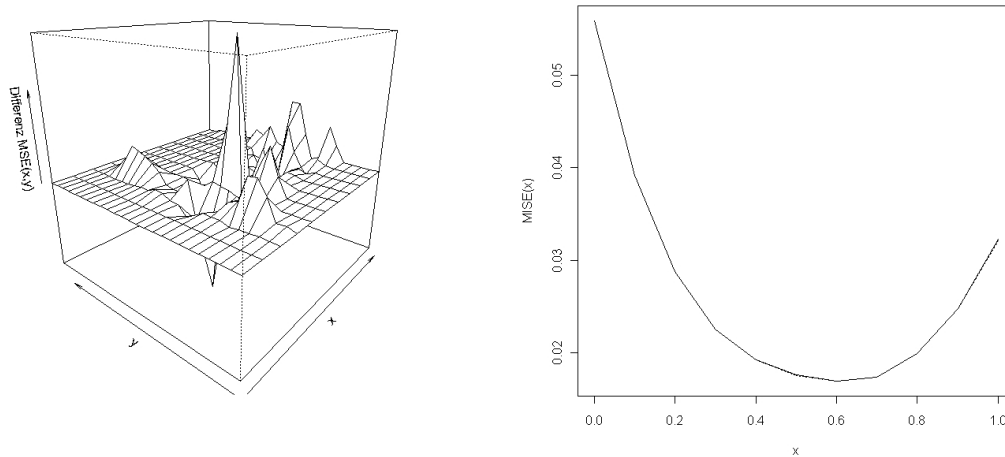


Abb.21: $\Delta \widehat{MSE}(x, y)$ bzw. $\widehat{MISE}(x)$

Bemerkung 8.2 (Zusammenfassung Beispiel 2) Es zeigt sich in Beispiel (a), dass auch hier wieder für kleine Werte von x , also kleine Varianzen eine Verbesserung des ELE im Vergleich zum KE im Bezug auf den MSE am Rand auftritt. In Beispiel (b) ist ein solches Muster nicht zu beobachten, die \widehat{MSE} -Unterschiede sind minimal (von der Größenordnung 10^{-5}) und die \widehat{MISE} -Kurven sind nahezu identisch.

8.6 Simulationsbeispiel 3 (Cauchy-verteilte Fehler)

X sei normalverteilt und ε gegeben $X = x$ Cauchy-verteilt mit Dichte $f_{\varepsilon|X}(u|x) = \frac{1}{\pi} \cdot \frac{1+3x}{(1+3x)^2+u^2}$, dabei ist $1 + 3x$ der Breitenparameter der Cauchyverteilung und beschreibe gleichzeitig die Abhängigkeit zwischen X und ε . Der Median ist 0. Es existieren weder der Erwartungswert noch andere Momente, die Voraussetzungen dieser Arbeit sind somit größtenteils nicht erfüllt. Diese Abhängigkeit, gegeben über den Breitenparameter, kann man in Form der Nebeninformation $g(\varepsilon, x) = I\{\varepsilon \leq 1 + 3x\} - 3/4$ schreiben. Dazu beachte, dass die Verteilungsfunktion der Residuen gegeben $X = x$ die folgende Gestalt

$$F_{\varepsilon|X}(u|x) = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{u}{1 + 3x}\right)$$

besitzt und damit folgt

$$\begin{aligned} E[g(\varepsilon, x) | X = x] &= E[I\{\varepsilon \leq 1 + 3x\} - 3/4 | X = x] \\ &= F_{\varepsilon|X}(1 + 3x|x) - 3/4 = \frac{1}{\pi} \arctan(1) - \frac{1}{4} = 0 \end{aligned}$$

Für die Erstellung der Schätzer $\hat{F}_n^*(y|x)$ und $\hat{F}_n(y|x)$ wurde mit obigen Verteilungen ein Datensatz $(X_1, Y_1), \dots, (X_n, Y_n)$ erzeugt und mittels des Nadaraya-Watson-Schätzers (vgl. Definition 2.2) und der Bandbreite $b = n^{-1/4}$ $(X_1, \hat{\varepsilon}_1), \dots, (X_n, \hat{\varepsilon}_n)$ gebildet. Ein Beispieldatensatz $(X_1, Y_1), \dots, (X_n, Y_n)$ bzw. $(X_1, \hat{\varepsilon}_1), \dots, (X_n, \hat{\varepsilon}_n)$ hat die Gestalt

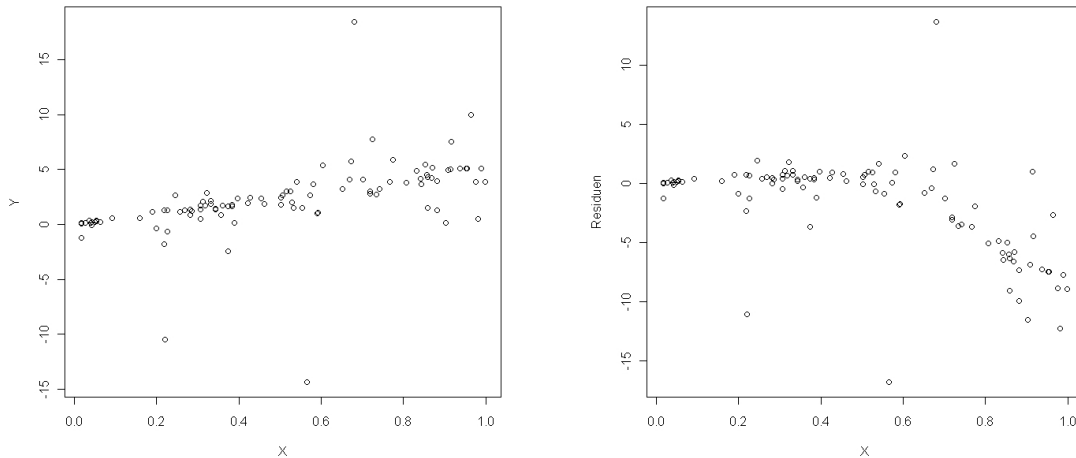


Abb.22: Beispieldatensatz

Cauchy-verteilte Zufallsvariablen haben die Eigenschaft häufiger größere Ausreißer aufzuweisen, vgl. den Beispieldatensatz Abb. 22, was auch die Schätzung der Modellfunktion m und damit die geschätzten Fehler sehr stark beeinflusst. Dies kann in obigem Datensatz sehr gut beobachtet werden, hier verzerrt eine große Beobachtung $(y, x) \approx (158, 0.9)$ die Schätzkurve so stark, dass die geschätzten Fehler für große Werte von x hauptsächlich ins Negative tendieren.

Es werden nun wieder $MSE(x, y)$ und $MISE(x)$ der beiden Schätzer verglichen, dabei ist $n = 100$, die Anzahl der wiederholten Datenerzeugungen D gleich 250, $b = n^{-1/4}$, $h = n^{-1/5}$ und die die Nebeninformation beschreibende Funktion g von der Gestalt $g(\varepsilon, x) = I\{\varepsilon \leq 1 + 3x\} - 3/4$, wie oben schon beschrieben.

Für $x = 0.8$ ergeben sich (beispielhaft) für einen Datensatz die Schätzkurven:

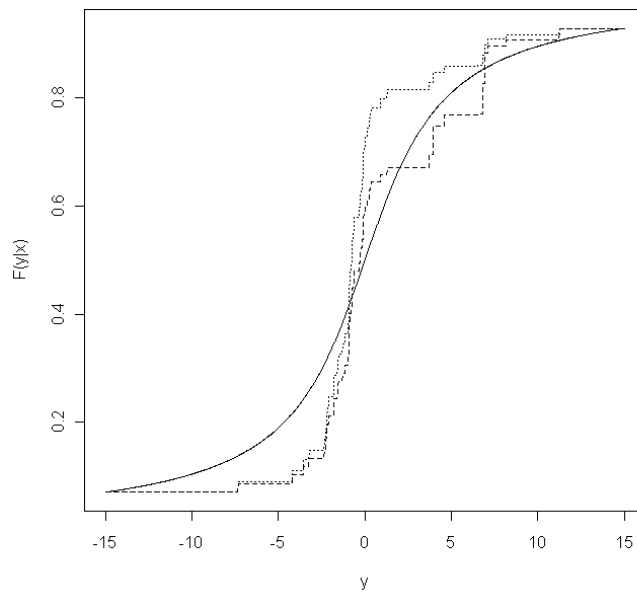


Abb.23: Schätzkurven für $x = 0.8$

Dabei ist $\hat{F}_n^*(y|x)$ die gestrichelte, $\hat{F}_n(y|x)$ die gepunktete und $F(y|x)$ die durchgezogene Kurve. Es ergeben sich die folgenden Graphen für $\widehat{\Delta MSE}(x, y)$ und $\widehat{MISE}(x)$:

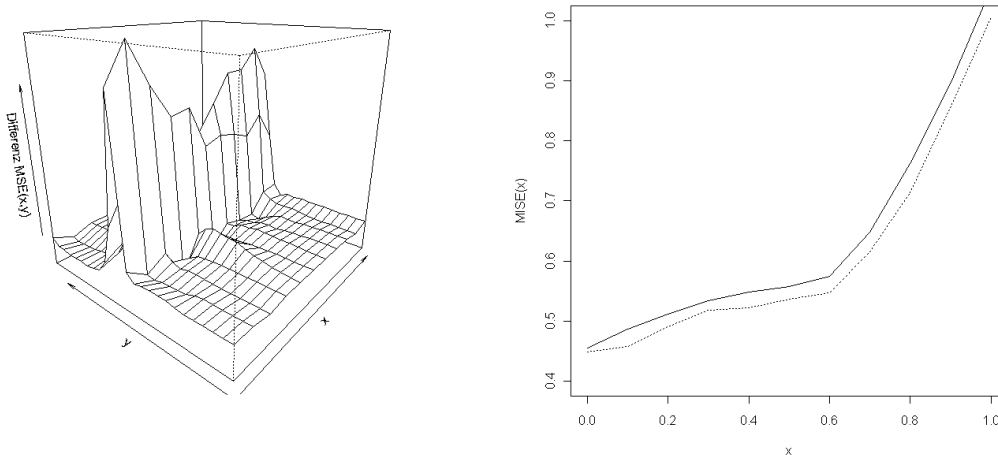


Abb.24: $\Delta \widehat{MSE}(x, y)$ bzw. $\widehat{MISE}(x)$

Bemerkung 8.3 (Zusammenfassung Beispiel 3) Es zeigt sich, dass es in der Umgebung $y = \frac{3}{4}$ eine deutliche Verbesserung des ELE im Vergleich zum KE im Hinblick auf den MSE gibt, ebenfalls liegt die MISE Kurve des ELE unterhalb der des KE. Es ist aber zu bemerken, dass bei wachsendem x die Fehlerverteilung sehr große Ausreißer generiert, wodurch die Schätzung des MISE sehr stark verzerrt wird (vgl. den Maßstab in obiger MISE-Grafik) und die Schätzung damit im Prinzip unbrauchbar ist.

8.7 Simulationsbeispiel 4 (geschätzte Nebenbedingungen)

Als vorletztes Beispiel wollen wir nochmals auf Beispiel 1 mit Nebenbedingung $g(\varepsilon, x) = \varepsilon^2 - (1 + \vartheta x)$ eingehen und es in soweit modifizieren, dass die Struktur der Nebenbedingung $g(\varepsilon, x) = \varepsilon^2 - (1 + \vartheta x)$ bekannt ist, aber nicht der exakte Wert von ϑ . Diesen ersetzen wir durch einen geeigneten Schätzer. Dazu nutzen wir, dass gilt $Var(\varepsilon) = E[\varepsilon^2]$ und bestimmen ϑ mithilfe der Beobachtungen $(\hat{\varepsilon}_i^2, X_i)$ und der linearen Regression $\varepsilon_i^2 = 1 + \vartheta \cdot X_i$. Wir nutzen den geschätzten Parameter $\hat{\vartheta}$ in der in den Schätzer $\hat{F}_n^*(y|x)$ eingehenden Nebeninformation.

Es ergeben sich die folgenden Graphen für $\Delta \widehat{MSE}(x, y)$ und $\widehat{MISE}(x)$:

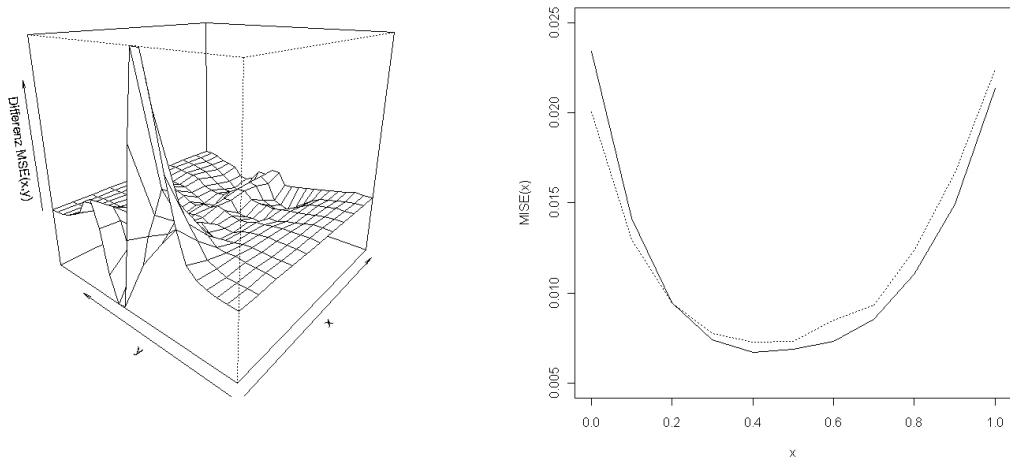


Abb.25: $\widehat{\Delta MSE}(x, y)$ bzw. $\widehat{MISE}(x)$

Erhöht sich der Stichprobenumfang von $n = 100$ auf $n = 500$ mit ebenfalls $D = 250$ Wiederholungen, so ergeben sich die folgenden Graphen für $\widehat{\Delta MSE}(x, y)$ und $\widehat{MISE}(x)$:

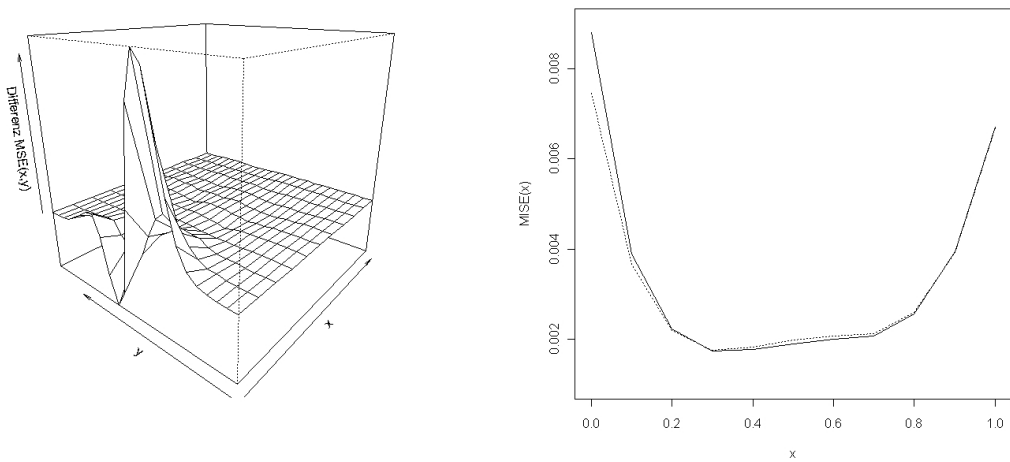


Abb.26: $\widehat{\Delta MSE}(x, y)$ bzw. $\widehat{MISE}(x)$

Verringert sich der Stichprobenumfang von $n = 100$ auf $n = 50$ mit ebenfalls $D = 250$ Wiederholungen, so ergeben sich die folgenden Graphen für $\widehat{\Delta MSE}(x, y)$ und $\widehat{MISE}(x)$:

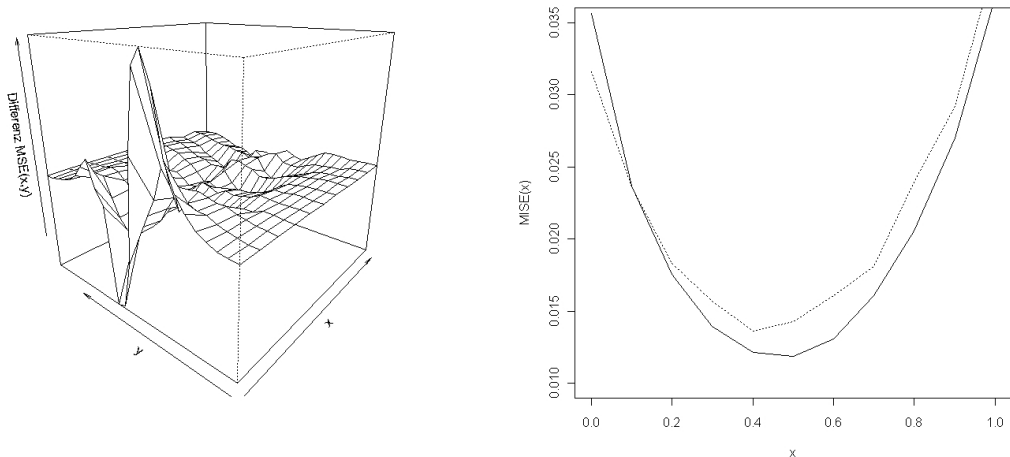


Abb.27: $\widehat{\Delta MISE}(x, y)$ bzw. $\widehat{MISE}(x)$

Bemerkung 8.4 (Zusammenfassung Beispiel 4) Es zeigt sich, dass in diesem Beispiel für alle drei Wahlen von n , ebenso wie in Beispiel 1, im Bereich kleiner Werte von x eine Verbesserung des ELE im Vergleich zum KE im Hinblick auf den MSE auftritt. Gleichmäßige Verbesserungen sind nicht zu beobachten. Für den MISE zeigt sich die Verbesserung besonders im Falle kleiner x -Werte, für größere Werte von x ergibt sich wiederum eine Verschlechterung. Allgemein ist aber zu bemerken, dass die obigen Kurven denen generiert unter der Annahme bekannter Nebeninformationen stark ähneln.

8.8 Simulationsbeispiel 5 (skew normal distribution)

Im letzten Beispiel wollen wir eine weitere schiefe Verteilung betrachten. Bevor wir mit diesem Beispiel beginnen definieren wir die genutzte schiefe Normalverteilung. Eine Zufallsvariable U heißt skew-Normalverteilt, falls sie die Dichte $f_U(u) = 2\phi_{\sigma^2}(u)\Phi(\alpha u)$ mit ϕ_{σ^2} die Dichte der Normalverteilung mit Erwartungswert Null und Varianz σ^2 , Φ die Verteilungsfunktion der standard-Normalverteilung, besitzt. Wir schreiben $U \sim SN(\sigma^2, \alpha)$, vergleiche zur skew-Normalverteilung auch Azzalini und Capitanò (1999). Es sei $\delta := \frac{\sigma^2 \alpha}{\sqrt{1 + \alpha^2 \sigma^2}}$. Dann ist $E[U] = \sqrt{\frac{2}{\pi}} \cdot \delta$ und $\text{Var}(U) = \sigma^2 - (\sqrt{\frac{2}{\pi}} \cdot \delta)^2$. Betrachtet man Zufallsvariablen U_1, \dots, U_n skew-normalverteilt, so kann man daraus einfach zentrierte Zufallsvariablen bilden: $(\varepsilon_1, \dots, \varepsilon_n) := (U_1 - E[U], \dots, U_n - E[U])$. Auf diese Weise konstruierte Zufallsvariablen sollen Grundlage des nachstehenden Beispiels sein.

Zu Beispiel 5:

X sei gleichverteilt auf $[0, 1]$ und ε gegeben $X = x$ sei skew-normalverteilt mit Dichte nach obiger Konstruktion und Parametern $\sigma^2(x) = 1 + 3x$ und $\alpha = 1$. ε sei zusätzlich durch Abzug des jeweiligen (von x abhängenden) Erwartungswertes zentriert. ε besitzt die Varianz

$$\text{Var}(\varepsilon|X = x) = (1 + 3x) - \frac{2(1 + 3x)^2}{\pi(2 + 3x)}.$$

Damit hat die Nebeninformation g die Gestalt

$$g(\varepsilon, x) = \varepsilon^2 - (1 + 3x) + \frac{2(1 + 3x)^2}{\pi(2 + 3x)},$$

eben diesen Zusammenhang zwischen ε und X werden wir betrachten.

Für die Erstellung der Schätzer $\hat{F}_n^*(y|x)$ und $\hat{F}_n(y|x)$ wurde mit obigen Verteilungen ein Datensatz $(X_1, Y_1), \dots, (X_n, Y_n)$ erzeugt und mittels des Nadaraya-Watson-Schätzers (vgl. Definition 2.2) und der Bandbreite $b = n^{-1/4}$ die Stichprobe $(X_1, \hat{\varepsilon}_1), \dots, (X_n, \hat{\varepsilon}_n)$ gebildet. Ein Beispieldatensatz $(X_1, Y_1), \dots, (X_n, Y_n)$ bzw. $(X_1, \hat{\varepsilon}_1), \dots, (X_n, \hat{\varepsilon}_n)$ mit $n = 100$ hat die Gestalt

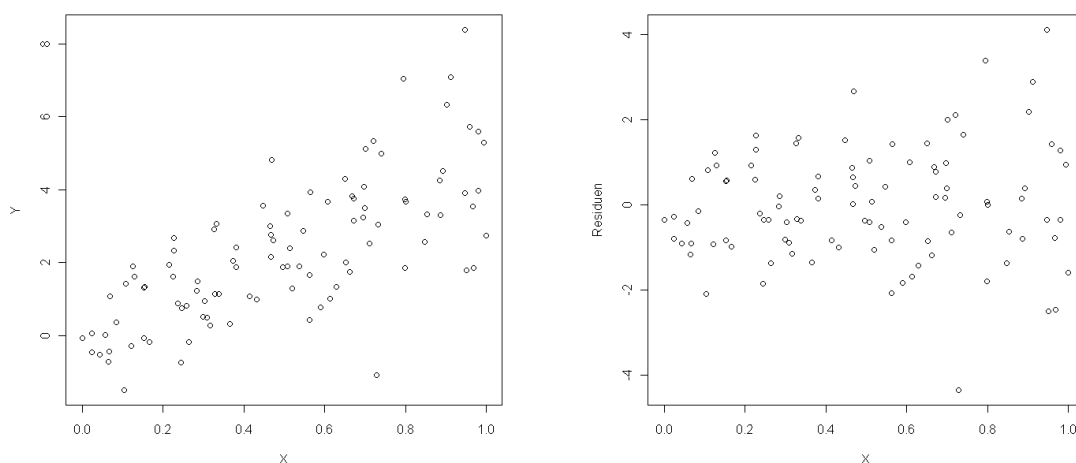


Abb.28: Beispieldatensatz

Dabei ist $\hat{F}_n^*(y|x)$ die gestrichelte, $\hat{F}_n(y|x)$ die gepunktete und $F(y|x)$ die durchgezogene Kurve. Es ergeben sich die folgenden Graphen für $\widehat{\Delta MSE}(x, y)$ und $\widehat{MISE}(x)$:

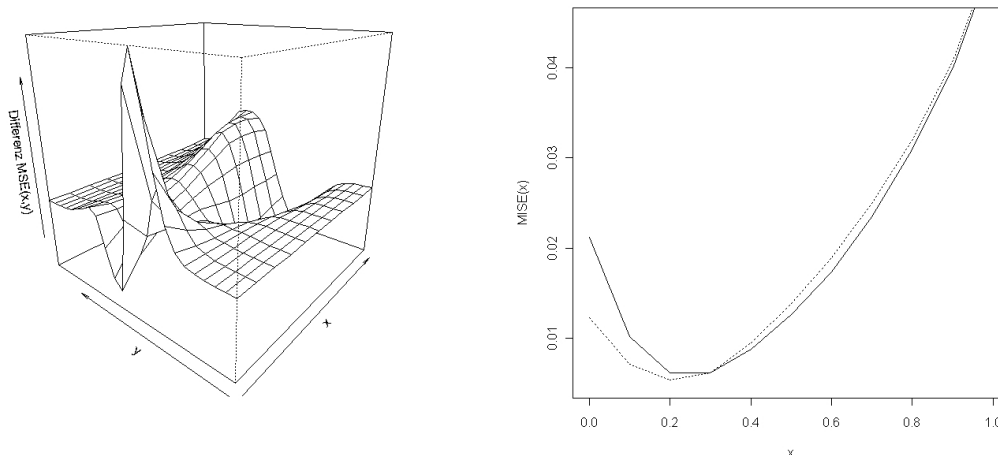


Abb.29: $\widehat{\Delta MSE}(x, y)$ bzw. $\widehat{MISE}(x)$

Bemerkung 8.5 (Zusammenfassung Beispiel 5) Es zeigt sich keine eindeutige Verbesserung. Für große bzw. kleine Werte von x ist für bestimmte Werte y eine Verbesserung, für andere eine Verschlechterung des MSE zu bemerken. Für den MISE ergibt sich eine Verbesserung nur hinsichtlich kleiner Werte von x .

8.9 Zusammenfassung der Simulationsergebnisse

Bemerkung 8.6 Zusammenfassend ist zu bemerken, dass sich in nahezu allen Beispielen eine Verbesserung des ELE im Vergleich zum KE aufgezeigt hat, speziell im Vergleich des MSE, welche aber meist nicht gleichmäßig für alle x zu beobachten waren. Im Falle von Nebenbedingungen in Form von Momenten der Verteilung zeigte sich eine Verbesserung meist an den Rändern von $[0, 1]$, bei Quantilsbedingungen gleichmäßig in x für das spezielle Quantil. Beispiel 1 (c) zeigt zusätzlich noch, dass die Verbesserung besonders für kleine Stichprobenumfänge gilt. Beispiel 1 zeigt sehr gut den Nutzen von Quantils-Nebenbedingungen.

Auch geschätzte Nebeninformationen und bekannte Nebeninformationen lieferten ähnliche Resultate, vgl. Beispiel 1(b) und 4. Insgesamt ist damit die Nutzung eines ELE für den Fall gegebener Quantils-Informationen zu empfehlen, für andere Nebeninformationen kann der resultierende MSE, auch für kleine Schwankungen in der Nebeninformation, (vgl. z.B. Bsp. 1 (b) und Bsp. 1 (c)) sehr unterschiedlich ausfallen, was deutlich gegen den Nutzen des ELE in diesem Fall spricht.

8.10 Informationen zum Simulationsprogramm

Die Simulationen wurden mithilfe der freien Statistik-Software *R* Version 2.9.1 durchgeführt. Zusätzlich zur Standardinstallation wurden die Packages *stats* (R Development Core Team (2009)) für den Nadaraya-Watson-Schätzer, *MASS* (W. N. Venables & B. D. Ripley (2002)), *mtvnorm* (A. Genz, F. Bretz and T. Hothorn (2009)) und *sn* (A. Azzalini (2009)) für die skew normal distribution genutzt.

9 Vorhersageintervalle und Testverfahren

Dieser Abschnitt soll einen kurzen Einblick in den weiteren Nutzen der vorgestellten Verteilungsfunktionsschätzer geben, ohne dabei die Theorie exakt auszuführen, meist wären wieder umfangreiche Entwicklungen nötig.

9.1 Vorhersageintervalle

In diesem Abschnitt wollen wir uns in Simulationen Konfidenzintervallen für Y gegeben $X = x$ zuwenden. Ein asymptotisches Vorhersageintervall für Y gegeben $X = x$ mit asymptotischem Niveau $(1 - \alpha)$ hat die Form

$$[\hat{m}(x) - \hat{F}_{\varepsilon|X}^{-1}(\alpha/2|x), \hat{m}(x) + \hat{F}_{\varepsilon|X}^{-1}(1 - \alpha/2|x)]$$

Es gilt $\hat{F}_{\varepsilon|X}^{-1}(\alpha|x) := \inf\{t | \hat{F}_{\varepsilon|X}(t|x) \geq \alpha\}$. Wir betrachten wiederum die beiden in dieser Arbeit im Mittelpunkt stehenden Schätzer $\hat{F}_n^*(y|x)$ (ELE) und $\hat{F}_n(y|x)$ (KE) der Verteilung $F_{\varepsilon|X}(y|x)$ (siehe Definition 4.2 und 4.6).

Unter recht allgemeinen Voraussetzungen kann die Konvergenzrate der schwachen Konvergenzaussage von $\hat{F}(x) - F(x)$ (geeignet standardisiert) übertragen werden auf diejenige von $\hat{F}^{-1}(x) - F^{-1}(x)$, vergleiche dazu van der Vaart (1996), S. 304 ff., genauer wird dieser Abschnitt nicht darauf eingehen.

Gezeigt werden soll exemplarisch in Simulationen, dass Konfidenzintervalle auf Basis des ELE zwar i. A. nicht kleiner sind als die mittels des KE erstellten, aber, bedingt durch die abweichende Lage, das $(1 - \alpha)$ -Niveau genauer einhalten. Dazu mitteln wir über $D = 1000$ Durchläufe.

Betrachtet wird ein Beispiel aus dem vorherigen Kapitel. Dazu wird $D = 1000$ Mal ein Beispieldatensatz $(X_1, Y_1), \dots, (X_n, Y_n)$ erzeugt und jeweils für beide Schätzer synchron Konfidenzintervalle für eine neue Beobachtung Y gegeben $X = x$, $x = 0, 0.1, 0.2, \dots, 1$ berechnet, eine solche Beobachtung (y, x) erzeugt und gezählt wie häufig die Beobachtung y im Konfidenzintervall liegt. Gemittelt über alle Durchläufe D wird zusätzlich der Durchmesser des Vorhersageintervalles angegeben. Der Stichprobenumfang n ist als 100 gewählt und es gilt $\alpha = 0.1$.

Es bezeichne im Weiteren

$$S_n^*(x) = [\hat{m}(x) - (\hat{F}_{n,\varepsilon|X}^*)^{-1}(\alpha/2|x), \hat{m}(x) + (\hat{F}_{n,\varepsilon|X}^*)^{-1}(1 - \alpha/2|x)]$$

und

$$S_n(x) = [\hat{m}(x) - (\hat{F}_{n,\varepsilon|X})^{-1}(\alpha/2|x), \hat{m}(x) + (\hat{F}_{n,\varepsilon|X})^{-1}(1 - \alpha/2|x)]$$

das Konfidenzintervall auf Basis des ELE bzw. des KE.

Beispiel

- (a) Es gelte das Normalverteilungsmodell aus Beispiel 1(a) des vorherigen Kapitels mit $n = 100$ (siehe Seite 146). Die Nebeninformation ist hier gegeben durch $g(\varepsilon, x) = \varepsilon$. Die Schätzer $\hat{F}_{\varepsilon|X}^{-1}(\alpha/2|x)$ und damit die Konfidenzintervalle werden für $h = c \cdot n^{-1/5}$ und $b = n^{-1/4}$ berechnet, dabei betrachte $c = 1, 0.5, 2$. Es ergeben sich die folgenden Graphen zur Breite der Konfidenzintervalle und zum Niveau, dabei bezieht sich die gestrichelte Kurve auf $S_n^*(x)$ und die durchgezogene Kurve auf $S_n(x)$:

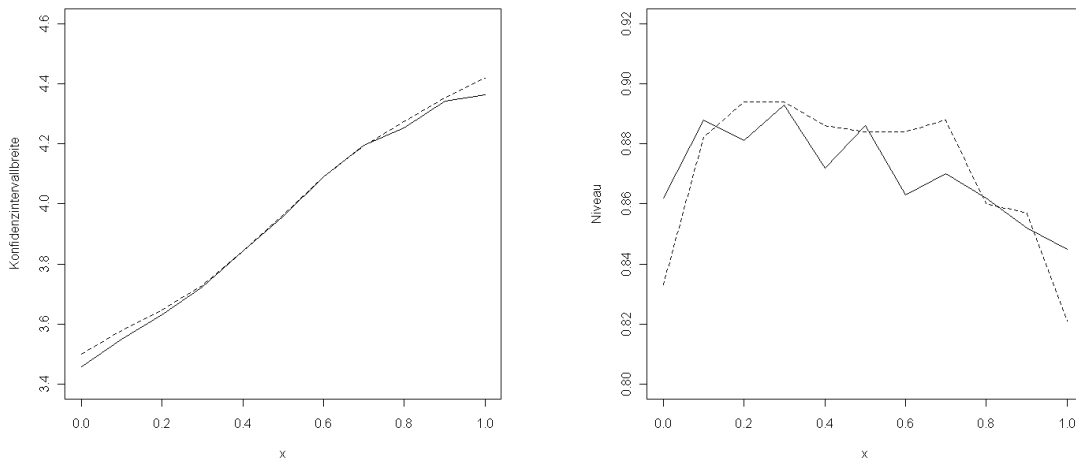


Abb.27a: Konfidenzintervallbreite und Niveaueurve mit $c = 1$

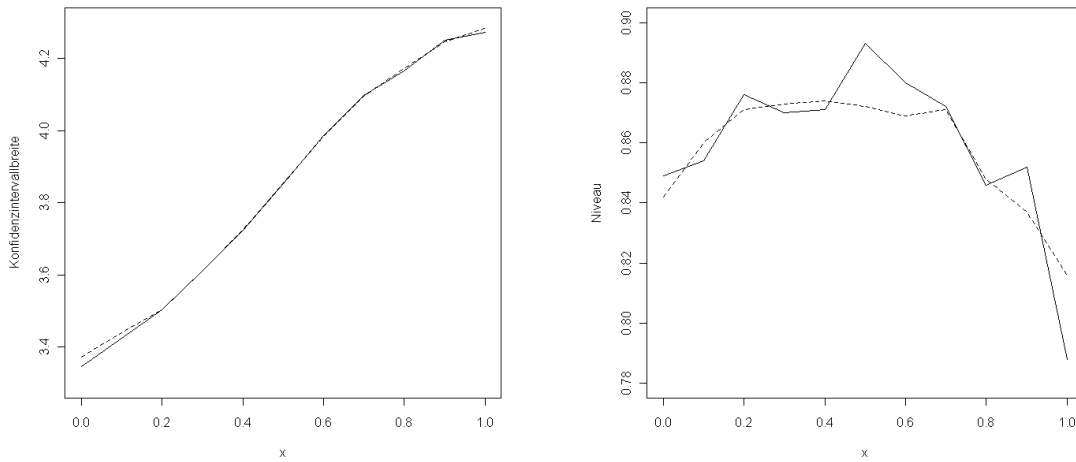


Abb.27b: Konfidenzintervallbreite und Niveaueurve mit $c = 0.5$

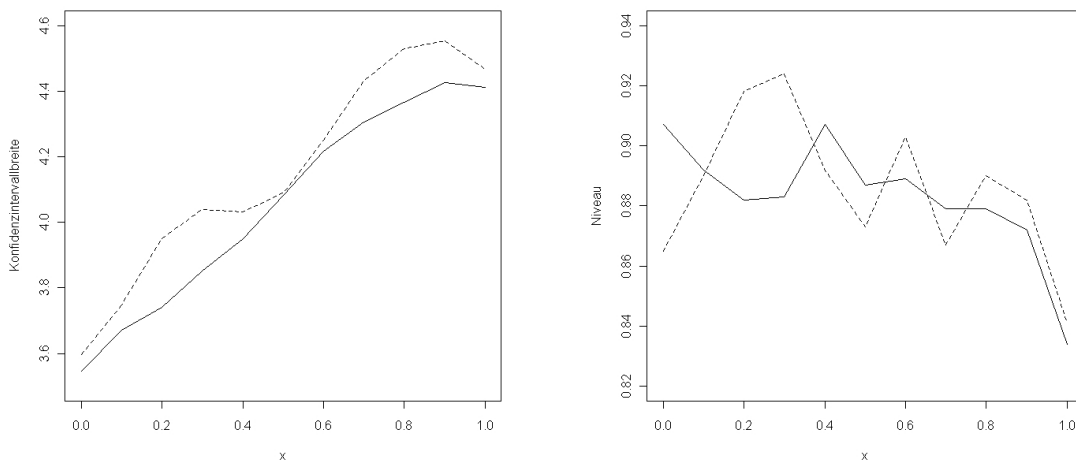


Abb.27c: Konfidenzintervallbreite und Niveaueurve mit $c = 2$

Bemerkung 9.1 (Zusammenfassung Beispiel (a)) Es zeigt sich, dass, unabhängig von den drei oben getroffenen Wahlen der Bandbreite, die Konfidenzintervalle auf Basis des ELE einen an den Rändern von $[0, 1]$ größeren Durchmesser aufweisen als diejenigen auf Basis des KE. Speziell im Bereich $x \in [0.2, 0.8]$ und Abb. 27a liegt das Niveau der Konfidenzintervalle auf Basis des ELE näher an $1 - \alpha$ als es für den KE der Fall ist. Zum anderen sieht man aber auch, dass das Niveau mit der Konstanten c schwankt. Dies gilt für beide Schätzer. Es zeigt sich zusätzlich, dass diese Konstante c von x abhängig zu wählen wäre, um auch an den Rändern das Niveau $1 - \alpha$ zu erhalten. Auf Grund der geringeren Menge von Datenpunkten an den Rändern ist dort die Schätzung für F^{-1} unzuverlässiger wodurch für kleine und große Werte von x das Konfidenzintervall für $m(x)$ ein schlechteres Niveau aufweist. Die drei Wahlen der Konstanten c in den

obigen Beispielen, 1, 0.5 und 2 beeinflusst für $n = 100$ die Bandbreite $b = n^{-1/4} = 0.316$ stark, das Niveau wird dadurch zwar beeinflusst, aber verhältnismäßig schwach. Die Grafiken zur Konfidenzintervallbreite geben auch deutlich die in diesem Beispiel vorliegende, mit x zunehmende (aber nicht als bekannt vorausgesetzte!) Varianz der Fehler ε wieder. Wählt man eine größeren Stichprobenumfang, $n = 1000$, so schwankt das Niveau mit $c = 1, 0.5, 2$ nicht in dem Maße wie es für $n = 100$ der Fall ist. Beispielhaft betrachte dazu die Kurven:

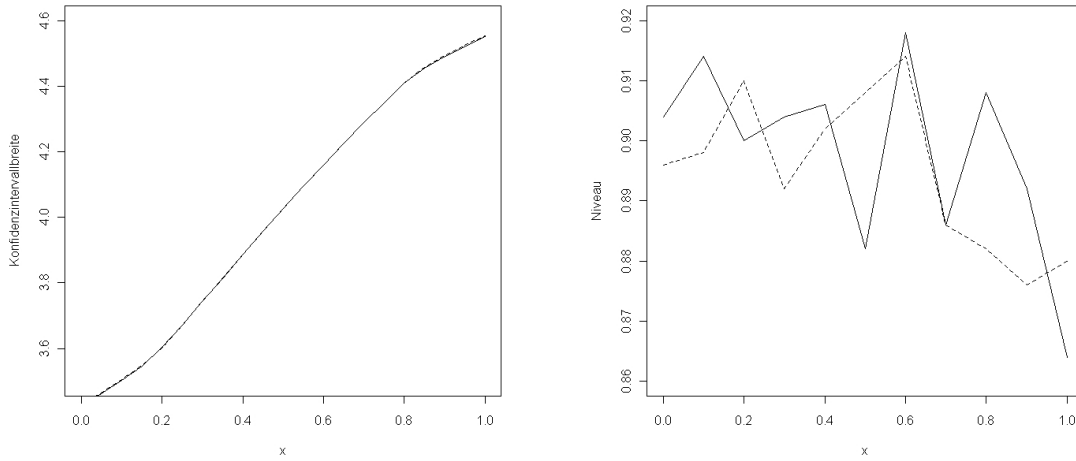


Abb.27d: Konfidenzintervallbreite und Niveaukurve für $n = 1000$ mit $c = 1$

Abb. 27d zeigt sehr gut, dass die deutliche Verschlechterung des Niveaus an den Rändern von $[0, 1]$ mit wachsenden Stichprobenumfang n nachlässt. In Simulationen mit $n = 10000$ war diese Verschlechterung nicht mehr zu beobachten. Insgesamt sprechen die Simulationen für den Nutzen der Konfidenzintervalle.

Die in dieser Arbeit getroffenen Bandbreitenannahmen (H1) und (H2) sind in Frage zu stellen, denn die optimalen Raten zur schwachen Konvergenz der Prozesse sind i. A. nicht diejenigen, welche die Konfidenzintervalle optimieren. Darauf wird diese Arbeit aber nicht genauer eingehen, die Rechnungen wären ähnlich umfangreich wie sie es zum Beweis der schwachen Konvergenz der Prozesse waren.

Ein ebensolches Bild wie es Beispiel (a) geliefert hat ergab sich auch für weitere aus dem vorherigen Kapitel entnommene Beispiele.

9.2 Testverfahren

In diesem Abschnitt werden wir drei Testideen für Tests auf Gültigkeit der Voraussetzung $E[g(\varepsilon, X) | X = x] = 0$ aus Nagel (2006) vorstellen und auf ihre Übertragbarkeit auf das Modell

dieser Arbeit hin überprüfen. Dabei nutzen wir die zuvor erbrachten Entwicklungen der Schätzer etc.

Der Empirical-Likelihood-Schätzer (ELE) nutzt aus, dass zusätzlich zu den beobachteten Daten eine Funktion g gegeben ist, für welche $E[g(\varepsilon, X) | X = x] = 0$ für alle x gilt, z.B. bekannte Varianz der Fehler ε , also $g(\varepsilon, x) = \varepsilon^2 - \sigma^2(x)$. Es ist also sinnvoll vor der Anwendung des ELE auf die Gültigkeit dieser Nebenbedingung zu prüfen.

Das erste Testverfahren basiert auf der in vorherigen Kapiteln hergeleiteten Entwicklung der Differenz des ELE und KE: $\hat{F}_n^*(y|x) - \hat{F}_n(y|x)$, das zweite Testverfahren nutzt ein asymptotisches Resultat der Empirical-Likelihood-Theorie. Das dritte Testverfahren hat eine deutlich einfachere Struktur, es verwendet $\frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) g(\hat{\varepsilon}_i, x)$ als Schätzer für $f_X(x) \cdot E[g(\varepsilon, X) | X = x]$ und damit auch einige der zuvor hergeleiteten Entwicklungen. Im weiteren sei g eindimensional; bei mehrdimensionalen Nebenbedingungen teste man die Gültigkeit komponentenweise.

Wir wollen mit der Vorstellung des Testproblems beginnen:

Getestet werden soll

$$H_0 : E[g(\varepsilon, X) | X = x] = 0 \text{ für alle } x \in [0, 1]$$

gegen

$$H_1 : E[g(\varepsilon, X) | X = x] \neq 0 \text{ für } x \text{ aus einem Intervall positiver Länge,}$$

dabei betrachten wir zwei unterschiedliche Alternativen, zum einen die feste Alternative

$$H_{f.a.} : E[g(\varepsilon, X) | X = x] = c(x), c(x) \neq 0 \text{ für } x \text{ aus einem Intervall positiver Länge}$$

und zum anderen die lokale Alternative

$$H_{l.a.} : E[g(\varepsilon, X) | X = x] = 0 + (nh)^{-1/2} c(x), c(x) \neq 0 \text{ für } x \text{ aus einem Intervall positiver Länge}$$

Bemerkung 9.2 (*Grenzverteilungen*)

Die drei bei Nagel (2006) hergeleiteten Testverfahren nutzen Statistiken deren Varianz unbe-

kannt ist und welche von Unbekannten wie der Dichte $f_{\varepsilon, X}$ abhängt. Bei Anwendung der Tests ist deshalb in der Arbeit von Nagel (2006) das zu nutzende Quantil geeignet (i. A. mittels Bootstrapverfahren) zu approximieren, was in unserem Fall bei möglicher Übertragung der Testidee ebenfalls der Fall wäre.

9.2.1 Erstes Testverfahren

In dieser Arbeit wurde die schwache Konvergenz von $\sqrt{nh}(\hat{F}_n^*(\cdot|x) - F_{\varepsilon|X}(\cdot|x) - b_n^*(\cdot, x))$ und $\sqrt{nh}(\hat{F}_n(\cdot|x) - F_{\varepsilon|X}(\cdot|x) - b_n^{**}(\cdot, x))$ bewiesen. Auf Basis dieser Resultate ergibt sich die Idee eines ersten Testverfahrens. Dieses baut auf der Differenz

$$\begin{aligned} \hat{F}_n^*(y|x) - \hat{F}_n(y|x) &= -\hat{\eta}_n^t(x)\mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = x] + o_p\left(\frac{1}{\sqrt{nh}}\right) \\ &= -\left[\Sigma^{-1}(x) \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) g(\varepsilon_i, x) \right. \\ &\quad - \Sigma^{-1}(x) \frac{1}{n} \sum_{i=1}^n \varepsilon_i \int \frac{1}{h} K_h(z - x) g'(y, x) \frac{1}{b} \tilde{K}_b(X_i - z) f_{\varepsilon|X}(y|z) d(y, z) \\ &\quad \left. - \Sigma^{-1}(x) \frac{1}{n} \sum_{i=1}^n \int \frac{1}{h} K_h(z - x) g'(y, x) \frac{1}{b} \tilde{K}_b(X_i - z) (m(X_i) - m(z)) f_{\varepsilon|X}(y|z) d(y, z) \right] \\ &\quad \cdot \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = x] + o_p\left(\frac{1}{\sqrt{nh}}\right) \end{aligned}$$

auf (vgl. Proposition 5.1 und die Entwicklung von $\hat{\eta}_n$, Proposition 4.12). Obige Darstellung ist diejenige für Funktionen g , die (G1) und (G2) erfüllen, der Fall g von Indikatorgestalt ist ebenso konstruierbar.

Der dritte Summand obiger Darstellung besitzt nach den Rechnungen zum Erwartungswert und Varianz die Rate $o_p(\frac{1}{\sqrt{nh}})$ und kann somit ebenfalls vernachlässigt werden, so dass sich für die Entwicklung ergibt:

$$\begin{aligned} V_n(y) := \hat{F}_n^*(y|x) - \hat{F}_n(y|x) &= -\hat{\eta}_n^t(x)\mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = x] + o_p\left(\frac{1}{\sqrt{nh}}\right) \\ &= -\left[\Sigma^{-1}(x) \frac{1}{nh} \sum_{i=1}^n K_h(X_i - x) g(\varepsilon_i, x) \right. \\ &\quad \left. - \Sigma^{-1}(x) \frac{1}{n} \sum_{i=1}^n \varepsilon_i \int \frac{1}{h} K_h(z - x) g'(y, x) \frac{1}{b} \tilde{K}_b(X_i - z) f_{\varepsilon|X}(y|z) d(y, z) \right] \end{aligned}$$

$$\cdot \mu_0^{K^2} E[g(\varepsilon, X) I\{\varepsilon \leq y\} | X = x] + o_p\left(\frac{1}{\sqrt{nh}}\right)$$

Für obige Entwicklung gilt unter H_0 , mit den Rechnungen zur Kovarianz in Kapitel 6 und den Aussagen zur schwachen Konvergenz die schwache Konvergenz von $\sqrt{nh}(V_n(\cdot) - E[V_n(\cdot)])$ gegen einen Prozess $V(\cdot)$ mit hier nicht näher angegebener Kovarianz $E[V(y)V(z)]$.

Zum Testen herangezogen werden könnten die Komogorow-Smirnov Teststatistik und die Cramér-von-Mises-Teststatistik:

$$S_{KS} := \sqrt{nh} \sup_{y \in \mathbb{R}, x \in [0,1]} |\hat{F}_n^*(y|x) - \hat{F}_n(y|x)|$$

$$S_{CM} := \sqrt{nh} \int \int (\hat{F}_n^*(y|x) - \hat{F}_n(y|x))^2 \hat{F}_n(dy|x) dx$$

Die in dieser Arbeit angegebenen Entwicklungen der Differenz $\hat{F}_n^*(y|x) - \hat{F}_n(y|x)$ sind nicht gleichmäßig in x bewiesen worden. Somit wäre die Komogorow-Smirnov Statistik nicht ohne weitere Entwicklungen und Modifizierungen, zum Beispiel Integration über x (was wiederum Entwicklungen erfordert) anwendbar. Zur zweiten Teststatistik ist zu überprüfen ob alle hergeleiteten Entwicklungen und Abschätzungen für $\hat{F}_n^*(y|x) - \hat{F}_n(y|x)$ übertragbar sind auf $\sup_{y \in \mathbb{R}} \int |\hat{F}_n^*(y|x) - \hat{F}_n(y|x)| dx$. Dies ist prinzipiell möglich, aber wegen der nötigen Überprüfung aller Restterme aufwändig. Ähnlich wie bei Nagel (2006) 6.2.2 ff. kann dann für die zweite Statistik neben obigem Verhalten unter H_0 auch (unter weiteren Zusatzannahmen) das Verhalten unter $H_{f.a.}$ und $H_{l.a.}$ untersucht und Konsistenz der Testverfahren nachgewiesen werden. Im Falle lokaler Alternativen $H_{l.a.}$ müssen dazu weitere Annahmen an g gestellt werden, welche eine Entwicklung der Form $g = g_{H_0} + r$ mit r hinreichend glatt ermöglichen, vgl. dazu auch die Rechnungen in Nagel (2006), 6.2.2 ff.

9.2.2 Zweites Testverfahren

Das zweite Testverfahren basiert auf dem in Kapitel 2, Seite 35 vorgestellten nichtparametrischen Likelihood-Quotienten. Es sei

$$P_1 := \left\{ (\hat{p}_1, \dots, \hat{p}_n)(x) \in \mathbb{R}^n : 0 < \hat{p}_i(x) < 1 \text{ für alle } i; \sum_{i=1}^n \hat{p}_i(x) = 1 \right\},$$

$$P_2 := \{ (\hat{p}_1, \dots, \hat{p}_n)(x) \in \mathbb{R}^n : 0 < \hat{p}_i(x) < 1 \text{ für alle } i; \sum_{i=1}^n \hat{p}_i(x) = 1, \sum_{i=1}^n \hat{p}_i(x) K\left(\frac{X_i - x}{h}\right) g(\hat{\varepsilon}_i, x) = 0 \}$$

und damit der nichtparametrische Likelihood-Quotient definiert als

$$\lambda(x) := \frac{\max_{(\hat{p}_1, \dots, \hat{p}_n) \in P_2} \left(\prod_{i=1}^n \hat{p}_i \right)}{\max_{(\hat{p}_1, \dots, \hat{p}_n) \in P_1} \left(\prod_{i=1}^n \hat{p}_i \right)} = \frac{\prod_{i=1}^n \frac{1}{1 + \hat{\eta}_n(x) K\left(\frac{X_i - x}{h}\right) g(\hat{\varepsilon}_i, x)}}{\frac{1}{n}} = \prod_{i=1}^n \frac{1}{1 + \hat{\eta}_n(x) K\left(\frac{X_i - x}{h}\right) g(\hat{\varepsilon}_i, x)}$$

Von Interesse ist nun die Statistik

$$\hat{W}_n(x) := -2 \log(\lambda)(x) = 2 \sum_{i=1}^n \log\left(1 + \hat{\eta}_n(x) K\left(\frac{X_i - x}{h}\right) g(\hat{\varepsilon}_i, x)\right).$$

Auch hier ist wieder eine Entwicklung gleichmäßig in x nötig, welche die vorliegende Arbeit nicht liefert. Mit weiteren Zusatzannahmen und umfangreicheren Rechnungen müsste es möglich sein für $\int \hat{W}_n^2(x) dx$ eine Entwicklung anzugeben, so dass für diese Statistik dann die Asymptotik bewiesen werden kann. Da diese Statistik nicht verteilungsfrei ist, ist bei der Anwendung wieder auf Bootstrapverfahren zurückzugreifen.

9.2.3 Drittes Testverfahren

Dieses Testverfahren ist das am einfachsten anzuwendende, es basiert auf

$$T_n(x) := \frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) g(\hat{\varepsilon}_i, x)$$

als Schätzer für $f_X(x) \cdot E[g(\varepsilon, X) | X = x]$. Ebenso wie zu beiden vorherigen Testverfahren kann wieder die Asymptotik der Statistik für festes x angegeben werden. Für die Konstruktion eines geeigneten Tests auf Basis von $\int T_n^2(x) dx$ ist aber die Asymptotik von $\int T_n^2(x) dx$, welche nicht aus den Entwicklungen dieser Arbeit allein herzuleiten ist, aber unter weiteren Voraussetzungen und Entwicklungen zu bestimmen wäre.

Bemerkung 9.3 *Insgesamt ist zu bemerken, dass, wenn denn jeweils die nötigen Entwicklungen gleichmäßig in x durchführbar sind, die resultierenden Varianzen der Verteilungen der Teststatistiken in allen drei Fällen unbekannt sind und somit geschätzt werden müssen. Wie schon im Falle von x fest werden die Varianzen aber selbst von der zu schätzenden Dichte abhängen. Es bietet sich an auf Basis von Bootstrapstichproben zu testen, dies könnte Bestandteil einer weiterführenden Arbeit sein.*

10 Zusammenfassung der Arbeit und Ausblick

10.1 Zusammenfassung

Die vorliegende Arbeit konnte für den betrachteten Empirical-Likelihood-Schätzer (ELE) sowie für den Kernschätzer (KE) (vgl. Definition 4.2 und 4.6), gesehen als stochastischer Prozess, die schwache Konvergenz gegen einen Gaußprozess nachgewiesen werden. Die jeweiligen Kovarianzen sowie der Bias wurden angegeben. Es zeigte sich, dass der ELE weder für große Klassen von Verteilungen, noch für betrachtete spezielle Klassen, wie der der Normalverteilungen, einen gleichmäßig kleineren MSE aufweist. Dabei wurde angenommen, dass Zusatzinformation zur Verteilung der Fehler in Form von $g(\varepsilon, x)$ mit $E[g(\varepsilon, x) | X = x] = 0$ vorlag. Für das Beispiel $g(\varepsilon, x) = \varepsilon$ ergab sich asymptotisch für beide Schätzer dieselbe Varianz und derselbe Bias. In weiteren Beispielen zeigte sich, dass die Struktur der Nebenbedingung großen Einfluss auf das Aussehen der Kovarianzstruktur des ELE hat. Diese Beobachtung an den theoretischen MSE -Kurven wurde anhand von Simulationen bestätigt. Da die zu schätzende Verteilung und die Struktur von g beide großen Einfluss auf das Verhalten des ELE im Vergleich zum KE bezogen auf den MSE haben spricht dies gegen den direkten Nutzen des ELE. Nebeninformationen von Indikatorgestalt, also $g(\varepsilon, x) = I\{\varepsilon \leq a(x)\} - d(x)$ haben sich in den Simulationen bewährt, speziell für y -Werte in der Umgebung des bekannten Quantils und dies gleichmäßig über alle Werte von x .

Die auf Basis der Konvergenzaussage vorgeschlagenen Konfidenzintervalle erwiesen sich als brauchbar, wären aber noch mittels genauerer theoretischer Überlegungen zu optimieren, speziell im Hinblick auf die Wahl der Bandbreiten. Vorgeschlagene Tests auf $E[g(\varepsilon, X) | X = x] = 0$ zeigen zum Ende der Arbeit noch den weiteren Nutzen der in der Arbeit geführten Entwicklungen auf, wären aber in der Theorie ebenfalls noch genauer zu betrachten, da sie aufgrund der unbekanntenen Kovarianzstrukturen nicht direkt anwendbar und teils die Entwicklungen dieser Arbeit für die nötigen Beweise noch nicht ausreichend sind.

10.2 Ausblick

Die im letzten Kapitel vorgestellten Verfahren zur Erstellung von Konfidenzintervallen und Anpassungstest auf die Modellvoraussetzungen könnten noch explizit herausgearbeitet werden, speziell für die Konfidenzintervalle und den KE $\hat{F}_n(y|x)$ wäre dies eine interessante Weiterführung dieser Arbeit. Für den ELE $\hat{F}_n^*(y|x)$ wären diese Rechnungen deutlich umfangreicher und auf Grund der in dieser Arbeit enthaltenen Ergebnisse im Hinblick auf den Nutzen schlecht abzuschätzen. Speziell die Untersuchung und Entwicklung eines geeigneten Bootstrapverfahrens

wäre wünschenswert.

Des Weiteren könnte man den in dieser Arbeit nur simulierten *MISE* auch noch theoretisch herleiten. Die Ergebnisse der Simulationen lassen darauf schließen, dass hier für größere Beispiellklassen eine Verringerung des *MISE* der ELE im Vergleich zum *MISE* des KE nachweisbar wäre.

Symbolverzeichnis

Symbol	Bedeutung	Seite
X, Y	Zufallsvariablen	5
f, F	Dichte und dazugehörige Verteilungsfunktion	5
(Ω, \mathcal{A}, P)	Wahrscheinlichkeitsraum	9
$(\mathbb{R}, \mathcal{B})$	Messraum im Fall reellwertiger Zufallsvariablen	9
$(X_n)_{n \in \mathbb{N}}$	Folge von Zufallsvariablen	9
$o(\cdot), O(\cdot)$	Landausche Symbole für reelle Zahlenfolgen	9
$o_P(\cdot), O_P(\cdot)$	stochastische Konvergenz und Beschränktheit	9
$X_n \xrightarrow{f.s.} X$	fast sichere Konvergenz	9
$X_n \xrightarrow{P} X$	P -stochastische Konvergenz	9
$(X(\cdot, t))_{t \in T}$	stochastischer Prozess mit Indexmenge T	10
B^0	Brownsche Brücke	10
F_n	empirische Verteilungsfunktion (EDF)	11
$D([a, b]), D([-\infty, \infty])$	cadlag-Raum auf $[a, b]$ bzw. auf $\bar{\mathbb{R}}$	10
$l^\infty(\mathcal{F})$	Raum der stetigen Funktionen auf $[a, b]$	10
$X_n \xrightarrow{D} X$	Konvergenz in Verteilung	12
$C([a, b])$	spezieller Funktionenraum	13
$\ \varphi\ _{P,r}$	$L_r(P)$ -Norm zu φ	13
Φ	Einhüllende einer Funktionenklasse	13
$VI(\mathcal{C})$	VC-Index zu einer Menge \mathcal{C} von Mengen	17
$C_\delta^{1+\alpha}[0, 1]$	bestimmte Funktionenklasse	19
K	Kernfunktion	26
\hat{f}_X	Kerndichteschätzer zur Dichte f_X	27
\hat{m}	Nadaraya-Watson-Schätzer für m	27
X_i	i te Designvariable ($i = 1, \dots, n$)	39
Y_i	i te Zielgrößenvariable ($i = 1, \dots, n$)	39
ε_i	i te Fehlervariable ($i = 1, \dots, n$)	39
m	Modellfunktion	39

Symbol	Bedeutung	Seite
n	Stichprobengröße	39
$f_X, f_\varepsilon, f_{\varepsilon X}, f_{\varepsilon,X}$	Dichten	39
$F_X, F_\varepsilon, F_{\varepsilon X}, F_{\varepsilon,X}$	Verteilungsfunktionen	39
$B(x)$	spezielle Funktion	39
\dot{f}, \ddot{f}, f'	Ableitungen	39
$\hat{\varepsilon}_i$	ites Residuum zu Modell	40
\tilde{K}, K	Kernfunktion	40
\tilde{K}_h, K_h	Kernfunktion	40
μ_2^K	Kurznotation für $\int K(u)u^2 du$	40
$\mu_0^{K^2}$	Kurznotation für $\int K^2(u) du$	40
b	Bandbreite	40
g	Die die Nebeninformation darstellende Funktion	41
$\hat{F}_n(y x); F_n(y x)$	Kernschätzer	46
h	Bandbreite	40
$F_n^*(y x)$	Empirical-Likelihood-Schätzer	45
p_1, \dots, p_n	Gewichte	45
L, l	Likelihood, Loglikelihood	49
$\hat{p}_1, \dots, \hat{p}_n$	Gewichte bzgl. Residuen	49
$\hat{\eta}_n$	Lagrange-Multiplikator bzgl. Residuen	49
\hat{F}_n^*	Empirical-Likelihood-Schätzer bzgl. Residuen	50
Σ	spezielle Matrix	52
G	spezieller Gaußprozess	128
\tilde{G}	spezieller Gaußprozess	133
λ	Verhältnis der Bandbreiten h und b	137
\widehat{MSE}	Schätzer des MSE	145
\widehat{MISE}	Schätzer des MISE	145
$\Delta\widehat{MSE}$	spezielle Differenz	148

Literaturverzeichnis

- M. G. Akritas and I. Van Keilegom** (2001). *Non-parametric Estimation of the Residual Distribution*. Scand. J. Stat. 28, 549–567.
- H. Amann and J. Escher** (2002). *Analysis I-III*. Birkhäuser.
- A. Azzalini and A. Capitanò** (1999). *Applications of the Multivariate Skew Normal Distribution*. J. R. Stat. Soc. B 61.
- P. Billingsley** (1995). *Probability and Measure*. Third Edition. Wiley.
- M. J. Buckley, G. K. Eagleson and B. W. Silverman** (1988). *The Estimation of Residual Variance in Nonparametric Regression*. Biometrika 75, 189–200.
- T. Duong and M. L. Hazelton** (2005). *Cross-Validation Bandwidth Matrices for Multivariate Density Estimation*. Scand. J. Statist. 40, 485–506.
- J. H. J. Einmahl and I. W. McKeague** (2003). *Empirical likelihood based hypothesis testing*. Bernoulli 9, 267–290.
- V. K. Epanechnikov** (1969). *Non-Parametric Estimation of a Multivariate Probability Density*. Theory Probab. Appl. 14, 153–158.
- P. Hall, R. C. L. Wolff and Q. Yao** (1999). *Methods for Estimating a Conditional Distribution Funktion*. J. Am. Stat. Assoc. 94, 154–163.
- W. Härdle** (1993). *Applied Nonparametric Regression*. Cambridge University Press.
- R. J. Karunamuni and T. Alberts** (2005). *A Generalized Reflection Method of Boundary Correction in Kernel Density Estimation*. Can. J. Stat. 33, 497–509.
- S. Kiwitt, E. Nagel and N. Neumeyer** (2009). *Empirical Likelihood Estimators for the Error Distribution in Nonparametric Regression Models*. Math. Methods Stat. 17, No. 3, 241–260.
- M. R. Kosorok** (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer Series in Statistics.
- R. M. Loynes** (1980). *The Empirical Distribution Function of Residuals from Generalized Regression*. Ann. Stat. 8, 285–298.

- E. Molanes-Lopez, I. van Keilegom and N. Veraverbeke** (2009). *Empirical Likelihood for Non-Smooth Criterion Functions*. Scand. J. Statist. (to appear).
- E. Nadaraya** (1964). *On estimating regression*. Theory Prob. Appl. 10, 186–190.
- E. Nagel** (2006). *Empirical-Likelihood-Schätzung der Fehlerverteilung im nichtparametrischen Regressionsmodell*. Dissertation. Ruhr-Universität-Bochum.
- N. Neumeyer** (2006). *Bootstrap procedures for empirical processes of nonparametric residuals*. Habilitationsschrift. Ruhr-Universität-Bochum.
- A. B. Owen** (2001). *Empirical Likelihood*. Chapman & Hall.
- D. Pollard** (1984). *Convergence of Stochastic Processes*. Springer Verlag, New York.
- D. Pollard** (1990). *Empirical Processes. Theory and Applications*. NSF-CBMS Regional Conference Series in Prob. and Stat. Vol. 2.
- J. Qin and J. Lawless** (1994). *Empirical likelihood and general estimation equations*. Ann. Statist. 22, 300–325.
- R Development Core Team** (2009). *R: A language and environment for statistical computing*. Vienna Austria: R Foundation for Statistical Computing.
- V. K. Rohatgi and E. Saleh** (2001). *An Introduction to Probability and Statistics*. Wileys series in Probability and Statistics.
- M. Rosenblatt** (1956). *Remarks on some nonparametric estimates of a density function*. Ann. Math. Statist. 27, 832–837.
- B. Seifert, T. Gasser and A. Wolf** (1993). *Nonparametric estimation of residual variance revisited*. Biometrika 80, 373–384.
- D. W. Scott** (1992). *Multivariate Density Estimation*. Wiley Series in Probability.
- W. Stute** (1986). *Conditional Empirical Processes*. Ann. Statist. 14, 638–647.
- A. W. van der Vaart** (1998). *Asymptotic Statistics*. Cambridge University Press.
- A. W. van der Vaart and J. A. Wellner** (2000). *Weak Convergence and Empirical Processes*. Springer Series in Statistics.
- V. N. Vapnik and A. Ya. Červonenkis** (1971). *On the uniform convergence of relative frequencies of events to their probabilities*. Theory Prop. Appl. 16, 264–280.

G. S. Watson (1964). *Smooth regression analysis*. Sankhyā, Series A, 26, 359–372.

Zusammenfassung

Häufig stellt sich in naturwissenschaftlichen Problemstellungen die Frage nach dem funktionalen Zusammenhang zweier Zufallsgrößen X und Y . Es wird angenommen, dass sich der Zusammenhang nicht allein funktional beschreiben lässt, sondern zusätzlich durch die Addition eines zufälligen (nicht beobachtbaren) Fehlertermes. Die vorliegende Arbeit gibt zwei Schätzer für die Verteilung der Fehler an, wobei angenommen wird, dass die Fehler und der Regressor X voneinander abhängen, was in vielen Fällen der Anwendung auch plausibel ist. Es werden zwei nichtparametrische Schätzer für die bedingte Verteilung der Fehler gegeben den Regressor X konstruiert. Die vorgestellten nichtparametrischen Schätzer wurden in diesem Modell bisher noch nicht betrachtet. Einer der Schätzer nutzt aus, dass Zusatzinformation zur Verteilung der Fehler gegeben ist. Beispiele hierfür sind die Zentriertheit oder auch bekannte Varianzstruktur der Fehler. Der die Zusatzinformation nutzende Schätzer wird über Empirical-Likelihood-Techniken motiviert. Für beide Schätzer werden stochastische Entwicklungen hergeleitet, wird schwache Konvergenz der (geeignet standardisierten) Schätzer gegen Gaußprozesse nachgewiesen und die asymptotische Kovarianzstruktur angegeben. Anhand einiger Beispiele wird die Asymptotik der Schätzer verglichen, speziell die mittlere quadratische Abweichung (MSE). Im Anschluss daran werden die Schätzer in Simulationen betrachtet. Es stellt sich in der Theorie sowie den Simulationen heraus, dass keiner der beiden Schätzer einen gleichmäßig kleineren MSE aufweist. Im Anschluss wird in einem Ausblick untersucht, auf welche Weise die in der Arbeit hergeleiteten Resultate im Bezug auf Vorhersageintervalle für Y und für Tests auf Gültigkeit der Zusatzinformation herangezogen werden können, dabei wird die nötige Theorie nur skizziert.

Lebenslauf

Persönliche Daten

Name	Sebastian Kiwitt
Geburtsdatum	16.08.1980
Geburtsort	Gelsenkirchen

Ausbildung

1987 – 1991	Besuch der Grundschule in Gelsenkirchen-Ückendorf
1991 – 2000	Besuch des Ricarda-Huch Gymnasiums, Gelsenkirchen
Juni 2000	Schulabschluss Abitur
WS 2001/02 bis SS 2007	Mathematikstudium im Diplomstudiengang der Ruhr-Universität Bochum
August 2007	Studienabschluss Diplom in Mathematik

Beschäftigung

Okt. 2000 – Sept. 2001	Zivildienst
Okt. 2008 – Sept. 2010	Wissenschaftlicher Mitarbeiter am Department Mathematik der Universität Hamburg