

Schätz- und Testverfahren in der nichtparametrischen Regression unter qualitativen Annahmen

MELANIE BIRKE

Dissertation

zur Erlangung des Doktorgrades
der Naturwissenschaften
an der Fakultät für Mathematik
der Ruhr-Universität Bochum

2006

Inhaltsverzeichnis

Einleitung	5
1 Grundlagen der Kernschätzung	11
1.1 Nichtparametrische Regressionsschätzung	11
1.2 Kernschätzer	12
1.3 Asymptotische Theorie	14
2 Monotone und konvexe Umordnungen	19
2.1 Isotone Umordnungen	19
2.2 Konvexe Umordnungen	26
3 Schätzen von isotonen Regressionsfunktionen	31
3.1 Asymptotik des isotonen Kernschätzers	31
3.2 Isotoner Kleinste-Quadrate Schätzer	40
4 Schätzen von konvexen Regressionsfunktionen	43
4.1 Asymptotik des konvexen Schätzers	43
4.2 Simulationen	46
4.2.1 Vergleich mit dem lokal linearen Schätzer	48
4.2.2 Vergleich mit einem konvexen Kleinste-Quadrate Schätzer	52
4.3 Datenbeispiele	54
4.3.1 Schätzen von Optionspreisen	55
4.3.2 Schätzen des Alters von australischen Kaninchen	56
4.4 Beweise	58
4.4.1 Beweis von Lemma 4.3	58
4.4.2 Beweis von Lemma 4.4	64
5 Testen auf strenge Isotonie	71
5.1 Definition der Teststatistik	71
5.2 Asymptotische Verteilung der Teststatistik	77
5.2.1 Asymptotische Verteilung des integrierten quadrierten Fehlers des Regressionsschätzers und dessen Ableitungen	78

5.2.2	Beweis von Satz 5.3	82
5.3	Bootstrapverfahren und Simulationen	88
A	Beweistechnische Details	93
A.1	Gleichmäßig fast sichere Konvergenzraten	93
A.2	Beweis von Satz 5.5	100
	Symbolverzeichnis	127
	Literaturverzeichnis	129

Einleitung

In vielen Bereichen der Wirtschaft, Physik, Biologie und den Ingenieurwissenschaften spielen funktionale Zusammenhänge zwischen zwei (reellwertigen) Messgrößen eine wichtige Rolle. Oft sind aus Experimenten anstatt des genauen Zusammenhangs jedoch nur einzelne Datenpaare bekannt, die zusätzlich mit Fehlern behaftet sein können. Die statistische Methode, von diesen Datenpaaren auf den zu Grunde liegenden, aber unbekanntem funktionalen Zusammenhang zu schließen nennt man Regression. Es gibt verschiedenste Ansätze, die wahre Regressionsfunktion zu schätzen. Eine Methode ist die parametrische Regression. Bei dieser wird angenommen, dass die Regressionsfunktion zu einer bestimmten parametrischen Klasse von Funktionen gehört. Zur genauen Bestimmung des Schätzers werden die unbekanntem Parameter aus den beobachteten Daten geschätzt. Die am häufigsten verwendete Funktionenklasse ist die der linearen Funktionen. In diesem Fall spricht man von einem allgemeinen linearen Modell (vgl. Searle, 1971 oder Seber, 1977). Der angenommene funktionale Zusammenhang ist hierbei $f(x) = x^T \beta$ für Vektoren $x, \beta \in \mathbb{R}^p$. Eine unterschiedliche Wahl des Designvektors x resultiert in verschiedenen parametrischen Modellen. So erhält man für $x = (1, t)^T$ die lineare Regression mit $f(t) = \beta_1 t + \beta_0$, ein quadratisches Regressionsmodell ergibt die Wahl $x = (1, t, t^2)^T$. Darüberhinaus lässt sich auch eine Abhängigkeit von mehreren Einstellgrößen modellieren ($x = (x_1, \dots, x_p)^T$, multiple Regression). Ein Problem der parametrischen Regression liegt darin, dass nicht in allen Zusammenhängen bekannt ist, welcher Funktionenklasse eine Regressionsfunktion angehört. Eine falsche Wahl der Funktionenklasse, eine sogenannte Missspezifikation hat, zumindest auf Teilbereichen schlechte Schätzungen der Regressionsfunktion zur Folge. Die Methode der nichtparametrischen Regression vermeidet Spezifikationsfehler. Als einzige Annahme an die Regressionsfunktion genügt hier die Glattheit. Zum Beispiel werden in Härdle (1995) Kernschätzer der Regressionsfunktion vorgestellt. Ein sehr verwandter Ansatz ist die Schätzung mit lokalen Polynomen, wie sie in Fan und Gijbels (1997) diskutiert wird. Darüber hinaus existieren Verfahren basierend auf Splines (Eubank, 1988) und orthogonalen Reihen (Efromovich, 1999). Bei allen genannten Methoden ist die Wahl eines Glättungsparameters nötig, der die Güte des Schätzers stark beeinflusst. In vielen Zusammenhängen ist der Typ der Regressionsfunktion zwar unbekannt, so dass sich die Anwendung eines nichtparametrischen Regressionsverfahrens anbietet, es existieren aber Informationen über die Form der Funktion. Solche Informationen werden auch qua-

litative Annahmen genannt und stammen häufig aus ökonomischen, physikalischen oder biologischen Zusammenhängen. Beispiele für die Form einer Regressionsfunktion sind Monotonie, Konvexität oder Konkavität. Es ist naheliegend, dass mit höherem Einkommen auch die Bereitschaft steigt, mehr Geld für Güter des täglichen Gebrauchs, wie etwa Nahrungsmittel oder Kleidung, auszugeben. Ebenfalls monoton wachsend sind das Verhältnis zwischen Investition und Nutzen (die sogenannte Nutzenfunktion), die Produktionsfunktion, die das Verhältnis zwischen Produktionsmenge und Produktionsfaktoren wie Arbeitskraft oder Anzahl der eingesetzten Maschinen angibt und Kostenfunktionen. Aït-Sahalia und Duarte (2003) begründen, dass die Preisfunktion einer Call-Option monoton fallend und konvex ist. Auch Volatilitäts-Funktionen sind häufig konvex. Im biologisch-medizinischen Zusammenhang zeigen oft höhere Dosierungen eines Wirkstoffs eine größere Wirkung bei den Probanden. Auch Wachstumskurven sind typischerweise monoton und/oder konvex oder konkav. Bakterienkulturen wachsen bekanntlich exponentiell über die Zeit. Der Verlauf dieser Wachstumskurve ist daher im Idealfall monoton wachsend und konvex. Das Gewicht von Lebewesen wächst zu Beginn ihres Lebens relativ stark, stagniert jedoch wenn sie ausgewachsen sind. Die zu Grunde liegende Wachstumsfunktion kann daher monoton wachsend und konkav sein.

Obwohl die üblichen nichtparametrischen Regressionsmethoden gute Schätzer für die unbekannte Regressionsfunktion liefern, erfüllen diese jedoch im Allgemeinen nicht die, sich durch die Substanzwissenschaften ergebenden Voraussetzungen an die Form der Regressionsfunktion. Aus diesem Grund wurden verschiedene Schätzer entwickelt, die die Form der Regressionsfunktion berücksichtigen. Die Arbeiten zu monotonen Regressionsschätzern lassen sich in zwei Kategorien einteilen. In der ersten Kategorie werden die Schätzer durch Minimierung von Kriteriumsfunktionen über eine eingeschränkte Menge von Funktionenklassen bestimmt. Erste Ansätze dazu und zu deren asymptotischer Theorie wurden von Brunk (1955, 1958) entwickelt. Dort wird eine Maximum-Likelihood beziehungsweise Kleinst-Quadrate Schätzung einer monoton wachsenden Folge von Parametern vorgeschlagen. Stellen die Parameter die Regressionsfunktion ausgewertet an festen Stützstellen dar, so kann dies als ein erster Ansatz zur monotonen Regressionsschätzung gedeutet werden. Die Asymptotik dieses Schätzers wird in Brunk (1970), Wright (1982) und Groeneboom (1985, 1989) behandelt. Ramsay (1998), Mammen und Thomas-Agnan (1999) und Mammen, Marron, Turlach und Wand (2001) stellen monotone Schätzer basierend auf der Minimierung über die Klasse der monotonen Splines vor. Letztere übertragen diese Vorgehensweise auf allgemeinere Vektorräume. Schließlich stellt Zhang (2004) eine Methode der monotonen Spline-Schätzung vor, in der er die Positivität der Ableitung der Regressionsfunktion ausnutzt. Eine für Spline-Regression übliche Kriteriumsfunktion wird darin mit Hilfe der Ableitung der Regressionsfunktion dargestellt und über die Menge der positiven Splines minimiert.

Die zweite Kategorie von Schätzverfahren zeichnet sich dadurch aus, dass ein uneingeschränkter Schätzer in einem zweiten Schritt so verändert wird, dass der resultierende Schätzer monoton ist. Friedman und Tibshirani (1984) wählen zunächst einen uneinge-

schränkten glatten Schätzer und wenden auf diesen die Monotonisierung von Brunk (1958) an. Den umgekehrten Weg wählt Mukerjee (1988). Nach der Monotonisierung der Daten wie in Brunk (1958) glättet er diese mit Spline-Regression und betrachtet das asymptotische Verhalten des so erhaltenen Schätzers. In Mammen (1991a) werden die beiden zuletzt genannten Methoden sowohl asymptotisch als auch für endlichen Stichprobenumfang verglichen. Hall und Huang (2001) stellen gewichtete Kernschätzer vor, deren Gewichte anschließend so gewählt werden sollen, dass sie möglichst nah an denen der Gleichverteilung liegen unter der Nebenbedingung der Monotonie des Schätzers. Die in Anevski und Hössjer (2006) beschriebene Methode verwendet die Ableitung der größten konvexen Minorante eines Kernschätzers als isotonen Schätzer der Regressionsfunktion. In der Arbeit von Dette, Neumeyer und Pilz (2006) wird ein vollständig auf Kernschätzung basierender monotoner Schätzer definiert. Die Autoren beginnen mit einem uneingeschränkten Kernschätzer wie zum Beispiel dem Nadaraya-Watson Schätzer oder lokalen Polynom-schätzern. Der monotone Schätzer ist die geglättete monotone Umordnung, das bedeutet, die Inverse der geglätteten Verteilungsfunktion, des uneingeschränkten Schätzers.

Die Literatur über konvexe oder konkave Regression beschränkt sich hauptsächlich auf die Methode der Kleinsten Quadrate. In Hildreth (1954) ist ein konkaver Kleinste-Quadrate Schätzer der Produktionsfunktion definiert, dessen Konsistenz in Hanson und Pledger (1976) nachgewiesen wird. Mammen (1991b) erhält einen konkaven oder konvexen Schätzer, indem er über die Klasse aller stückweise monoton fallenden beziehungsweise monoton wachsenden Funktionen minimiert und bestimmt dafür die Konvergenzrate. Durch Minimierung über die Klasse aller zwischen den Datenpunkten stückweise linearen konvexen Funktionen erhalten Groeneboom, Jongbloed und Wellner (2001) einen Schätzer für den sie neben der Konvergenzrate auch die asymptotische Verteilung bestimmen können. Aït-Sahalia und Duarte (2003) kombinieren eine Kleinste-Quadrate Schätzung auf der Menge der konvexen und monoton fallenden Funktionen mit einer anschließenden Glättung durch einen lokal linearen Schätzer und erhalten auf diese Weise einen glatten konvexen und monoton fallenden Schätzer der Call-Preis Funktion. Schließlich verwenden Anevski und Hössjer (2006) die größte konvexe Minorante eines Kernschätzers als konvexen Regressionsschätzer.

Oft erhält man durch Betrachtung der Daten Hinweise darauf, dass der Regressionsfunktion eine bestimmte Form zu Grunde liegt, hat dafür jedoch keine theoretische Erklärung. Will man dennoch, zum Beispiel für die weitere Analyse der Daten, von einem monotonen Zusammenhang ausgehen, ist es naheliegend, diese Hypothese zunächst durch einen statistischen Test zu überprüfen. Bisher existieren relativ wenige Testverfahren dieser Art. Schlee (1982) stellt einen Test vor, der auf dem Abstand eines Ableitungsschätzers zur 0 beruht. Der in Bowman, Jones und Gijbels (1998) entwickelte Bootstrap-Test verwendet das in Silverman (1981) vorgestellte Verfahren. Dieses beruht auf dem Grad der Glättung, also der Größe der Bandbreite, die nötig ist, um einen monotonen lokal linearen Schätzer zu erhalten. Dieser Test hat jedoch niedrige Power wenn die wahre Regressionsfunktion

flache Stellen besitzt (Hall und Heckman, 2000). Letztere Autoren entwickeln daher einen Test, der auf den Steigungen einer linearen Regression auf kleinen Ausschnitten der Daten basiert und besser mit solchen Situationen umgehen kann. Gijbels, Hall, Jones und Koch (2000) betrachten zum Testen auf Monotonie die Anzahl der positiven Vorzeichen in einer Folge vorgegebener Länge von aufeinanderfolgenden Differenzen der y -Werte. Die Teststatistik, die in Goshal, Sen und van der Vaart (2000) vorgeschlagen wird ist ein Funktional eines stochastischen Prozesses und kann als lokale Version von Kedall's τ angesehen werden. Ein exakter Test auf Monotonie ist in Baraud, Huet und Laurent (2003) gegeben. Der Kolmogorov-Test von Durot (2003) misst den größten Abstand eines empirischen Prozesses zu seiner kleinsten konkaven Majorante. Zuletzt sei hier der Ansatz von Domínguez-Mencheró, González-Rodríguez und López-Palomo (2005) genannt. Dort wird für den Abstand bezüglich des empirischen Maßes eines Kernschätzers zu der Menge aller monoton wachsenden Funktionen ein Bootstrap-Verfahren vorgeschlagen.

Viele der hier genannten Schätz- und Testverfahren, insbesondere diese, die allein auf der Minimierung über eine Funktionenklasse beruhen, sind zeitaufwendig und daher nur für relativ kleine Datenmengen geeignet. Darüber hinaus sind einige dieser Verfahren nur auf homoskedastische Regressionsmodelle anwendbar. Die Bedingung der zweimaligen Differenzierbarkeit wird in vielen Methoden für asymptotische Untersuchungen vorausgesetzt. So basiert zum Beispiel die asymptotische Theorie in Dette, Neumeyer und Pilz (2006) darauf, dass die Regressionsfunktion zweimal stetig differenzierbar ist mit positiver erster Ableitung.

Die vorliegende Arbeit diskutiert die drei oben genannten Bereiche der Regression unter Einschränkung der Form der Regressionsfunktion. Alle hier vorgestellten Methoden basieren ausschließlich auf Kernschätzung und gehen wie in Dette, Neumeyer und Pilz (2006) zunächst von einer uneingeschränkten Schätzung aus. Diese wird in einem zweiten Schritt entsprechend der Einschränkungen an die Form verändert. Eine zentrale Rolle dabei spielt die Theorie der monotonen Umordnungen. In Kapitel 1 wird das Regressionsmodell spezifiziert und es werden uneingeschränkte Kernschätzer und deren grundlegende Eigenschaften vorgestellt. Die zur Definition der eingeschränkten Schätzer notwendige Theorie der monotonen Umordnung von Funktionen wird in Kapitel 2 kurz beschrieben und das Prinzip verwendet um eine sogenannte "konvexe Umordnung" zu definieren.

Für den bereits von Dette, Neumeyer und Pilz (2006) definierten monotonen Schätzer wird in Kapitel 3 Asymptotik für eine nur einmal stetig differenzierbare Regressionsfunktion entwickelt. Anders als in Dette, Neumeyer und Pilz (2006) unterscheidet sich die Konvergenzrate des monotonen von der des uneingeschränkten Schätzers, kann jedoch beliebig nahe an dieser gewählt werden.

Der zweite Teil der Arbeit widmet sich der Regressionsschätzung unter der Voraussetzung, dass die wahre Regressionsfunktion konvex oder konkav ist. Die ausführliche Darstellung der Methode für den Fall der Konvexität erfolgt in Abschnitt 4.1. Der konvexe Schätzer ist punktweise asymptotisch normal und erster Ordnung äquivalent zu dem uneingeschränk-

ten Schätzer. Daraus folgt, dass beide Schätzer dieselben Konvergenzraten und auch dieselbe Grenzverteilung besitzen. Der Fall der Konkavität kann analog hergeleitet werden. Das Vorgehen dazu wird kurz in einer Bemerkung am Ende von Abschnitt 4.1 beschrieben. An Hand von Simulationen wird in Abschnitt 4.2 das Verhalten des Schätzers in endlich dimensionalen Stichproben illustriert, indem zunächst für drei verschiedene Regressionsfunktionen die entsprechenden lokal linearen und konvexen Schätzer berechnet und im Vergleich zu der wahren Regressionsfunktion dargestellt werden. Diesem ersten visuellen Eindruck folgen Simulationen zu dem mittleren quadratischen Fehler (MSE) des konvexen Schätzers im Vergleich zu dem des lokal linearen Schätzers. Der mittlere quadratische Fehler misst den zu erwartenden Abstand des Schätzers zu der wahren Regressionsfunktion und ist daher ein Qualitätsmerkmal. Je kleiner der MSE ist, desto besser ist der Schätzer. Von besonderem Interesse sind Simulationen zum Vergleich des in dieser Arbeit definierten Schätzers mit dem konvexen Kleinste-Quadrate Schätzer von Groeneboom, Jonbloed und Wellner (2001) in Abschnitt 4.2.2. Zum Abschluss des zweiten Teils werden zwei Datenbeispiele präsentiert bei denen es sinnvoll ist konvexe oder konkave Schätzer zu verwenden. Das erste Beispiel zeigt die Call-Preisfunktion einer DAX Option, die nach den Ausführungen in Ait-Sahalia und Duarte (2003) konvex und monoton fallend sein muss. Als zweites Beispiel dient das Verhältnis zwischen dem Alter von australischen Kaninchen und dem Trockengewicht ihrer Augenlinse. Hierbei handelt es sich um einen konkaven Zusammenhang.

Basierend auf der Inversen des monotonen und des uneingeschränkten Schätzers kann eine Teststatistik zum Testen auf strenge Monotonie definiert werden. Es wird in Kapitel 5 gezeigt, dass diese Teststatistik strenge Monotonie eindeutig erkennt und in Verteilung gegen eine Normalverteilung konvergiert. Dazu notwendige, aber in der Literatur bisher noch nicht vorhandene Sätze über die asymptotische Verteilung des L^2 -Abstands von Ableitungsschätzern zu der wahren Ableitung werden in Abschnitt 5.2.1 vorgestellt und lehnen sich an den von Hall (1984) bewiesenen Satz für den L^2 -Abstand des Regressionschätzers zu der wahren Regressionsfunktion an. Wie in diesem Zusammenhang üblich, hängt die asymptotische Verteilung der Teststatistik von der wahren aber unbekanntem Regressionsfunktion ab. Aus diesem Grund kann sie nicht verwendet werden, um einen asymptotischen Test auf Monotonie der Regressionsfunktion zu definieren. Für Anwendungen wird daher ein Bootstrap-Test vorgeschlagen. Dies ist eine alternative Methode um Tests zu definieren, ohne unbekannt Größen wie etwa Regressionsfunktionen in der asymptotischen Varianz erneut schätzen zu müssen. Das Verhalten des Tests in endlichen Stichproben wird abschließend mit Simulationen illustriert.

Kapitel 1

Nichtparametrisches Regressionsmodell und Grundlagen der Kernschätzung

1.1 Nichtparametrische Regressionssschätzung

Wie schon in der Einleitung erwähnt, gibt es in vielen Bereichen der Wirtschaft, der Natur- und Ingenieurwissenschaften funktionale Zusammenhänge zwischen zwei Messgrößen. In der nichtparametrischen Regression stellt man diesen Zusammenhang durch das Regressionsmodell

$$Y_i = m(X_i) + \sigma(X_i)\varepsilon_i, \quad i = 1, \dots, n \quad (1.1)$$

dar. Die Größen X_i heißen Designpunkte und sind die Punkte, in denen beobachtet wird. Liegen deterministische Designpunkte vor, so spricht man von einem festen Design. Das Experiment ist dann wiederholbar. Sind sie zufällig in dem Sinne, dass sie unabhängig identisch verteilt sind, so liegt ein zufälliges Design vor. In diesem Fall kann ein Experiment nicht wiederholt werden, da jedes Mal an unterschiedlichen Stellen beobachtet wird. In dieser Arbeit wird von einem zufälligen Design auf dem Intervall $[0, 1]$ ausgegangen. Als Stichprobengröße bezeichnet man die Anzahl n der Designpunkte. Die Zufallsvariablen ε_i stellen Fehlergrößen dar. Darunter fallen etwa zufällig auftretende Messfehler oder externe Störungen des Experiments. Sie bewirken also zufällige Abweichungen von dem wahren funktionalen Zusammenhang. Üblicherweise werden diese Fehler als unabhängig und identisch verteilt und unabhängig von den Designpunkten angenommen. Desweiteren nimmt man an, dass die Fehler sich im Mittel aufheben, das heißt es gilt $E[\varepsilon_1] = 0$, und die Variabilität normiert ist ($E[\varepsilon_i^2] = 1$). Die Funktion $m : [0, 1] \rightarrow \mathbb{R}$ bezeichnet die Regressionsfunktion, $\sigma : [0, 1] \rightarrow \mathbb{R}^+$ beschreibt die Standardabweichung der Fehler und damit der Beobachtungen Y_i . Ist σ konstant auf $[0, 1]$, so handelt es sich um ein homoskedastisches Regressionsmodell, andernfalls spricht man von heteroskedastischer Regression. Im weiteren Verlauf wird von dem allgemeineren Fall der Heteroskedastizität ausgegangen.

1.2 Kernschätzer

Da die Regressionsfunktion im Allgemeinen unbekannt ist, benötigt man Schätzer für diese. In diesem Abschnitt wird die Methode der Kernschätzung an Hand der drei gängigsten Kernschätzer, dem Nadaraya-Watson Schätzer, dem lokalen Polynomschätzer und dem Gasser-Müller Schätzer vorgestellt. Alle Schätzer verwenden einen Kern K_r und einen Glättungsparameter $h_r \in \mathbb{R}$, die sogenannte Bandbreite. Ein Kern der Ordnung $q \in \mathbb{N}$ ist eine Funktion mit den Eigenschaften

$$\int_{\mathbb{R}} u^j K_r(u) du = \begin{cases} 1 & \text{für } j = 0 \\ 0 & \text{für } j = 1, \dots, q-1 \\ c \neq 0 & \text{für } j = q \end{cases},$$

$$\int_{\mathbb{R}} K_r^2(u) du < \infty \text{ und } \lim_{|u| \rightarrow \infty} |u| K_r(u) = 0.$$

Nadaraya-Watson Schätzer

Dieser Schätzer wurde unabhängig voneinander von Nadaraya (1964) und Watson (1964) vorgeschlagen und ist gegeben durch

$$\hat{m}(x) = \frac{\frac{1}{nh_r} \sum_{i=1}^n K_r\left(\frac{x-X_i}{h_r}\right) Y_i}{\frac{1}{nh_r} \sum_{i=1}^n K_r\left(\frac{x-X_i}{h_r}\right)}. \quad (1.2)$$

Er ist also ein, abhängig von der Stelle x gewichtetes Mittel der Daten Y_i normiert mit dem Kerndichteschätzer der Designpunkte X_i . Datenpaare (X_i, Y_i) , für die X_i nahe bei x liegt erhalten mehr Gewicht als solche, die weiter entfernt sind. Man erhält diesen Schätzer unter anderem durch die gewichtete Kleinste-Quadrate Schätzung

$$\hat{m}(x) = \arg \min_a \sum_{i=1}^n (Y_i - a)^2 \frac{1}{h_r} K_r\left(\frac{x - X_i}{h_r}\right).$$

Für jeden Punkt x wird die Regressionsfunktion lokal mit einer konstanten Funktion approximiert. Aus diesem Grund wird der Nadaraya-Watson Schätzer auch als lokal konstanter Schätzer bezeichnet. Ein Nachteil dieses Schätzers ist der relativ große Bias, insbesondere am Rand des Beobachtungsbereichs $[0, 1]$ (Härdle, 1995; Kapitel 4.4).

Lokale Polynomschätzer

Diese Schätzer verallgemeinern das zuletzt genannte Prinzip des Nadaraya-Watson Schätzers. Ist die Regressionsfunktion k -mal differenzierbar, so kann sie in einer Umgebung von x durch ein Taylorpolynom p -ter Ordnung, $p \leq k$, approximiert werden,

$$m(x) \approx \sum_{j=0}^p \frac{m^{(j)}(z)}{j!} (x-z)^j = \sum_{j=0}^p \beta_j (x-z)^j$$

mit $\beta_j = m^{(j)}(z)/j!$. Anstatt eine lokal konstante Funktion an die Daten anzupassen, verwendet man nun lokal ein Polynom p -ten Grades ($p \leq k$). Ein lokaler Polynomschätzer ergibt sich daher durch die Minimierung

$$\hat{m}_p(x) = \arg \min_{\beta_0, \dots, \beta_p} \sum_{i=1}^n \left(Y_i - \sum_{j=0}^p \beta_j (x - X_i)^j \right)^2 \frac{1}{h_r} K_r \left(\frac{x - X_i}{h_r} \right).$$

Den Nadaraya-Watson Schätzer erhält man durch die Wahl $p = 0$. Gegenüber diesem besitzen lokale Polynomschätzer höherer Ordnung ($p > 0$) den Vorteil, dass die Anpassung am Rand des Intervalls $[0, 1]$ gewöhnlich besser ist (Fan und Gijbels, 1997). Lokale Polynomschätzer eignen sich außerdem gut um Ableitungen der Regressionsfunktion zu schätzen. Ein Schätzer der ν -ten Ableitung ($\nu \leq p$) ergibt sich aus der Taylorapproximation als $\hat{m}_{\nu,p}(x) = \nu! \hat{\beta}_\nu$. Es ist üblich, die Differenz $p - \nu$ zwischen Ordnung p und Grad der Ableitung ν ungerade zu wählen. In diesem Fall erhält man einen einfacheren Bias als für eine gerade Differenz. Der gebräuchlichste lokale Polynomschätzer der Regressionsfunktion ist der lokal lineare Schätzer ($p = 1$ und $\nu = 0$). Dieser hat die geschlossene Darstellung

$$\hat{m}_1(x) = \frac{S_{n,2}(x)T_{n,0}(x) - S_{n,1}(x)T_{n,1}(x)}{S_{n,2}(x)S_{n,0}(x) - S_{n,1}^2(x)} \quad (1.3)$$

mit

$$\begin{aligned} S_{n,j}(x) &= \frac{1}{h_r} \sum_{i=1}^n K_r \left(\frac{x - X_i}{h_r} \right) (X_i - x)^j \text{ und} \\ T_{n,j}(x) &= \frac{1}{h_r} \sum_{i=1}^n K_r \left(\frac{x - X_i}{h_r} \right) (X_i - x)^j Y_i. \end{aligned}$$

Gasser-Müller Schätzer

Gasser und Müller (1979) schlagen den Schätzer

$$\hat{m}(x) = \frac{1}{h_r} \sum_{i=1}^n \int_{S_{i-1}}^{S_i} K_r \left(\frac{x - u}{h_r} \right) du Y_i$$

mit $S_i = (X_i + X_{i+1})/2$ und $X_0 = -\infty$ vor. Im Vergleich zu dem Nadaraya-Watson Schätzer besitzt dieser Schätzer zwar einen kleineren Bias aber eine größere Varianz.

In der vorliegenden Arbeit werden im Weiteren der Nadaraya-Watson Schätzer und der lokale Polynomschätzer behandelt, die in den folgenden Kapiteln beschriebenen Methoden können jedoch auch auf andere Schätzer wie etwa den Gasser-Müller Schätzer übertragen werden.

1.3 Asymptotische Theorie

Geeignete Schätzer konvergieren gegen die wahre Regressionsfunktion, sie sind also konsistent. Der Bias und die Varianz sollten aber auch schon für einen endlichen Stichprobenumfang nicht zu groß sein. Für Aussagen über Konsistenz, Bias und Varianz eines Regressionsschätzers werden üblicherweise die folgenden Annahmen getroffen.

- (R1) $X_i, i = 1, \dots, n$ sind unabhängig identisch verteilt auf dem Intervall $[0, 1]$ mit Dichte f
- (R2) $\varepsilon_i, i = 1, \dots, n$ sind unabhängig identisch verteilt und unabhängig von der Folge $\{X_i\}_{i=1}^n$ mit $E[\varepsilon_1] = 0, E[\varepsilon_1^2] = 1$ und $E[\varepsilon_1^4] < \infty$
- (R3) Die Dichte f ist $q - 1$ -mal stetig differenzierbar ($f \in \mathcal{C}^{q-1}([0, 1])$)
- (R4) Die Regressionsfunktion ist q -mal stetig differenzierbar ($m \in \mathcal{C}^q([0, 1])$)
- (R5) Die Funktion der Standardabweichung ist stetig ($\sigma \in \mathcal{C}([0, 1])$)
- (K) K_r sei ein Kern der Ordnung q

Gleichmäßige Konsistenz von Kernschätzern wird unter anderem in Mack und Silverman (1982) oder Stone (1982) behandelt. Das Resultat von Mack und Silverman (1982) für den Nadaraya-Watson Schätzer sei hier exemplarisch genannt.

Satz 1.1 [Mack und Silverman (1982)] *Es gelten die Voraussetzungen (R1) - (R5) und (K) und darüberhinaus*

(i) *sei die gemeinsame Dichte $f_{X,Y}$ von X und Y stetig und erfülle*

$$\sup_x \int |y|^s f_{X,Y}(x, y) dx dy < \infty$$

für ein $s \geq 2$,

(ii) *erfülle der Kern $\int |x \log |x||^{1/2} |dK(x)| < \infty$ und sei symmetrisch,*

(iii) *seien die zweiten Ableitungen von m und f beschränkt und*

(iv) *gelte für die Folge $h_r = h_r(n)$ von Bandbreiten für $n \rightarrow \infty$ $h_r \rightarrow 0$ $n^{2\eta-1} h_r \rightarrow \infty$ für ein $0 < \eta < 1 - s^{-1}$, $h_r^2 = o(\log h_r^{-1}/nh_r)^{1/2}$ und $\sum_n h_r^\lambda$ für ein $\lambda > 0$.*

Dann ist der Nadaraya-Watson Schätzer auf jedem beschränkten Intervall J , auf dem f positiv ist konsistent und es gilt

$$\sup_{x \in J} |\hat{m}(x) - m(x)| = O\left(\frac{\log h_r^{-1}}{nh_r}\right)^{1/2} \text{ f.s.}$$

Der folgende Satz trifft Aussagen über den Bias und die Varianz des Nadaraya-Watson und des lokal linearen Schätzers (vgl. Härdle, 1995 bzw. Fan und Gijbels, 1997).

Satz 1.2 Die Voraussetzungen (R1)-(R5) und (K) seien erfüllt für $r = 2$ und es sei $x \in [h, 1 - h]$. Weiter bezeichne

$$\kappa_q(K_r) = \frac{(-1)^q}{q!} \int_{-1}^1 u^q K_r(u) du \quad \text{und} \quad b_{K_r} = \int_{-1}^1 K_r^2(u) du.$$

(i) Für den Bias und die Varianz des Nadaraya-Watson Schätzers (lokal konstanten Schätzers) gilt dann

$$\begin{aligned} \mathbb{E}[\hat{m}(x)] - m(x) &= h_r^q \kappa_q(K_r) \frac{(mf)^{(q)}(x) - mf^{(q)}(x)}{f(x)} + o(h_r^q) \\ \text{Var}(\hat{m}(x)) &= \frac{\sigma^2(x)}{nh_r f(x)} b_{K_r} + o\left(\frac{1}{nh_r}\right). \end{aligned}$$

(ii) Der lokal lineare Schätzer besitzt den Bias und die Varianz

$$\begin{aligned} \mathbb{E}[\hat{m}_1(x)] - m(x) &= h_r^q \kappa_q(K_r) m^{(q)}(x) + o(h_r^q) \\ \text{Var}(\hat{m}_1(x)) &= \frac{\sigma^2(x)}{nh_r f(x)} b_{K_r} + o\left(\frac{1}{nh_r}\right). \end{aligned}$$

Eine Schwierigkeit in der Methode der Kernschätzung besteht in der richtigen Wahl der Bandbreite. Zwei gebräuchliche Kriterien dazu werden nun vorgestellt.

Aus Satz 1.2 ergibt sich der mittlere quadratische Fehler (MSE) der Schätzer als Summe aus Varianz und quadriertem Bias,

$$\begin{aligned} \mathbb{E}[(\hat{m}(x) - m(x))^2] &= \frac{\sigma^2(x)}{nh_r f(x)} b_{K_r} + h_r^{2q} \kappa_q^2(K_r) \left(\frac{(mf)^{(q)}(x) - mf^{(q)}(x)}{f(x)} \right)^2 \\ &\quad + o(h_r^{2q}) + o\left(\frac{1}{nh_r}\right) \\ \mathbb{E}[(\hat{m}_1(x) - m(x))^2] &= \frac{\sigma^2(x)}{nh_r f(x)} b_{K_r} + h_r^{2q} \kappa_q^2(K_r) (m^{(q)}(x))^2 + o(h_r^{2q}) + o\left(\frac{1}{nh_r}\right). \end{aligned}$$

Ein Optimalitätskriterium für die Wahl der Bandbreite h_r besteht darin, dass der dominierende Term des MSE minimiert werden soll. Mit diesem Kriterium ist die lokal optimale Bandbreite des Nadaraya-Watson Schätzers gegeben durch

$$h_{r,opt} = \left(\frac{\sigma^2(x) b_{K_r}}{2q n ((mf)^{(q)}(x) - mf^{(q)}(x)) \kappa_q^2(K_r)} \right)^{1/2q+1}$$

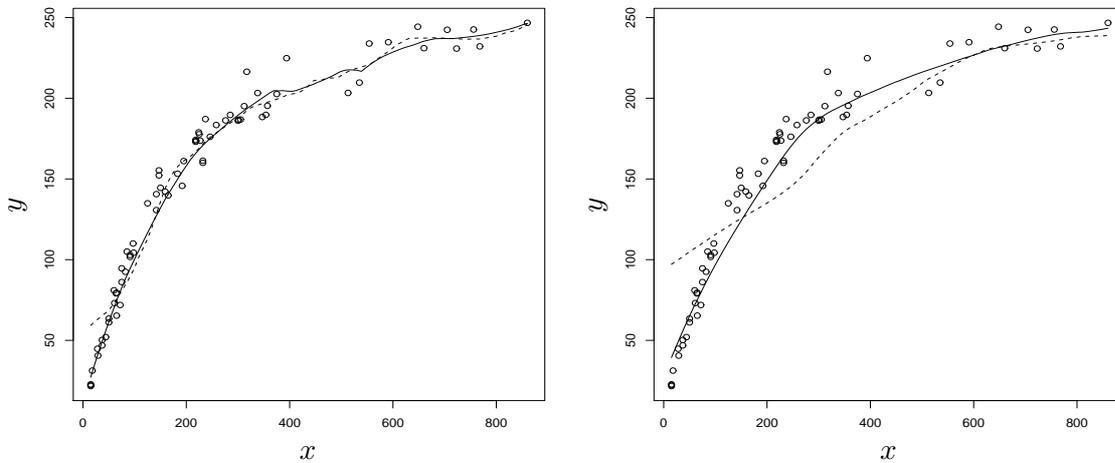


Abbildung 1.1: Daten aus Kapitel 4.3.2 mit dem lokal linearen Schätzer (durchgezogene Linie) und dem Nadaraya-Watson Schätzer (gestrichelte Linie). Die Bandbreite ist durch Least Squares Cross-Validation (linkes Bild) und die Methode von Rice (1984) (rechtes Bild) gewählt.

und für den lokal linearen Schätzer durch

$$h_{r,opt} = \left(\frac{\sigma^2(x) b_{K_r}}{2q n m^{(q)}(x) \kappa_q^2(K_r)} \right)^{1/2q+1}.$$

Die optimale Bandbreite hängt von den unbekannt Funktionen m , σ und f ab und muss daher in Anwendungen geschätzt werden. Dazu gibt es verschiedene Ansätze. Rice (1984) schlägt vor, zunächst die Varianzfunktion durch

$$\hat{\sigma}^2 = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} (Y_{(i+1)} - Y_{(i)})^2$$

zu schätzen. Dabei bezeichnen $Y_{(1)}, \dots, Y_{(n)}$ die zu der Ordnung $X_{(1)}, \dots, X_{(n)}$ gehörenden y -Daten. Der Bandbreitenschätzer wird proportional zu der optimalen Bandbreite als

$$\hat{h}_{r,opt} = \left(\frac{\hat{\sigma}^2}{n} \right)^{1/(2q+1)}$$

gewählt.

Ein weiterer Ansatz ist die Methode der Least Squares Cross-Validation. Hierbei bezeichne $\hat{m}_{p,-i}(x)$, $p = 0, 1$ den Nadaraya-Watson ($p = 0$) oder lokal linearen Schätzer ($p = 1$) der

aus den Daten ohne die i -te Beobachtung berechnet wird. Die Bandbreite wird nun so gewählt, dass sie den Ausdruck

$$CV(h_r) = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{m}_{p,-i}(X_i))^2$$

minimiert. Diese und weitere Methoden zur Wahl der Bandbreite werden ausführlich in Härdle (1995) beschrieben.

Abbildung 1.1 zeigt zu den Daten aus Abschnitt 4.3.2 sowohl den lokal linearen (durchgezogene Linie) als auch den Nadaraya-Watson Schätzer (gestrichelte Linie). Diese Daten sind Beobachtungen zum Verhältnis zwischen Alter und Trochengewicht der Augenlinse von wildlebenden Kaninchen in Australien. Im linken Bild wurde die Bandbreite mit Least Squares Cross-Validation gewählt. Beide Schätzer passen sich auf weiten Teilen des Intervalls gut den Daten an und zeigen ein ähnliches Verhalten. Der Effekt der schlechteren Anpassung des Nadaraya-Watson Schätzers am Rand wird am linken Rand des linken Bildes deutlich. Während der lokal lineare Schätzer dort sehr nah an den Daten liegt, überschätzt der Nadaraya-Watson Schätzer die Funktion erheblich. Wählt man die Bandbreite nach der Methode von Rice (1984), so ergibt sich das rechte Bild. Die Anpassung ist in diesem Beispiel mit relativ wenigen Daten schlechter als bei der Least Squares Cross-Validation. Wiederum zeigt der lokal lineare Schätzer am linken Rand des Beobachtungsbereichs ein besseres Verhalten als der Nadaraya-Watson Schätzer. Es wird an diesem Beispiel sehr deutlich, wie stark die Güte der Schätzung von der Wahl der Bandbreite abhängt.

Aus Satz 1.2 erhält man mit einem zentralen Grenzwertsatz für Dreieckschemata die asymptotische Verteilung der Schätzer. Das genaue Vorgehen wird zum Beispiel in Müller (1987) beschrieben. Im Folgenden bezeichne $Z_n \xrightarrow{\mathcal{D}} Z$ die schwache Konvergenz der Zufallsvariablen Z_n gegen die Zufallsvariable Z und $\mathcal{N}(\mu, \tau^2)$ eine Normalverteilung mit Erwartungswert μ und Varianz τ^2 .

Satz 1.3 *Es gelte für die Bandbreite $h_r \rightarrow 0$, $nh_r^{q+1} \rightarrow \infty$ und $h_r = O(n^{-1/(2q+1)})$ und die Bedingungen aus Satz 1.2 seien erfüllt.*

(i) *Für den Nadaraya-Watson Schätzer gilt*

$$\sqrt{nh_r} \left(\hat{m}(x) - m(x) - h_r^q \kappa_q(K_r) \frac{(mf)^{(q)}(x) - mf^{(q)}(x)}{f(x)} \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left(0, \frac{\sigma^2(x)}{f(x)} b_{K_r} \right).$$

(ii) *Für den lokal linearen Schätzer gilt*

$$\sqrt{nh_r} (\hat{m}(x) - m(x) - h_r^q \kappa_q(K_r) m^{(q)}(x)) \xrightarrow{\mathcal{D}} \mathcal{N} \left(0, \frac{\sigma^2(x)}{f(x)} b_{K_r} \right).$$

Kapitel 2

Monotone und konvexe Umordnungen von Funktionen

In diesem Kapitel wird ein Verfahren vorgestellt, mit dem eine beliebige Funktion monotonisiert wird, unter der Voraussetzung, dass ihre L^2 -Norm gleich bleibt (vgl. Bennett und Sharpley, 1988). Aus dieser Methode wird anschließend ein Verfahren hergeleitet, mit dem aus glatten Funktionen eine ebenfalls glatte Monotonisierung entsteht. Um die Übersicht zu bewahren, wird der Fall der isotonen Umordnung ausführlich behandelt, die analoge Methode der antitonen Umordnung anschließend in Bemerkung 2.6 berücksichtigt. Die hier beschriebene Methode findet in den folgenden Kapiteln unterschiedliche Anwendungen in der Regressionsschätzung. In Kapitel 3 wird sie verwendet um einen isotonen Schätzer einer einmal stetig differenzierbaren Regressionsfunktion zu konstruieren. Kapitel 4 beschäftigt sich mit der Anwendung von monotonen Umordnungen auf Ableitungen von Regressionsschätzern zur Konstruktion von konvexen und konkaven Schätzern. Schließlich wird aus dem inversen isotonen Schätzer und dem zugehörigen uneingeschränkten Schätzer in Kapitel 5 eine Teststatistik zum Testen auf strenge Isotonie der Regressionsfunktion hergeleitet.

2.1 Isotone Umordnungen

Definition 2.1 [Bennett and Sharpley (1988)]

(i) Die Verteilungsfunktion einer Funktion g auf $[0, 1]$ bezüglich des Lebesgue-Maßes ist definiert als

$$\phi(g)(t) = \int_{[0,1]} I\{g(v) \leq t\} dv, \quad t \in g([0, 1])$$

(ii) Die isotone Umordnung einer Funktion g ist definiert als

$$\phi(g)^{-1}(x) = \inf\{t \in g([0, 1]) \mid \phi(g)(t) \geq x\}$$

Bemerkung 2.2 (i) Fixpunkte des isotonen Umordnungsoperators sind die isotonen Funktionen, das heißt, es gilt für isotone Funktionen $\phi(g)^{-1} = g$. Diese Eigenschaft sieht man wie folgt. Ist g isoton, so ist die Verteilungsfunktion gegeben durch

$$\phi(g)(t) = \int_{[0,1]} I\{g(x) \leq t\} dx = \sup\{x \in [0,1] | g(x) \leq t\}.$$

Wendet man darauf die verallgemeinerte Inverse an, so ergibt sich die Gleichung

$$\phi(g)^{-1}(x) = \inf\{t \in g([0,1]) | \sup\{z \in [0,1] | g(z) \leq t\} \leq x\},$$

man wählt demnach das kleinste t , so dass das größte z mit $g(z) \leq t$ immer noch kleiner oder gleich einem vorgegebenen x ist. Der Wert t ist dadurch eindeutig bestimmt als $t = g(x)$.

(ii) Die isotope Umordnung erhält die L^p -Norm einer Funktion für alle $p > 0$, das bedeutet

$$\left(\int_0^1 |g(x)|^p dx \right)^{1/p} = \left(\int_0^1 |\phi(g)^{-1}(x)|^p dx \right)^{1/p}.$$

Zum Beweis dieser Aussage sei g ohne Beschränkung der Allgemeinheit eine Treppenfunktion, die allgemeineren Fälle folgen durch algebraische Induktion. Die Verteilungsfunktion einer Treppenfunktion

$$g(x) = \sum_{j=1}^k a_j I_{E_j}$$

mit Konstanten $-\infty < a_1 < \dots < a_k < \infty$ und einer disjunkten Zerlegung E_1, \dots, E_k des Intervalls $[0, 1]$ ist gegeben durch

$$\phi(g)(t) = \sum_{j=1}^k m_j I_{[a_j, a_{j+1})}(t),$$

$m_0 = 0$, $m_j = \sum_{i=1}^j \lambda(E_i)$ und λ bezeichne das Lebesgue-Maß. Daraus ergibt sich für die isotope Umordnung die Darstellung

$$\phi(g)^{-1}(x) = \sum_{j=1}^k a_j I_{(m_{j-1}, m_j]}(x).$$

Beachtet man, dass E_j , $j = 1, \dots, k$ disjunkte Mengen sind und nutzt die Definition der m_j aus, so sind die Normen der wahren Funktion und ihrer isotonen Umordnung gleich,

$$\begin{aligned} \left(\int_0^1 |g(x)|^p dx \right)^{1/p} &= \left(\sum_{j=1}^k |a_j|^p \lambda(E_j) \right)^{1/p} = \left(\sum_{j=1}^k |a_j|^p \lambda([m_j, m_{j-1})) \right)^{1/p} \\ &= \left(\int_0^1 |\phi(g)^{-1}(x)|^p dx \right)^{1/p}. \end{aligned}$$

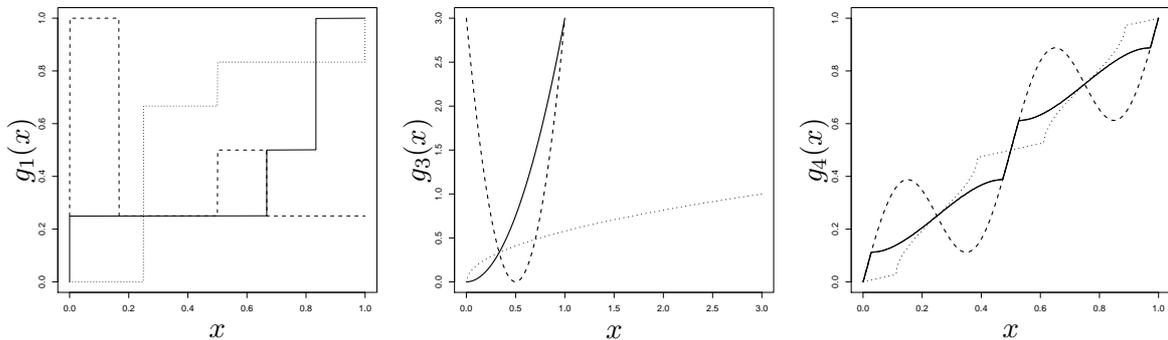


Abbildung 2.1: Die Funktionen g_1 , g_3 und g_4 (gestrichelte Linie) mit ihren Verteilungsfunktionen (gepunktete Linie) und den isotonen Umordnungen (durchgezogene Linie)

Beispiel 2.3 Dieses Beispiel veranschaulicht an verschiedenen Funktionen das Verhalten der isotonen Umordnung.

(i) Sehr gut ist das Prinzip der Umordnung bei einer Treppenfunktion zu sehen. Zu der Treppenfunktion

$$g_1(x) = \frac{1}{4}I_{[\frac{1}{6}, \frac{1}{2}] \cup [\frac{2}{3}, 1]} + \frac{1}{2}I_{[\frac{1}{2}, \frac{2}{3}]} + 1I_{[0, \frac{1}{6}]}$$

gehört die isotone Umordnung

$$\phi(g_1)^{-1}(x) = \frac{1}{4}I_{[0, \frac{2}{3}]} + \frac{1}{2}I_{(\frac{2}{3}, \frac{5}{6}]} + 1I_{(\frac{5}{6}, 1]}.$$

Die Blöcke der ursprünglichen Funktion werden in der isotonen Umordnung ihrer Höhe nach angeordnet. Dadurch werden die beiden Blöcke der Höhe $1/4$ zusammengefasst zu einem Block der Länge $2/3$ und an den Anfang des Intervalls gesetzt. Es folgt der Block der Höhe $1/2$ und zum Schluss der Block der Höhe 1 . Dies ist gut im linken Bild von Abbildung 2.1 zu erkennen. Dort ist die Funktion g_1 als gestrichelte Linie zusammen mit ihrer Verteilungsfunktion (gepunktete Linie) und der isotonen Umordnung (durchgezogene Linie) dargestellt. Auch das Prinzip der Normerhaltung wird hieran deutlich. Da es sich um eine reine Umordnung der Blöcke handelt, bleibt die Fläche unter dem Graphen der Funktion und somit auch die L^p -Norm gleich.

(ii) Die Funktion

$$g_2(x) = (x - 1)^2, \quad x \in [0, 1]$$

ist streng monoton fallend und ihre Verteilungsfunktion daher gegeben durch

$$\phi(g_2)(t) = \int_0^1 I\{v \geq -\sqrt{t} + 1\} dv = \sqrt{t}.$$

Daraus erhält man die isotone Umordnung als

$$\phi(g_2)^{-1}(x) = x^2.$$

(iii) Die Funktionen

$$g_3(x) = 3(2x - 1)^2$$

und

$$g_4(x) = x + \frac{1}{4} \sin(4\pi x)$$

sind weder monoton wachsend noch monoton fallend. Die Verteilungsfunktion von g_3 ist gegeben durch

$$\phi(g_3)(t) = \int_0^1 I\{3(2x - 1)^2 \leq t\} = \int_0^1 I\left\{\frac{1}{2}\left(1 - \sqrt{\frac{t}{3}}\right) \leq x \leq \frac{1}{2}\left(1 + \sqrt{\frac{t}{3}}\right)\right\} dx = \sqrt{\frac{t}{3}},$$

die isotone Umordnung von g_3 lautet demnach

$$\phi(g_3)^{-1}(x) = 3x^2.$$

Das mittlere und rechte Bild in Abbildung 2.1 zeigen die wahren Funktionen g_3 und g_4 als gestrichelte Linie mit den zugehörigen Verteilungsfunktionen, dargestellt als gepunktete Linie und den isotonen Umordnungen als durchgezogene Linie. Es fällt auf, dass die isotone Umordnung von g_4 auf den Intervallen mit der wahren Funktion übereinstimmt, auf denen diese injektiv ist. Weiter ist anzumerken, dass die isotone Umordnung einer nicht isotonen aber differenzierbaren Funktion nicht mehr notwendig differenzierbar ist. Diese Eigenschaft ist im Zusammenhang der Regressionsschätzung nicht wünschenswert. Der Schätzer einer glatten Funktion sollte ebenfalls glatt sein. Es ist jedoch möglich, Definition 2.1 mit Hilfe einer Kern-Glättung der Verteilungsfunktion so zu verändern, dass die isotone Umordnung ebenfalls glatt ist.

Definition 2.4 Sei K_d ein positiver, differenzierbarer Kern auf $[-1, 1]$ und h_d die zugehörige Bandbreite.

(i) Die geglättete Verteilungsfunktion von g bezüglich des Lebesgue-Maßes ist definiert als

$$\phi_{h_d}(g)(t) = \int_{[0,1]} \int_{-\infty}^t \frac{1}{h_d} K_d\left(\frac{g(v) - u}{h_d}\right) dudv$$

(ii) Die geglättete isotone Umordnung von g ist definiert als

$$\phi_{h_d}(g)^{-1}(x) = \inf\{t \in \mathbb{R} | \phi_{h_d}(g)(t) \geq x\}$$

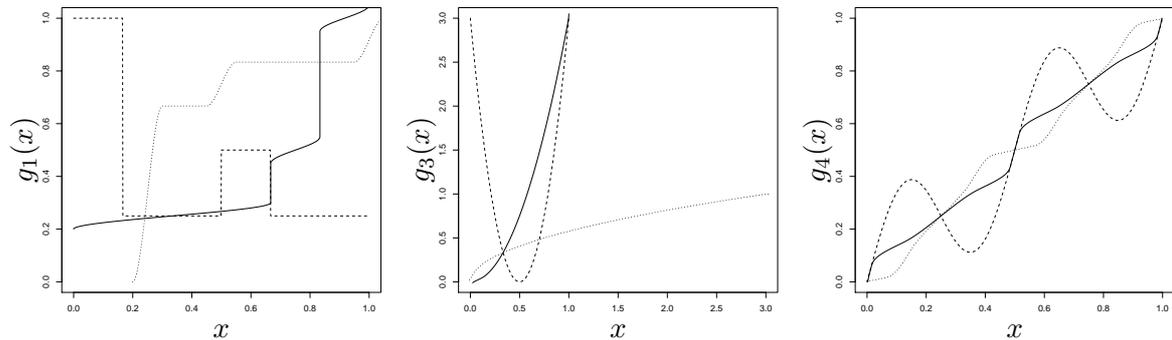


Abbildung 2.2: Die Funktionen g_1 (links), g_3 (Mitte) und g_4 (rechts) (gestrichelte Linie) mit ihren geglätteten Verteilungsfunktionen (gepunktete Linie) und den geglätteten isotonen Umordnungen (durchgezogene Linie) zu der Bandbreite $h_d = 0.05$

Das Vorgehen in Definition 2.4 ist dadurch begründet, dass die Glättung der Verteilungsfunktion im Gegensatz zu der wahren Verteilungsfunktion durch Differenzierbarkeitsannahmen an die Kernfunktion immer hinreichend glatt gewählt werden kann und darüberhinaus eine geeignete Approximation der Verteilungsfunktion ist. Aus der geglätteten Verteilungsfunktion wird im zweiten Schritt die isotone Umordnung berechnet. Diese ist nun ebenfalls glatt, approximiert die wahre isotone Umordnung jedoch nur. Durch geeignete Wahl der Bandbreite h_d kann diese Approximation verbessert werden. Die genaue Wahl der Bandbreite wird in den folgenden Kapiteln ausführlicher diskutiert. Die Approximation hat jedoch nun zur Folge, dass die isotone Umordnung einer bereits isotonen Funktion nicht mehr exakt mit dieser übereinstimmt. Wie weit die geglättete isotone Umordnung in diesem Fall von der wahren Funktion abweicht, hängt wiederum von der Bandbreitenwahl ab. Aus demselben Grund stimmt die L^p -Norm der Umordnung nicht mehr exakt mit der L^p -Norm der ursprünglichen Funktion überein. Die erste Ableitung der Verteilungsfunktion ergibt

$$\frac{1}{h_d} \int_{[0,1]} K_d\left(\frac{g(v) - t}{h_d}\right) dv.$$

Dieser Ausdruck ist positiv, da der Kern K_d positiv gewählt wird. Die geglättete Verteilungsfunktion ist aus diesem Grund monoton wachsend und sogar streng monoton wachsend, falls h_d klein genug gewählt wird.

Abbildung 2.2 zeigt die Funktionen g_1 , g_3 und g_4 (gestrichelte Linie) aus Beispiel 2.3 zusammen mit ihren geglätteten isotonen Umordnungen als durchgezogene Linie und den zugehörigen geglätteten Verteilungsfunktionen (gepunktete Linie) für die Bandbreite $h_d = 0.05$. Die isotone Umordnung der stückweise konstanten Funktion g_1 in Beispiel 2.3 ist wieder eine stückweise konstante Funktion. Auf der rechten Seite von Abbildung 2.2

erkennt man, dass im Gegensatz dazu die geglättete isotone Umordnung streng monoton wachsend ist. Zwischen der isotonen und der geglätteten isotonen Umordnung der Funktion g_3 in der Mitte von Abbildung 2.2 lassen sich keine wesentlichen Unterschiede entdecken. Die isotone Umordnung dieser Funktion ist bereits glatt, eine zusätzliche Glättung ändert diese daher kaum. An der geglätteten isotonen Umordnung von g_4 ist zu sehen, dass sie auch auf den Bereichen, auf denen die wahre Funktion streng monoton wachsend und injektiv ist, nicht mehr exakt mit dieser übereinstimmt. Die Stellen, an denen die isotone Umordnung nicht differenzierbar ist, werden hier geglättet.

Das folgende Lemma zeigt, dass die geglättete isotone Umordnung einer Funktion für eine monoton fallende Folge von Bandbreiten gegen die isotone Umordnung konvergiert.

Lemma 2.5 *Unter der Voraussetzung $h_d \rightarrow 0$ gilt für jede Funktion g und für alle $x \in [0, 1]$*

$$\phi_{h_d}(g)^{-1}(x) \rightarrow \phi(g)^{-1}(x).$$

Beweis Zunächst wird gezeigt, dass die geglättete Verteilungsfunktion ϕ_{h_d} der Funktion g gegen die Verteilungsfunktion $\phi(g)$ konvergiert.

$$\begin{aligned} \phi_{h_d}(g)(t) &= \frac{1}{h_d} \int_0^1 \int_{-\infty}^t K_d\left(\frac{g(v) - u}{h_d}\right) dudv \\ &= \frac{1}{h_d} \int_0^1 I\{g(v) \leq t + h_d\} \int_{g(v)-h_d}^t K_d\left(\frac{g(v) - u}{h_d}\right) dudv \\ &= \int_0^1 I\{g(v) \leq t + h_d\} \int_{(g(v)-t)/h_d}^1 K_d(u) dudv \\ &= \int_0^1 I\{g(v) \leq t + h_d\} I\{g(v) \leq t - h_d\} dv \\ &\quad + \int_0^1 I\{t - h_d \leq g(v) \leq t + h_d\} \int_{(g(v)-t)/h_d}^1 K_d(u) dudv \\ &= \int_0^1 I\{g(v) \leq t - h_d\} dv \\ &\quad + \int_0^1 I\{t - h_d \leq g(v) \leq t + h_d\} \int_{(g(v)-t)/h_d}^1 K_d(u) dudv. \end{aligned}$$

Der erste Term konvergiert für $h_d \rightarrow 0$ gegen die Verteilungsfunktion $\phi(g)$, der zweite Term konvergiert gegen 0. Wegen der Stetigkeit des Funktional, das eine Funktion auf ihre verallgemeinerte Inverse abbildet, konvergiert auch die geglättete isotone Umordnung gegen die isotone Umordnung.

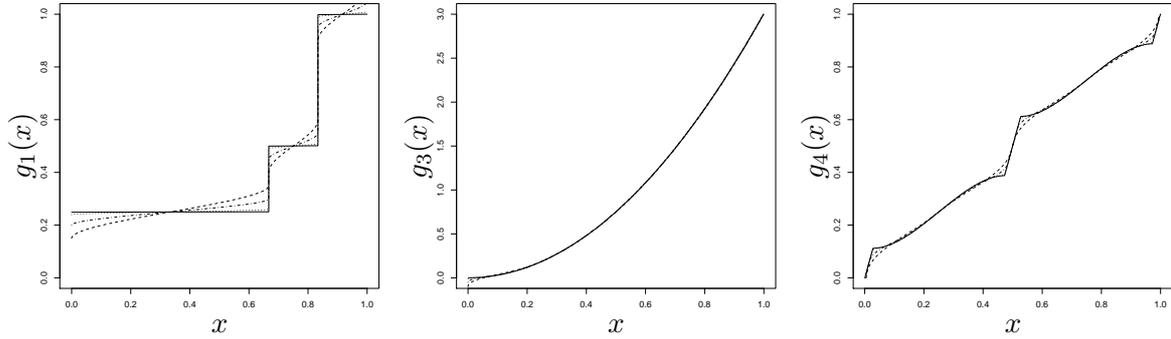


Abbildung 2.3: Die isotonen Umordnungen der Funktionen g_1 (links), g_3 (Mitte) und g_4 (rechts) (durchgezogene Linie) und ihre geglätteten isotonen Umordnungen zu den Bandbreiten $h_d = 0.1$ (gestrichelte Linie) $h_d = 0.05$ (Strich-Punkt Linie) und $h_d = 0.01$ (gepunktete Linie)

In Abbildung 2.3 sind die isotonen Umordnungen der Funktionen g_1 , g_3 und g_4 und ihre geglätteten isotonen Umordnungen für die Bandbreiten $h_d = 0.1$, 0.05 und 0.01 dargestellt. Man erkennt gut die Konvergenz der letzteren gegen die isotonen Umordnungen. Während bei den Funktionen g_1 und g_4 an den Stellen, an denen die isotone Umordnung nicht differenzierbar ist, noch Unterschiede, bedingt durch die Wahl von h_d , zu erkennen sind, zeigen die geglätteten isotonen Umordnungen der Funktion g_3 für verschiedene Bandbreiten h_d ein nahezu identisches Verhalten.

Bemerkung 2.6 (i) Analog zu dem Vorgehen bei der isotonen Umordnung ist die anti-tone Umordnung einer Funktion g gegeben durch

$$\varphi(g)^{-1}(x) = \inf\{t \in \mathbb{R} | \varphi(g)(t) \leq x\}$$

mit

$$\varphi(g)(t) = \int_0^1 I\{g(v) \geq t\} dv$$

und auch die geglättete Version lässt sich hier definieren. Es ist dann

$$\varphi_{h_d}(g)^{-1}(x) = \inf\{t \in \mathbb{R} | \varphi_{h_d}(g)(t) \leq x\}$$

mit

$$\varphi_{h_d}(g)(t) = \int_0^1 \int_t^\infty \frac{1}{h_d} K_d\left(\frac{g(v) - u}{h_d}\right) dudv. \quad (2.1)$$

(ii) Die isotonen Umordnungen sind in Definition 2.1 nur für das Lebesgue-Maß formuliert. Sie können jedoch auch für jedes beliebige Maß μ bestimmt werden. Anstatt in der Definition der Verteilungsfunktion bezüglich des Lebesgue-Maßes zu integrieren, wird dann bezüglich des Maßes μ integriert. Dieser allgemeinere Fall wird in Bennett und Sharpley (1988) detailliert beschrieben.

2.2 Konvexe Umordnungen

Die Theorie der isotonen Umordnung aus dem vorhergehenden Abschnitt wird hier verwendet, um konvexe Umordnungen von differenzierbaren Funktionen zu definieren. Eine differenzierbare Funktion m ist genau dann (strikt) konvex, wenn ihre Ableitung (streng) monoton wachsend ist. Daher wird die konvexe Umordnung einer Funktion definiert über die Stammfunktion der isotonen Umordnung ihrer Ableitung. Nach Bemerkung 2.2 (ii) besitzt die Ableitung der konvexen Umordnung somit dieselbe L^p -Norm wie die Ableitung der ursprünglichen Funktion.

Definition 2.7 Sei m eine auf dem offenen Intervall $(0, 1)$ differenzierbare Funktion.

(i) Für $a \in [0, 1]$ ist eine konvexe Umordnung von m gegeben durch

$$\rho(m)(x, a) = \int_a^x \phi(m')^{-1}(z) dz + m(a).$$

(ii) Eine geglättete konvexe Umordnung von m ist definiert als

$$\rho_{h_d}(m)(x, a) = \int_a^x \phi_{h_d}(m')^{-1}(z) dz + m(a).$$

Die konvexe Umordnung einer Funktion ist nicht eindeutig bestimmt, sondern hängt von der Wahl der Integralgrenze a ab. Je nach Zusammenhang und Funktion erweisen sich verschiedene Grenzen als sinnvoll. Um nicht in jeder Situation neu über die Wahl von a entscheiden zu müssen, ist die Definition einer optimalen konvexen Umordnung hilfreich. Als Optimalitätskriterium dient hier der L^2 -Abstand der Umordnung zu der wahren Funktion. Es sind jedoch auch andere Abstandsmaße denkbar. Folgender Satz liefert eine geschlossene Formel für die L^2 -optimale konvexe Umordnung.

Satz 2.8 Es existiert eine L^2 -optimale (geglättete) konvexe Umordnung, die unabhängig von der Wahl von a ist und sich berechnen lässt als

$$\rho(m)(x) = \rho(m)(x, a^*) = \int_0^1 \rho(m)(x, a) da \quad (2.2)$$

bzw.

$$\rho_{h_d}(m)(x) = \rho_{h_d}(m)(x, a^*) = \int_0^1 \rho_{h_d}(m)(x, a) da. \quad (2.3)$$

Beweis Hier wird nur der Fall der konvexen Umordnung betrachtet. Die Aussage für die geglättete konvexe Umordnung folgt analog. Es ist nachzuweisen, dass eine konvexe Umordnung existiert, die den L^2 -Abstand

$$\int_0^1 (\rho(m)(x, a) - m(x))^2 dx$$

minimiert. Dazu bezeichne Φ die Stammfunktion von $\phi(m')^{-1}$. Mit dieser Bezeichnung lässt sich die konvexe Umordnung schreiben als

$$\rho(m)(x, a) = \Phi(x) - \Phi(a) + m(a)$$

und mit der Bezeichnung $\Delta(x) := m(x) - \Phi(x)$ ist das Minimum des L^2 -Abstands

$$\min_{a \in [0,1]} \int_0^1 (\rho(m)(x, a) - m(x))^2 dx = \min_{a \in [0,1]} \int_0^1 (\Delta(a) - \Delta(x))^2 dx.$$

Wegen der Stetigkeit von Δ kann in obiger Gleichung anstatt über alle $a \in [0, 1]$ auch über alle $b \in B = \Delta([0, 1])$ minimiert werden. Es ist dann

$$\min_{a \in [0,1]} \int_0^1 (\rho(m)(x, a) - m(x))^2 dx = \min_{b \in B} \int_0^1 (b - \Delta(x))^2 dx$$

und die Minimalstelle ist offensichtlich

$$b^* = \int_0^1 \Delta(x) dx.$$

Erneut wegen der Stetigkeit von Δ existiert nun mindestens eine Stelle a^* , mit $\Delta(a^*) = b^*$. Somit gibt es eine konvexe Umordnung, die den L^2 -Abstand minimiert. Zum Beweis der restlichen Aussage beachte man, dass nach obiger Diskussion

$$\begin{aligned} \rho(m)(x, a^*) &= \Phi(x) - \Phi(a^*) + m(a^*) = \Delta(a^*) + \Phi(x) \\ &= \int_0^1 \Delta(a) da + \Phi(x) = \int_0^1 (\Delta(a) + \Phi(x)) da \\ &= \int_0^1 \rho(m)(x, a) da \end{aligned}$$

gilt. Diese Gleichung beweist zum einen, dass die L^2 -optimale Umordnung unabhängig ist von der Wahl von a , und liefert zum anderen die Darstellung in (2.2). \square

Bemerkung 2.9 Ist eine differenzierbare Funktion bereits konvex, so implizieren die Eigenschaften der isotonen Umordnung, dass jede konvexe Umordnung und somit auch die L^2 -optimale Umordnung der ursprünglichen Funktion gleicht. Ähnlich wie schon bei den geglätteten isotonen Umordnungen gilt diese Aussage für geglättete konvexe Umordnungen nur noch approximativ.

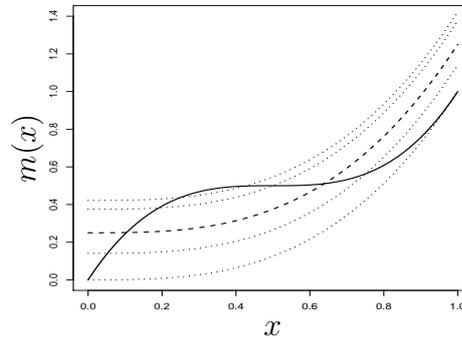


Abbildung 2.4: Die Funktion $m(x) = (2x - 1)^3/2 + 1/2$ (durchgezogene Linie) mit vier konvexen Umordnungen (gepunktete Linien) für $a = 0, 0.25, 0.5$ und 0.75 und der L^2 -optimalen Umordnung (gestrichelte Linie)

Beispiel 2.10 Die Funktion

$$m(x) = \frac{1}{2}(2x - 1)^3 + \frac{1}{2}$$

ist nicht konvex. Die isotone Umordnung ihrer Ableitung $m'(x) = 3(2x - 1)^2$ ist nach Beispiel 2.3 gegeben durch $\phi(m')^{-1}(x) = 3x^2$. Eingesetzt in die Definition der konvexen Umordnung ergibt dies

$$\rho(m)(x, a) = x^3 - a^3 + \frac{1}{2}(2a - 1)^3 + \frac{1}{2}.$$

Nach (2.2) erhält man für die L^2 -optimale Umordnung

$$\rho(m)(x) = x^3 + \frac{1}{4}.$$

Die optimale Umordnung erhält man auch durch die Integralgrenzen $a_1^* \approx 0.103741$ und $a_2^* \approx 0.638824$.

Abbildung 2.4 zeigt die Funktion m als durchgezogene Linie mit vier konvexen Umordnungen für $a = 0, 0.25, 0.5$ und 0.75 dargestellt als gepunktete Linien und der L^2 -optimalen konvexen Umordnung als gestrichelte Linie.

Bemerkung 2.11 Um eine (geglättete) konkave Umordnung zu erhalten wird an Stelle der isotonen Umordnung die (geglättete) antitone Umordnung aus Bemerkung 2.6 auf die Ableitung der Funktion m angewandt. Es ist dann

$$\varrho(m)(x, a) = \int_a^x \varphi(m')^{-1}(z) dz + m(a),$$

beziehungsweise

$$\varrho_{h_d}(m)(x, a) = \int_a^x \varphi_{h_d}(m')^{-1}(z) dz + m(a).$$

In Analogie zu Satz 2.8 existiert eine L^2 -optimale konkave Umordnung gegeben durch

$$\varrho(m)(x) = \varrho(m)(x, a^*) = \int_0^1 \varrho(m)(x, a) da$$

beziehungsweise

$$\varrho_{h_d}(m)(x) = \varrho_{h_d}(m)(x, a^*) = \int_0^1 \varrho_{h_d}(m)(x, a) da.$$

Kapitel 3

Schätzen von isotonen Regressionsfunktionen

In diesem Kapitel wird die Theorie der isotonen Umordnung aus Kapitel 2.1 auf einen uneingeschränkten Schätzer der Regressionsfunktion angewandt, um einen isotonen Schätzer der Regressionsfunktion zu erhalten. Bezeichnet \hat{m} den ursprünglichen Schätzer der Regressionsfunktion, so ist mit den Bezeichnungen aus dem vorhergehenden Kapitel die (geglättete) Verteilungsfunktion von \hat{m} gegeben durch

$$\hat{\phi}(t) = \phi(\hat{m})(t) \quad \text{bzw.} \quad \hat{\phi}_{h_d}(t) = \phi(\hat{m})(t). \quad (3.1)$$

Der (geglättete) isotone Schätzer wird im Folgenden geschrieben als

$$\hat{\phi}^{-1}(x) = \phi(\hat{m})^{-1}(x) \quad \text{bzw.} \quad \hat{\phi}_{h_d}^{-1}(x) = \phi_{h_d}(\hat{m})^{-1}(x). \quad (3.2)$$

Die gerade definierten isotonen Schätzer sind daher (geglättete) Umordnungen des uneingeschränkten Schätzers. Als uneingeschränkte Schätzer der Regressionsfunktion kommen alle Arten von Schätzern der Regressionsfunktion in Frage. Beispiele dafür sind Kern-Schätzer wie der Nadaraya-Watson Schätzer, der Gasser-Müller Schätzer oder lokale Polynomschätzer, aber auch Schätzer basierend auf Splines, Fourier-Reihen Entwicklung oder Wavelets sind als Anfangsschätzung denkbar.

3.1 Asymptotik des isotonen Kernschätzers

Im Folgenden wird das asymptotische Verhalten des isotonen geglätteten Schätzers unter verschiedenen Annahmen an den Grad der Differenzierbarkeit der wahren Regressionsfunktion untersucht. Dabei beschränken sich die Aussagen auf den Nadaraya-Watson Schätzer als uneingeschränkte Schätzung der Regressionsfunktion. Liegt das Regressionsmodell (1.1) vor, so haben Dette, Neumeyer und Pilz (2006) für eine streng monoton wachsende und zweimal stetig differenzierbare Regressionsfunktion m gezeigt, dass der isotone Schätzer in Verteilung gegen eine Normalverteilung konvergiert.

Satz 3.1 [Dette, Neumeyer, Pilz (2006)] *Es gelte $m, f \in C^2([0, 1])$, $\sigma^2 \in C([0, 1])$, und die Bandbreiten erfüllen $\log h_r^{-1}/(nh_r h_d^3) = o(1)$, $h_d, h_r \rightarrow 0$ für $n \rightarrow \infty$.*

(a) *Existiert darüberhinaus $\lim_{n \rightarrow \infty} \frac{h_d}{h_r} = c \in (0, \infty]$, so gilt für alle $x \in (0, 1)$ mit $m'(x) > 0$*

$$\begin{aligned} & \sqrt{nh_d} \left(\hat{\phi}_{h_d}^{-1}(x) - m(x) - \kappa_2(K_d) h_d^2 \frac{m''(x)}{(m'(x))^2} - \kappa_2(K_r) h_r^2 \left(\frac{m'' f + 2m' f'}{f} \right)(x) \right) \\ & \xrightarrow{\mathcal{D}} \mathcal{N}(0, s^2(x)), \end{aligned}$$

für $\kappa_2(K_r) = \frac{1}{2} \int_{-1}^1 u^2 K_r(u) du$ und die asymptotische Varianz ist gegeben durch

$$s^2(x) = \frac{\sigma^2(x) m'(x)}{f(x)} \int \int \int K_d(w + c m'(x)(v - u)) K_d(w) K_r(u) K_r(v) dw du dv. \quad (3.3)$$

(b) *Ist $\lim_{n \rightarrow \infty} \frac{h_d}{h_r} = 0$, so folgt für jedes $x \in (0, 1)$ mit $m'(x) > 0$*

$$\sqrt{nh_r} \left(\hat{\phi}_{h_d}^{-1}(x) - m(x) - \kappa_2(K_r) h_r^2 \left(\frac{m'' f + 2m' f'}{f} \right)(x) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tilde{s}^2(x))$$

und die asymptotische Varianz ist gegeben durch

$$\tilde{s}^2(x) = \frac{\sigma^2(x)}{f(x)} \int K_r^2(u) du. \quad (3.4)$$

Dieser Satz zeigt, dass der isotone Schätzer in der hier betrachteten Situation unter der Voraussetzung $h_d/h_r \rightarrow 0$ in erster Ordnung asymptotisch äquivalent ist zu dem uneingeschränkten Schätzer, da der Bias der beiden Schätzer bis auf einen vernachlässigbaren Term gleich ist und die asymptotischen Verteilungen dieselbe Varianz besitzen. In der Situation von Teil a) des Satzes ist dies nicht der Fall. Sowohl die Grenzverteilung als auch die Konvergenzrate der beiden Schätzer unterscheiden sich in diesem Fall. Die Annahme der zweimaligen stetigen Differenzierbarkeit der Regressionsfunktion und der Dichte wird in dem Beweis von Satz 3.1 an mehreren Stellen verwendet. Das bedeutet, dass auf Grundlage dieses Satzes asymptotische Verteilungsaussagen des isotonen Schätzers nur für Regressionsfunktionen getroffen werden können, die zweimal oder öfter differenzierbar sind. Für den Nadaraya-Watson Schätzer existieren asymptotische Verteilungsaussagen auch, falls m nur einmal stetig differenzierbar ist. Im Folgenden wird nun das asymptotische Verhalten des isotonen Schätzers untersucht, falls m einmal stetig differenzierbar ist, und mit der asymptotischen Verteilung des Nadaraya-Watson Schätzers unter dieser Voraussetzung verglichen. In dieser Situation sei K_r ein Kern der Ordnung 1 und ebenfalls

einmal stetig differenzierbar. Der Kern K_d sei weiterhin von der Ordnung zwei, symmetrisch und zweimal stetig differenzierbar auf dem Intervall $[-1, 1]$. Darüberhinaus gelte nun noch $K_d(\pm 1) = 0$ und K_d'' sei von 0 weg beschränkt. Gilt für die Bandbreiten

$$h_r, h_d = o(1), nh_r, nh_d \rightarrow \infty \quad (3.5)$$

$$h_r = o(h_d) \quad (3.6)$$

$$nh_r^3 = O(1) \quad (3.7)$$

$$nh_d^{3/(1-3\delta)} = O(1), 0 < \delta < \frac{1}{12} \quad (3.8)$$

$$\frac{\log h_r^{-1}}{nh_r h_d^2} = o(1), \quad (3.9)$$

so kann für den isotonen Schätzer asymptotische Normalität nachgewiesen werden. Das Resultat ist in den folgenden beiden Sätzen zusammengefasst. Es gilt für den inversen isotonen Schätzer $\hat{\phi}_{h_d} = \phi_{h_d}(\hat{m})$

Satz 3.2 *Erfüllen die Bandbreiten die Bedingungen (3.5)-(3.9) und ist die Dichte $f \in \mathcal{C}^1([0, 1])$, so gilt für eine streng monoton wachsende Regressionsfunktion $m \in \mathcal{C}^1([0, 1])$, und für alle $t \in (m(0), m(1))$ mit $m'(m^{-1}(t)) > 0$*

$$\sqrt{nh_d}(\hat{\phi}_{h_d}(t) - m^{-1}(t) + h_r a_{K_d, K_r}(t) - h_d b_{K_d}(t)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tau^2(t))$$

mit

$$\begin{aligned} a_{K_d, K_r}(t) &= \int_{-1}^1 K_d(v) \int_{-1}^1 u K_r(u) \frac{m'(m^{-1}(t + h_d v) + h_r \mu u)}{m'(m^{-1}(t + h_d v))} dudv \\ b_{K_d}(t) &= \int_{-1}^1 u K_d(u) (m^{-1})'(t + h_d \lambda u) du \\ \tau^2(t) &= \frac{\sigma^2(m^{-1}(t))}{f(m^{-1}(t))m'(m^{-1}(t))} \int_{-1}^1 K_d^2(y) dy \end{aligned}$$

und Konstanten $\lambda, \mu \in [-1, 1]$.

Beweis Es sei

$$g(z) = \int_z^\infty K_d(u) du.$$

Dann gilt $g'(z) = -K_d(z)$ und $g''(z) = -K_d'(z)$ und die Differenz $\hat{\phi}_{h_d}(t) - m^{-1}(t)$ kann wie in Dette, Neumeyer und Pilz (2006) mit Hilfe einer Taylorentwicklung geschrieben werden als

$$\hat{\phi}_{h_d}(t) - m^{-1}(t) = -m^{-1}(t) + \int_0^1 \int_{\frac{\hat{m}(v)-t}{h_d}}^\infty K_d(u) dudv = -m^{-1}(t) + \int_0^1 g\left(\frac{\hat{m}(v)-t}{h_d}\right) dv$$

$$\begin{aligned}
&= -m^{-1}(t) + \phi_{h_d}(m)(t) - \frac{1}{h_d} \int_0^1 K_d\left(\frac{m(v) - t}{h_d}\right) (\hat{m}(v) - m(v)) dv \\
&\quad - \frac{1}{2h_d^2} \int_0^1 K'_d\left(\frac{\xi(v) - t}{h_d}\right) (\hat{m}(v) - m(v))^2 dv \\
&= \Delta_n^{(0)}(t) + \Delta_n^{(1)}(t) + \frac{1}{2} \Delta_n^{(2)}(t)
\end{aligned} \tag{3.10}$$

mit

$$\Delta_n^{(0)}(t) = \phi_{h_d}(m)(t) - m^{-1}(t) \tag{3.11}$$

$$\Delta_n^{(1)}(t) = -\frac{1}{h_d} \int_0^1 K_d\left(\frac{m(v) - t}{h_d}\right) (\hat{m}(v) - m(v)) dv \tag{3.12}$$

$$\Delta_n^{(2)}(t) = -\frac{1}{h_d^2} \int_0^1 K'_d\left(\frac{\xi(v) - t}{h_d}\right) du (\hat{m}(v) - m(v))^2 dv \tag{3.13}$$

und $|\xi(v) - m(v)| \leq |\hat{m}(v) - m(v)|$. In einem ersten Schritt wird die Differenz $\phi_{h_d}(m)(t) - m^{-1}(t)$ berechnet. Wie in Dette, Neumeyer und Pilz (2006) erhält man

$$\begin{aligned}
\Delta_n^{(0)}(t) &= \phi_{h_d}(m)(t) - m^{-1}(t) \\
&= m^{-1}(t - h_d) + h_d \int_{-1}^1 (m^{-1})'(t + h_d z) \int_z^1 K_d(v) dv dz - m^{-1}(t).
\end{aligned}$$

Mit partieller Integration und einer Taylorentwicklung ergibt dies

$$\begin{aligned}
\Delta_n^{(0)}(t) &= m^{-1}(t - h_d) + \left(m^{-1}(t + h_d z) \int_z^1 K_d(v) dv \right) \Big|_{-1}^1 \\
&\quad + \int_{-1}^1 K_d(z) m^{-1}(t + h_d z) dz - m^{-1}(t) \\
&= h_d \int_{-1}^1 z K_d(z) (m^{-1})'(t + h_d \lambda z) dz \\
&= b_{K_d}(t)
\end{aligned} \tag{3.14}$$

für ein $\lambda \in [0, 1]$. Dies ist ein Teil des Bias in Satz 3.2.

Der Term $\Delta_n^{(1)}(t)$ liefert die asymptotische Verteilung. Um die Aussage zu beweisen, wird er weiter zerlegt in

$$\Delta_n^{(1)}(t) = \left(\Delta_n^{(1.1)}(t) + \Delta_n^{(1.2)}(t) \right) (1 + o_p(1)) \tag{3.15}$$

mit

$$\begin{aligned}
\Delta_n^{(1.1)}(t) &= -\frac{1}{nh_r h_d} \sum_{i=1}^n \int_0^1 K_d\left(\frac{m(v) - t}{h_d}\right) K_r\left(\frac{v - X_i}{h_r}\right) \frac{m(X_i) - m(v)}{f(v)} dv \\
\Delta_n^{(1.2)}(t) &= -\frac{1}{nh_r h_d} \sum_{i=1}^n \int_0^1 K_d\left(\frac{m(v) - t}{h_d}\right) K_r\left(\frac{v - X_i}{h_r}\right) \frac{\sigma(X_i) \varepsilon_i}{f(v)} dv.
\end{aligned}$$

Der Erwartungswert von $\Delta_n^{(1.1)}(t)$ ist

$$\begin{aligned}
\mathbb{E}[\Delta_n^{(1.1)}(t)] &= -\frac{1}{nh_r h_d} \sum_{i=1}^n \mathbb{E} \left[\int_0^1 K_d \left(\frac{m(v) - t}{h_d} \right) K_r \left(\frac{v - X_i}{h_r} \right) \frac{m(X_i) - m(v)}{f(v)} dv \right] \\
&= -\frac{1}{h_r h_d} \int_0^1 \int_0^1 K_d \left(\frac{m(v) - t}{h_d} \right) K_r \left(\frac{v - y}{h_r} \right) \frac{m(y) - m(v)}{f(v)} dv f(y) dy \\
&= -\frac{1}{h_r h_d} \int_{m^{-1}(t-h_d)}^{m^{-1}(t+h_d)} \int_{v-h_r}^{v+h_r} K_d \left(\frac{m(v) - t}{h_d} \right) K_r \left(\frac{v - y}{h_r} \right) \frac{m(y) - m(v)}{f(v)} f(y) dy dv \\
&= -h_r \int_{-1}^1 \int_{-1}^1 K_d(v) y K_r(y) \frac{m'(m^{-1}(t + h_d v) - h_r \mu y)}{(m^{-1})'(t + h_d v)} f(m^{-1}(t + h_d v)) \\
&\quad \times \left\{ 1 + h_r y \frac{f'(m^{-1}(t + h_d v) - h_r \nu y)}{f(m^{-1}(t + h_d v))} \right\} dv \\
&= -h_r \int_{-1}^1 K_d(v) \int_{-1}^1 y K_r(y) \frac{m'(m^{-1}(t + h_d v) - h_r \mu y)}{m'(m^{-1}(t + h_d v))} f(m^{-1}(t + h_d v)) dy dv \\
&\quad + o\left(\frac{1}{\sqrt{nh_d}}\right). \tag{3.16}
\end{aligned}$$

Die Abschätzung der Varianz von $\Delta_n^{(1.1)}(t)$ kann auch bei nur einmaliger Differenzierbarkeit der Regressionsfunktion ähnlich wie in Dette, Neumeyer und Pilz (2006) durchgeführt werden. Man erhält wie dort mit den Substitutionen $v = y - h_r u$ im dritten Schritt und $y = m^{-1}(t - h_d z) + h_r u$ im vierten Schritt und einer Taylorentwicklung der Ordnung 1

$$\begin{aligned}
\text{Var}(\Delta_n^{(1.1)}(t)) &= \frac{1}{nh_d^2 h_r^2} \text{Var} \left(\int_0^1 K_d \left(\frac{m(v) - t}{h_d} \right) K_r \left(\frac{X_1 - v}{h_d} \right) \frac{m(X_1) - m(v)}{f(v)} dv \right) \\
&\leq \frac{1}{nh_d^2 h_r^2} \mathbb{E} \left[\left(\int_0^1 K_d \left(\frac{m(v) - t}{h_d} \right) K_r \left(\frac{X_1 - v}{h_d} \right) \frac{m(X_1) - m(v)}{f(v)} dv \right)^2 \right] \\
&= \frac{h_r^2}{nh_d^2} \int_0^1 \left(\int_{-1}^1 K_d \left(\frac{m(y - h_r u) - t}{h_d} \right) K_r(u) \frac{t_m(y, u)}{f(y - h_r u)} du \right)^2 f(y) dy \\
&= \frac{h_r^2}{nh_d} \int_{-1}^1 \int_{-1}^1 \int_{-1}^1 K_d(z) K_d \left(\frac{m(m^{-1}(t + h_d z) + h_r(u - w)) - t}{h_d} \right) K_r(u) K_r(w) \\
&\quad \times f(m^{-1}(t + h_d z) + h_r u) \frac{t_m(m^{-1}(t + h_d z) + h_r u, u) t_m(m^{-1}(t + h_d z) + h_r u, w)}{f(m^{-1}(t + h_d z)) f(m^{-1}(t + h_d z) + h_r(u - w))} \\
&\quad \times (m^{-1})'(t + h_d z) dudwdz \\
&= \frac{h_r^2}{nh_d} \frac{(m^{-1})'(t)}{f(m^{-1}(t))} \int_{-1}^1 K_d^2(z) dz \left(\int_{-1}^1 K_r(u) t_m(m^{-1}(t), u) du \right)^2 (1 + o(1)) \\
&= o\left(\frac{1}{nh_d}\right)
\end{aligned}$$

mit $t_m(y, u) = um'(y) + o(1)$. Im fünften Schritt wird die Stetigkeit von m , m' und f und die Bedingung (3.6) verwendet. Somit gilt

$$\begin{aligned}\Delta_n^{(1.1)}(t) &= -h_r \int_{-1}^1 K_d(v) \int_{-1}^1 y K_r(y) \frac{m'(m^{-1}(t + h_d v)) - h_r \mu y}{m'(m^{-1}(t + h_d v))} dy dv + o_P\left(\frac{1}{\sqrt{nh_d}}\right) \\ &= a_{K_d, K_r}(t) + o_P\left(\frac{1}{\sqrt{nh_d}}\right).\end{aligned}$$

Dieser Term liefert den zweiten Teil des Bias in Satz 3.2, hat jedoch keinen Einfluss auf die Varianz der asymptotischen Verteilung. Diese wird allein durch den Term $\Delta_n^{(1.2)}(t)$ bestimmt. Der Erwartungswert von $\Delta_n^{(1.2)}(t)$ ist 0, daher erhält man die Varianz als

$$\begin{aligned}\text{Var}(\Delta_n^{(1.2)}(t)) &= \mathbb{E}[(\Delta_n^{(1.2)}(t))^2] \\ &= \frac{1}{nh_d^2 h_r^2} \int_0^1 \left(\int_0^1 K_d\left(\frac{m(v) - t}{h_d}\right) K_r\left(\frac{y - v}{h_r}\right) \frac{dv}{f(v)} \right)^2 \sigma^2(y) f(y) dy \\ &= \frac{1}{nh_d} \int_{-1}^1 \int_{-1}^1 \int_{-1}^1 K_d(y) K_d\left(\frac{m(m^{-1}(t + h_d z) + h_r(u - w)) - t}{h_d}\right) K_r(u) K_r(w) \\ &\quad \times (m^{-1})'(t + h_d z) \frac{\sigma^2 f(m^{-1}(t + h_d z) + h_r u)}{f(m^{-1}(t + h_d z)) f(m^{-1}(t + h_d z) + h_r(u - w))} dudwdz \\ &= \frac{1}{nh_d} \int_{-1}^1 \int_{-1}^1 \int_{-1}^1 K_d(z) K_d\left(z + \frac{h_r}{h_d}(u - w) m'(m^{-1}(t + h_d z) + o(h_r))\right) \\ &\quad \times K_r(u) K_r(w) \frac{\sigma^2 f(m^{-1}(t + h_d z) + h_r u)}{f(m^{-1}(t + h_d z)) f(m^{-1}(t + h_d z) + h_r(u - w))} dudwdz \\ &= \frac{1}{nh_d} \tau^2(t) (1 + o(1)).\end{aligned}$$

Im folgenden Schritt wird eine Ljapunov-Bedingung für $\sqrt{nh_d} \Delta_n^{(1.2)}(t)$ nachgewiesen. Mit dem zentralen Grenzwertsatz folgt dann die Konvergenz in Verteilung gegen eine Normalverteilung mit Erwartungswert 0 und Varianz $\tau^2(t)$. Es ist mit $x_t(v) = m^{-1}(t + h_d z)$

$$\begin{aligned}&\sum_{i=1}^n \mathbb{E} \left[\left(\frac{1}{n^{1/2} h_d^{1/2} h_r} \sigma(X_i) \varepsilon_i \int_0^1 K_d\left(\frac{m(v) - t}{h_d}\right) K_r\left(\frac{y - v}{h_r}\right) \frac{dv}{f(v)} \right)^4 \right] \\ &= \frac{\mathbb{E}[\varepsilon_1^4]}{nh_d^2 h_r^4} \int_0^1 \left(\prod_{k=1}^4 \int_0^1 K_d\left(\frac{m(v_k) - t}{h_d}\right) K_r\left(\frac{y - v_k}{h_r}\right) \frac{dv_k}{f(v_k)} \right) \sigma^4(y) f(y) dy \\ &= \frac{\mathbb{E}[\varepsilon_1^4]}{nh_d} \int_{-1}^1 K_d(y) \int_{-1}^1 K_r(u_4) \frac{x'_t(v)}{f(x_t(v))} \sigma^4 f(x_t(v) + h_r u_4) \\ &\quad \times \left[\prod_{k=1}^3 \int_{-1}^1 K_d\left(\frac{m(x_t(v) + h_r(u_4 - u_k)) - t}{h_d}\right) \frac{K_r(u_k) du_k}{f(x_t(v) + h_r(u_4 - u_k))} \right] du_4 dz \\ &= O\left(\frac{1}{nh_d}\right) = o(1).\end{aligned}$$

Zum Abschluss des Beweises wird der verbleibende Term $\Delta_n^{(2)}(t)$ betrachtet,

$$\begin{aligned}
\Delta_n^{(2)}(t) &= -\frac{1}{h_d^2} \int_0^1 K_d' \left(\frac{\xi(v) - t}{h_d} \right) (\hat{m}(v) - m(v))^2 dv \\
&= \frac{1}{h_d} \frac{\partial}{\partial t} \int_0^1 K_d \left(\frac{m(v) - t}{h_d} \right) du (\hat{m}(v) - m(v))^2 \\
&\quad \times \left\{ 1 + \left(K_d \left(\frac{m(v) - t}{h_d} \right) \right)^{-1} \left[K_d \left(\frac{\xi(v) - t}{h_d} \right) - K_d \left(\frac{m(v) - t}{h_d} \right) \right] \right\} dv \\
&= \frac{1}{h_d} \frac{\partial}{\partial t} \int_0^1 K_d \left(\frac{m(v) - t}{h_d} \right) du (\hat{m}(v) - m(v))^2 dv (1 + o_P(1)) \\
&= -\frac{1}{h_d^2} \int_0^1 K_d' \left(\frac{m(v) - t}{h_d} \right) (\hat{m}(v) - m(v))^2 dv (1 + o_P(1)) \\
&= \Delta_n^{(2.1)}(t) (1 + o_P(1)). \tag{3.17}
\end{aligned}$$

Die Bezeichnung $\Delta_n^{(2.1)}(t)$ ergibt sich aus der vorhergehenden Gleichung. Die dritte Identität folgt aus der gleichmäßig fast sicheren Abschätzung der Argumente in K_d

$$\left| \frac{\xi(v) - t}{h_d} - \frac{m(v) - t}{h_d} \right| \leq \left| \frac{\hat{m}(v) - m(v)}{h_d} \right| = O\left(\frac{\log h_r^{-1}}{nh_r h_d}\right) = o(1).$$

Da K_d gleichmäßig stetig ist, gilt auch fast sicher gleichmäßig

$$K_d \left(\frac{\xi(v) - t}{h_d} \right) - K_d \left(\frac{m(v) - t}{h_d} \right) = o(1),$$

und dieser Term kann aus dem Integral herausgezogen werden. Die vorletzte Identität in (3.17) folgt nun durch Ableiten nach t .

Beachtet man, dass für eine einmal stetig differenzierbare Regressionsfunktion der mittlere quadrierte Fehler des Nadaraya-Watson Schätzers von der Ordnung $O(1/nh_r + h_r^2)$ ist, so erhält man für $\Delta_n^{(2.1)}(t)$ ebenfalls mit der Bezeichnung $x_t(v) = m^{-1}(t + h_d z)$

$$\begin{aligned}
\mathbb{E}[|\Delta_n^{(2.1)}(t)|] &\leq -\frac{1}{h_d} \int_{-1}^1 |K_d'(v)| \mathbb{E} \left[(\hat{m}(x_t(v)) - m(x_t(v)))^2 \right] x_t'(v) dv \\
&= O\left(\frac{1}{nh_r h_d} + \frac{h_r^2}{h_d}\right) = o\left(\frac{1}{\sqrt{nh_d}}\right).
\end{aligned}$$

Aus dieser L_2 -Konvergenz folgt die stochastische Konvergenz $\Delta_n^{(2.1)}(t) = o_P(1/\sqrt{nh_d})$ und Satz 3.2 ist bewiesen. \square

Um von der Verteilung des inversen isotonen Schätzers bei einmaliger Differenzierbarkeit auf die asymptotische Verteilung des monotonen Schätzers zu schließen, kann hier nicht

so vorgegangen werden wie in Dette, Neumeyer und Pilz (2006). Dort wird an mehreren Stellen des Beweises die zweimalige Differenzierbarkeit verwendet. Im Folgenden wird daher der Mittelwertsatz der Differentialrechnung für $\hat{\phi}_{h_d}$ und die fast sichere Konvergenz des inversen monotonen Schätzers verwendet, um auf die asymptotische Verteilung des monotonen Schätzers zu schließen. Es sei $J(\delta) = [m(0) + \delta, m(1) - \delta]$ mit $\delta > 0$ derart, dass für alle $t \in J(\delta)$ gilt $t + h_d v \in [m(0), m(1)]$ für alle $v \in [-1, 1]$ und

$$I(\eta) := [m^{-1}(m(0) + \delta) + \eta, m^{-1}(m(1) - \delta) - \eta].$$

Dabei sei $\eta := \eta(h_d, h_r) > 0$ so gewählt, dass $\hat{\phi}_{h_d}^{-1}(x)$ für alle $x \in I(\eta)$ und $n \geq n_0$ fast sicher in $J(\delta)$ liegt. Eine solche Wahl ist möglich, da nach Lemma A.1 im Anhang die Statistik $\hat{\phi}_{h_d}$ fast sicher gleichmäßig gegen m^{-1} und somit auch $\hat{\phi}_{h_d}^{-1}$ gegen m konvergiert. Eine Darstellung des isotonen Schätzers mit Hilfe des inversen isotonen Schätzers erhält man unter Verwendung des Mittelwertsatzes der Differentialrechnung für $\hat{\phi}$ an den Stellen $\hat{\phi}_{h_d}^{-1}(x)$ und $m(x)$, $x \in I(\eta)$.

$$\hat{\phi}_{h_d}(\hat{\phi}_{h_d}^{-1}(x)) - \hat{\phi}_{h_d}(m(x)) = (\hat{\phi}_{h_d}^{-1}(x) - m(x))(\hat{\phi}_{h_d})'(\xi_{h_d}(x)) \quad (3.18)$$

mit $|\xi_{h_d}(x) - m(x)| \leq |\hat{\phi}_{h_d}^{-1}(x) - m(x)|$. (Die Zwischenstelle $\xi_{h_d}(x)$ liegt als Konvexkombination von $m(x)$ und $\hat{\phi}_{h_d}^{-1}(x)$ ebenfalls fast sicher in $J(\delta)$). Da $(m^{-1})'$ auf $[m(0), m(1)]$ nach unten beschränkt ist durch ein $\alpha > 0$ und nach Lemma A.1 $(\hat{\phi}_{h_d})'$ auf $J(\delta)$ gleichmäßig fast sicher gegen $(m^{-1})'$ konvergiert, ist dort $(\hat{\phi}_{h_d})'$ für alle $n \geq n_0$ nach unten beschränkt durch ein $\beta > 0$. Beachtet man nun die Identität

$$\hat{\phi}_{h_d}(\hat{\phi}_{h_d}^{-1}(x)) = m^{-1}(m(x)),$$

so kann man die Gleichung (3.18) umformen zu

$$\hat{\phi}_{h_d}^{-1}(x) - m(x) = -\frac{\hat{\phi}_{h_d}(m(x)) - m^{-1}(m(x))}{(\hat{\phi}_{h_d})'(\xi_{h_d}(x))}. \quad (3.19)$$

Satz 3.3 *Erfüllen die Bandbreiten die Bedingungen (3.5)-(3.9) und ist die Dichte $f \in \mathcal{C}^1([0, 1])$, so gilt für eine streng monoton wachsende Regressionsfunktion $m \in \mathcal{C}^1([0, 1])$ und $x \in I(\eta)$ die schwache Konvergenz*

$$\sqrt{nh_d}(\hat{\phi}_{h_d}^{-1}(x) - m(x) - h_r a_{K_d, K_r}(m(x))m'(x) + h_d b_{K_d}(m(x))m'(x)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, s^2(x))$$

mit

$$a_{K_d, K_r}(t) = \int_{-1}^1 K_d(v) \int_{-1}^1 u K_r(u) \frac{m'(m^{-1}(t + h_d v) + h_r \mu u)}{m'(m^{-1}(t + h_d v))} du dv, \quad \mu \in [0, 1]$$

$$b_{K_d}(t) = \int_{-1}^1 u K_d(u) (m^{-1})'(t + h_d \lambda u) du, \quad \lambda \in [0, 1]$$

und

$$s^2(t) = \frac{\sigma^2(x)m'(x)}{f(x)} \int_{-1}^1 K_d^2(y) dy.$$

Beweis Konvergiert der Nenner in (3.19) fast sicher und damit stochastisch gegen $(m^{-1})'(m(x))$, so folgt dieser Satz aus Satz 3.2 und dem Lemma von Slutsky. Die Differenz dieser beiden Ausdrücke lässt sich durch

$$\begin{aligned} |(\hat{\phi}_{h_d})'(\xi_{h_d}(x)) - (m^{-1})'(m(x))| &\leq |(\hat{\phi}_{h_d})'(\xi_{h_d}(x)) - (m^{-1})'(\xi_{h_d}(x))| \\ &\quad + |(m^{-1})'(\xi_{h_d}(x)) - (m^{-1})'(m(x))| \end{aligned}$$

abschätzen. Nach Lemma A.1 konvergieren sowohl

$$T^{(0)}(t) = |\hat{\phi}_{h_d}(t) - m^{-1}(t)|$$

als auch

$$T^{(1)}(t) = |(\hat{\phi}_{h_d})'(t) - (m^{-1})'(t)|$$

gleichmäßig fast sicher gegen null. Aus der gleichmäßig fast sicheren Konvergenz von $T^{(0)}(t)$ auf $J(\delta)$ folgt auf $I(\eta)$ die gleichmäßig fast sichere Konvergenz von $\hat{\phi}_{h_d}^{-1}(x)$ und somit von $\xi_{h_d}(x)$ gegen $m(x)$. Unter Beachtung der Stetigkeit von $(m^{-1})'$ erhält man nun die fast sichere Konvergenz von $(m^{-1})'(\xi_{h_d}(x))$ gegen $(m^{-1})'(m(x))$.

Da nach Voraussetzung $\xi_{h_d}(x) \in J(\delta)$ ist, konvergiert $T^{(1)}(\xi_{h_d}(x))$ fast sicher gegen null und es ist gezeigt, dass der Nenner aus (3.19) stochastisch gegen $(m^{-1})'(m(x))$ konvergiert. Mit Satz 3.2 und dem Lemma von Slutsky folgt die Aussage von Satz 3.3. \square

Satz 3.3 zeigt, dass auch für eine einmal stetig differenzierbare Regressionsfunktion die Differenz zwischen isotonem Schätzer und wahrer Regressionsfunktion gegen eine Normalverteilung konvergiert. Aufgrund der Bandbreitenbedingungen (3.5) - (3.9) unterscheidet sich die Konvergenzrate geringfügig von der des uneingeschränkten Schätzers. Der Nadaraya-Watson Schätzer besitzt eine Konvergenzrate von $\sqrt{nh_r}$. Wählt man für h_r die optimale Größenordnung $h_r = \gamma_r n^{-1/3}$, so beträgt die Konvergenzrate $n^{-1/3}$. Der isotone Schätzer hingegen besitzt eine Konvergenzrate von $\sqrt{nh_d}$. Die optimale Wahl von $h_d = \gamma_d n^{-1/3}$ ist wegen den Bedingungen (3.8) und (3.9) nicht möglich. Die Bandbreite h_d kann jedoch beliebig nahe an der optimalen Bandbreite gewählt werden. Nach Bedingung (3.8) ist $h_d = \gamma_d n^{-1/3+\delta}$ für ein $0 < \delta < 1/12$ möglich. Das ergibt eine Konvergenzrate von $n^{1/3+\delta/2}$. Diese weicht nur um ein beliebig kleines δ von der des Nadaraya-Watson Schätzers ab.

Ein analoges Resultat zu Teil b) von Satz 3.1 kann nicht gezeigt werden. Falls die Bandbreiten die Bedingung $h_d/h_r \rightarrow 0$ erfüllen, ist die Varianz von $\Delta_n^{(1,2)}(t)$ von der Größenordnung $O(1/nh_r)$. Das legt in diesem Fall eine Standardisierung der Differenz zwischen isotonem Schätzer und wahrer Regressionsfunktion mit $\sqrt{nh_r}$ nahe. Multipliziert man den Restterm $\Delta_n^{(2,1)}(t)$ mit $\sqrt{nh_r}$, so erhält man

$$\sqrt{nh_r} \Delta_n^{(2,1)}(t) = O_P\left(\frac{1}{\sqrt{nh_r h_d^2}} + \frac{\sqrt{nh_r^5}}{h_d^2}\right).$$

Dieser Ausdruck konvergiert unter der Bedingung $h_d/h_r \rightarrow 0$ nicht gegen 0.

3.2 Vergleich mit dem isotonen Kleinste-Quadrate Schätzer

In diesem Abschnitt wird ein weiterer konsistenter isotoner Schätzer beschrieben und dessen Asymptotik angegeben. Den schon in der Einleitung erwähnten Kleinste-Quadrate Schätzer von Brunk (1955, 1958) erhält man unter anderem durch die Minimierung

$$\hat{m}_I(X_i) = \arg \min_{\mathcal{Z}} \sum_{i=1}^n (Y_i - Z_i)^2$$

mit $\mathcal{Z} = \{Z \in \mathbb{R}^n | Z_1 \leq \dots \leq Z_n\}$. Zwischen den Datenpunkten wird der Schätzer linear interpoliert. In Brunk (1970) wird die asymptotische Verteilung dieses Schätzers angegeben. Ist die Regressionsfunktion m einmal stetig differenzierbar und $x \in (0, 1)$, so gilt

$$n^{1/3} \left(\sigma^2(x) m'(x) \right)^{1/3} (\hat{m}_I(x) - m(x)) \xrightarrow{\mathcal{D}} H'(0),$$

wobei $H(t)$ die größte konvexe Minorante von $W(t) + t^2$ ist, und $\{W(t) | t \in \mathbb{R}\}$ ist eine zweiseitige Brownsche Bewegung mit $E[W(t)] = 0$, $\text{Var}(W(t)) = |t|$ und $W(0) = 0$ (vgl. Prakasa Rao, 1969).

Somit besitzt der isotone Kleinste-Quadrate Schätzer \hat{m}_I nach Brunk (1970) dieselbe Konvergenzrate wie der Nadaraya-Watson Schätzer im Fall der einmaligen Differenzierbarkeit und die Rate des isotonen Kernschätzers $\hat{\phi}_{h_d}^{-1}$ weicht nur um den Faktor $n^{\delta/2}$, $0 < \delta < 1/12$, von dieser ab. Letzterer hat den Vorteil, dass er mit der Normalverteilung eine viel einfachere Grenzverteilung besitzt als der isotone Kleinste-Quadrate Schätzer. Ein Nachteil des Schätzers \hat{m}_I ist auch, dass für die Gültigkeit der Asymptotik die wahre Regressionsfunktion einmal stetig differenzierbar sein muss, der Schätzer \hat{m}_I jedoch im Allgemeinen wegen der stückweisen Linearität nicht überall differenzierbar ist.

Sowohl zu dem isotonen Kernschätzer $\hat{\phi}_{h_d}^{-1}$ als auch zu dem Kleinste-Quadrate Schätzer \hat{m}_I wurden Simulationen durchgeführt, um das Verhalten der beiden Schätzer in endlichen Stichproben zu vergleichen. Ihr Verhalten hängt sehr von der wahren Regressionsfunktion und der Standardabweichung der Daten ab. Die Abbildung 3.1 zeigt den isotonen Kernschätzer $\hat{\phi}_{h_d}^{-1}$ als durchgezogene Linie zusammen mit \hat{m}_I (gestrichelte Linie) und der wahren Regressionsfunktion m (gepunktete Linie). Es wurden $n = 100$ Daten mit einer Standardabweichung von $\sigma = 0.1$ zu den Regressionsfunktionen

$$\begin{aligned} m_1(x) &= 4 \left(x - \frac{1}{2} \right)^3 + \frac{1}{2} \\ m_2(x) &= \begin{cases} \frac{12}{7} x^3 - \frac{24}{7} x^2 + \frac{16}{7} x & \text{für } x \in [0, \frac{1}{2}] \\ \frac{12}{7} x^3 - \frac{12}{7} x^2 + \frac{4}{7} x + \frac{3}{7} & \text{für } x \in (\frac{1}{2}, 1] \end{cases} \\ m_3(x) &= \begin{cases} \frac{1}{15} \exp(3x) - \frac{1}{15} & \text{für } x \in [0, \frac{1}{2}] \\ \frac{1}{10} \exp(2x + \frac{1}{2}) - \frac{1}{30} \exp(\frac{3}{2}) - \frac{1}{15} & \text{für } x \in (\frac{1}{2}, 1] \end{cases} \end{aligned}$$

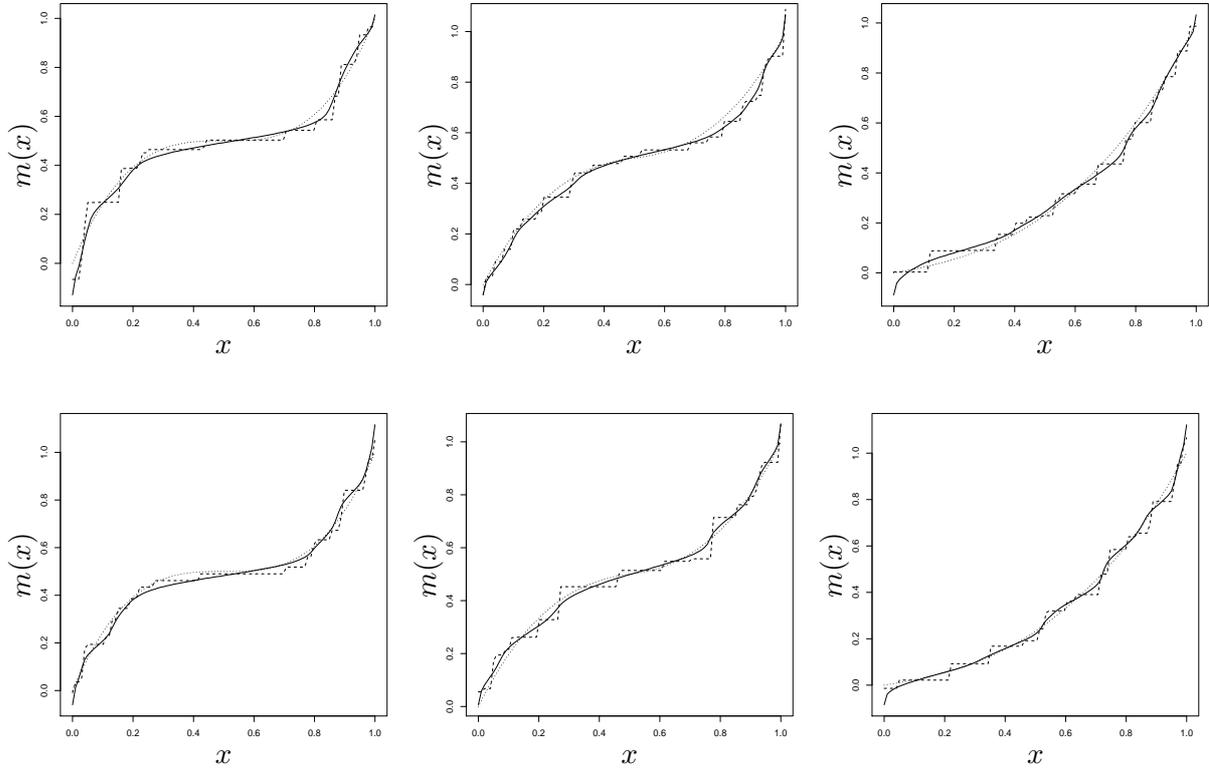


Abbildung 3.1: Monotoner Kernschätzer $\hat{\phi}_{h_d}^{-1}$ (durchgezogene Linie) und monotoner Kleinste-Quadrate Schätzer \hat{m}_I (gestrichelte Linie) zusammen mit den wahren Regressionsfunktionen (gepunktete Linie) m_1 (links), m_2 (Mitte) und m_3 (rechts). Die Stichprobengröße beträgt $n = 100$ und die Standardabweichung ist $\sigma = 0.1$. Die obere und untere Zeile entsprechen 2 verschiedenen Simulationsläufen.

erzeugt. Die Regressionsfunktion m_1 ist als Polynom dritten Grades beliebig oft differenzierbar, erfüllt jedoch trotzdem nicht die Voraussetzungen der Asymptotik des konvexen Kernschätzers, da die erste Ableitung in $x = 1/2$ den Wert 0 hat. Die Regressionsfunktionen m_2 und m_3 sind nur einmal stetig differenzierbar, die Voraussetzungen der Asymptotik sind dennoch erfüllt, da einmalige Differenzierbarkeit nach Satz 3.3 genügt. Man erkennt gut, dass der Schätzer $\hat{\phi}_{h_d}^{-1}$ im Gegensatz zu dem Kleinste-Quadrate Schätzer glatt ist. Darüberhinaus kann man an dieser Abbildung jedoch keine weiteren Aussagen über generelle Vor- oder Nachteile bezüglich der Approximation der Regressionsfunktion durch einen der beiden Schätzer treffen. Einen besseren Vergleich liefert Abbildung 3.2 in der der MSE beider Schätzer dargestellt ist. Die obere Zeile zeigt zu den Regressionsfunktionen m_1 (links), m_2 (Mitte) und m_3 (rechts) und einer Standardabweichung von $\sigma = 0.1$ den MSE des isotonen Kernschätzers $\hat{\phi}_{h_d}^{-1}$ (durchgezogene Linie) und des isotonen Kleinste-Quadrate

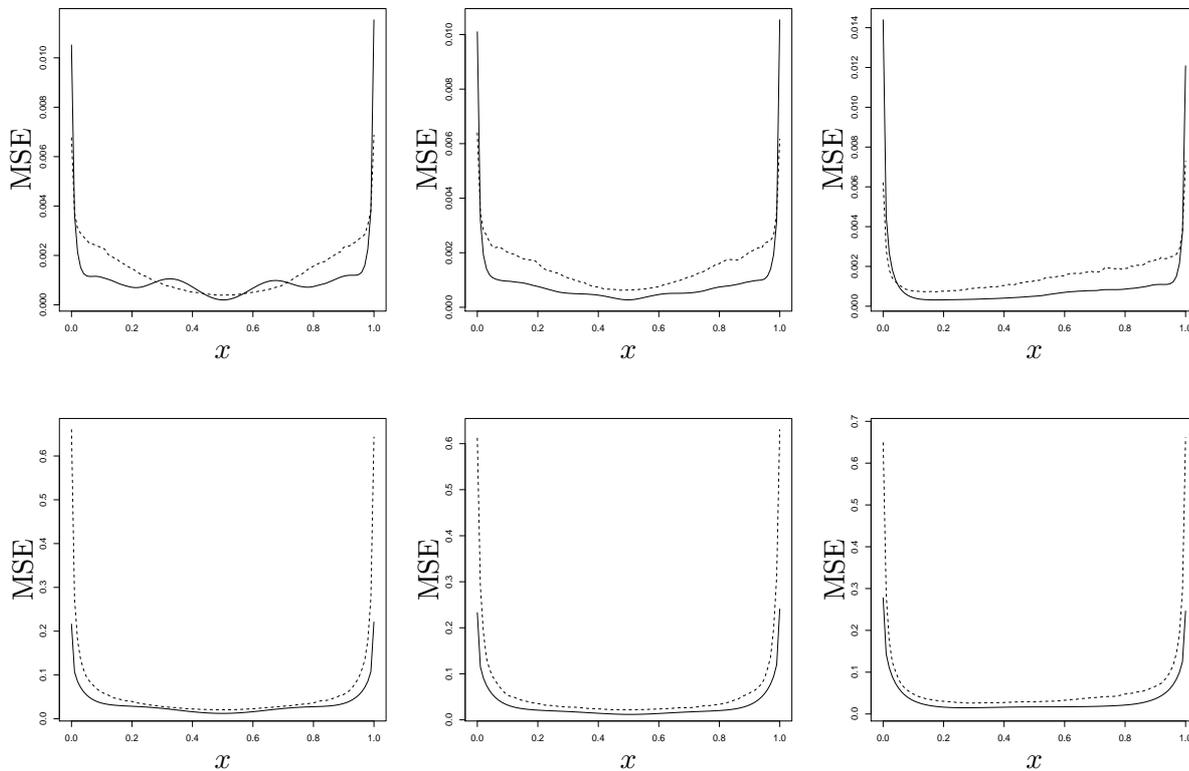


Abbildung 3.2: MSE des isotonen Schätzers $\hat{\phi}_{h_d}^{-1}$ (durchgezogene Linie) und des isotonen Kleinste-Quadrate Schätzers \hat{m}_I (gestrichelte Linie) aus 2000 Wiederholungen zu den Regressionsfunktionen m_1 (rechts), m_2 (Mitte) und m_3 (rechts). Die Stichprobengröße beträgt $n = 100$. Die obere Zeile entspricht einer Standardabweichung von $\sigma = 0.1$, die untere einer Standardabweichung von $\sigma = 1$.

Schätzers \hat{m}_I (gestrichelte Linie). Während der MSE des Schätzers $\hat{\phi}_{h_d}^{-1}$ zu der Regressionsfunktion m_1 in dem Bereich $[0.29, 0.42] \cup [0.59, 0.71]$ und am Rand des Intervalls $[0, 1]$ schlechter ist als der des Schätzers \hat{m}_I , zeigt er für die Regressionsfunktionen m_2 und m_3 fast auf dem ganzen Intervall Vorteile gegenüber letzterem. Das legt die Vermutung nahe, dass eine Verletzung der Voraussetzung $m'(x) > 0$ nicht nur in der Asymptotik, sondern auch in endlichen Stichproben Einfluss auf das Verhalten des Schätzers $\hat{\phi}_{h_d}^{-1}$ hat. Anders ist die Situation falls die Standardabweichung größer ist, hier $\sigma = 1$. Diese Resultate sind in der zweiten Zeile von Abbildung 3.2 dargestellt. Nun ist der MSE des Schätzers $\hat{\phi}_{h_d}^{-1}$ für alle Regressionsfunktionen auf dem ganzen Intervall $[0, 1]$ kleiner als der des Schätzers \hat{m}_I . Für die Regressionsfunktion m_1 fällt die Differenz auf dem Intervall $[0.2, 0.8]$ jedoch deutlich kleiner aus als für die Regressionsfunktionen m_2 und m_3 .

Kapitel 4

Schätzen von konvexen Regressionsfunktionen

In dem vorhergehenden Kapitel wurden die isotonen Umordnungen verwendet, um einen isotonen Schätzer für die Regressionsfunktion zu konstruieren. In analoger Weise wird in diesem Kapitel ein konvexer Schätzer der Regressionsfunktion definiert, indem die konvexe Umordnung aus Kapitel 2.2 auf einen uneingeschränkten Schätzer angewandt wird. Im Folgenden wird gezeigt, dass der konvexe Schätzer konsistent, asymptotisch normalverteilt und in erster Ordnung asymptotisch äquivalent zu dem uneingeschränkten Schätzer ist, falls die Regressionsfunktion m tatsächlich konvex ist. Analoge Resultate können für konkave Schätzer gezeigt werden. Dies wird in Bemerkung 4.5 angedeutet. Anschließend wird im Abschnitt 4.2 das Verhalten des Schätzers für endliche Stichproben in einer Simulationsstudie untersucht und dieser mit dem Kleinste-Quadrate Schätzer von Groeneboom, Jongbloed und Wellner (2001) verglichen. Zum Abschluss werden in Abschnitt 4.3 zwei Anwendungsbeispiele aus der Finanzmathematik und der Biologie präsentiert.

4.1 Definition und asymptotisches Verhalten des konvexen Schätzers

Die Definition der konvexen Umordnung legt das folgende Vorgehen zur Konstruktion eines konvexen Schätzers nahe. Zunächst wird die Regressionsfunktion mit einem uneingeschränkten differenzierbaren Schätzer geschätzt. Dazu sei K_r ein differenzierbarer Kern der Ordnung 2. Wie schon im Kapitel zuvor dient hier zur Demonstration der Methode der Nadaraya-Watson Schätzer \hat{m} , aber sie kann auf jeden anderen differenzierbaren Kurvenschätzer angewandt werden. Im zweiten Schritt wird nun die geglättete konvexe Umordnung auf diesen uneingeschränkten Schätzer angewandt, so dass sich der konvexe

Schätzer

$$\hat{\rho}_{h_d}(x, a) = \rho_{h_d}(\hat{m})(x, a) = \int_a^x \phi_{h_d}(\hat{m}')^{-1}(z) dz + \hat{m}(a) \quad (4.1)$$

ergibt. Nach Satz 2.8 existiert darüberhinaus auch ein L^2 -optimaler konvexer Schätzer

$$\hat{\rho}_{h_d}(x) = \hat{\rho}_{h_d}(x, a^*) = \int_0^1 \hat{\rho}_{h_d}(x, a) da \quad (4.2)$$

der Regressionsfunktion. Der folgenden Satz zeigt, dass für jede Wahl von a der konvexe Schätzer, und daher auch der L^2 -optimale Schätzer für $n \rightarrow \infty$ gegen die wahre Regressionsfunktion konvergiert. Ausserdem folgt, dass er dieselbe Grenzverteilung besitzt wie der uneingeschränkte Schätzer. Zusätzlich zu den Annahmen an das Regressionsmodell (1.1) seien die Regressionsfunktion m und die Dichte f dreimal stetig differenzierbar. Die Funktion K_r sei hier ein symmetrischer, dreimal stetig differenzierbarer Kern der Ordnung 3. Der symmetrische Kern K_d der Ordnung 2 sei zweimal differenzierbar mit Lipschitz stetiger und von 0 weg beschränkter zweiter Ableitung. Beide Kerne haben wie bisher den kompakten Träger $[-1, 1]$. Für die zugehörigen Folgen von Bandbreiten gelte

$$h_d, h_r \rightarrow 0, \quad (4.3)$$

$$nh_d, nh_r \rightarrow \infty, \quad (4.4)$$

$$h_d/h_r^{3/2} \rightarrow 0, \quad (4.5)$$

$$nh_r^7 = O(1), \quad (4.6)$$

$$(\log h_r^{-1})^{3/2}/nh_r^5 h_d = o(1). \quad (4.7)$$

Diese Bedingungen sind zum Beispiel alle erfüllt für $h_r = \gamma_r n^{-1/7}$ und $h_d = \gamma_d n^{-a}$ mit $3/14 < a < 2/7$.

Satz 4.1 *Es sei m strikt konvex und dreimal stetig differenzierbar. Erfüllen die Bandbreiten h_r und h_d die Bedingungen (4.3)-(4.7), so gilt für alle $x \in (0, 1)$ mit $m''(x) > 0$ und jedes $a \in (0, 1)$*

$$\hat{\rho}_{h_d}(x, a) - m(x) = \hat{m}(x) - m(x) + o_p\left(\frac{1}{\sqrt{nh_r}}\right).$$

Mit dem Lemma von Slutsky erhält man als direkte Folgerung aus Satz 1.3 und 4.1 die Verteilungskonvergenz.

Folgerung 4.2 *Unter den Voraussetzungen von Satz 4.1 gilt*

$$\sqrt{nh_r}(\hat{\rho}_{h_d}(x, a) - m(x) - b_n(x)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \gamma^2(x))$$

mit

$$b_n(x) = h_r^3 \kappa_3(K_r) \frac{(mf)^{(3)} - mf^{(3)}}{f}(x),$$

$$\gamma^2(x) = \frac{\sigma^2(x)}{f(x)} \int_{-1}^1 K_r^2(y) dy$$

und $\kappa_3(K_r) = 1/3! \int_0^1 u^3 K_r(u) du$.

Beweis von Satz 4.1 Es sei \mathcal{M} die Menge aller Funktionen $g \in C^2(\mathbb{R})$ mit positiver Ableitung, so dass t im Inneren des Bildes von g enthalten ist. Das Funktional

$$\Psi : \begin{cases} \mathcal{M} & \rightarrow \mathbb{R} \\ g & \rightarrow g^{-1}(t) \end{cases}$$

bildet die strikt isotonen Funktionen aus \mathcal{M} auf die zugehörige Inverse ab. Weiter sei die Funktion Q definiert durch

$$Q : \begin{cases} [0, 1] & \rightarrow \mathbb{R} \\ \lambda & \rightarrow \Psi(g_1 + \lambda(g_2 - g_1)). \end{cases}$$

Wie in Dette, Neumeyer und Pilz (2006), Lemma A.1 lässt sich mit dem Satz über implizite Funktionen zeigen, dass Q differenzierbar ist mit

$$Q'(\lambda) = -\frac{g_2 - g_1}{g_1' + \lambda(g_2' - g_1')} \circ (g_1 + \lambda(g_2 - g_1))^{-1}(t)$$

$$Q''(\lambda) = Q'(\lambda) \left\{ \frac{-2(g_2' - g_1')}{g_1' + \lambda(g_2' - g_1')} + \frac{(g_2 - g_1)(g_1'' + \lambda(g_2'' - g_1''))}{(g_1' + \lambda(g_2' - g_1'))^2} \right\} \circ Q(\lambda).$$

Damit erhält man mit $g_1 = m'^{-1}$ und $g_2 = \phi_{h_d}(\hat{m}')$ für den konvexen Schätzer die von-Mises Entwicklung

$$\hat{\rho}_{h_d}(x, a) - m(x) = \int_a^x A_n(z) dz + \frac{1}{2} \int_a^x B_n(z) dz + (\hat{m}(a) - m(a)) \quad (4.8)$$

mit

$$A_n(z) = \frac{1}{(m'^{-1})'(m'(z))} (\hat{\phi}_{h_d}(\hat{m}')(m'(z)) - m'^{-1}(m(z))), \quad (4.9)$$

$$B_n(z) = \frac{2(\phi_{h_d}(\hat{m}') - m'^{-1})(\phi_{h_d}(\hat{m}') - m'^{-1})'}{\{(m'^{-1} + \lambda^*(\phi_{h_d}(\hat{m}') - m'^{-1}))'\}^2} \circ \lambda_n(z) \quad (4.10)$$

$$- \frac{(\phi_{h_d}(\hat{m}') - m'^{-1})^2 (m'^{-1} + \lambda^*(\phi_{h_d}(\hat{m}') - m'^{-1}))''}{\{(m'^{-1} + \lambda^*(\phi_{h_d}(\hat{m}') - m'^{-1}))'\}^3} \circ \lambda_n(z)$$

und $\lambda_n(z) = (m'^{-1} + \lambda^*(\phi_{h_d}(\hat{m}') - m'^{-1}))^{-1}(z)$. Die Betrachtung der Einzelterme in der Zerlegung (4.8) erfolgt getrennt in zwei Lemmata, die am Ende dieses Kapitels bewiesen werden.

Lemma 4.3 *Unter den Voraussetzungen von Satz 4.1 gilt*

$$\int_a^x A_n(z) dz = \hat{m}(x) - m(x) + o_P\left(\frac{1}{\sqrt{nh_r}}\right).$$

Lemma 4.4 *Unter den Voraussetzungen von Satz 4.1 gilt*

$$\int_a^x B_n(z) dz = o_P\left(\frac{1}{\sqrt{nh_r}}\right).$$

Mit Lemma 4.3 und 4.4 und der Zerlegung (4.8) ergibt sich nun die Behauptung des Satzes. \square

Bemerkung 4.5 Wendet man anstatt der konvexen Umordnung die konkave Umordnung aus Bemerkung 2.11 auf den Schätzer der Regressionsfunktion an, so erhält man einen konkaven Schätzer von der Form

$$\hat{\varrho}_{h_d}(x, a) = \varrho_{h_d}(\hat{m})(x, a) = \int_a^x \varphi_{h_d}(\hat{m}')^{-1}(z) dz + \hat{m}(a)$$

und den L^2 -optimalen konkaven Schätzer

$$\hat{\varrho}_{h_d}(x) = \hat{\varrho}_{h_d}(x, a^*) = \int_0^1 \hat{\varrho}_{h_d}(x, a) da.$$

Mit ähnlichen Methoden wie im Beweis von Satz 4.1 kann man für alle $x \in (0, 1)$ mit $m''(x) < 0$ und jedes feste $a \in (0, 1)$ zeigen, dass

$$\hat{\varrho}_{h_d}(x, a) - m(x) = \hat{m}(x) - m(x) + o_P\left(\frac{1}{\sqrt{nh_r}}\right),$$

gilt, falls m streng konkav und dreimal stetig differenzierbar ist und die Bandbreiten die Bedingungen (4.3)-(4.7) erfüllen.

4.2 Simulationen

In diesem Abschnitt wird das Verhalten des konvexen Schätzers für verschiedene konvexe Regressionsfunktionen in endlichen Stichproben untersucht und mit dem des lokal linearen und des konvexen Kleinste-Quadrate Schätzers von Groeneboom, Jongbloed und Wellner (2001) verglichen. Dazu werden Daten basierend auf dem Regressionsmodell

$$Y_i = m_j(X_i) + \sigma \varepsilon_i, \quad i = 1, \dots, n, \quad j = 1, 2, 3$$

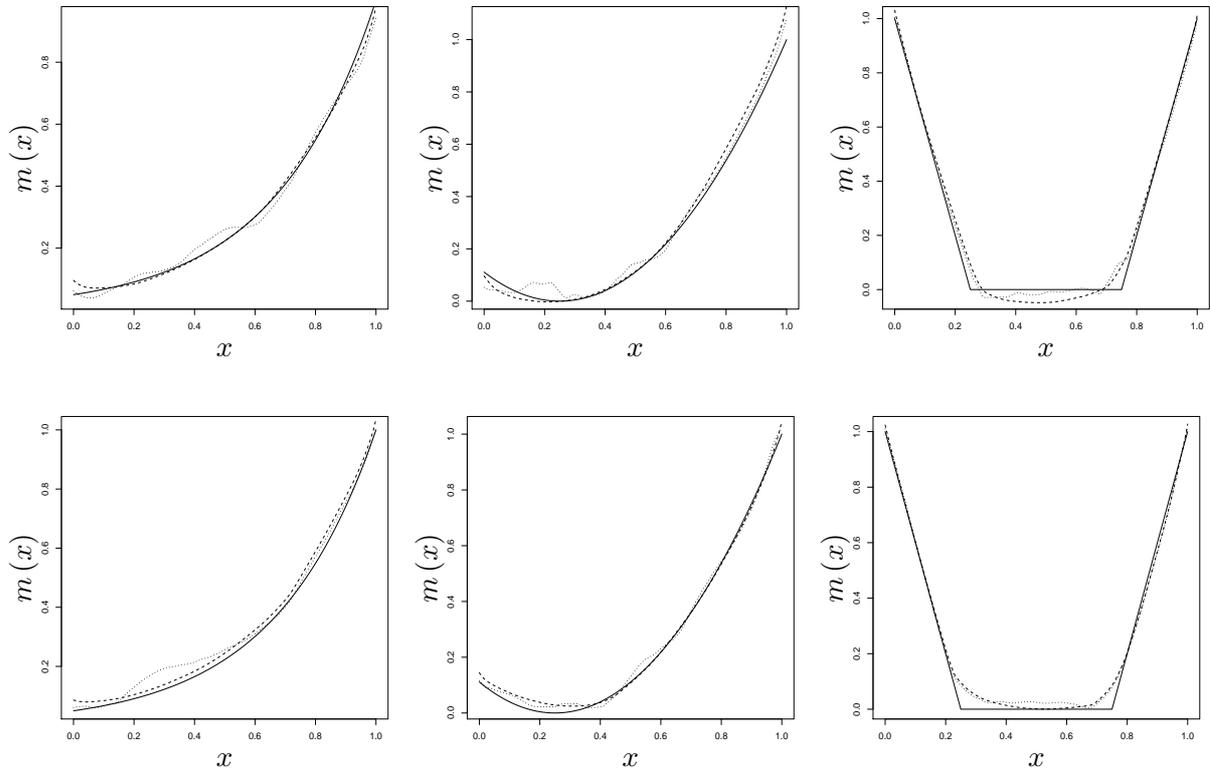


Abbildung 4.1: Die Regressionsfunktionen (durchgezogene Linie) m_1 (linke Spalte), m_2 (mittlere Spalte) und m_3 (rechte Spalte), der lokal lineare Schätzer (gepunktete Linie) und die zugehörigen konvexen Schätzer (gestrichelte Linie) aus zwei verschiedenen Simulationsläufen (obere bzw. untere Zeile) mit Stichprobengröße $n = 100$ und Standardabweichung $\sigma = 0.1$.

erzeugt. Die Zufallsvariablen X_i sind gleichverteilt auf dem Intervall $[0, 1]$ und die Fehler ε_i sind standardnormalverteilt. Für die Standardabweichung σ wird je nach Situation der Wert 0.1 oder 1 gewählt und die Regressionsfunktionen sind gegeben durch

$$\begin{aligned}
 m_1(x) &= \exp(3(x-1)) \\
 m_2(x) &= \frac{16}{9} \left(x - \frac{1}{4}\right)^2 \\
 m_3(x) &= \begin{cases} -4x + 1 & \text{falls } 0 \leq x \leq \frac{1}{4} \\ 0 & \text{falls } \frac{1}{4} \leq x \leq \frac{3}{4} \\ 4x - 3 & \text{falls } \frac{3}{4} \leq x \leq 1. \end{cases}
 \end{aligned}$$

Alle drei Funktionen sind konvex, im Gegensatz zu den ersten beiden ist die dritte an den Stellen $x = 1/4$ und $x = 3/4$ nicht differenzierbar und erfüllt somit nicht die Annahmen

der asymptotischen Theorie in Abschnitt 4.1. Als uneingeschränkter Regressionsschätzer dient hier im Gegensatz zum vorhergehenden Abschnitt der lokal lineare Schätzer \hat{m}_1 , definiert in (1.3), um Randeffekte wie sie beim Nadaraya-Watson Schätzer auftreten zu vermeiden. Die Kernfunktionen K_r und K_d werden beide als Epanechnikov Kern gewählt, das heißt

$$K_j(x) = 3(1 - x^2)/4, \quad j = r, d.$$

Außerdem wird in allen Simulationen der L^2 -optimale konvexe Schätzer verwendet, um Verzerrungen der Ergebnisse bedingt durch die Wahl von a zu vermeiden.

4.2.1 Vergleich mit dem lokal linearen Schätzer

Abbildung 4.1 zeigt für jede der drei Regressionsfunktionen den konvexen Schätzer als gestrichelte Linie zusammen mit dem zugehörigen uneingeschränkten Schätzer (gepunktete Linie) und der wahren Regressionsfunktion (durchgezogene Linie) aus zwei verschiedenen Simulationsdurchläufen. Die Standardabweichung ist hier als $\sigma = 0.1$ gewählt und die Schätzungen basieren auf $n = 100$ Daten. Die zur uneingeschränkten Schätzung der Regressionsfunktion nötige Bandbreite h_r wird durch Least Squares-Cross Validation bestimmt und die zur Monotonisierung benötigte Bandbreite ist gemäß der Bedingungen (4.5) und (4.6) $h_d = h_r^{1.6}$. Während der konvexe Schätzer die Regressionsfunktionen m_1 und m_2 sehr gut approximiert, wird m_3 in einer Umgebung der Punkte, in denen diese Funktion nicht differenzierbar ist, schlechter geschätzt. Die Schätzung in den restlichen Bereichen ist wieder sehr gut. In allen Fällen approximiert der uneingeschränkte Schätzer die Regressionsfunktion bereits gut, es wird jedoch deutlich, dass er nicht unbedingt konvex ist.

Im Folgenden wird der mittlere quadratische Fehler (MSE) aus 2000 Simulationswiederholungen geschätzt, um die Qualität des konvexen mit der des lokal linearen Schätzers zu vergleichen. Wiederum werden in jeder Wiederholung $n = 100$ Daten erzeugt. Die Bandbreite h_r wird nun so gewählt, dass sie die Bedingung (4.6) erfüllt. Dazu bezeichne $\hat{\sigma}^2$ den Varianzschätzer

$$\hat{\sigma}^2 = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} (Y_{(i+1)} - Y_{(i)})^2$$

mit den zu der Ordnung $X_{(1)}, \dots, X_{(n)}$ gehörenden $Y_{(1)}, \dots, Y_{(n)}$, wie er in Rice (1984) vorgeschlagen wird. Die Bandbreite erhält man daraus als $h_r = (\hat{\sigma}^2/n)^{1/7}$ (vgl. Kapitel 1). Zur Monotonisierung der Ableitung wird die Bandbreite $h_d = h_r^{1.6}$ gewählt, so dass die Bedingungen (4.3) - (4.7) der asymptotischen Theorie erfüllt sind. In den Abbildungen 4.2 und 4.3 ist der geschätzte MSE zusammen mit dem quadrierten Bias und der Varianz, aus denen er sich zusammensetzt, für die drei Regressionsfunktionen und $\sigma = 0.1$ bzw. $\sigma = 1$ dargestellt.

Ist $\sigma = 0.1$, so erkennt man aus Abbildung 4.2, dass der konvexe Schätzer (durchgezogene Linie) und der lokal lineare Schätzer (gestrichelte Linie) auch in endlichen Stichproben ein sehr ähnliches Verhalten aufweisen wie es auf Grund des asymptotischen Resultats in Satz 4.1 zu erwarten ist. In der ersten Zeile zeigt der konvexe Schätzer der Regressionsfunktion m_1 bezüglich der Varianz für $x \in [0.2, 0.6]$ und am Rand leichte Vorteile gegenüber dem lokal linearen Schätzer, aber insgesamt sind beide Varianzen sehr ähnlich. Deutlichere Unterschiede kann man bei dem quadrierten Bias erkennen. Dieser ist für den konvexen Schätzer für $x \in [0.15, 0.4] \cup [0.9, 1]$ kleiner und auf dem restlichen Bereich größer als der quadrierte Bias des lokal linearen Schätzers. Im wesentlichen dominiert er die Varianz im MSE. Letzterer lässt daher Vorteile des konvexen Schätzers im Bereich $[0.2, 0.4]$ und am Rand erkennen. Auf dem restlichen Intervall ist der lokal lineare Schätzer etwas besser.

Ein ähnliches Verhalten lässt sich für die Regressionsfunktion m_2 in der mittleren Zeile von Abbildung 4.2 beobachten. Die Varianzen beider Schätzer sind sehr ähnlich mit leichten Vorteilen für den konvexen Schätzer am Rand und auf $[0.2, 0.8]$. Doch in großen Teilen des Intervalls $[0, 1]$ ist der quadrierte Bias des konvexen Schätzers größer als der des lokal linearen Schätzers. Da auch hier dieser den MSE dominiert, lassen sich mit diesem Kriterium lediglich Vorteile für den konvexen Schätzer auf dem Intervall $[0.6, 0.9]$ und am Rand erkennen.

Für die Regressionsfunktion m_3 kommt zusätzlich der Effekt hinzu, der schon bei den einzelnen Schätzern beobachtet werden konnte. Sowohl die Varianz als auch der quadrierte Bias weisen in der Umgebung der Punkte, in denen die Funktion nicht differenzierbar ist, höhere Werte auf als im restlichen Intervall. Hier hat der konvexe Schätzer jedoch geringe Vorteile gegenüber dem lokal linearen Schätzer. Da wiederum der quadrierte Bias die Varianz im MSE dominiert, zeigt auch dieser das eben beschriebene Verhalten. Dies belegt die schlechtere Approximation an Stellen, in denen die Regressionsfunktion nicht differenzierbar ist.

Der Effekt, dass der quadrierte Bias die Varianz im MSE dominiert, kann zum Teil dadurch erklärt werden, dass die Standardabweichung sehr klein gewählt wurde. Abbildung 4.3 zeigt die Resultate für $\sigma = 1$. Hier ist die Situation beinahe umgekehrt. Für die Regressionsfunktionen m_1 und m_2 dominiert die Varianz offensichtlich den quadrierten Bias im MSE. Sowohl die Varianz als auch der quadrierte Bias des konvexen Schätzers sind in großen Teilen des Intervalls kleiner als bei dem lokal linearen Schätzer. Daher zeigt hier der MSE eindeutig leichte Vorteile für den konvexen Schätzer. Etwas anders ist die Situation für die Regressionsfunktion m_3 . Die Varianzen der beiden Schätzer sind hier nahezu nicht unterscheidbar. Wie schon im Fall $\sigma = 0.1$ weist der quadrierte Bias jedoch erheblich höhere Werte in den Punkten auf, in denen die Funktion nicht differenzierbar ist. Der konvexe Schätzer zeigt in diesen Punkten wiederum ein etwas besseres Verhalten als der lokal lineare Schätzer. Am Rand des Intervalls $[0, 1]$ wird der MSE durch die Varianz dominiert, im Inneren des Intervalls kommen jedoch die extrem großen Werte des quadrierten Bias zum Tragen. Insgesamt verhalten sich die beiden Schätzer nach dem

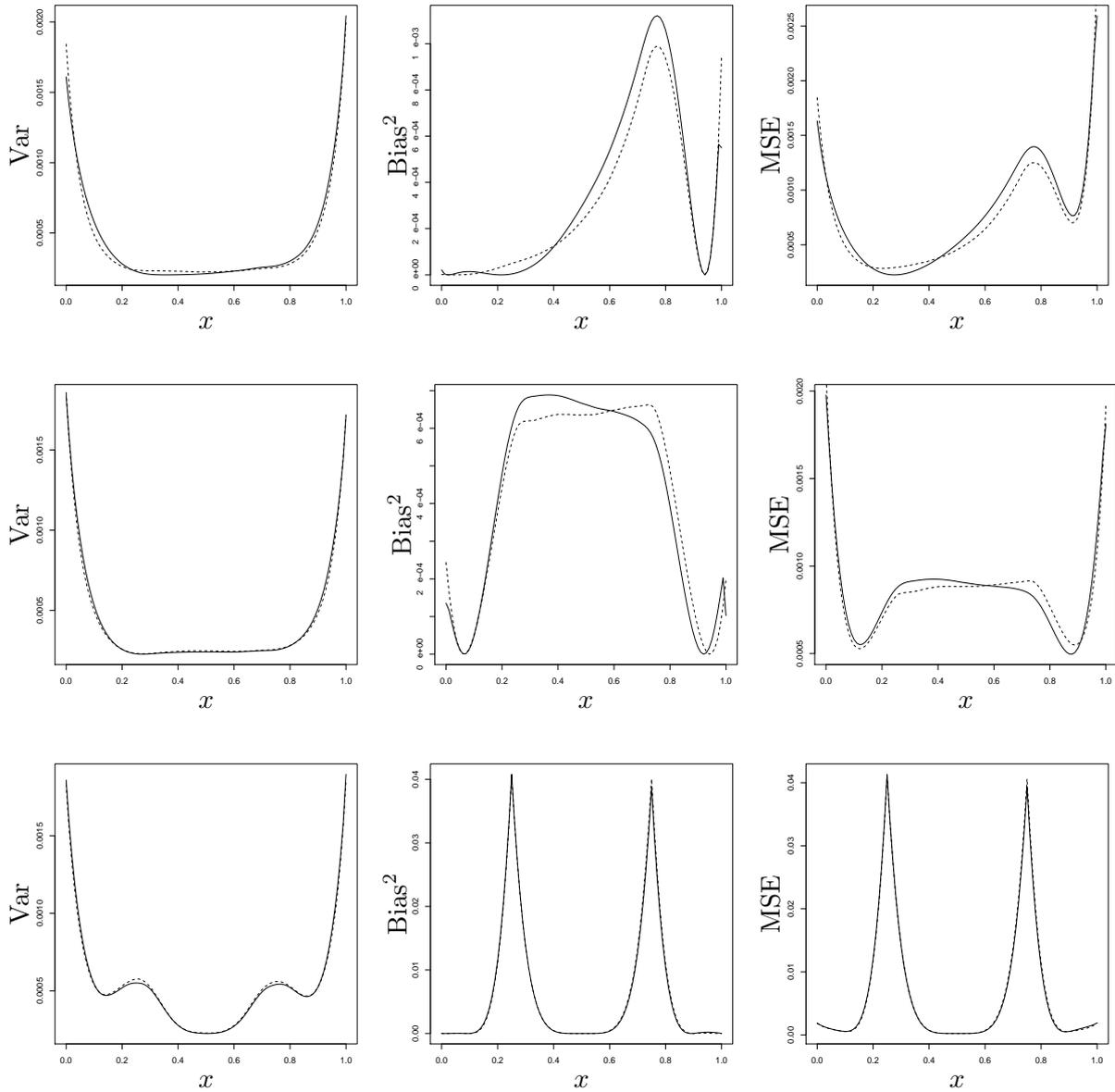


Abbildung 4.2: Geschätzte Varianz, quadrierter Bias und MSE des konvexen Schätzers $\hat{\rho}_{h_d}$ (durchgezogene Linie) und des lokal linearen Schätzers \hat{m} (gestrichelte Linie) aus 2000 Wiederholungen für die Regressionsfunktionen m_1 (obere Zeile), m_2 (mittlere Zeile) und m_3 (untere Zeile). Die Stichprobengröße ist $n = 100$ und die Standardabweichung $\sigma = 0.1$.

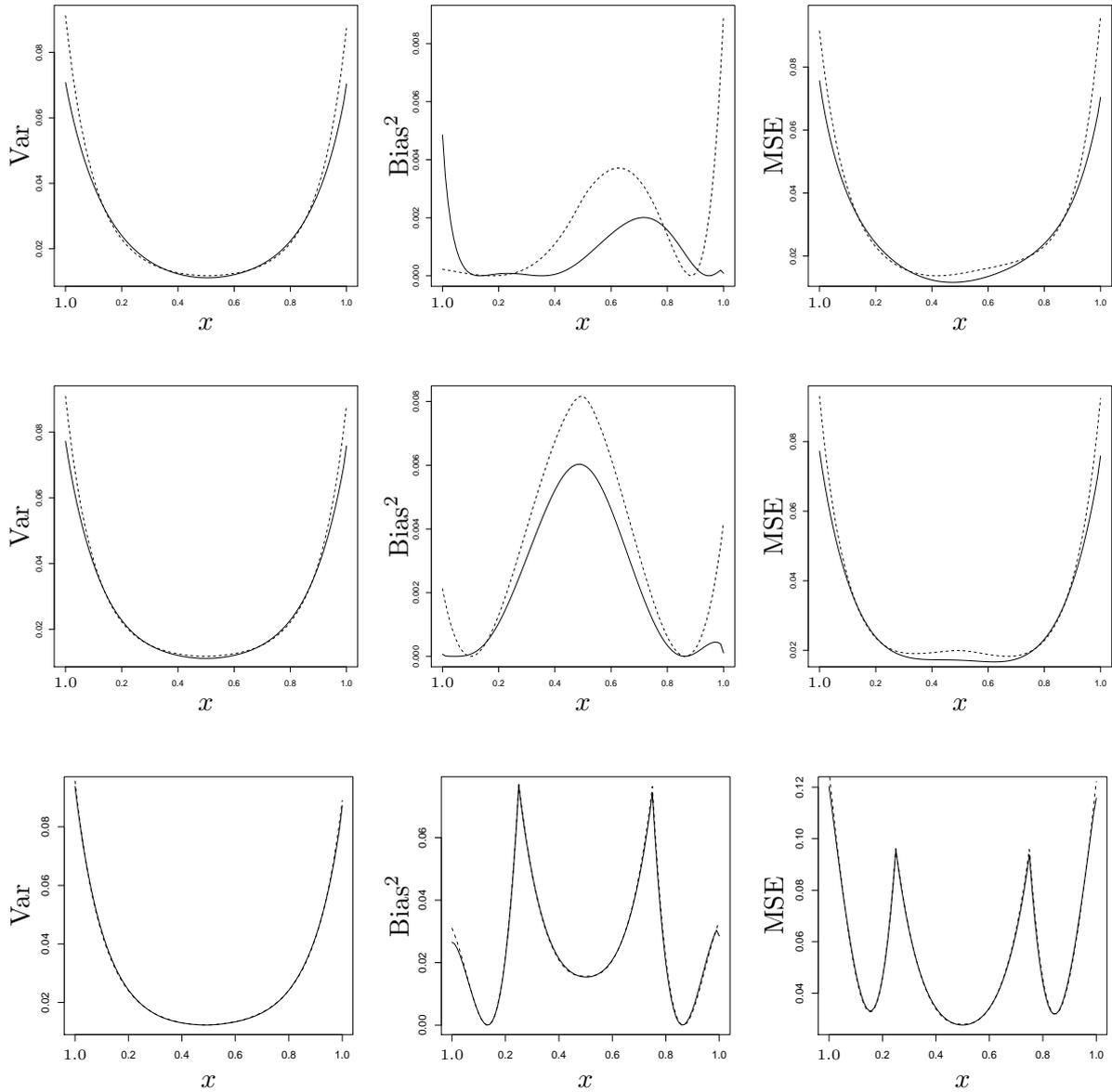


Abbildung 4.3: Geschätzte Varianz, quadrierter Bias und MSE des konvexen Schätzers $\hat{\rho}_{h_d}$ (durchgezogene Linie) und des lokal linearen Schätzers \hat{m} (gestrichelte Linie) aus 2000 Wiederholungen zu den Regressionsfunktionen m_1 (obere Zeile), m_2 (mittlere Zeile) und m_3 (untere Zeile). Die Stichprobengröße ist $n = 100$ und die Standardabweichung $\sigma = 1$.

MSE-Kriterium gleich mit leichten Vorteilen für den konvexen Schätzer am Rand und in der Umgebung der Punkte, in denen die Funktion nicht differenzierbar ist.

4.2.2 Vergleich mit einem konvexen Kleinste-Quadrate Schätzer

Eine weitere Möglichkeit eine konvexe Funktion zu schätzen ist durch den konvexen Kleinste-Quadrate Schätzer von Groeneboom, Jongbloed und Wellner (2001) gegeben. Um diesen Schätzer zu erhalten wird die Summe

$$\sum_{i=1}^n (Y_i - r(X_i))^2$$

minimiert über alle konvexen Funktionen r , die zwischen den Datenpunkten X_i , links von $X_{(1)}$ und rechts von $X_{(n)}$ linear sind. Die Funktion r , die die Summe minimiert, wird als konvexer Schätzer \hat{m}_{LS} der Regressionsfunktion verwendet. Nach Konstruktion kann dieser Schätzer im Allgemeinen nicht überall differenzierbar sein da er stückweise linear ist. Theoretisch ist es möglich, über eine andere Klasse von konvexen Funktionen zu minimieren, so dass der resultierende Schätzer glatt ist. Je größer aber die Klasse der betrachteten Funktionen ist, desto aufwendiger wird die Minimierung. Der konvexe Kleinste-Quadrate Schätzer \hat{m}_{LS} besitzt unter der Voraussetzung, dass die Regressionsfunktion m zweimal stetig differenzierbar ist, eine Konvergenzrate von $n^{2/5}$. Genauer gilt

$$n^{2/5} d_1(m)(\hat{m}_{LS}(x) - m(x)) \xrightarrow{\mathcal{D}} H''(0)$$

mit $d_1(m) = (24/\sigma^4 m''(x))^{1/5}$. Die zufällige Funktion H , deren zweite Ableitung in 0 die Grenzverteilung darstellt, ist bestimmt durch

- $H(t) \geq \int_0^t W(s) ds + t^4 = Y(t)$ für alle $t \in \mathbb{R}$,
- H besitzt eine konvexe zweite Ableitung und ist fast sicher dreimal differenzierbar in 0, sowie
- $\int_{\mathbb{R}} (H(t) - Y(t)) dH^{(3)}(t) = 0$,

wobei $\{W(t)\}_{t \in \mathbb{R}}$ eine zweiseitige Brownsche Bewegung bezeichnet mit $E[W(t)] = 0$, $\text{Var}(W(t)) = |t|$ und $W(0) = 0$ (vgl. Prakasa Rao, 1969).

Wählt man für den konvexen Schätzer $\hat{\rho}_{h_d}$ die optimale Bandbreite $h_r = O(n^{-1/7})$, so unterscheidet sich dessen Konvergenzrate lediglich um den Faktor $n^{1/35}$ von der des Kleinste-Quadrate Schätzers \hat{m}_{LS} . Dieser leichte Nachteil des konvexen Kernschätzers $\hat{\rho}_{h_d}$ wird durch die erheblich einfachere Grenzverteilung teilweise aufgewogen. Während der Kleinste-Quadrate Schätzer in Groeneboom, Jongbloed und Wellner (2001) nur für homoskedastische Regressionsmodelle definiert wird, ist ein weiterer Vorteil von $\hat{\rho}_{h_d}$ dadurch gegeben, dass er auch in heteroskedastischen Regressionsmodellen verwendbar ist.

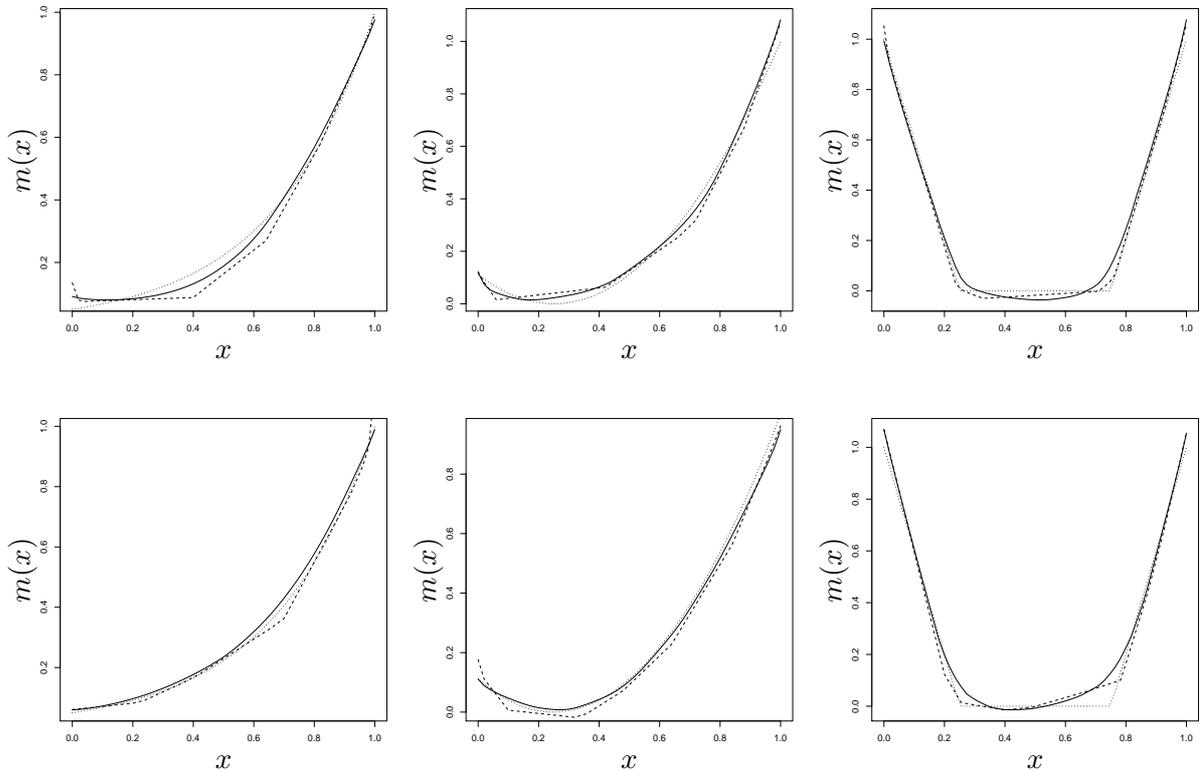


Abbildung 4.4: Die Regressionsfunktionen m_1 (linke Spalte, gepunktete Linie), m_2 (mittlere Spalte, gepunktete Linie) und m_3 (rechte Spalte, gepunktete Linie) mit dem konvexen Kernschätzer $\hat{\rho}_{h_d}$ (durchgezogene Linie) und dem konvexen Kleinste-Quadrate Schätzer \hat{m}_{LS} (gestrichelte Linie). Die Stichprobengröße ist $n = 51$, die Standardabweichung beträgt $\sigma = 0.1$.

Es ist nun naheliegend, das Verhalten des Kleinste-Quadrate Schätzers mit dem des in dieser Arbeit definierten konvexen Schätzers zu vergleichen. Da die Simulationen des Kleinste-Quadrate Schätzers zeitaufwendig sind, wird hier eine kleinere Simulationstudie präsentiert als in Abschnitt 4.2.1. Die Stichprobengröße wird mit $n = 51$ gewählt und es werden lediglich 100 Simulationsläufe durchgeführt. Abbildung 4.4 zeigt beide Schätzer für die schon in Abschnitt 4.2.1 verwendeten Regressionsfunktionen m_1 in der linken Spalte, m_2 in der mittleren Spalte und m_3 in der rechten Spalte. Die obere und untere Zeile entsprechen wiederum zwei verschiedenen Simulationsläufen. Schon diese Abbildung zeigt, dass die Qualität der beiden Schätzer sehr von der zu Grunde liegenden Regressionsfunktion abhängt. Für die Regressionsfunktionen m_1 (linke Spalte) und m_2 (mittlere Spalte) verhalten sich beide Schätzer sehr ähnlich. Man erkennt jedoch einige Nachteile

des Kleinste-Quadrate Schätzers an den Stellen, an denen sich dessen Steigung ändert und am Rand. Dort ist \hat{m}_{LS} oft deutlich weiter von der wahren Regressionsfunktion entfernt als $\hat{\rho}_{h_d}$. Anders stellt sich die Situation bei der stückweise linearen konvexen Funktion m_3 dar. Hier kann der Kleinste-Quadrate Schätzer die Punkte, an denen die Steigung wechselt besser approximieren, da über die Klasse der stückweise linearen Funktionen minimiert wird, in der auch die wahre Regressionsfunktion liegt. Diesen ersten Eindruck bestätigt auch der MSE der beiden Schätzer. Abbildung 4.5 zeigt den aus 100 Simulationsläufen geschätzten MSE beider Schätzer für die drei verschiedenen Regressionsfunktionen. Man erkennt nun deutlich, dass der Kleinste-Quadrate Schätzer am Rand die wahre Funktion erheblich schlechter schätzt als der Kernschätzer $\hat{\rho}_{h_d}$. Liegt die wahre Regressionsfunktion m_1 zu Grunde, so ist der Kernschätzer im Bereich $[0, 0.45] \cup [0.9, 1]$ besser und auf dem Rest lediglich etwas schlechter als der Kleinste-Quadrate Schätzer. Das wird durch die Dominanz des Bias, wie sie schon im Abschnitt zuvor beschrieben wurde verursacht. Aus dem gleichen Grund ist für m_2 der Kernschätzer auf dem Bereich $[0.2, 0.85]$ deutlich schlechter als der Kleinste-Quadrate Schätzer, zeigt am Rand jedoch leichte Vorteile gegenüber diesem. Anders ist die Situation für die Funktion m_3 . Hier ist der Kleinste-Quadrate Schätzer fast auf dem ganzen Intervall $[0, 1]$ besser als der Kernschätzer. Dieser Vorteil ist in den Punkten, in denen die Regressionsfunktion nicht differenzierbar ist, besonders deutlich ausgeprägt. Das lässt sich wie schon oben erwähnt dadurch erklären, dass die wahre Regressionsfunktion in der Klasse von Funktionen liegt, über die minimiert wird, und nicht in der, auf die der Kernschätzer abgestimmt ist.

Eine deutlich andere Situation ergibt sich bei größer gewählter Standardabweichung. Wählt man $\sigma = 0.5$, so ist der MSE des konvexen Kernschätzers $\hat{\rho}_{h_d}$ für die Regressionsfunktionen m_1 und m_2 fast auf dem ganzen Intervall kleiner als der des Kleinste-Quadrate Schätzers. Für die Regressionsfunktion m_3 ergibt sich wieder ein anderes Bild. Wie schon im Fall $\sigma = 0.1$ zeigt der konvexe Kernschätzer Nachteile in der Umgebung der Punkte, in denen die Steigung wechselt und Vorteile am Rand des Intervalls. Anders als im oben beschriebenen Fall sind die Randeffekte des Kleinste-Quadrate Schätzers nun erheblich größer als die Effekte des Kernschätzers an den Stellen, an denen die Regressionsfunktion m_3 nicht differenzierbar ist.

4.3 Datenbeispiele

Nachdem das Verhalten des konvexen Kernschätzers für realistische Stichprobenumfänge durch eine Simulationsstudie untersucht wurde, werden nun zwei Anwendungsmöglichkeiten in der Praxis vorgestellt. Ein konvexer beziehungsweise konkaver Schätzer ist immer dann sinnvoll anzuwenden, wenn schon vorher aus dem Sachzusammenhang deutlich wird, dass die wahre Funktion konvex oder konkav ist.

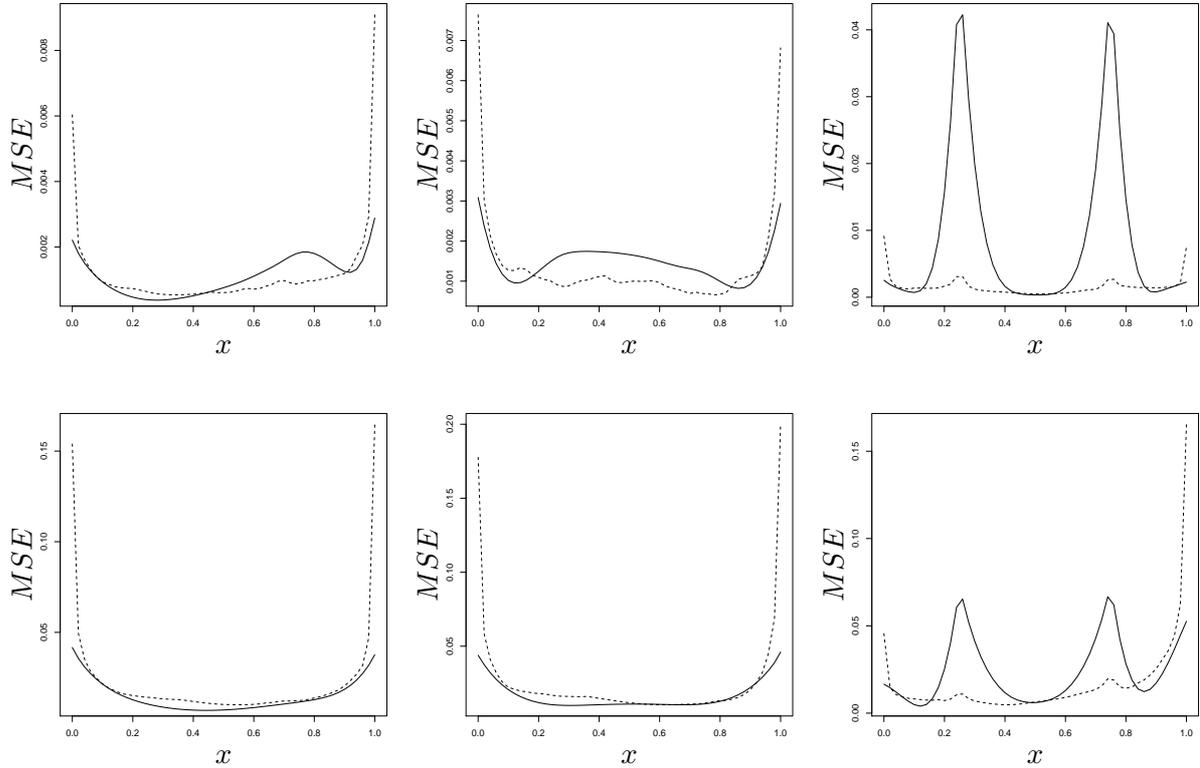


Abbildung 4.5: Der MSE von $\hat{\rho}_{h_d}$ (durchgezogene Linie) und \hat{m}_{LS} (gestrichelte Linie) für $\sigma = 0.1$ (obere Zeile) und $\sigma = 0.5$ (untere Zeile) zu den Regressionsfunktionen m_1 (linke Spalte), m_2 (mittlere Spalte) und m_3 (rechte Spalte), geschätzt aus 100 Simulationsläufen mit einer Stichprobengröße von $n = 51$.

4.3.1 Schätzen von Optionspreisen

Das erste Beispiel stammt aus der Finanzmathematik. Betrachtet werden europäische Optionen, das heißt Optionen, die nur am Fälligkeitstermin ausgeübt werden können. Die Call-Preis Funktion m solch einer Option zum Zeitpunkt t ist bestimmt durch den Wert S_t des zu Grunde liegenden Wertpapiers zum Zeitpunkt t , den Basispreis X , die Restlaufzeit τ , die risikofreie Zinsrate $r_{t,\tau}$ und die Dividendenrendite $\delta_{t,\tau}$ des Wertpapiers. Genauer lässt sich die Call-Preis Funktion darstellen als

$$m(S_t, X, \tau, r_{t,\tau}, \delta_{t,\tau}) = e^{-r_{t,\tau}\tau} \int_0^\infty \max\{S_T - X, 0\} p^*(S_T | S_t, \tau, r_{t,\tau}, \delta_{t,\tau}) dS_T \quad (4.11)$$

wobei $T = t + \tau$ den Ausübungszeitpunkt und p^* die State-Price Dichte bezeichnet (vgl. Breeden und Litzenberger, 1979 oder Black und Scholes, 1973). Ait-Sahalia und Duarte (2003) zeigen mit partiellen Ableitungen von m in Richtung X , dass die Call-Preis

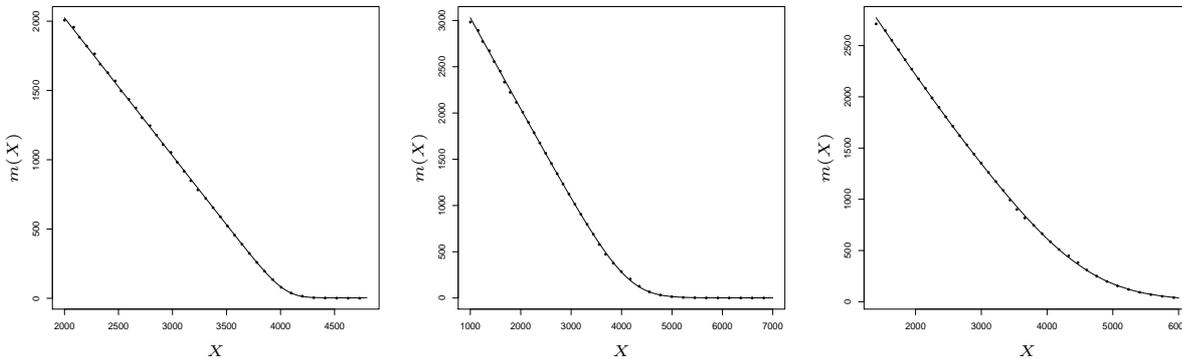


Abbildung 4.6: Der lokal lineare Schätzer (gepunktete Linie) und der L^2 -optimale konvexe Kernschätzer (durchgezogene Linie) der Call-Preis Funktion von drei DAX-Optionen aufgefasst als Funktion des Basispreises. Links: 2. Januar 2004, Restlaufzeit 14 Tage; Mitte: 2. Januar 2004, Restlaufzeit 168 Tage; rechts: 9. February 2004, Restlaufzeit 494 Tage.

Funktion in X konvex und monoton fallend sein muss um Arbitrage-Möglichkeiten zu vermeiden. Ist m lokal nicht konvex in X , resultiert dies in negativen State Preisen, die der “no arbitrage” Bedingung widersprechen. Ait-Sahalia und Duarte (2003) verwenden eine Kombination aus einer Kleinste-Quadrate Schätzung auf der Menge aller konvexen und monoton fallenden Funktionen und einer anschließenden Kernglättung um die State-Preis Dichte p^* zu schätzen und erhalten anschließend durch Einsetzen in (4.11) einen geeigneten Schätzer der Call-Preis Funktion. Mit $\hat{\rho}_{h_d}$ erhält man hier auch ohne vorheriges Schätzen der State-Preis Dichte einen konvexen Schätzer der Call-Preis Funktion. Als Beispiel dienen Daten von DAX-Optionen, die von K. Pilz (Sal. Oppenheim Bank) zur Verfügung gestellt wurden. Die Bandbreite h_r der uneingeschränkten lokal linearen Schätzung wird durch Least-Squares Cross Validation und die zweite Bandbreite als $h_d = h_r^{1.6}$ gewählt. Der Epanechnikov-Kern wird sowohl für K_r als auch für K_d verwendet. In Abbildung 4.6 sind lokal lineare (gepunktete Linie) und L^2 -optimale konvexe Schätzer (durchgezogene Linie) zu drei verschiedenen Zeitpunkten beziehungsweise Restlaufzeiten gegeben. Der lokal lineare und der konvexe Schätzer unterscheiden sich praktisch nicht, da die Daten bereits arbitragefrei sind, der lokal lineare Schätzer dadurch schon nahezu konvex ist und die konvexe Umordnung diesen somit kaum verändert.

4.3.2 Schätzen des Alters von australischen Kaninchen

Das zweite Beispiel stammt aus der Biologie und erlaubt die Anwendung des konkaven Kernschätzers. Dudzinski und Mykytowycz (1961) haben durch ein Experiment mit freilebenden australischen Kaninchen herausgefunden, dass das Trockengewicht der Augenlinse ein besserer Indikator für das Alter eines Tieres ist als sein Körpergewicht, da letzteres

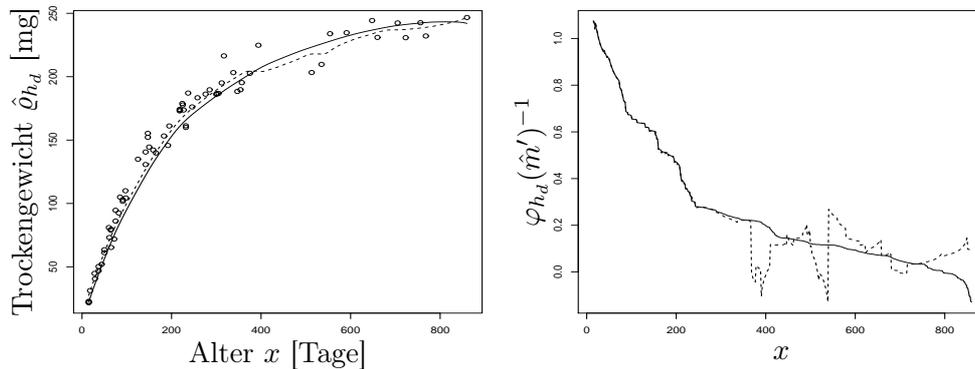


Abbildung 4.7: Links: Der lokal lineare Schätzer \hat{m} (gestrichelte Linie) und der L^2 -optimale konkave Kernschätzer $\hat{\varrho}_{h_d}$ (durchgezogene Linie) des Trockengewichts der Augenlinse in Milligramm als Funktion des Alters in Tagen, geschätzt aus den Daten von 71 frei lebenden Wildkaninchen. Rechts: Die Ableitung \hat{m}' des lokal linearen Schätzers und ihre antitone Umordnung $\varphi_{h_d}(\hat{m}')^{-1}$ gemäß Definition (2.1).

stärker von Umwelteinflüssen beeinflusst wird. In diesem Experiment wurden 71 Tiere, deren Geburtsdatum bekannt war in einem abgetrennten Bereich unter natürlichen Bedingungen gehalten. Nachdem die Tiere gestorben waren oder getötet wurden, bestimmte man das Trockengewicht in Milligramm ihrer Augenlinsen. Das zugehörige Alter wurde in Tagen notiert. Die so erhaltenen Daten verwendet Ratkowsky (1983) zur Schätzung des Trockengewichts der Augenlinse m als Funktion des Alters x , wobei er das parametrische Modell

$$m(x) = \alpha \exp\left(\frac{-\beta}{x + \gamma}\right)$$

benutzt. Aus dem dort beschriebenen Zusammenhang wird deutlich, dass es sich um eine konkave Funktion handeln muss. Es macht daher Sinn, den in Bemerkung 4.5 definierten konkaven Kernschätzer auf die Daten anzuwenden. Der Datensatz sowie eine genaue Beschreibung des Experiments und weitere Literaturhinweise sind im Internet unter <http://www.statsci.org/data/oz/rabbit.html> zu finden. Abbildung 4.7 zeigt auf der linken Seite die Datenpunkte zusammen mit dem L^2 -optimalen konkaven Schätzer $\hat{\varrho}_{h_d}$ (durchgezogene Linie) und dem lokal linearen Schätzer (gestrichelte Linie). Der lokal lineare Schätzer ist im Bereich $[400, 860]$, bedingt durch die wenigen Datenpunkte dort, noch nicht konkav. In diesem Beispiel stellt die Verwendung des konkaven Kernschätzers daher eine Verbesserung im Vergleich zu dem lokal linearen Schätzer dar. Die rechte Seite von Abbildung 4.7 zeigt zusätzlich die Ableitung des lokal linearen Schätzers als gestrichelte Linie und ihre antitone Umordnung (durchgezogene Linie).

4.4 Beweise

4.4.1 Beweis von Lemma 4.3

Ohne Einschränkung sei $a < x$. Der Beweis für $a > x$ erfolgt analog. Aus Darstellung (4.9) erhält man mit

$$\begin{aligned} A_{n1}(x) &= - \int_{m'(a)}^{m'(x)} (\phi_{h_d}(\hat{m}')(t) - \phi_{h_d}(m')(t)) dt, \\ A_{n2}(x) &= - \int_{m'(a)}^{m'(x)} (\phi_{h_d}(m')(t) - m'^{-1}(t)) dt \end{aligned} \quad (4.12)$$

die Zerlegung

$$\int_a^x A_n(z) dz = A_{n1}(x) + A_{n2}(x).$$

Setzt man die Definition der isotonen Umordnung ein, so ergibt $A_{n1}(x)$ mit einer ähnlichen Rechnung wie in (3.10)

$$\begin{aligned} A_{n1}(x) &= - \int_{m'(a)}^{m'(x)} \left[\frac{1}{h_d} \int_0^1 \int_{-\infty}^t K_d\left(\frac{\hat{m}'(v) - u}{h_d}\right) dudv - \frac{1}{h_d} \int_0^1 \int_{-\infty}^t K_d\left(\frac{m'(v) - u}{h_d}\right) dudv \right] dt \\ &= \tilde{\Delta}_n^{(1)}(x) + \frac{1}{2} \tilde{\Delta}_n^{(2)}(x) \end{aligned} \quad (4.13)$$

mit

$$\tilde{\Delta}_n^{(1)}(x) = \frac{1}{h_d} \int_{m'(a)}^{m'(x)} \int_0^1 K_d\left(\frac{m'(v) - t}{h_d}\right) (\hat{m}'(v) - m'(v)) dv dt, \quad (4.15)$$

$$\tilde{\Delta}_n^{(2)}(x) = \frac{1}{h_d^2} \int_{m'(a)}^{m'(x)} \int_0^1 K_d'\left(\frac{\xi(v) - t}{h_d}\right) (\hat{m}'(v) - m'(v))^2 dv dt. \quad (4.16)$$

Diese beiden Ausdrücke werden nun einzeln betrachtet. Dabei erhält man aus $\tilde{\Delta}_n^{(1)}(x)$ die asymptotische Normalität, $\tilde{\Delta}_n^{(2)}(x)$ ist asymptotisch vernachlässigbar. Vertauschen der Integrationsreihenfolge und eine Substitution in (4.15) ergibt für den zuerst genannten Term

$$\begin{aligned} \tilde{\Delta}_n^{(1)}(x) &= \frac{1}{h_d} \int_0^1 \int_{m'(a)}^{m'(x)} K_d\left(\frac{m'(v) - t}{h_d}\right) dt (\hat{m}'(v) - m'(v))' dv \\ &= \int_0^1 \int_{\frac{m'(v) - m'(x)}{h_d}}^{\frac{m'(v) - m'(a)}{h_d}} K_d(t) dt (\hat{m}'(v) - m'(v))' dv. \end{aligned}$$

Nun wird der Integrationsbereich des äußeren Integrals in drei Bereiche aufgeteilt, so dass in dem ersten beide Grenzen des inneren Integrals für hinreichend großes n auf dem Rand oder außerhalb des Trägers von K_d liegen und in den anderen beiden Bereichen jeweils eine Grenze des inneren Integrals im Inneren des Trägers von K_d liegt. Es ist für $v < m'^{-1}(m'(x) - h_d)$

$$\frac{m'(v) - m'(x)}{h_d} < -1$$

und für $v > m'^{-1}(m'(a) + h_d)$

$$\frac{m'(v) - m'(a)}{h_d} > 1.$$

Damit erhält man

$$\begin{aligned} \tilde{\Delta}_n^{(1)}(x) &= \int_{m'^{-1}(m'(a)+h_d)}^{m'^{-1}(m'(x)-h_d)} \int_{-1}^1 K_d(t) dt (\hat{m}(v) - m(v))' dv \\ &\quad + \int_0^{m'^{-1}(m'(a)+h_d)} \int_{-1}^{\frac{m'(v)-m'(a)}{h_d}} K_d(t) dt (\hat{m}(v) - m(v))' dv \\ &\quad + \int_{m'^{-1}(m'(x)-h_d)}^1 \int_{\frac{m'(v)-m'(x)}{h_d}}^1 K_d(t) dt (\hat{m}(v) - m(v))' dv \\ &= \tilde{\Delta}_n^{(1.1)}(x) + \tilde{\Delta}_n^{(1.2)}(x) + \tilde{\Delta}_n^{(1.3)}(x) \end{aligned}$$

Der erste Term in dieser Zerlegung ergibt

$$\begin{aligned} \tilde{\Delta}_n^{(1.1)}(x) &= \int_{m'^{-1}(m'(a)+h_d)}^{m'^{-1}(m'(x)-h_d)} (\hat{m}(v) - m(v))' dv & (4.17) \\ &= (\hat{m} - m)(m'^{-1}(m'(x) - h_d)) - (\hat{m} - m)(m'^{-1}(m'(a) + h_d)) \\ &= (\hat{m} - m)(x) - h_d(m'^{-1})'(m'(x))(\hat{m} - m)'(x) + \frac{h_d^2}{2}((\hat{m} - m) \circ m'^{-1})''(\zeta_x) \\ &\quad - (\hat{m} - m)(a) - h_d(m'^{-1})'(m'(a))(\hat{m} - m)'(a) - \frac{h_d^2}{2}((\hat{m} - m) \circ m'^{-1})''(\zeta_a) \end{aligned}$$

mit Zwischenstellen ζ_z , $z = x, a$, für die $|\zeta_z - m'(z)| \leq h_d$ gilt. Man beachte, dass für kleines h_d und $0 < a, x < 1$ die Ungleichungen

$$\frac{m'(0) - m'(a)}{h_d} \leq -1$$

und

$$\frac{m'(1) - m'(x)}{h_d} \geq 1$$

gelten. Für $\tilde{\Delta}_n^{(1.2)}$ und $\tilde{\Delta}_n^{(1.3)}$ bedeutet dies

$$\begin{aligned}
\tilde{\Delta}_n^{(1.2)}(x) &= \int_0^{m'^{-1}(m'(a)+h_d)} \int_{-1}^{\frac{m'(v)-m'(a)}{h_d}} K_d(t) dt (\hat{m} - m)'(v) dv & (4.18) \\
&= h_d \int_{-1}^1 (m'^{-1})'(m'(a) + h_d v) \int_{-1}^v K_d(t) dt (\hat{m} - m)'(m'^{-1}(m'(a) + h_d v)) dv \\
&= h_d \int_{-1}^1 K_d(t) \int_t^1 (m'^{-1})'(m'(a) + h_d v) (\hat{m} - m)'(m'^{-1}(m'(a) + h_d v)) dv dt \\
&= \int_{-1}^1 K_d(t) \int_{m'^{-1}(m'(a)+h_d t)}^{m'^{-1}(m'(a)+h_d)} (\hat{m} - m)'(v) dv dt \\
&= \int_{-1}^1 K_d(t) [(\hat{m} - m)(m'^{-1}(m'(a) + h_d)) - (\hat{m} - m)(m'^{-1}(m'(a) + h_d t))] dt \\
&= (\hat{m} - m)(a) + h_d (m'^{-1})'(m'(a)) (\hat{m} - m)'(a) + \frac{h_d^2}{2} ((\hat{m} - m) \circ m'^{-1})''(\zeta_a) \\
&\quad - \int_{-1}^1 K_d(t) \left\{ (\hat{m} - m)(a) + h_d t (m'^{-1})'(m'(a)) (\hat{m} - m)'(a) \right. \\
&\quad \quad \quad \left. + \frac{h_d^2 t^2}{2} ((\hat{m} - m) \circ m'^{-1})''(\zeta_{a,t}) \right\} dt \\
&= h_d (m'^{-1})'(m'(a)) (\hat{m} - m)'(a) + \frac{h_d^2}{2} ((\hat{m} - m) \circ m'^{-1})''(\zeta_a) \\
&\quad + \frac{h_d^2}{2} \int_{-1}^1 t^2 K_d(t) ((\hat{m} - m) \circ m'^{-1})''(\zeta_{a,t})
\end{aligned}$$

mit einer Zwischenstelle $\zeta_{a,t}$, für die $|\zeta_{a,t} - m'(a)| \leq h_d |t| \leq h_d$ gilt und

$$\begin{aligned}
\tilde{\Delta}_n^{(1.3)}(x) &= \int_{m'^{-1}(m'(x)-h_d)}^1 \int_{\frac{m'(v)-m'(x)}{h_d}}^1 K_d(t) dt (\hat{m} - m)'(v) dv & (4.19) \\
&= h_d \int_{-1}^1 (m'^{-1})'(m'(x) + h_d v) \int_v^1 K_d(t) dt (\hat{m} - m)'(m'^{-1}(m'(x) + h_d v)) dv \\
&= h_d \int_{-1}^1 K_d(t) \int_{-1}^t (m'^{-1})'(m'(x) + h_d v) (\hat{m} - m)'(m'^{-1}(m'(x) + h_d v)) dv dt \\
&= \int_{-1}^1 K_d(t) \int_{m'^{-1}(m'(x)-h_d)}^{m'^{-1}(m'(x)+h_d t)} (\hat{m} - m)'(v) dv dt \\
&= \int_{-1}^1 K_d(t) \{ (\hat{m} - m)(m'^{-1}(m'(x) + h_d t)) - (\hat{m} - m)(m'^{-1}(m'(x) - h_d)) \} dt \\
&= \int_{-1}^1 K_d(t) \left\{ (\hat{m} - m)(x) + h_d t (m'^{-1})'(m'(x)) (\hat{m} - m)'(x) \right.
\end{aligned}$$

$$\begin{aligned}
& + \frac{h_d^2 t^2}{2} ((\hat{m} - m) \circ m'^{-1})''(\zeta_{x,t}) \} dt \\
& - (\hat{m} - m)(x) + h_d (m'^{-1})'(m'(x)) (\hat{m} - m)'(x) - \frac{h_d^2}{2} ((\hat{m} - m) \circ m'^{-1})''(\zeta_x) \\
= & h_d (m'^{-1})'(m'(x)) (\hat{m} - m)'(x) - \frac{h_d^2}{2} \int_{-1}^1 t^2 K_d(t) ((\hat{m} - m) \circ m'^{-1})''(\zeta_{x,t}) \\
& - \frac{h_d^2}{2} ((\hat{m} - m) \circ m'^{-1})''(\zeta_x)
\end{aligned}$$

mit einer Zwischenstelle $\zeta_{x,t}$, für die $|\zeta_{x,t} - m'(x)| \leq h_d |t| \leq h_d$ gilt. Der Term $\tilde{\Delta}_n^{(1)}(x)$ hat damit insgesamt die Darstellung

$$\begin{aligned}
\tilde{\Delta}_n^{(1)}(x) &= (\hat{m} - m)(x) - (\hat{m} - m)(a) \\
&+ h_d^2 \kappa_2(K_d) ((\hat{m} - m) \circ m'^{-1})''(\zeta_{a,t}) - h_d^2 \kappa_2(K_d) ((\hat{m} - m) \circ m'^{-1})''(\zeta_{x,t})
\end{aligned} \tag{4.20}$$

Die beiden Restterme in (4.20) sind von der Ordnung $o(1/\sqrt{nh_r})$, denn es gilt für $z = a, x$

$$\begin{aligned}
((\hat{m} - m) \circ m'^{-1})''(\zeta_{z,t}) &= (m'^{-1})'^2(\zeta_{z,t}) (\hat{m} - m)''(m'^{-1}(\zeta_{z,t})) \\
&+ (m'^{-1})''(\zeta_{z,t}) (\hat{m} - m)'(m'^{-1}(\zeta_{z,t})).
\end{aligned}$$

Für die k -te Ableitung des Nadaraya-Watson Schätzers ist der mittlere quadratische Fehler von der Ordnung $O(1/nh_r^{2k+1})$ (siehe Härdle (1995)). Aus diesem Grund gilt für die beiden Restterme

$$\begin{aligned}
\mathbb{E} \left[\left(\sqrt{nh_r} h_d^2 \kappa_2(K_d) [(\hat{m} - m) \circ m'^{-1}]''(\zeta_{z,t}) \right)^2 \right] &= nh_r h_d^4 O \left(\frac{1}{nh_r^5} + \frac{1}{nh_r^3} \right) \\
&= O \left(\frac{h_d^4}{h_r^4} \right) = o(1)
\end{aligned}$$

mit $z = a, x$. Für $\tilde{\Delta}_n^{(2)}(x)$ ist eine genauere Betrachtung notwendig. Da dieser Ausdruck von der Zwischenstelle $\xi(v)$, die im Gegensatz zu $m'(v)$ nicht streng monoton wachsend sein muss abhängt, kann hier nicht so leicht substituiert werden wie bei der Abschätzung von $\tilde{\Delta}_n^{(1)}(x)$. Die Aussagen für $\tilde{\Delta}_n^{(2)}(x)$ sind im folgenden Lemma zusammengefasst.

Lemma 4.6 *Unter den Voraussetzungen von Satz 4.1 gilt*

$$\tilde{\Delta}_n^{(2)}(x) = \tilde{\Delta}_n^{(2.1)}(x) (1 + o_P(1))$$

mit

$$\tilde{\Delta}_n^{(2.1)}(x) = \frac{1}{h_d^2} \int_0^1 \int_{m'(a)}^{m'(x)} K_d' \left(\frac{m'(v) - t}{h_d} \right) dt (\hat{m}'(v) - m'(v))^2 dv.$$

Außerdem gilt fast sicher

$$\tilde{\Delta}_n^{(2.1)}(x) = o \left(\frac{1}{\sqrt{nh_r}} \right).$$

Beweis Der Kern K_d ist nach Voraussetzung symmetrisch und differenzierbar. Diese Eigenschaft kann wie schon im Beweis von Satz 3.2 verwendet werden, um den Term $\tilde{\Delta}_n^{(2)}(x)$ umzuschreiben. Es ist

$$\begin{aligned}\tilde{\Delta}_n^{(2)}(x) &= \frac{1}{h_d^2} \int_{m'(a)}^{m'(x)} \int_0^1 K_d' \left(\frac{\xi(v) - t}{h_d} \right) (\hat{m}'(v) - m'(v))^2 dv dt \\ &= -\frac{1}{h_d} \int_{m'(a)}^{m'(x)} \frac{\partial}{\partial t} \int_0^1 K_d \left(\frac{\xi(v) - t}{h_d} \right) (\hat{m}'(v) - m'(v))^2 dv dt \\ &= -\frac{1}{h_d} \int_{m'(a)}^{m'(x)} \frac{\partial}{\partial t} \int_0^1 K_d \left(\frac{m'(v) - t}{h_d} \right) (\hat{m}'(v) - m'(v))^2 \\ &\quad \times \left[1 + \left(K_d \left(\frac{m'(v) - t}{h_d} \right) \right)^{-1} \left[K_d \left(\frac{\xi(v) - t}{h_d} \right) - K_d \left(\frac{m'(v) - t}{h_d} \right) \right] \right] dv dt.\end{aligned}$$

Aufgrund der gleichmäßigen Stetigkeit des Kerns K_d und der gleichmäßig fast sicheren Konvergenz

$$\left| \frac{\xi(v) - t}{h_d} - \frac{m'(v) - t}{h_d} \right| \leq \left| \frac{\hat{m}'(v) - m'(v)}{h_d} \right| = O\left(\frac{\log h_r^{-1}}{nh_r^3 h_d}\right) = o_P(1)$$

ist die Differenz

$$K_d \left(\frac{\xi(v) - t}{h_d} \right) - K_d \left(\frac{m'(v) - t}{h_d} \right)$$

ebenfalls gleichmäßig fast sicher von der Ordnung $o(1)$. Der Term $\tilde{\Delta}_n^{(2)}(x)$ kann damit weiter vereinfacht werden zu

$$\begin{aligned}\tilde{\Delta}_n^{(2)}(x) &= -\frac{1}{h_d} \int_{m'(a)}^{m'(x)} \int_0^1 \frac{\partial}{\partial t} K_d \left(\frac{m'(v) - t}{h_d} \right) (\hat{m}'(v) - m'(v))^2 dv dt (1 + o(1)). \\ &= \frac{1}{h_d^2} \int_{m'(a)}^{m'(x)} \int_0^1 K_d' \left(\frac{m'(v) - t}{h_d} \right) (\hat{m}'(v) - m'(v))^2 dv dt (1 + o(1)) \\ &= \tilde{\Delta}_n^{(2.1)}(x) (1 + o(1)) \text{ f.s.}\end{aligned}$$

□

Für $\tilde{\Delta}_n^{(2.1)}(x)$ erhält man mit Substitution

$$\begin{aligned}\tilde{\Delta}_n^{(2.1)}(x) &= -\frac{1}{h_d} \int_0^1 \int_{\frac{m'(v)-m'(a)}{h_d}}^{\frac{m'(v)-m'(x)}{h_d}} K_d'(t) dt (\hat{m}'(v) - m'(v))^2 dv \\ &= -\frac{1}{h_d} \int_0^1 \left[K_d \left(\frac{m'(v) - m'(x)}{h_d} \right) - K_d \left(\frac{m'(v) - m'(a)}{h_d} \right) \right] (\hat{m}'(v) - m'(v))^2 dv \\ &= -\int_{-1}^1 K_d(v) (m'^{-1})'(m'(x) + h_d v) (\hat{m}' - m')^2 (m'^{-1}(m'(x) + h_d v)) dv\end{aligned}$$

$$+ \int_{-1}^1 K_d(v)(m'^{-1})'(m'(a) + h_d v)(\hat{m}' - m')^2(m'^{-1}(m'(a) + h_d v))dv.$$

Dies bedeutet für den L^2 -Abstand von $\tilde{\Delta}_n^{(2,1)}(x)$ zur 0

$$\begin{aligned} \mathbb{E}|\tilde{\Delta}_n^{(2,1)}(x)| &\leq \int_{-1}^1 K_d(v)|(m'^{-1})'(m'(x) + h_d v)| \mathbb{E}[(\hat{m}' - m')^2(m'^{-1}(m'(x) + h_d v))]dv \\ &\quad + \int_{-1}^1 K_d(v)|(m'^{-1})'(m'(a) + h_d v)| \mathbb{E}[(\hat{m}' - m')^2(m'^{-1}(m'(a) + h_d v))]dv \\ &= O\left(\frac{1}{nh_r^3}\right), \end{aligned}$$

da für den MSE von \hat{m}' gleichmäßig bezüglich v gilt

$$\mathbb{E}[(\hat{m}'(v) - m'(v))^2] = O\left(\frac{1}{nh_r^3}\right).$$

Somit erhält man die Ordnung

$$\tilde{\Delta}_n^{(2,1)}(x) = o_p\left(\frac{1}{\sqrt{nh_r}}\right).$$

□

Durch die Darstellung in Lemma 4.6 überträgt sich die Konvergenzrate von $\tilde{\Delta}_n^{(2,1)}(x)$ auf $\tilde{\Delta}_n^{(2)}(x)$. Eine Abschätzung des deterministische Terms $A_{n,2}(x)$ beendet den Beweis von Lemma 4.3. Eine ähnliche Rechnung wie im Beweis von Satz 3.2 zeigt

$$\begin{aligned} A_{n,2}(x) &= - \int_{m'(a)}^{m'(x)} \left\{ m'^{-1}(t - h_d) + h_d \int_{-1}^1 (m'^{-1})'(t + h_d z) \int_z^1 K_d(v) dv dz - m'^{-1}(t) \right\} dt \\ &= - \int_{m'(a)}^{m'(x)} \left\{ m'^{-1}(t - h_d) + \left(m'^{-1}(t + h_d z) \int_z^1 K_d(v) dv \right) \Big|_{-1}^1 \right. \\ &\quad \left. + \int_{-1}^1 K_d(z) m'^{-1}(t + h_d z) dz - m'^{-1}(t) \right\} dt \\ &= -h_d^2 \left(\kappa_2(K_d) \int_{m'(a)}^{m'(x)} (m'^{-1})''(t) dt + o(1) \right) \\ &= -h_d^2 \left(\kappa_2(K_d) \left(\frac{1}{m''(x)} - \frac{1}{m''(a)} \right) + o(1) \right) \end{aligned}$$

und somit

$$\sqrt{nh_r} A_{n,2}(x) = -\frac{h_d^2}{h_r^3} \sqrt{nh_r^7} O(1) = o(1)$$

nach den Voraussetzungen (4.5) und (4.6). Damit ist der Beweis von Lemma 4.3 abgeschlossen. □

4.4.2 Beweis von Lemma 4.4

Das Integral über $B_n(z)$ lässt sich nach (4.10) und mit den Bezeichnungen

$$\lambda_n(z) = (m'^{-1} + \lambda^*(\hat{\psi}_{h_d} - m'^{-1}))^{-1}(z)$$

und $\hat{\psi}_{h_d} = \phi_{h_d}(\hat{m}')$ darstellen als

$$\int_a^x B_n(z) dz = 2 \int_a^x \frac{(\hat{\psi}_{h_d} - m'^{-1})(\hat{\psi}_{h_d} - m'^{-1})'}{(\lambda_n^{-1})'^2} \circ \lambda_n(z) dz \quad (4.21)$$

$$\begin{aligned} & - \int_a^x \frac{(\hat{\psi}_{h_d} - m'^{-1})^2 (\lambda_n^{-1})''}{(\lambda_n^{-1})'^3} \circ \lambda_n(z) dz \\ & = 2B_n^{(1)}(x) - B_n^{(2)}(x). \end{aligned} \quad (4.22)$$

Es genügt daher zu zeigen, dass die Terme $\sqrt{nh_r} B_n^{(i)}(x)$, $i = 1, 2$, fast sicher gegen 0 konvergieren. Wie schon in Kapitel 3 im Beweis von Satz 3.2 gezeigt wurde, lässt sich die Differenz $\hat{\psi}_{h_d} - m'^{-1}$ schreiben als

$$\hat{\psi}_{h_d}(\lambda_n(z)) - m'^{-1}(\lambda_n(z)) = \Delta_n^{(0)}(\lambda_n(z)) + \Delta_n^{(1)}(\lambda_n(z)) + \Delta_n^{(2)}(\lambda_n(z))(1 + o_p(1)) \quad (4.23)$$

mit

$$\begin{aligned} \Delta_n^{(0)}(t) &= \phi_{h_d}(m')(t) - m'^{-1}(t), \\ \Delta_n^{(1)}(t) &= \frac{1}{h_d} \int_0^1 K_d\left(\frac{m'(v) - t}{h_d}\right) (\hat{m}'(v) - m'(v)) dv, \\ \Delta_n^{(2)}(t) &= \frac{1}{h_d^2} \int_0^1 K_d'\left(\frac{m'(v) - t}{h_d}\right) (\hat{m}'(v) - m'(v))^2 dv. \end{aligned}$$

Der Unterschied zu der Zerlegung im Beweis von Satz 3.2 besteht darin, dass die Funktion m hier durch ihre Ableitung m' ersetzt wird. Mit diesen Bezeichnungen kann $B_n^{(1)}(x)$ geschrieben werden als

$$\begin{aligned} B_n^{(1)}(x) &= \int_a^x \frac{\Delta_n^{(0)}(\lambda_n(z)) \Delta_n^{(0)' }(\lambda_n(z))}{(\lambda_n^{-1})' ^2(\lambda_n(z))} dz + \int_a^x \frac{\Delta_n^{(1)}(\lambda_n(z)) \Delta_n^{(0)' }(\lambda_n(z))}{(\lambda_n^{-1})' ^2(\lambda_n(z))} dz \\ &+ \int_a^x \frac{\Delta_n^{(2)}(\lambda_n(z)) \Delta_n^{(0)' }(\lambda_n(z))}{(\lambda_n^{-1})' ^2(\lambda_n(z))} dz + \int_a^x \frac{\Delta_n^{(0)}(\lambda_n(z)) \Delta_n^{(1)' }(\lambda_n(z))}{(\lambda_n^{-1})' ^2(\lambda_n(z))} dz \\ &+ \int_a^x \frac{\Delta_n^{(1)}(\lambda_n(z)) \Delta_n^{(1)' }(\lambda_n(z))}{(\lambda_n^{-1})' ^2(\lambda_n(z))} dz + \int_a^x \frac{\Delta_n^{(2)}(\lambda_n(z)) \Delta_n^{(1)' }(\lambda_n(z))}{(\lambda_n^{-1})' ^2(\lambda_n(z))} dz \\ &+ \int_a^x \frac{\Delta_n^{(0)}(\lambda_n(z)) \Delta_n^{(2)' }(\lambda_n(z))}{(\lambda_n^{-1})' ^2(\lambda_n(z))} dz + \int_a^x \frac{\Delta_n^{(1)}(\lambda_n(z)) \Delta_n^{(2)' }(\lambda_n(z))}{(\lambda_n^{-1})' ^2(\lambda_n(z))} dz \\ &+ \int_a^x \frac{\Delta_n^{(2)}(\lambda_n(z)) \Delta_n^{(2)' }(\lambda_n(z))}{(\lambda_n^{-1})' ^2(\lambda_n(z))} dz. \end{aligned} \quad (4.24)$$

Eine ähnliche Rechnung wie bei der Abschätzung des Terms $A_{n,2}$ im Beweis von Lemma 4.3 zeigt, dass der Term $\Delta_n^{(0)}(\lambda_n(z))$ gleichmäßig abgeschätzt werden kann als

$$\Delta_n^{(0)}(\lambda_n(z)) = O(h_d^2), \quad (4.25)$$

und man erhält für dessen Ableitung

$$\begin{aligned} \Delta_n^{(0)'}(\lambda_n(z)) &= \int_{-1}^1 K_d(v)(m'^{-1})'(\lambda_n(z) + h_d v) dv - (m'^{-1})'(\lambda_n(z)) \\ &= h_d \int_{-1}^1 v K_d(v)(m'^{-1})''(\lambda_n(z) h_d \mu v) dv = O(h_d). \end{aligned} \quad (4.26)$$

Ein Problem in der Abschätzung der Terme von $B_n^{(1)}(x)$ in (4.24) ist die Zwischenstelle $\lambda_n(z)$, die im Gegensatz zu ihrem Grenzwert m' nicht streng monoton wachsend ist. Ähnlich wie schon in der Abschätzung von $\tilde{\Delta}_n^{(2)}(x)$ in Lemma 4.6 gilt hier gleichmäßig fast sicher

$$\Delta_n^{(j)}(\lambda_n(z)) = \tilde{\Delta}_n^{(j)}(m'(z))(1 + o(1)) \text{ und } \Delta_n^{(j)'}(\lambda_n(z)) = \Delta_n^{(j)'}(m'(z))(1 + o(1)) \quad (4.27)$$

für $j = 1, 2$. Diese Darstellung wird nun hergeleitet. Wegen der Positivität von K_d ist

$$\begin{aligned} \Delta_n^{(1)}(\lambda_n(z)) &= \frac{1}{h_d} \int_0^1 K_d\left(\frac{m'(v) - m'(z)}{h_d}\right) (\hat{m}'(v) - m'(v)) \\ &\quad \times \left\{ 1 + \left(K_d\left(\frac{m'(v) - m'(z)}{h_d}\right) \right)^{-1} \left[K_d\left(\frac{m'(v) - \lambda_n(z)}{h_d}\right) - K_d\left(\frac{m'(v) - m'(z)}{h_d}\right) \right] \right\} dv. \end{aligned} \quad (4.28)$$

Nach Lemma A.3 im Anhang gilt gleichmäßig fast sicher in t

$$|\hat{\psi}_{h_d}(t) - m'^{-1}(t)| = O\left(\frac{\log h_r^{-1}}{nh_r^3}\right)^{1/2} + O(h_d^2), \quad (4.29)$$

$$|\hat{\psi}'_{h_d}(t) - (m'^{-1})'(t)| = O\left(\frac{\log h_r^{-1}}{nh_r^5}\right)^{1/2} + o(h_d). \quad (4.30)$$

Mit einer Taylorentwicklung für die Inverse einer Funktion erhält man die Darstellung der Zwischenstelle

$$\lambda_n(z) - m'^{-1}(z) = \frac{\lambda^*(\hat{\psi}_{h_d} - m'^{-1})}{(m'^{-1})' + \mu(\hat{\psi}'_{h_d} - (m'^{-1})')} ((m'^{-1} + \mu(\hat{\psi}_{h_d} - m'^{-1}))^{-1}(z)). \quad (4.31)$$

Eine Kombination von (4.29) - (4.31) mit den Bedingungen (4.5) - (4.7) liefert fast sicher gleichmäßig die Ordnung

$$|\lambda_n(z) - m'^{-1}(z)| = O\left(\frac{\log h_r^{-1}}{nh_r^3}\right)^{1/2} + O(h_d^2) = o(h_d). \quad (4.32)$$

Daraus ergibt sich mit der gleichmäßigen Stetigkeit von K_d

$$K_d\left(\frac{m'(v) - \lambda_n(z)}{h_d}\right) - K_d\left(\frac{m'(v) - m'(z)}{h_d}\right) = o(1),$$

und $\Delta_n^{(1)}(\lambda_n(z))$ hat fast sicher die Darstellung

$$\Delta_n^{(1)}(\lambda_n(z)) = \frac{1}{h_d} \int_0^1 K_d\left(\frac{m'(v) - m'(z)}{h_d}\right) (\hat{m}'(v) - m'(v)) dv (1 + o(1)).$$

Das Vorgehen für $\Delta_n^{(1)'}$ ist ähnlich, jedoch ist die erste Ableitung des Kerns nicht überall von 0 verschieden, so dass eine Darstellung wie in (4.28) nicht direkt möglich ist. Statt dessen wird hier ein Ausweg wie in der Abschätzung von $\Delta_n^{(2)}(x)$ in Lemma 4.6 gewählt. Es gilt

$$\begin{aligned} \Delta_n^{(1)' }(\lambda_n(z)) &= \frac{1}{h_d^2} \int_0^1 K_d'\left(\frac{m'(v) - \lambda_n(z)}{h_d}\right) (\hat{m}'(v) - m'(v)) dv \\ &= \frac{1}{h_d} \int_0^1 \frac{\partial}{\partial u} \left(K_d\left(\frac{u - \lambda_n(z)}{h_d}\right) \Big|_{u=m'(v)} \right) (\hat{m}'(v) - m'(v)) dv \\ &= \frac{1}{h_d} \int_0^1 (\hat{m}'(v) - m'(v)) \frac{\partial}{\partial u} \left(K_d\left(\frac{u - m'(z)}{h_d}\right) \right. \\ &\quad \left. \times \left\{ 1 + \left(K_d\left(\frac{u - m'(z)}{h_d}\right) \right)^{-1} \left[K_d\left(\frac{u - \lambda_n(z)}{h_d}\right) - K_d\left(\frac{u - m'(z)}{h_d}\right) \right] \right\} \Big|_{u=m'(v)} \right) dv. \end{aligned}$$

Nun folgt aus (4.32) und der gleichmäßigen Stetigkeit von K_d fast sicher die Darstellung

$$\Delta_n^{(1)' }(\lambda_n(z)) = \frac{1}{h_d^2} \int_0^1 K_d'\left(\frac{m'(v) - m'(z)}{h_d}\right) (\hat{m}'(v) - m'(v)) dv (1 + o(1)).$$

Mit derselben Rechnung erhält man fast sicher

$$\Delta_n^{(2)}(\lambda_n(z)) = \frac{1}{h_d^2} \int_0^1 K_d''\left(\frac{m'(v) - m'(z)}{h_d}\right) (\hat{m}'(v) - m'(v)) dv (1 + o(1)).$$

Für den Term $\Delta_n^{(2)' }(\lambda_n(z))$ wird ausgenutzt, dass K_d'' auf $[-1, 1]$ von Null weg beschränkt und gleichmäßig stetig ist. Es gilt dann fast sicher

$$\begin{aligned} \Delta_n^{(2)' }(\lambda_n(z)) &= \frac{1}{h_d^3} \int_0^1 K_d''\left(\frac{m'(v) - \lambda_n(z)}{h_d}\right) (\hat{m}'(v) - m'(v))^2 dv \\ &= \frac{1}{h_d^3} \int_0^1 K_d''\left(\frac{m'(v) - m'(z)}{h_d}\right) (\hat{m}'(v) - m'(v))^2 \\ &\quad \times \left\{ 1 + \left(K_d''\left(\frac{m'(v) - m'(z)}{h_d}\right) \right)^{-1} \left[K_d''\left(\frac{m'(v) - \lambda_n(z)}{h_d}\right) - K_d''\left(\frac{m'(v) - m'(z)}{h_d}\right) \right] \right\} dv \\ &= \frac{1}{h_d^3} \int_0^1 K_d''\left(\frac{m'(v) - m'(z)}{h_d}\right) (\hat{m}'(v) - m'(v))^2 dv (1 + o(1)). \end{aligned}$$

Außerdem gilt nach obiger Diskussion der Zwischenstelle $\lambda_n(z)$

$$\begin{aligned} \frac{1}{(\lambda_n^{-1})'^2(\lambda_n(z))} &= \frac{1}{(m'^{-1})'^2(m'(z))} \left(1 + \frac{(m'^{-1})'^2(m'(z)) - (\lambda_n^{-1})'^2(\lambda_n(z))}{(\lambda_n^{-1})'^2(\lambda_n(z))(m'^{-1})'^2(m'(z))} \right) \\ &= \frac{1}{(m'^{-1})'^2(m'(z))} (1 + o(1)) \text{ f.s.} \end{aligned} \quad (4.33)$$

Mit den Gleichungen (4.25), (4.26), (4.27), (4.33) und der Substitution $z = m'^{-1}(t)$ hat der Term $B_n^{(1)}(x)$ die Darstellung

$$\begin{aligned} B_n^{(1)}(x) &= \left\{ O(h_d^3) + O(h_d) \int_{m'(a)}^{m'(x)} \frac{\Delta_n^{(1)}(t)}{(m'^{-1})'(t)} dt + O(h_d) \int_{m'(a)}^{m'(x)} \frac{\Delta_n^{(2)}(t)}{(m'^{-1})'(t)} dt \right. \\ &\quad + O(h_d^2) \int_{m'(a)}^{m'(x)} \frac{\Delta_n^{(1)'}(t)}{(m'^{-1})'(t)} dt + \int_{m'(a)}^{m'(x)} \frac{\Delta_n^{(1)}(t)\Delta_n^{(1)'}(t)}{(m'^{-1})'(t)} dt \\ &\quad + \int_{m'(a)}^{m'(x)} \frac{\Delta_n^{(2)}(t)\Delta_n^{(1)'}(t)}{(m'^{-1})'(t)} dt + O(h_d^2) \int_{m'(a)}^{m'(x)} \frac{\Delta_n^{(2)'}(t)}{(m'^{-1})'(t)} dt \\ &\quad \left. + \int_{m'(a)}^{m'(x)} \frac{\Delta_n^{(1)}(t)\Delta_n^{(2)'}(t)}{(m'^{-1})'(t)} dt + \int_{m'(a)}^{m'(x)} \frac{\Delta_n^{(2)}(t)\Delta_n^{(2)'}(t)}{(m'^{-1})'(t)} dt \right\} (1 + o(1)). \end{aligned} \quad (4.34)$$

Die Substitution $v = m'^{-1}(t + h_d w)$ ergibt

$$\Delta_n^{(1)}(t) = \int_{-1}^1 K_d(w) (m'^{-1})'(t + h_d w) d^{(1)}(m'^{-1}(t + h_d w)) dw = O\left(\frac{\log h_r^{-1}}{nh_r^3}\right)^{1/2} \quad (4.35)$$

mit der Bezeichnung $d(x) = \hat{m}(x) - m(x)$, und mit einer partiellen Integration gilt

$$\begin{aligned} \Delta_n^{(1)'}(t) &= \frac{1}{h_d} \int_{-1}^1 K_d'(w) (m'^{-1})'(t + h_d w) d^{(1)}(m'^{-1}(t + h_d w)) dw \\ &= - \int_{-1}^1 K_d(w) \{ (m'^{-1})''(t + h_d w) d^{(1)}(m'^{-1}(t + h_d w)) \\ &\quad + (m'^{-1})'^2(t + h_d w) d^{(2)}(m'^{-1}(t + h_d w)) \} dw \\ &= O\left(\frac{\log h_r^{-1}}{nh_r^5}\right)^{1/2}. \end{aligned} \quad (4.36)$$

Eine genauere Betrachtung des Integrals über das Produkt $\Delta_n^{(1)}(t)\Delta_n^{(1)'}(t)$ ergibt nun mit den Bezeichnungen $z(t, v) = m'^{-1}(t + h_d v)$ und $z^{(k)}(t + h_d v) = (m'^{-1})^{(k)}(t + h_d v)$

$$\int_{m'(a)}^{m'(x)} \frac{\Delta_n^{(1)}(t)\Delta_n^{(1)'}(t)}{(m'^{-1})'(t)} dt$$

$$\begin{aligned}
&= - \int_{-1}^1 \int_{-1}^1 K_d(v)K_d(w) \int_{m'(a)}^{m'(x)} \frac{z'(t,v)z''(t,w)}{(m'^{-1})'(t)} d^{(1)}(z(t,v))d^{(1)}(z(t,w))dt dv dw \\
&\quad - \int_{-1}^1 \int_{-1}^1 K_d(v)K_d(w) \int_{m'(a)}^{m'(x)} \frac{z'(t,v)(z'(t,w))^2}{(m'^{-1})'(t)} d^{(1)}(z(t,v))d^{(2)}(z(t,w))dt dv dw \\
&= - \int_{-1}^1 \int_{-1}^1 K_d(v)K_d(w) \int_{m'(a)}^{m'(x)} d^{(1)}(z(t,v))d^{(1)}(z(t,w))z''(t,w)dt dv dw O(1) \\
&\quad - \int_{-1}^1 \int_{-1}^1 K_d(v)K_d(w) \int_{m'(a)}^{m'(x)} d^{(1)}(z(t,v))d^{(2)}(z(t,w))(z'(t,w))^2 dt dv dw O(1) \\
&= O\left(\frac{\log h_r^{-1}}{nh_r^3}\right) - R_n O(1) \tag{4.37}
\end{aligned}$$

mit

$$R_n = \int_{-1}^1 \int_{-1}^1 K_d(v)K_d(w) \int_{m'(a)}^{m'(x)} d^{(1)}(z(t,v))d^{(2)}(z(t,w))(z'(t,w))^2 dt dv dw.$$

Der Restterm R_n lässt sich mit einer Taylorentwicklung weiter abschätzen zu

$$\begin{aligned}
R_n &= \int_{-1}^1 \int_{-1}^1 K_d(v)K_d(w) \int_{m'(a)}^{m'(x)} \{d^{(1)}(m'^{-1}(t)) + h_d v d^{(2)}(\xi_1)\} \\
&\quad \times \{(m'^{-1})'^2(t)d^{(2)}(m'^{-1}(t)) + h_d w [(m'^{-1})'^2 d^{(2)}]'(\xi_2)\} dt dv dw \\
&= \int_a^x d^{(1)}(z)d^{(2)}(z)dz O(1) \\
&+ h_d \int_{-1}^1 \int_{-1}^1 v K_d(v)K_d(w) \int_{m'(a)}^{m'(x)} [(m'^{-1})'^2 d^{(2)} \circ m'^{-1}](t) [(m'^{-1})' d^{(2)} \circ m'^{-1}](\xi_1) dt dv dw \\
&+ h_d \int_{-1}^1 \int_{-1}^1 w K_d(v)K_d(w) \int_{m'(a)}^{m'(x)} [2(m'^{-1})'(m'^{-1})'' d^{(2)} \circ m'^{-1} + (m'^{-1})'^3(t)d^{(3)} \circ m'^{-1}] \\
&\quad \times d^{(1)}(m'^{-1}(t)) dt dv dw \\
&+ h_d^2 \int_{-1}^1 \int_{-1}^1 v w K_d(v)K_d(w) \int_{m'(a)}^{m'(x)} [(m'^{-1})' d^{(2)} \circ m'^{-1}](\xi_1) \\
&\quad \times [2(m'^{-1})'(m'^{-1})'' d^{(2)} \circ m'^{-1} + (m'^{-1})'^3 d^{(3)} \circ m'^{-1}](\xi_2) dt dv dw.
\end{aligned}$$

Weiter führt eine partielle Integration fast sicher zu

$$\int_a^x d^{(1)}(z)d^{(2)}(z)dz O(1) = (d^{(1)}(z))^2 \Big|_a^x O(1) = O\left(\frac{\log h_r^{-1}}{nh_r^3}\right)$$

und

$$R_n = O\left(\frac{\log h_r^{-1}}{nh_r^3}\right) + h_d O\left(\frac{\log h_r^{-1}}{nh_r^5}\right) + h_d O\left(\frac{\log h_r^{-1}}{nh_r^5}\right) + O\left(h_d^2 \frac{\log h_r^{-1}}{nh_r^6}\right)$$

$$= O\left(\frac{\log h_r^{-1}}{nh_r^3}\right) + O\left(h_d \frac{\log h_r^{-1}}{nh_r^5}\right).$$

Eingesetzt in (4.37) erhält man

$$\int_{m'(a)}^{m'(x)} \frac{\Delta_n^{(1)}(t)\Delta_n^{(1)'}(t)}{(m'^{-1})'(t)} dt = O\left(\frac{\log h_r^{-1}}{nh_r^3}\right) + O\left(h_d \frac{\log h_r^{-1}}{nh_r^5}\right) = o\left(\frac{1}{\sqrt{nh_r}}\right). \quad (4.38)$$

Für den Term $\Delta_n^{(2)}(t)$ ergibt sich

$$\begin{aligned} \Delta_n^{(2)}(t) &= - \int_{-1}^1 K_d(w) \{(m'^{-1})''(t + h_d w) d^{(1)2}(m'^{-1}(t + h_d w)) \\ &\quad + 2(m'^{-1})'2(t + h_d w) d^{(1)}(m'^{-1}(t + h_d w)) d^{(2)}(m'^{-1}(t + h_d w))\} dw \\ &= O\left(\frac{\log h_r^{-1}}{nh_r^4}\right) \end{aligned} \quad (4.39)$$

und schließlich

$$\begin{aligned} \Delta_n^{(2)'}(t) &= \frac{1}{h_d} \int_{-1}^1 K_d'(w) \{(m'^{-1})''(t + h_d w) d^{(1)2}(m'^{-1}(t + h_d w)) \\ &\quad + 2(m'^{-1})'2(t + h_d w) d^{(1)}(m'^{-1}(t + h_d w)) d^{(2)}(m'^{-1}(t + h_d w))\} dw \\ &= O\left(\frac{\log h_r^{-1}}{nh_r^4 h_d}\right). \end{aligned} \quad (4.40)$$

Eingesetzt in (4.34) ergeben (4.28) - (4.40) für $B_n^{(1)}(x)$ unter den Voraussetzungen (4.3) - (4.7) die Ordnung

$$\begin{aligned} B_n^{(1)}(x) &= O(h_d^3) + O\left(\frac{h_d^2 \log h_r^{-1}}{nh_r^3}\right)^{1/2} + O\left(\frac{h_d \log h_r^{-1}}{nh_r^4}\right) + O\left(\frac{h_d^4 \log h_r^{-1}}{nh_r^5}\right)^{1/2} \\ &\quad + O\left(\frac{\log h_r^{-1}}{nh_r^4}\right) + O\left(\frac{\log h_r^{-1}}{nh_r^3}\right) + O\left(\frac{(\log h_r^{-1})^{3/2}}{n^{3/2} h_r^{13/2}}\right) + O\left(\frac{h_d \log h_r^{-1}}{nh_r^4}\right) \\ &\quad + O\left(\frac{(\log h_r^{-1})^{3/2}}{n^{3/2} h_r^{11/2} h_d}\right) + O\left(\frac{(\log h_r^{-1})^2}{nh_r^8 h_d}\right) \\ &= \left\{ O\left(\sqrt{nh_r^7} \left(\frac{h_d}{h_r^{3/2}}\right) h_r^{3/2}\right) + O\left(\frac{h_d}{h_r^{3/2}} \left(\frac{\log h_r^{-1}}{nh_r^6}\right)^{1/2} \sqrt{nh_r^7}\right) + O\left(\frac{h_d \log h_r^{-1}}{h_r^{3/2} n^{1/2} h_r^2}\right) \right. \\ &\quad + O\left(\left(\frac{h_d}{h_r^{3/2}}\right)^2 \left(\frac{\log h_r^{-1}}{nh_r^5}\right)^{1/2} \sqrt{nh_r^7}\right) + O\left(\frac{\log h_r^{-1}}{n^{1/2} h_r^{5/2}}\right) + O\left(\frac{(\log h_r^{-1})^{3/2}}{nh_r^6}\right) \\ &\quad + O\left(\frac{\log h_r^{-1}}{n^{1/2} h_r^{1/2}} \frac{h_d}{h_r^{3/2}}\right) + O\left(\frac{\log h_r^{-1}}{nh_r^5 h_d} (h_r \log h_r^{-1})^{1/2}\right) \\ &\quad \left. + O\left(\frac{\log h_r^{-1}}{nh_r^5 h_d} \left(\frac{\log h_r^{-1}}{nh_r^6}\right)^{1/2} (h_r \log h_r^{-1})^{1/2}\right) \right\} O\left(\frac{1}{\sqrt{nh_r}}\right) \\ &= o\left(\frac{1}{\sqrt{nh_r}}\right). \end{aligned}$$

Für den verbleibenden Restterm $B_n^{(2)}(x)$ erhält man mit denselben Argumenten wie bei der Herleitung der Darstellung (4.34)

$$\begin{aligned}
B_n^{(2)}(x) &= \int_{m'(a)}^{m'(x)} (\Delta_n^{(0)}(t))^2 \frac{(m'^{-1})''(t)}{(m'^{-1})'^3(t)} dt + \int_{m'(a)}^{m'(x)} (\Delta_n^{(1)}(t))^2 \frac{(m'^{-1})''(t)}{(m'^{-1})'^3(t)} dt \\
&+ \int_{m'(a)}^{m'(x)} (\Delta_n^{(2)}(t))^2 \frac{(m'^{-1})''(t)}{(m'^{-1})'^3(t)} dt + 2 \int_{m'(a)}^{m'(x)} \Delta_n^{(0)}(t) \Delta_n^{(1)}(t) \frac{(m'^{-1})''(t)}{(m'^{-1})'^3(t)} dt \\
&+ 2 \int_{m'(a)}^{m'(x)} \Delta_n^{(0)}(t) \Delta_n^{(2)}(t) \frac{(m'^{-1})''(t)}{(m'^{-1})'^3(t)} dt + 2 \int_{m'(a)}^{m'(x)} \Delta_n^{(1)}(t) \Delta_n^{(2)}(t) \frac{(m'^{-1})''(t)}{(m'^{-1})'^3(t)} dt.
\end{aligned} \tag{4.41}$$

Mit den Abschätzungen (4.25), (4.28) und (4.39) ergibt sich dann fast sicher

$$\begin{aligned}
B_n^{(2)}(x) &= O(h_d^4) + O\left(\frac{\log h_r^{-1}}{nh_r^3}\right) + O\left(\frac{\log h_r^{-1}}{nh_r^4}\right)^2 + O\left(\frac{h_d^4 \log h_r^{-1}}{nh_r^3}\right)^{1/2} \\
&+ O\left(\frac{h_d^2 \log h_r^{-1}}{nh_r^4}\right) + O\left(\frac{(\log h_r^{-1})^{3/2}}{n^{3/2} h_r^{11/2}}\right) \\
&= \left\{ O\left(\left(\frac{h_d}{h_r^{3/2}}\right)^4 \sqrt{nh_r^7} h_r^3\right) + O\left(\frac{\log h_r^{-1}}{n^{1/2} h_r^{5/2}}\right) + O\left(\frac{(\log h_r^{-1})^2}{n^{3/2} h_r^{15/2}}\right) \right. \\
&+ O\left(\left(\frac{h_d}{h_r^{3/2}}\right)^2 \sqrt{nh_r^7} \left(\frac{\log h_r^{-1}}{nh_r^3}\right)^{1/2}\right) + O\left(\left(\frac{h_d}{h_r^{3/2}}\right)^2 \sqrt{nh_r^7} \frac{\log h_r^{-1}}{nh_r^4}\right) \\
&+ O\left(\frac{(\log h_r^{-1})^{3/2}}{nh_r^5}\right) \left. \right\} O\left(\frac{1}{\sqrt{nh_r}}\right) \\
&= o\left(\frac{1}{\sqrt{nh_r}}\right).
\end{aligned}$$

Damit ist der Beweis von Lemma 4.4 abgeschlossen. □

Kapitel 5

Testen auf strenge Isotonie der Regressionsfunktion

Auf strenge Isotonie der Regressionsfunktion sollte immer dann getestet werden, falls man aufgrund der Daten Hinweise auf eine solche Form der Regressionsfunktion erhält ohne diese aus dem Sachzusammenhang erschließen zu können. Entscheidet der Test für strenge Isotonie der Regressionsfunktion, kann zur weiteren Datenanalyse ein streng monoton wachsender Schätzer der Regressionsfunktion verwendet werden. Im Folgenden wird ein neues Testverfahren basierend auf der Verteilungsfunktion des uneingeschränkten Schätzers \hat{m} vorgestellt und mit einem bereits existierenden von Bowman, Jones und Gijbels (1998) verglichen.

5.1 Definition der Teststatistik

In dem Regressionsmodell (1.1) sollen die Hypothesen

$$H_0 : m \text{ ist streng monoton wachsend} \quad \text{gegen} \quad H_1 : m \text{ ist von anderer Form}$$

getestet werden. Die Teststatistik basiert auf dem Nadaraya-Watson Schätzer

$$\hat{m}(x) = \frac{1}{nh_r} \sum_{i=1}^n K_r\left(\frac{x - X_i}{h_r}\right) Y_i / \hat{f}_n(x)$$

als ursprünglichen Schätzer der Regressionsfunktion und der Verteilungsfunktion von \hat{m} bezüglich des Lebesgue-Maßes

$$\hat{\phi}(t) = \phi(\hat{m})(t) = \int_0^1 I\{\hat{m}(v) \leq t\} dv$$

beziehungsweise ihrer geglätteten Version

$$\hat{\phi}_{h_d}(t) = \phi_{h_d}(\hat{m})(t) = \frac{1}{h_d} \int_0^1 \int_{-\infty}^t K_d\left(\frac{\hat{m}(v) - u}{h_d}\right) dudv.$$

Letztere kann auch als inverser isotoner Regressionsschätzer betrachtet werden (vgl. Definition 2.4 und Satz 3.2). In obigen Definitionen bezeichne

$$\hat{f}_n(x) = \frac{1}{nh_r} \sum_{i=1}^n K_r\left(\frac{x - X_i}{h_r}\right)$$

den Dichteschätzer basierend auf der Stichprobe X_1, \dots, X_n , K_r und K_d seien Kernfunktionen der Ordnung 2 und h_r und h_d seien die zugehörigen Bandbreiten.

Als Teststatistik wird die Statistik

$$\tilde{T}_n = \int_0^1 (\hat{\phi}(\hat{m}(x)) - x)^2 dx$$

bzw. ihre geglättete Form

$$T_n = \int_0^1 (\hat{\phi}_{h_d}(\hat{m}(x)) - x)^2 dx$$

verwendet. Ist die wahre Regressionsfunktion bereits monoton, so ist zu erwarten, dass $\hat{\phi} \circ \hat{m}$ bzw. $\hat{\phi}_{h_d} \circ \hat{m}$ ungefähr die Identität ergeben, die Teststatistik somit nahe bei 0 liegt. In folgendem Lemma wird gezeigt, dass die geglättete Teststatistik T_n asymptotisch äquivalent ist zu \tilde{T}_n .

Lemma 5.1 *Es sei K_d ein positiver Kern mit kompaktem Träger $[-1, 1]$ und der Nadaraya-Watson Schätzer sei gleichmäßig konsistent. Dann gilt $T_n = \tilde{T}_n + o_P(1)$.*

Beweis Die Differenz der Teststatistiken kann abgeschätzt werden durch

$$\begin{aligned} T_n - \tilde{T}_n &= \int_0^1 \left((\hat{\phi}_{h_d}(\hat{m}(x)) - x)^2 - (\hat{\phi}(\hat{m}(x)) - x)^2 \right) dx \\ &= \int_0^1 \left(\hat{\phi}_{h_d}^2(\hat{m}(x)) - \hat{\phi}^2(\hat{m}(x)) - 2x(\hat{\phi}_{h_d}(\hat{m}(x)) - \hat{\phi}(\hat{m}(x))) \right) dx \\ &= \int_0^1 (\hat{\phi}_{h_d}(\hat{m}(x)) + \hat{\phi}(\hat{m}(x)) - 2x)(\hat{\phi}_{h_d}(\hat{m}(x)) - \hat{\phi}(\hat{m}(x))) dx \\ &= O_P(1) \int_0^1 (\hat{\phi}_{h_d}(\hat{m}(x)) - \hat{\phi}(\hat{m}(x))) dx. \end{aligned}$$

Daher reicht es zu zeigen, dass die Differenz $\hat{\phi}_{h_d}(\hat{m}(x)) - \hat{\phi}(\hat{m}(x))$ stochastisch gleichmäßig gegen 0 konvergiert. Es ist

$$\hat{\phi}_{h_d}(\hat{m}(x)) = \frac{1}{h_d} \int_0^1 \int_{-\infty}^{\hat{m}(x)} K_d\left(\frac{\hat{m}(v) - u}{h_d}\right) dudv$$

$$\begin{aligned}
&= \frac{1}{h_d} \int_0^1 I\{\hat{m}(v) \leq \hat{m}(x) + h_d\} \int_{\hat{m}(v)-h_d}^{\hat{m}(x)} K_d\left(\frac{\hat{m}(v) - u}{h_d}\right) dudv \\
&= \int_0^1 I\{\hat{m}(v) \leq \hat{m}(x) + h_d\} \int_{(\hat{m}(v)-\hat{m}(x))/h_d} K_d(u) dudv \\
&= \int_0^1 I\{\hat{m}(v) \leq \hat{m}(x) + h_d\} I\{\hat{m}(v) \leq \hat{m}(x) - h_d\} dv \\
&\quad + \int_0^1 I\{\hat{m}(x) - h_d \leq \hat{m}(v) \leq \hat{m}(x) + h_d\} \int_{(\hat{m}(v)-\hat{m}(x))/h_d}^1 K_d(u) dudv \\
&= \int_0^1 I\{\hat{m}(v) \leq \hat{m}(x) - h_d\} dv \\
&\quad + \int_0^1 I\{\hat{m}(x) - h_d \leq \hat{m}(v) \leq \hat{m}(x) + h_d\} \int_{(\hat{m}(v)-\hat{m}(x))/h_d}^1 K_d(u) dudv
\end{aligned}$$

und somit gilt für die Differenz

$$\begin{aligned}
\hat{\phi}_{h_d}(\hat{m}(x)) - \hat{\phi}(\hat{m}(x)) &= \int_0^1 I\{\hat{m}(x) - h_d \leq \hat{m}(v) \leq \hat{m}(x)\} dv \\
&\quad - \int_0^1 I\{\hat{m}(x) - h_d \leq \hat{m}(v) \leq \hat{m}(x) + h_d\} \int_{(\hat{m}(v)-\hat{m}(x))/h_d}^1 K_d(u) dudv.
\end{aligned}$$

Das erste Integral ist von der Ordnung $o_P(1)$, da die Indikatorfunktion im Integral wegen der gleichmäßigen Konsistenz des Nadaraya-Watson Schätzers gegen 0 konvergiert. Das zweite Integral lässt sich abschätzen durch

$$\begin{aligned}
&\int_0^1 I\{\hat{m}(x) - h_d \leq \hat{m}(v) \leq \hat{m}(x) + h_d\} \int_{(\hat{m}(v)-\hat{m}(x))/h_d}^1 K_d(u) dudv \\
&\leq \int_0^1 I\{\hat{m}(x) - h_d \leq \hat{m}(v) \leq \hat{m}(x) + h_d\}
\end{aligned}$$

und ist daher ebenfalls von der Ordnung $o_P(1)$. \square

Wiederum wegen der gleichmäßigen Konsistenz des Nadaraya-Watson Schätzers konvergiert die Teststatistik \tilde{T}_n fast sicher gegen ihre deterministische Version

$$\tilde{T} = \int_0^1 \left(\int_0^1 I\{m(v) \leq m(x)\} dv - x \right)^2 dx.$$

Um zu zeigen, dass die Teststatistik strenge Isotonie eindeutig erkennt, reicht es daher \tilde{T} für streng monoton wachsende und nicht streng monoton wachsende Funktionen m zu betrachten. Lemma 5.1 garantiert dann, dass die geglättete Version T_n für hinreichend großes $n \in \mathbb{N}$ ebenfalls diese Eigenschaft erfüllt.

Lemma 5.2 *Es sei m stetig. Dann gilt*

$$\tilde{T} = \int_0^1 \left(\int_0^1 I\{m(v) \leq m(x)\} dv - x \right)^2 dx = 0$$

genau dann wenn m streng monoton wachsend ist.

Beweis Da der Integrand nichtnegativ ist, ist \tilde{T} genau dann null, wenn der Integrand für fast alle x aus dem Intervall $[0, 1]$ null ergibt. Es reicht daher zu zeigen, dass fast überall

$$\int_0^1 I\{m(v) \leq m(x)\} dv = x \quad (5.1)$$

gilt genau dann, wenn m streng monoton wachsend ist.

Zunächst sei m streng monoton wachsend. Dann existiert die Inverse m^{-1} von m und

$$\int_0^1 I\{m(v) \leq m(x)\} dv = \int_0^1 I\{v \leq x\} dv = x \text{ für } x \in [0, 1].$$

Es gelte nun die Bedingung (5.1). Im Folgenden werden in einer Fallunterscheidung die unterschiedlichen Formen von nicht streng monoton wachsenden Funktionen betrachtet und es wird gezeigt, dass eine Menge $A \subset [0, 1]$ mit positivem Lebesgue-Maß existiert, so dass für alle $x \in A$ die Gleichung (5.1) nicht erfüllt ist.

a) Sei m monoton wachsend, aber nicht streng monoton wachsend. Dann gibt es Intervalle A_i , $i = 1, \dots, r$, $r \in \mathbb{N}$, auf denen m konstant ist und Intervalle B_j , $j = 1, \dots, s$ auf denen m streng monoton wachsend ist. $\{\{A_i\}_{i=1}^r, \{B_j\}_{j=1}^s\}$ ist eine disjunkte Zerlegung des Intervalls $[0, 1]$. Dann kann m geschrieben werden als

$$m(x) = \sum_{i=1}^r m_i I_{A_i}(x) + \sum_{j=1}^s m_j^\nearrow(x) I_{B_j}(x) \quad (5.2)$$

mit Konstanten $m_i = m(x)$ für $x \in A_i$, $i = 1, \dots, r$ und streng monoton wachsenden Funktionen $m_j^\nearrow(x) = m(x)$ für $x \in B_j$, $j = 1, \dots, s$.

Für monoton wachsende Funktionen m und t aus dem Bildbereich von m ist

$$\phi(t) = \int_0^1 I\{m(v) \leq t\} dv = \sup\{v \in [0, 1] \mid m(v) \leq t\}$$

eine Art verallgemeinerte Inverse. Ist $x \in A_i$, $i = 1, \dots, r$, so gilt $\phi(m(x)) \geq x$ wegen der Definition der verallgemeinerten Inversen und $\phi(m(x)) > x$ für x aus dem Inneren A_i° von A_i . Somit gilt $\phi(m(x)) - x > 0$ für $x \in A_i^\circ$. Diese Menge besitzt positives Lebesgue-Maß

und \tilde{T} ist größer als 0. Das schließt den Spezialfall einer auf $[0,1]$ konstanten Regressionsfunktion ein.

b) Sei m monoton fallend. Dann existieren wie oben Intervalle A_i , $i = 1, \dots, r$, auf denen m konstant ist und Intervalle B_j , $j = 1, \dots, s$, auf denen m streng monoton fallend ist. Für m ergibt sich die Darstellung

$$m(x) = \sum_{i=1}^r m_i I_{A_i}(x) + \sum_{j=1}^s m_j^{\searrow}(x) I_{B_j}(x)$$

mit Konstanten $m_i = m(x)$ für $x \in A_i$ und streng monoton fallenden Funktionen $m_j^{\searrow}(x) = m(x)$ für $x \in B_j$.

$$\phi(m(x)) = \int_0^1 I\{m(v) \leq m(x)\} dv = 1 - \inf\{v \in [0, 1] \mid m(v) \leq m(x)\}$$

Ist $s \geq 1$, also mindestens ein B_j nicht leer, so gilt auf B_j , $j = 1, \dots, s$ $\phi(m(x)) = 1 - x \neq x$ für fast alle x . Somit ist $\phi(m(x)) \neq x$ auf einer Menge $A \subset B_j$ mit positivem Lebesgue-Maß und \tilde{T} ist größer als 0.

c) Es sei nun m weder monoton wachsend noch fallend. Dann existiert eine disjunkte Zerlegung $\{\{A_i\}_{i=1}^r, \{B_j\}_{j=1}^s, \{C_k\}_{k=1}^t\}$, $r, s \geq 1$ von $[0, 1]$, so dass m auf A_i streng monoton wachsend, auf B_j streng monoton fallend und auf C_k konstant ist (m hat mindestens ein streng monoton wachsendes und ein streng monoton fallendes Stück). Wie oben lässt sich die Funktion m darstellen als

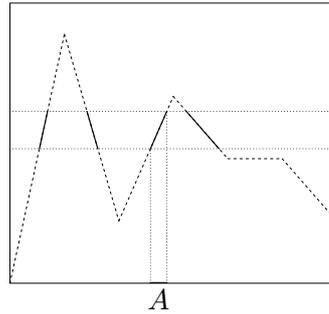
$$m(x) = \sum_{i=1}^r m_i^{\nearrow}(x) I_{A_i}(x) + \sum_{j=1}^s m_j^{\searrow}(x) I_{B_j}(x) + \sum_{k=1}^t m_k I_{C_k}(x)$$

und

$$\phi(t) = \sum_{i=1}^r \int_{A_i} I\{m_i^{\nearrow}(v) \leq t\} dv + \sum_{j=1}^s \int_{B_j} I\{m_j^{\searrow}(v) \leq t\} dv + \sum_{k=1}^t \int_{C_k} I\{m_k \leq t\} dv$$

Seien $a_i^{(l)}$, $b_j^{(l)}$ bzw. $c_k^{(l)}$ die unteren Grenzen der Intervalle A_i , B_j bzw. C_k und $a_i^{(u)}$, $b_j^{(u)}$ bzw. $c_k^{(u)}$ die oberen Grenzen. Dann ist

$$\begin{aligned} \phi(t) &= \sum_{i=1}^r \left((m_i^{\nearrow^{-1}}(t) - a_i^{(l)}) I\{m(a_i^{(l)}) < t \leq m(a_i^{(u)})\} + (a_i^{(u)} - a_i^{(l)}) I\{m(a_i^{(u)}) < t\} \right) \\ &+ \sum_{j=1}^s \left((b_j^{(u)} - m_j^{\searrow^{-1}}(t)) I\{m(b_j^{(u)}) < t \leq m(b_j^{(l)})\} + (b_j^{(u)} - b_j^{(l)}) I\{m(b_j^{(l)}) < t\} \right) \\ &+ \sum_{k=1}^t (c_k^{(u)} - c_k^{(l)}) I\{m_k \leq t\} \end{aligned}$$

Abbildung 5.1: Eine mögliche Wahl der Menge A

Sei A_l für ein $l \in \{1, \dots, r\}$ so gewählt, dass $m(A_l) \cap m(B_k)$ nicht leer ist für ein $k \in \{1, \dots, s\}$, d.h. es wird ein Bereich betrachtet, auf dem die Funktion m streng monoton wachsend ist ($m(x) = m_l^\nearrow(x)$) und dessen Bildbereich einen nichtleeren Schnitt mit mindestens einem der Bildbereiche der streng monoton fallenden Stücke hat.

Sei A ein nichtleeres Intervall aus der Menge $m^{-1}(m(A_l) \cap m(B_k)) \cap A_l$ so gewählt, dass die Regressionsfunktion m auf $m^{-1}(m(A))$ weder Extremstellen noch flache Stücke besitzt (solch eine Menge A existiert immer, vgl. Abbildung 5.1). Dann lässt sich $\phi(m(x))$ auf A darstellen als eine Summe von streng monoton wachsenden Funktionen und einer Konstanten κ ,

$$\begin{aligned} \phi(m(x)) &= \sum_{i \in \mathcal{I}} m_i^\nearrow^{-1}(m(x)) - \sum_{j \in \mathcal{J}} m_j^\searrow^{-1}(m(x)) + \kappa \\ &= \sum_{i \in \mathcal{I}} m_i^\nearrow^{-1}(m_l^\nearrow(x)) - \sum_{j \in \mathcal{J}} m_j^\searrow^{-1}(m_l^\nearrow(x)) + \kappa \\ &= x + \sum_{i \in \mathcal{I}, i \neq l} m_i^\nearrow^{-1}(m_l^\nearrow(x)) - \sum_{j \in \mathcal{J}} m_j^\searrow^{-1}(m_l^\nearrow(x)) + \kappa. \end{aligned}$$

Summiert wird über alle i und j , für die gilt $m(A) \subset m(A_i)$ bzw. $m(A) \subset m(B_j)$ und zumindest die Summe über $j \in \mathcal{J}$ ist von null verschieden, da nach Voraussetzung $m(A) \subset m(B_k)$ ist. Somit kann $\phi \circ m$ auf der Menge A nicht die Identität sein, denn eine Summe von streng monoton wachsenden Funktionen gleicht nicht einer Konstanten. \square

Bemerkung Zwei ebenfalls plausible Teststatistiken sind durch die Statistiken

$$\int_{\hat{R}} (\hat{m}(\hat{\phi}_{h_d}(t)) - t)^2 dt$$

und

$$\int_0^1 (\hat{\phi}_{h_d}^{-1}(x) - \hat{m}(x))^2 dx$$

gegeben, wobei \hat{R} den Bildbereich von \hat{m} bezeichne. Beide sind aber bei näherer Betrachtung für obige Fragestellung nicht geeignet. Zum einen hängt der Integrationsbereich der ersten von der zufälligen Funktion \hat{m} ab. Zum anderen unterscheiden beide Teststatistiken nicht zwischen Monotonie und strenger Monotonie denn für eine monotone aber nicht streng monotone Funktion m und alle t aus dem Bildbereich von m gilt

$$m(\phi(t)) = m(\sup\{v \in [0, 1] \mid m(v) \leq t\}) = t$$

und für alle $x \in [0, 1]$

$$\phi^{-1}(x) = m(x).$$

5.2 Asymptotische Verteilung der Teststatistik

In diesem Abschnitt gelten zusätzlich die folgenden Bedingungen an die Kernfunktionen und Bandbreiten.

- (K1) Der Kern K_r sei ein dreimal stetig differenzierbarer Kern der Ordnung 2 mit kompaktem Träger $[-1, 1]$ und $K_r(\pm 1) = K_r'(\pm 1) = 0$
- (K2) Der Kern K_d sei ein zweimal stetig differenzierbarer Kern der Ordnung 2 mit kompaktem Träger $[-1, 1]$ und $K_d(\pm 1) = K_d'(\pm 1) = 0$
- (B) Für $n \rightarrow \infty$ gelte für die Bandbreiten: $h_r, h_d \rightarrow 0$, $nh_r, nh_d \rightarrow \infty$, $h_r = O(n^{-1/5})$, $h_d^2 \log h_r^{-1} / h_r^{5/2} \rightarrow 0$ und $h_r^{1/2} (\log h_r^{-1})^2 / nh_d^4 = o(1)$

Satz 5.3 *Es sei m viermal stetig differenzierbar. Die Dichte f sei dreimal stetig differenzierbar, auf $[0, 1]$ von 0 weg beschränkt und m sei streng monoton wachsend mit $m'(x) > 0$ für alle $x \in (0, 1)$. Die Kerne und Bandbreiten erfüllen die Bedingungen (K1), (K2) und (B). Weiter sei das vierte Moment $\mu_4(X_1) = \mathbb{E}[(Y_1 - m(X_1))^4 | X_1]$ wohldefiniert mit $\mathbb{E}[\mu_4(X_1)] < \infty$.*

Dann gilt

$$\frac{nh_r^{9/2}}{h_d^4} \left(T_n - h_d^4 \kappa_2^2(K_d)(B_n^{[1]} + B_n^{[2]}) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, V)$$

mit

$$B_n^{[1]} = \int_0^1 \mathbb{E} \left[\frac{1}{n^2 h_r^6} \sum_{i=1}^n \sigma^2(X_i) K_r''^2 \left(\frac{x - X_i}{h_r} \right) \right] (f(x))^{-2} (m'(x))^{-6} dx$$

$$B_n^{[2]} = \int_0^1 (m''(x))^2 / (m'(x))^6 dx$$

und asymptotischer Varianz

$$V = 4\kappa_2^4(K_d) \left(\int_0^1 \sigma^2(y) f^2(y) (m'(y))^{-12} dy \right) \left(\int_0^1 \left(\int_0^1 K_r''(x) K_r''(x+z) dx \right)^2 dz \right).$$

Bemerkung 5.4 Gelten die Voraussetzungen von Satz 5.3 und ist σ^2 darüberhinaus einmal stetig differenzierbar auf $[0, 1]$, so gilt

$$\frac{nh_r^{9/2}}{h_d^4} \left(T_n - h_d^4 \kappa_2^2(K_d) (\tilde{B}_n^{[1]} + B_n^{[2]}) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, V)$$

mit

$$\tilde{B}_n^{[1]} = \frac{1}{nh_r^5} \int_0^1 \frac{\sigma^2(x)}{f(x)(m'(x))^6} dx \int_{-1}^1 K_r''^2(y) dy.$$

In diesem Fall kann bei der Berechnung des Erwartungswerts eine Taylorentwicklung erster Ordnung durchgeführt werden. Man erhält

$$\begin{aligned} B_n^{[1]} &= \frac{1}{nh_r^6} \int_0^1 \int_0^1 \sigma^2(y) f(y) K_r''^2 \left(\frac{x-y}{h_r} \right) dy (f(x))^{-2} (m'(x))^{-6} dx \\ &= \frac{1}{nh_r^5} \int_0^1 \frac{\sigma^2(x)}{f(x)(m'(x))^6} dx \int_{-1}^1 K_r''^2(y) dy \\ &\quad + \frac{1}{nh_r^4} \int_{-1}^1 \int_0^1 y K_r''^2(y) \frac{\sigma^{2'}(\xi_y) f(\xi_y) + \sigma^2(\xi_y) f'(\xi_y)}{(f(x))^2 (m'(x))^6} dx dy \\ &= \frac{1}{nh_r^5} \int_0^1 \frac{\sigma^2(x)}{f(x)m^6(x)} dx \int_{-1}^1 K_r''^2(y) dy + o\left(\frac{1}{nh_r^{9/2}}\right). \end{aligned}$$

Das zeigt obige Behauptung.

Bevor Satz 5.3 bewiesen werden kann, benötigt man Aussagen über das asymptotische Verhalten des integrierten quadrierten Fehlers für den Schätzer der Regressionsfunktion und dessen ersten beiden Ableitungen. Diese werden im nächsten Unterabschnitt gegeben.

5.2.1 Asymptotische Verteilung des integrierten quadrierten Fehlers des Regressionsschätzers und dessen Ableitungen

Ein Grenzwertsatz für den integrierten quadrierten Fehler des Schätzers der Regressionsfunktion ist in Hall (1984) gegeben. Analoge Aussagen für Ableitungsschätzer existieren noch nicht in der Literatur und werden daher an dieser Stelle bewiesen. Dazu wird der Nadaraya-Watson Schätzer in der Form

$$\hat{m}(x) - m(x) = \frac{\frac{1}{nh_r} \sum_{i=1}^n (Y_i - m(X_i)) K_r \left(\frac{x-X_i}{h_r} \right)}{\hat{f}(x)} + \frac{\hat{g}(x)}{\hat{f}(x)} \quad (5.3)$$

verwendet. Die Schätzer \hat{f} und \hat{g} sind definiert als

$$\begin{aligned}\hat{f}(x) &= \frac{1}{nh_r} \sum_{i=1}^n K_r\left(\frac{x - X_i}{h_r}\right) \\ \hat{g}(x) &= \frac{1}{nh_r} \sum_{i=1}^n (m(X_i) - m(x)) K_r\left(\frac{x - X_i}{h_r}\right)\end{aligned}$$

mit Bandbreite h_r und zweimal stetig differenzierbarem Kern K_r mit kompaktem Träger $[-\lambda, \lambda]$. Für den Kern und dessen Ableitung gelte $K_r(\pm\lambda) = K_r'(\pm\lambda) = 0$. Der Schätzer der k -ten Ableitung der Regressionsfunktion sei definiert durch

$$\hat{m}^{(k)}(x) = \frac{\partial^k}{\partial x^k} \hat{m}(x).$$

Die Bezeichnung $T_n^{(k)}$ stehe für den integrierten quadrierten Fehler des Ableitungsschätzers,

$$T_n^{(k)} = \int_A (\hat{m}^{(k)}(x) - m^{(k)}(x))^2 w(x) dx.$$

Mit

$$S_k(x) = \left[\frac{1}{nh_r^{k+1}} \sum_{i=1}^n K_r^{(k)}\left(\frac{x - X_i}{h_r}\right) (Y_i - m(X_i)) + \hat{g}^{(k)}(x) \right] / \hat{f}(x) \quad (5.4)$$

hat $\hat{m}^{(k)}(x) - m^{(k)}(x)$ für $k = 0, 1, 2$ eine Darstellung als Linearkombination der $S_j(x)$, $j = 0, \dots, k$,

$$\hat{m}^{(k)}(x) - m^{(k)}(x) = \sum_{l=0}^k a_{k,l}(x) S_l(x), \quad (5.5)$$

und es ist

$$T_n^{(k)} = \sum_{l=0}^k \int_A a_{k,l}^2(x) w(x) S_l^2(x) dx + 2 \sum_{l < m} \int_A a_{k,l}(x) a_{k,m}(x) w(x) S_l(x) S_m(x) dx \quad (5.6)$$

mit stochastisch konvergenten Faktoren

$$\begin{aligned}a_{k,k}(x) &= 1, \quad k = 0, 1, 2, \\ a_{1,0}(x) &= -\hat{f}'(x)/\hat{f}(x), \\ a_{2,1}(x) &= -2\hat{f}'(x)/\hat{f}(x) \text{ und} \\ a_{2,0}(x) &= -(\hat{f}'(x)/\hat{f}(x))^2 + \hat{f}''(x)/\hat{f}(x).\end{aligned}$$

Bevor eine Aussage über die asymptotische Verteilung von $T_n^{(k)}$ getroffen wird, wird in dem folgenden Satz die asymptotische Verteilung von $\int_0^1 S_k^2(x) w(x) dx$ untersucht.

Satz 5.5 *Es sei $A \subset \mathbb{R}$ kompakt und $A^\varepsilon := \{x \in \mathbb{R} \mid \inf_{a \in A} |x - a| < \varepsilon\}$ die Menge aller $x \in \mathbb{R}$ deren Abstand zu A kleiner als ein $\varepsilon > 0$ ist. Die Varianzfunktion σ^2 sei auf A^ε stetig und beschränkt und die Gewichtsfunktion w sei stetig, beschränkt und nichtnegativ. Ist die Regressionsfunktion m auf A^ε $(k+2)$ -mal differenzierbar, $k \in \{0, 1, 2\}$, die Dichte f auf A^ε von 0 weg beschränkt, gleichmäßig stetig und dort $(k+1)$ -mal gleichmäßig stetig differenzierbar und erfüllt die Bandbreite h_r die Bedingungen $h_r \rightarrow 0$, $nh_r^{3/2+k} \rightarrow \infty$ und $nh_r^{2k+5} = O(1)$ für $n \rightarrow \infty$, so gilt für $k = 0$*

$$(n^{-2}h_r^{-1}\alpha_{1,0} + n^{-1}h_r^4\alpha_{2,0})^{-1/2} \left(\int_A S_0^2(x)w(x)dx - B_{n,0}^{[1]} - B_{n,0}^{[2]} \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$$

und für $k = 1, 2$

$$nh_r^{2k+1/2}\alpha_{1,k}^{-1/2} \left(\int_A S_k^2(x)w(x)dx - B_{n,k}^{[1]} - B_{n,k}^{[2]} \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$$

mit

$$B_{n,k}^{[1]} = \frac{1}{nh_r^{2k+2}} \int_A \int_{-1}^1 \mathbb{E} \left[K_r^{(k)^2} \left(\frac{x - X_1}{h_r} \right) \sigma^2(X_1) \right] (f(x))^{-2} w(x) dx$$

$$B_{n,k}^{[2]} = h_r^4 \int_A (\mathbb{E}[\hat{g}^{(k)}(x)])^2 (f(x))^{-2} w(x) dx$$

$$\alpha_{1,k} = 2 \left(\int_A \sigma^4(x)w^2(x)f^{-2}(x)dx \right) \left(\int \left(\int K_r^{(k)}(x)K_r^{(k)}(x+y)dx \right)^2 dy \right)$$

$$\alpha_{2,0} = 4 \int_A \sigma^2(x)\gamma_0^2(x)w^2(x)f^{-4}(x)dx$$

$$\begin{aligned} \gamma_k(x) &= \kappa_2(K_r) \left(m^{(k+2)}(x)f(x) + 2m^{(1)}(x)f^{(k+1)}(x) \right. \\ &\quad \left. + \sum_{j=0}^{k-1} \binom{k}{j+1} \frac{k+2+j}{k-j} m^{(k+2-j)}(x)f^{(j)}(x) \right) \end{aligned}$$

Ein Beweis von Satz 5.5 wird in Anhang A.2 gegeben.

Folgerung 5.6 *Unter den Voraussetzungen von Satz 5.5 und der Bedingung $nh_r^5 = O(1)$ gilt für $0 \leq l < k$*

$$nh_r^{2k+1/2} \int_0^1 S_l^2(x)w(x)dx = O_P(h_r^{2(k-l)-1/2})$$

Beweis Durch Einfügen von $B_{n,l}^{[j]}$, $j = 1, 2$ in obigen Ausdruck erhält man

$$nh_r^{2k+1/2} \int_0^1 S_l^2(x)w(x)dx = nh_r^{2k+1/2} \left(\int_0^1 S_l^2(x)w(x)dx - B_{n,l}^{[1]} - B_{n,l}^{[2]} \right) + nh_r^{2k+1/2} (B_{n,l}^{[1]} + B_{n,l}^{[2]})$$

Der erste Term auf der rechten Seite ist nach Satz 5.5 von der Ordnung $O_P(h_r^{2(k-l)})$. Weiter gilt

$$nh_r^{2k+1/2}B_{n,l}^{[1]} = O(h_r^{2(k-l)-1/2}) = o(1)$$

und

$$nh_r^{2k+1/2}B_{n,l}^{[2]} = O(nh_r^{2k+9/2}) = o(1),$$

da $k \geq 1$ und $h_r = O(n^{-1/5})$ ist. □

Satz 5.7 *Unter den Voraussetzungen von Satz 5.5 gilt für $k = 0$*

$$(n^{-2}h_r^{-1}\alpha_{1,0} + n^{-1}h_r^4\alpha_{2,0})^{-1/2} \left(\int_A (\hat{m}(x) - m(x))^2 w(x) dx - B_{n,0}^{[1]} - B_{n,0}^{[2]} \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$$

und für $k \geq 1$ falls $nh_r^5 = O(1)$

$$nh_r^{2k+1/2}\alpha_{1,k}^{-1/2} \left(\int_A (\hat{m}^{(k)}(x) - m^{(k)}(x))^2 w(x) dx - B_{n,k}^{[1]} \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

Beweis Für $k = 0$ ist dies Theorem 4 aus Hall (1984) für den eindimensionalen Fall und folgt direkt aus Satz 5.5. Ist $k > 0$ und $nh_r^5 = O(1)$, so gilt $B_{n,0}^{[2]} = O(h_r^4) = o(1/nh_r^{2k+1/2})$ und der Beweis von Satz 5.7 folgt aus Satz 5.5 und Folgerung 5.6 mit der Zerlegung

$$\begin{aligned} T_n^{(k)} - B_{n,k}^{[1]} &= \left(\int_A S_k^2(x)w(x)dx - B_{n,k}^{[1]} \right) + \sum_{l=0}^{k-1} \int_A S_l^2(x)a_{k,l}(x)w(x)dx \\ &\quad + 2 \sum_{l < m} \int_A S_l(x)a_{k,l}(x)S_m(x)a_{k,m}(x)w(x)dx. \end{aligned} \quad (5.7)$$

Multipliziert man den ersten Term der rechten Seite mit $nh_r^{2k+1/2}$, so konvergiert er nach Satz 5.5 in Verteilung gegen die in Satz 5.7 angegebene Grenzverteilung. In der ersten Summe auf der rechten Seite ist der Index immer kleiner als k . Daher haben die Summanden nach Folgerung 5.6 die Größenordnung $o_P(1/nh_r^{2k+1/2})$ wenn als Gewichtsfunktion $a_{k,l}(x)w(x)$ verwendet wird. Die zweite Summe wird mit der Cauchy-Schwarz Ungleichung abgeschätzt. Die einzelnen Faktoren sind nach Folgerung 5.6 von der Größenordnung $o_P(1/nh_r^{2k+1/2})^{1/2}$, das Produkt daher von der Ordnung $o_P(1/nh_r^{2k+1/2})$. □

Satz 5.7 bzw. Satz 5.5 können auch losgelöst von diesem Zusammenhang von Interesse sein und lassen sich auf andere Ableitungsschätzer als den hier verwendeten übertragen. Ist bekannt, dass es sich um eine auf dem Intervall A konstante Dichte handelt, ihre erste Ableitung demnach 0 ist, so macht es Sinn, Ableitungsschätzer der Form

$$\hat{m}^{(k)}(x) = \frac{1}{nh_r^{2k+1}} \sum_{i=1}^n K_r^{(k)}\left(\frac{x - X_i}{h_r}\right) Y_i / \hat{f}(x)$$

zu betrachten. Satz 5.7 kommt in diesem Fall ohne die Voraussetzung $h_r = O(n^{-1/5})$ aus und ist äquivalent zu Satz 5.5.

5.2.2 Beweis von Satz 5.3

Die Teststatistik T_n lässt sich als Funktional

$$T_n = \Psi(\hat{\phi}_{hd}, \hat{m})$$

mit

$$\Psi(f, g) = \int_0^1 (f(g(x)) - x)^2 dx$$

schreiben. Weiter sei P eine Abbildung von $[0, 1]$ nach \mathbb{R} mit

$$P(\lambda) = \Psi(f + \lambda\tilde{f}, g + \lambda\tilde{g}).$$

Sie ist bezüglich λ dreimal stetig differenzierbar mit

$$\begin{aligned} P^{(1)}(\lambda) &= 2 \int_0^1 \left\{ [f + \lambda\tilde{f}]([g + \lambda\tilde{g}](x)) - x \right\} \\ &\quad \times \left\{ \tilde{g}(x)[f^{(1)} + \lambda\tilde{f}^{(1)}]([g + \lambda\tilde{g}](x)) + \tilde{f}([g + \lambda\tilde{g}](x)) \right\} dx, \\ P^{(2)}(\lambda) &= 2 \int_0^1 \left\{ \tilde{g}(x)[f^{(1)} + \lambda\tilde{f}^{(1)}]([g + \lambda\tilde{g}](x)) + \tilde{f}([g + \lambda\tilde{g}](x)) \right\}^2 dx \\ &\quad + 2 \int_0^1 \left\{ [f + \lambda\tilde{f}]([g + \lambda\tilde{g}](x)) - x \right\} \\ &\quad \times \left\{ \tilde{g}^2(x)[f^{(2)} + \lambda\tilde{f}^{(2)}]([g + \lambda\tilde{g}](x)) + 2\tilde{g}(x)\tilde{f}^{(1)}([g + \lambda\tilde{g}](x)) \right\} dx, \\ P^{(3)}(\lambda) &= 6 \int_0^1 \left\{ \tilde{g}(x)[f^{(1)} + \lambda\tilde{f}^{(1)}]([g + \lambda\tilde{g}](x)) + \tilde{f}([g + \lambda\tilde{g}](x)) \right\} \\ &\quad \times \left\{ \tilde{g}^2(x)[f^{(2)} + \lambda\tilde{f}^{(2)}]([g + \lambda\tilde{g}](x)) + 2\tilde{g}(x)\tilde{f}^{(1)}([g + \lambda\tilde{g}](x)) \right\} dx \\ &\quad + 2 \int_0^1 \left\{ [f + \lambda\tilde{f}]([g + \lambda\tilde{g}](x)) - x \right\} \\ &\quad \times \left\{ \tilde{g}^3(x)[f^{(3)} + \lambda\tilde{f}^{(3)}]([g + \lambda\tilde{g}](x)) + 2\tilde{g}^2(x)\tilde{f}^{(2)}([g + \lambda\tilde{g}](x)) \right\} dx. \end{aligned}$$

Damit ergibt sich für T_n mit $f = m^{-1}$, $g = m$, $\tilde{f} = \hat{\phi}_{hd} - m^{-1}$ und $\tilde{g} = \hat{m} - m$ die von Mises Entwicklung in $\lambda = 0$

$$T_n = \int_0^1 \left\{ (\hat{m} - m)(x)(m^{-1})'(m(x)) + (\hat{\phi}_{hd} - m^{-1})(m(x)) \right\}^2 dx + \frac{1}{6}P^{(3)}(\lambda^*)$$

mit $\lambda^* \in [0, 1]$. Auch $\hat{\phi}_{h_d}(m(x))$ lässt sich als ein Funktional $\phi_{h_d}(\hat{m})$ an der Stelle $m(x)$ schreiben mit

$$\phi_{h_d}(f)(\cdot) = \frac{1}{h_d} \int_0^1 \int_{-\infty}^{\cdot} K_d\left(\frac{f(v) - u}{h_d}\right) dudv$$

und die von Mises Entwicklung ergibt hier wie in (3.10) die Darstellung

$$\hat{\phi}_{h_d}(m(x)) - m^{-1}(m(x)) = \Delta_n^{(0)}(m(x)) + \Delta_n^{(1)}(m(x)) + \frac{1}{2}\Delta_n^{(2)}(m(x))(1 + o_P(1))$$

mit

$$\begin{aligned} \Delta_n^{(0)}(m(x)) &= \phi_{h_d}(m)(m(x)) - m^{-1}(m(x)), \\ \Delta_n^{(1)}(m(x)) &= - \int_0^1 K_d(v)(m^{-1})'(m(x) + h_d v)(\hat{m} - m)(m^{-1}(m(x) + h_d v))dv, \\ \Delta_n^{(2)}(m(x)) &= -\frac{1}{h_d} \int_0^1 K_d'(v)(m^{-1})'(m(x) + h_d v)(\hat{m} - m)^2(m^{-1}(m(x) + h_d v))dv. \end{aligned}$$

Der Ausdruck $\Delta_n^{(1)}(m(x))$ wird mit Hilfe einer Taylorentwicklung von

$$[(m^{-1})'(\hat{m} - m) \circ m^{-1}](m(x) + h_d v)$$

um $m(x)$ weiter vereinfacht zu

$$\begin{aligned} \Delta_n^{(1)}(m(x)) &= -(m^{-1})'(m(x))(\hat{m} - m)(x) - h_d^2 \kappa_2(K_d)[(m^{-1})'(m(x))]^3(\hat{m} - m)''(x) \\ &\quad - R_n(x) \end{aligned}$$

(K_r und damit \hat{m} sind nach Voraussetzung (K1) dreimal stetig differenzierbar) wobei der Restterm durch

$$\begin{aligned} R_n(x) &= h_d^2 \kappa_2(K_d)[(m^{-1})^{(3)}(m(x))(\hat{m} - m) + 3[(m^{-1})''(m^{-1})'](m(x))(\hat{m} - m)'(x)] \\ &\quad + \frac{h_d^3}{6}[(m^{-1})'(\hat{m} - m) \circ m^{-1}]^{(3)}(\xi_n(x)) \end{aligned}$$

definiert ist. Eingesetzt in die Teststatistik T_n führt dies zu der Zerlegung

$$T_n = h_d^4 \kappa_2^2(K_d) \int_0^1 [m'(x)]^{-6} (\hat{m}^{(2)}(x) - m^{(2)}(x))^2 dx + \int_0^1 \Delta_n^{(0)2}(x) dx + Q_n \quad (5.8)$$

mit

$$\begin{aligned} Q_n &= \int_0^1 R_n^2(x) dx + \frac{1}{4} \int_0^1 (\Delta_n^{(2)}(m(x)))^2 dx \\ &\quad + 2 \left\{ -h_d^2 \kappa_2(K_d) \int_0^1 [m'(x)]^{-3} (\hat{m}^{(2)}(x) - m^{(2)}(x)) \Delta_n^{(0)}(x) dx \right. \end{aligned}$$

$$\begin{aligned}
& -h_d^2 \kappa_2(K_d) \int_0^1 [m'(x)]^{-3} (\hat{m}^{(2)}(x) - m^{(2)}(x)) R_n(x) dx \\
& -h_d^2 \kappa_2(K_d) \int_0^1 [m'(x)]^{-3} (\hat{m}^{(2)}(x) - m^{(2)}(x)) \Delta_n^{(2)}(m(x)) dx \\
& + \int_0^1 \Delta_n^{(0)}(x) R_n(x) dx + \frac{1}{2} \int_0^1 \Delta_n^{(0)}(x) \Delta_n^{(2)}(m(x)) dx \\
& + \frac{1}{2} \int_0^1 R_n(x) \Delta_n^{(2)}(m(x)) dx \Big\} + \frac{1}{6} P^{(3)}(\lambda^*).
\end{aligned}$$

Der erste Term der rechten Seite von (5.8) konvergiert nach Satz 5.7 mit $w(x) = [m'(x)]^{-6}$ in Verteilung gegen eine Normalverteilung mit Konvergenzrate $nh_r^{9/2}/h_d^4$, genauer

$$\frac{nh_r^{9/2}}{h_d^4} \cdot h_d^4 \kappa_2^2(K_d) \left(\int_0^1 [m'(x)]^{-6} (\hat{m}^{(2)}(x) - m^{(2)}(x))^2 dx - B_n^{[1]} \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, V).$$

Der zweite Term ist nicht von der Ordnung $o(h_d^4/nh_r^{9/2})$, ergibt aber

$$\int_0^1 A_{h_d}^2(x) dx = h_d^4 \kappa_2^2(K_d) B_n^{[2]} + o(h_d^6).$$

Der erste Teil wird von der Teststatistik abgezogen und der verbleibende Fehler $o(h_d^6)$ ist von der Ordnung $o(h_d^4/nh_r^{9/2})$.

Zum Abschluss des Beweises wird nun gezeigt, dass $nh_r^{9/2}/h_d^4 Q_n$ stochastisch gegen null konvergiert, der Restterm Q_n also keinen Beitrag zur asymptotischen Verteilung liefert. Dazu werden die neun Terme von Q_n einzeln betrachtet. Die Abschätzung von $\int_0^1 R_n^2(x) dx$ erfolgt ähnlich wie die Abschätzung der Restterme im Beweis von Satz 5.7. Mit der Bezeichnung $d(x) = \hat{m}(x) - m(x)$ ist für eine hinreichend große Konstante C und Gewichtsfunktionen $w_1(x) = [(m^{-1})^{(3)}(m(x))]^2$ und $w_2(x) = [(m^{-1})^{(2)}(m(x))(m^{-1})'(m(x))]^2$

$$\begin{aligned}
\frac{1}{Ch_d^4} \int_0^1 R_n^2(x) dx & \leq \int_0^1 w_1(x) d^2(x) dx + \int_0^1 w_2(x) (d'(x))^2 dx \\
& + h_d^2 \int_0^1 \left([(m^{-1})'d \circ m^{-1}]^{(3)}(\xi_n(x)) \right)^2 dx.
\end{aligned}$$

Die ersten beiden Integrale sind nach Folgerung 5.6 von der Ordnung $o_P(1/nh_r^{9/2})$. Der Integrand des dritten Integrals ist gleichmäßig von der Ordnung $O_P(\log h_r^{-1}/nh_r^7)$ und zusammen mit dem Vorfaktor h_d^2 ergibt das für den dritten Term die Ordnung

$$h_d^2 \int_0^1 \left([(m^{-1})'d \circ m^{-1}]^{(3)}(\xi) \right)^2 dx = O_P \left(h_r^{1/2} \frac{h_d^2 \log h_r^{-1}}{h_r^3} \frac{1}{nh_r^{9/2}} \right) = o_P \left(\frac{1}{nh_r^{9/2}} \right).$$

Fasst man die vorhergehenden Abschätzungen für den ganzen Restterm $\int_0^1 R_n(x)dx$ zusammen, so hat dieser die Ordnung

$$\int_0^1 R_n^2(x)dx = o_P\left(\frac{h_d^4}{nh_r^{9/2}}\right).$$

Die Abschätzung des zweiten Summanden von Q_n ,

$$\begin{aligned} & \int_0^1 (\Delta_n^{(2)}(m(x)))^2 dx \\ &= \int_0^1 \left(\frac{1}{h_d} \int_{-1}^1 K_d'(v)(m^{-1})'(m(x) + h_d v) d^2(m^{-1}(m(x) + h_d v)) dv \right)^2 dx \\ &= \int_0^1 \left(\int_{-1}^1 K_d(v) \left[(m^{-1})''(m(x) + h_d v) d^2(m^{-1}(m(x) + h_d v)) \right. \right. \\ & \quad \left. \left. + 2[(m^{-1})'(m(x) + h_d v)]^2 d'(m^{-1}(m(x) + h_d v)) d(m^{-1}(m(x) + h_d v)) \right] dv \right)^2 dx \\ &= O_P\left(\frac{(\log h_r^{-1})^2}{n^2 h_r^4}\right) = O_P\left(\frac{h_d^4}{nh_r^{9/2}} \frac{(\log h_r^{-1})^2}{nh_d^4} h_r^{1/2}\right) \\ &= o_P\left(\frac{h_d^4}{nh_r^{9/2}}\right), \end{aligned}$$

erfolgt über partielle Integration unter Beachtung, dass $K_d(\pm 1) = 0$ ist.

Auch der erste gemischte Term von $Q_n(x)$ wird mit Hilfe der partiellen Integration abgeschätzt,

$$\begin{aligned} & \left| -h_d^2 \kappa_2(K_d) \int_0^1 [m'(x)]^{-3} (\hat{m}^{(2)}(x) - m^{(2)}(x)) \Delta_n^{(0)}(x) dx \right| \\ &= \left| h_d^4 \kappa_2^2(K_d) \int_0^1 [m'(x)]^{-3} (m^{-1})^{(2)}(m(x)) (\hat{m} - m)^{(2)}(x) dx + o(h_d^4) \int_0^1 (\hat{m} - m)^{(2)}(x) dx \right| \\ &= \left| h_d^4 \kappa_2^2(K_d) \left([m'(x)]^{-3} (m^{-1})^{(2)}(m(x)) (\hat{m}' - m')(x) \right) \Big|_0^1 \right. \\ & \quad \left. - \int_0^1 \{ [m'(x)]^{-3} (m^{-1})^{(2)}(m(x)) \}' (\hat{m}' - m')(x) dx \right) + o(1) (\hat{m}' - m')(x) \Big|_0^1 \Big| \\ &\leq O(h_d^4) \sup |\hat{m}'(x) - m'(x)| = O_P\left(h_d^4 \left(\frac{\log h_r^{-1}}{nh_r^3}\right)^{1/2}\right) \\ &= O_P\left(\frac{h_d^4}{nh_r^{9/2}} (nh_r^5)^{1/2} (h_r \log h_r^{-1})^{1/2}\right) = o_P\left(\frac{h_d^4}{nh_r^{9/2}}\right). \end{aligned}$$

Auf die restlichen fünf gemischten Terme wird die Cauchy-Schwarz Ungleichung angewandt. Es ist

$$\left| h_d^2 \int_0^1 w^{1/2}(x) (\hat{m}^{(2)}(x) - m^{(2)}(x)) \Lambda(x) dx \right|$$

$$\begin{aligned} &\leq \left(h_d^4 \left(\int_0^1 w(x) (\hat{m}^{(2)} - m^{(2)})^2(x) dx - B_{n,2}^{[1]} \right) \int_0^1 \Lambda^2(x) dx \right)^{1/2} \\ &\quad + \left(h_d^4 B_{n,2}^{[1]} \int_0^1 \Lambda^2(x) dx \right)^{1/2}, \end{aligned}$$

wobei $\Lambda(x)$ für die Restterme $R_n(x)$ und $\Delta_n^{(2)}(m(x))$ stehe und $w(x)$ die Gewichtfunktion aus (5.8) sei. Aus den vorhergehenden Abschätzungen und dem Beweis von Satz 5.7 ist ersichtlich, dass gilt

$$\begin{aligned} h_d^4 \left(\int_0^1 w(x) (\hat{m}^{(2)} - m^{(2)})^2(x) dx - B_{n,2}^{[1]} \right) &= O_P \left(\frac{h_d^4}{nh_r^{9/2}} \right), \\ \int_0^1 \Lambda^2(x) dx &= O_P(h_r^{1/2}) o_P \left(\frac{h_d^4}{nh_r^{9/2}} \right) \end{aligned}$$

und $B_{n,2}^{[1]} = O(1)$. Daher folgt

$$\begin{aligned} \left| h_d^2 \int_0^1 w^{1/2}(x) (\hat{m}^{(2)} - m^{(2)})(x) \Lambda(x) dx \right| &= O_P(h_r^{1/4}) o_P \left(\frac{h_d^4}{nh_r^{9/2}} \right) + O(1) o_P \left(\frac{h_d^4}{nh_r^{9/2}} \right) \\ &= o_P \left(\frac{h_d^4}{nh_r^{9/2}} \right). \end{aligned}$$

Die folgenden beiden gemischten Terme sind von der Ordnung

$$\left| \int_0^1 \Delta_n^{(0)}(x) \Lambda(x) dx \right| \leq O(h_d^2) O_P(h_r^{1/4}) o_P \left(\frac{h_d^2}{n^{1/2} h_r^{9/4}} \right) = o_P \left(\frac{h_d^4}{nh_r^{9/2}} \right)$$

und der letzte der gemischten Terme in $Q_n(x)$ ist offensichtlich von der Ordnung $o_P(h_d^4/nh_r^{9/2})$, da sowohl $\int_0^1 R_n^2(x) dx$, als auch $\int_0^1 \Delta_n^{(2)}(x) dx$ von dieser Ordnung sind.

Abschließend muss noch der Restterm $P^{(3)}(\lambda^*)$ diskutiert werden. Dazu werden die fast sicher gleichmäßigen Konvergenzraten für den Schätzer $\hat{\phi}_{h_d}(t)$ und dessen erster bis dritter Ableitung aus Lemma A.2 im Anhang benötigt.

Mit den Bezeichnungen $d(x) = \hat{m}(x) - m(x)$ und $d_{I,-1}(y) = \hat{\phi}_{h_d}(y) - m^{-1}(y)$ hat $P^{(3)}(\lambda^*)$ die Darstellung

$$\begin{aligned} P^{(3)}(\lambda^*) &= 6 \int_0^1 \left\{ d(x) [(m^{-1})^{(1)} + \lambda d_{I,-1}^{(1)}] ([m + \lambda d](x)) + d_{I,-1}([m + \lambda d](x)) \right\} \\ &\quad \times \left\{ d^2(x) [(m^{-1})^{(2)} + \lambda d_{I,-1}^{(2)}] ([m + \lambda d](x)) + 2d(x) d_{I,-1}^{(1)}([m + \lambda d](x)) \right\} dx \\ &\quad + 2 \int_0^1 \left\{ [m^{-1} + \lambda d_{I,-1}] ([m + \lambda d](x)) - x \right\} \\ &\quad \times \left\{ d^3(x) [(m^{-1})^{(3)} + \lambda d_{I,-1}^{(3)}] ([m + \lambda d](x)) + 2d^2(x) d_{I,-1}^{(2)}([m + \lambda d](x)) \right\} dx \end{aligned}$$

$$\begin{aligned}
&= 6 \int_0^1 \left\{ d(x)[(m^{-1})^{(1)} + \lambda d_{I,-1}^{(1)}]([m + \lambda d](x)) + d_{I,-1}([m + \lambda d](x)) \right\} \\
&\quad \times \left\{ d^2(x)[(m^{-1})^{(2)} + \lambda d_{I,-1}^{(2)}]([m + \lambda d](x)) + 2d(x)d_{I,-1}^{(1)}([m + \lambda d](x)) \right\} dx \\
&\quad + 2 \int_0^1 \left\{ d(x)(m^{-1})'(\hat{\xi}(x)) + \lambda d_{I,-1}([m + \lambda d](x)) \right\} \\
&\quad \times \left\{ d^3(x)[(m^{-1})^{(3)} + \lambda d_{I,-1}^{(3)}]([m + \lambda d](x)) + 3d^2(x)d_{I,-1}^{(2)}([m + \lambda d](x)) \right\} dx.
\end{aligned}$$

Für die oben definierten Differenzen gelten nach Mack und Silverman (1982) und Lemma A.2 im Anhang die folgenden gleichmäßig fast sicheren Konvergenzraten

$$\begin{aligned}
d(x) &= O\left(\frac{\log h_r^{-1}}{nh_r}\right)^{1/2} \\
d_{I,-1}^{(k)}(y) &= O\left(\frac{\log h_r^{-1}}{nh_r^{2k+1}}\right)^{1/2} + O(h_d^2) \text{ für } k = 0, 1, 2 \\
d_{I,-1}^{(3)}(y) &= O\left(\frac{\log h_r^{-1}}{nh_r^7}\right)^{1/2} + o(h_d).
\end{aligned}$$

Damit erhält man als Abschätzung für $P^{(3)}(\lambda)$

$$\begin{aligned}
P^{(3)}(\lambda) &= \left\{ O_P\left(\frac{\log h_r^{-1}}{nh_r}\right)^{1/2} \left[O_P(1) + O_P\left(\frac{\log h_r^{-1}}{nh_r^3}\right)^{1/2} + O(h_d^2) \right] + O(h_d^2) \right\} \\
&\quad \times \left\{ O_P\left(\frac{\log h_r^{-1}}{nh_r}\right) \left[O_P(1) + O_P\left(\frac{\log h_r^{-1}}{nh_r^5}\right)^{1/2} + O(h_d^2) \right] \right. \\
&\quad \left. + O_P\left(\frac{\log h_r^{-1}}{nh_r}\right)^{1/2} \left[O_P\left(\frac{\log h_r^{-1}}{nh_r^3}\right)^{1/2} + O(h_d^2) \right] \right\} \\
&+ \left\{ O_P\left(\frac{\log h_r^{-1}}{nh_r}\right)^{1/2} + O(h_d^2) \right\} \left\{ O_P\left(\frac{\log h_r^{-1}}{nh_r}\right)^{3/2} \left[O_P(1) + O_P\left(\frac{\log h_r^{-1}}{nh_r^7}\right)^{1/2} + o(h_d) \right] \right. \\
&\quad \left. + O_P\left(\frac{\log h_r^{-1}}{nh_r}\right) \left[O_P\left(\frac{\log h_r^{-1}}{nh_r^5}\right)^{1/2} + O(h_d^2) \right] \right\} \\
&= O_P\left(\frac{h_d^4}{nh_r^{9/2}}\right) \left\{ \frac{(h_r^{1/2} \log h_r^{-1})^2}{nh_d^4} \frac{nh_r^5}{\log h_r^{-1}} + \frac{(\log h_r^{-1})^2}{nh_d^4} h_r^{1/2} + \left(\frac{\log h_r^{-1}}{nh_d^4}\right)^{1/2} (nh_r^5)^{1/2} h_r \right. \\
&\quad + (h_r^3 \log h_r^{-1})^{1/2} (nh_r^5)^{1/2} + \frac{(h_r^{1/2} \log h_r^{-1})^2}{nh_d^4} h_r^2 + \frac{h_r^{1/2} (\log h_r^{-1})^2}{nh_d^4} \left(\frac{\log h_r^{-1}}{nh_r^3}\right)^{1/2} \\
&\quad + \frac{(\log h_r^{-1})^2}{nh_d^4} h_r^{1/2} + \left(\frac{h_r^{1/2} (\log h_r^{-1})^2}{nh_d^4}\right)^{1/2} (h_r^{11/2} \log h_r^{-1})^{1/2} + \frac{h_r^{1/2} (\log h_r^{-1})^2 h_d^2}{nh_d^4 h_r} \\
&\quad \left. + \left(\frac{h_r^{1/2} (\log h_r^{-1})^2}{nh_d^4}\right)^{1/2} (h_r^{3/2} \log h_r^{-1})^{1/2} + h_r^{7/2} \log h_r^{-1} \right\}
\end{aligned}$$

$$= o_P\left(\frac{h_d^4}{nh_r^{9/2}}\right).$$

Somit ist gezeigt, dass Q_n von der Ordnung $o_P(h_d^4/nh_r^{9/2})$ ist und zusammen mit der vorhergehenden Argumentation folgt die Aussage von Satz 5.3 \square

5.3 Bootstrapverfahren und Simulationen

In diesem Kapitel wird das sogenannte ‘‘Wild-Bootstrap’’-Verfahren verwendet, das erstmals von Wu (1986) vorgeschlagen und in Mammen (1993) beschrieben wird um einen Bootstrap-Test zum Testen auf strenge Isotonie der Regressionsfunktion zu konstruieren. Zuerst wird aus den Daten die Regressionsfunktion unter der Nullhypothese der strengen Isotonie geschätzt, das heißt, es wird der isotone Schätzer der Regressionsfunktion $\hat{\phi}_{h_d}^{-1}(t)$ aus (3.2) bestimmt. Anschließend werden die Residuen mit Hilfe des uneingeschränkten Regressionschätzers \hat{m} geschätzt. Man erhält sie als

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

Weiter seien c_1, \dots, c_n unabhängig identisch verteilte Zufallsvariablen mit $E[c_i] = 0$, $E[c_i^2] = 1$ und $E[c_i^3] = 1$. Hier wird die Verteilung verwendet, die Gewicht $(\sqrt{5} + 1)/2\sqrt{5}$ auf dem Punkt $(1 - \sqrt{5})/2$ und Gewicht $(\sqrt{5} - 1)/2\sqrt{5}$ auf $(1 + \sqrt{5})/2$ legt. Die Bootstrap-Stichprobe erhält man als

$$Y_i^* = \hat{\phi}^{-1}(X_i) + c_i \hat{\varepsilon}_i.$$

Daraus wird die Bootstrap-Teststatistik T_n^* berechnet. Wiederholt man dieses Verfahren B -mal, so erhält man B verschiedene Versionen $T_n^{*,1}, \dots, T_n^{*,B}$ der Bootstrap-Statistik. Die $\lfloor (1 - \alpha)B \rfloor$ -te Ordnungsstatistik wird als $(1 - \alpha)$ -Quantil der wahren Teststatistik T verwendet.

Die folgenden Ergebnisse stammen aus Simulationen mit 501 Simulationsdurchläufe. In jedem Simulationsdurchlauf wird der Bootstrap 500 mal wiederholt. An verschiedenen Stellen dieses Verfahrens müssen Bandbreiten gewählt werden. Zur uneingeschränkten Schätzung der Regressionsfunktion aus den wahren Daten wird die Bandbreite h_r benötigt. Die zur Monotonisierung verwendete Bandbreite h_d wird in Abhängigkeit von h_r als $h_d = h_r^{1.2}$ gewählt. Im Bootstrap werden die Daten mit Hilfe einer isotonen Schätzung der Regressionsfunktion erzeugt. Dazu müssen ebenfalls Bandbreiten bestimmt werden. Diese seien mit $h_{r,b}$ und $h_{d,b} = h_{r,b}^{1.5}$ bezeichnet. Zur Bestimmung der Bootstrap-Teststatistiken werden wieder die Bandbreiten h_r und h_d verwendet. Da es sehr zeitaufwendig ist die Bandbreiten h_r und $h_{r,b}$ in jedem Durchlauf mit Least Squares Cross-Validation zu berechnen, wird vor der eigentlichen Simulation eine Simulation mit 200 Durchläufen zur Bestimmung dieser Bandbreiten durchgeführt. Der Mittelwert der so erhaltenen Bandbreiten wird in der darauffolgenden Simulation der Teststatistik verwendet. Die Simulationen wurden hier zu vier verschiedenen Regressionsfunktionen vom Typ

Tabelle 5.1: Simuliertes Niveau und Power des Tests auf strenge Isotonie basierend auf $\hat{\phi}_{h_a}$ (Tabelle (a)) und des Tests von Bowman, Jones and Gijbels (1998) (Tabelle (b)) zum Niveau $\alpha = 0.05$. Die Regressionsfunktionen sind gegeben durch (5.9). Verschiedene Stichprobenumfänge n und Standardabweichungen σ werden berücksichtigt.

(a) Test basierend auf T_n				(b) Bowman, Jones und Gijbels			
σ	n			σ	n		
	25	50	100		25	50	100
$a = 0$				$a = 0$			
0.025	0.006	0.009	0.006	0.025	0.002	0.006	0.002
0.05	0.004	0.008	0.004	0.05	0.018	0	0.008
0.1	0.013	0.006	0.006	0.1	0.022	0.008	0.004
$a = 0.15$				$a = 0.15$			
0.025	0.012	0.064	0.122	0.025	0.058	0	0.08
0.05	0.034	0.044	0.066	0.05	0.042	0.022	0.052
0.1	0.025	0.034	0.045	0.1	0.026	0.018	0.008
$a = 0.25$				$a = 0.25$			
0.025	0.497	0.842	0.982	0.025	0.816	0.926	0.994
0.05	0.275	0.522	0.74	0.05	0.31	0.482	0.748
0.1	0.131	0.211	0.351	0.1	0.088	0.1	0.174
$a = 0.45$				$a = 0.45$			
0.025	0.97	1	1	0.025	1	1	1
0.05	0.924	1	1	0.05	0.966	1	1
0.1	0.714	0.914	0.996	0.1	0.374	0.544	0.874

$$m(x) = 1 + x - a \exp(-50(x - 0.5)^2), x \in [0, 1] \quad (5.9)$$

mit $a = 0, 0.15, 0.25$ und 0.45 durchgeführt, um zu verdeutlichen, dass die vorgeschlagene Teststatistik strenge Isotonie eindeutig erkennt. Für $a \in \{0, 0.25\}$ sind die zugehörigen Regressionsfunktionen streng isoton, in den Fällen $a = 0.25$ und $a = 0.45$ ist die Nullhypothese der strengen Isotonie hingegen nicht erfüllt. Die Ergebnisse werden in Tabelle 5.1(a) dargestellt. Zum Vergleich mit einem anderen Testverfahren sind hier in Tabelle 5.1(b) die Ergebnisse von Bowman, Jones und Gijbels (1998) gegeben. Diese Autoren schlagen einen Bootstrap-Test vor basierend auf der kleinsten Bandbreite, so dass der Regressionsschätzer monoton ist. Der in dieser Arbeit vorgestellte Test verhält sich sehr

Tabelle 5.2: Simuliertes Niveau und Power des Tests auf strenge Isotonie basierend auf $\hat{\phi}_{hd}$ zum Niveau $\alpha = 0.05$. Die Regressionsfunktionen sind gegeben durch (5.10). Verschiedene Stichprobenumfänge n und Standardabweichungen σ werden berücksichtigt.

σ	n			σ	n		
	25	50	100		25	50	100
$c = 0$				$c = 1/4$			
0.025	0.006	0.004	0	0.025	0.623	0.647	0.721
0.05	0.014	0.01	0	0.05	0.453	0.513	0.583
0.1	0.014	0.008	0	0.1	0.355	0.453	0.501
$c = 1/8$				$c = 1/2$			
0.025	0.246	0.369	0.475	0.025	0.932	0.95	0.964
0.05	0.152	0.287	0.339	0.05	0.725	0.776	0.912
0.1	0.112	0.144	0.212	0.1	0.699	0.82	0.84

ähnlich zu dem von Bowman, Jones und Gijbels (1998). Beide unterschätzen das vorgegebene Niveau von $\alpha = 0.05$ für die streng monoton wachsende Funktion $m(x) = 1 + x$. Bei beiden Tests fallen die Ergebnisse für verschiedene Standardabweichungen sehr unterschiedlich aus. Während für die Regressionsfunktion zu $a = 0.15$, die ebenfalls streng monoton wachsend ist, das Niveau für $\sigma = 0.05$ und $n = 100$ relativ gut eingehalten wird, wird es für $\sigma = 0.025$ bei beiden Tests überschätzt. Für $\sigma = 0.1$ wird das Niveau mit dem hier vorgeschlagenen Testverfahren besser eingehalten als mit dem von Bowman, Jones und Gijbels (1998). Die Power des Tests basierend auf T_n ist in beiden Fällen einer nicht isotonen Regressionsfunktion vergleichbar mit der des Test von Bowman, Jones und Gijbels (1998) oder sogar besser als diese. Anders als der zuletzt genannte Test, ist der Test basierend auf T_n ein Test auf strenge Isotonie der Regressionsfunktion. Während der Test von Bowman Jones und Gijbels die Nullhypothese der Isotonie für lokal konstante isotone Regressionsfunktionen beibehält, verwirft der hier vorgestellte Test die Nullhypothese für solche Funktionen. Das Verhalten des Tests basierend auf T_n für lokal konstante isotone Regressionsfunktionen wird hier mit Simulationen zu vier Regressionsfunktionen vom Typ

$$m(x) = \begin{cases} \frac{1}{1-2c}x & \text{für } x \in [0, \frac{1}{2} - c] \\ \frac{1}{2} & \text{für } x \in [\frac{1}{2} - c, \frac{1}{2} + c] \\ \frac{1}{1-2c}x - \frac{2c}{1-2c} & \text{für } x \in [\frac{1}{2} + c, 1] \end{cases} \quad (5.10)$$

verdeutlicht. Für $c = 1/2$ ist hierdurch die auf dem ganzen Intervall konstante Funktion $m(x) = 1/2$ gegeben. Die Ergebnisse sind in Tabelle 5.2 dargestellt. Ist $c = 0$, so handelt es sich um die streng monoton wachsende Funktion $m(x) = x$. Hier entscheidet der Test in den meisten Fällen für die Nullhypothese, das Niveau wird jedoch wie schon in den oben dargestellten Fällen erheblich unterschätzt. Je größer der konstante Bereich der Funktion m ist ($c = 0.125, 0.25$ und 0.5), desto besser erkennt der Test die Verletzung der strengen Isotonie. Zuletzt sei hier angemerkt, dass für den Test auf strenge Isotonie der Regressionsfunktion kein schlechtester Fall der Monotonie, wie er zum Beispiel in Durot (2003) diskutiert wird existiert. Unter dem schlechtesten Fall der Monotonie versteht man dort den Fall einer konstanten Regressionsfunktion. Die meisten, in der Literatur angegebenen Testverfahren approximieren in Simulationen das vorgegebene Niveau in diesem Fall relativ gut, unterschätzen es jedoch für andere monotone Regressionsfunktionen. Bei dem in dieser Arbeit vorgestellten Test auf strenge Isotonie fällt der sogenannte schlechteste Fall der Isotonie unter die Alternative. Letzterer weist für konstante Regressionsfunktionen eine sehr hohe Power auf (vgl. Tabelle 5.2 für $c = 1/2$).

Anhang A

Beweistechnische Details

A.1 Gleichmäßig fast sichere Konvergenzraten des isotonen Schätzers

In diesem Kapitel werden gleichmäßig fast sichere Konvergenzraten des inversen isotonen Schätzers und seiner ersten beiden Ableitungen unter verschiedenen Annahmen an den Grad der Differenzierbarkeit gezeigt.

Konvergenzraten des inversen isotonen Schätzers bei einmaliger Differenzierbarkeit der Regressionsfunktion

In diesem Abschnitt sei die Regressionsfunktion einmal stetig differenzierbar, K_d sei ein zweimal stetig differenzierbarer Kern der Ordnung 2, dessen zweite Ableitung von 0 weg beschränkt ist und K_r sei ein einmal stetig differenzierbarer Kern der Ordnung 1. Die Bandbreiten h_r und h_d erfüllen die Bedingungen $nh_r^3 = O(1)$, $h_d/h_r \rightarrow \infty$ und $\log h_r^{-1}/nh_r^2h_d = o(1)$

Lemma A.1 *Unter den oben genannten Voraussetzungen sei*

$$J := J(\delta) = [m(0) + \delta, m(1) - \delta]$$

mit $\delta := \delta(h_d) > 0$ derart, dass für alle $t \in J(\delta)$ gilt $t + h_d v \in [m(0), m(1)]$ für alle $v \in [-1, 1]$. Dann gilt

$$\begin{aligned} \sup_{t \in J} \left| \hat{\phi}_{h_d}(t) - m^{-1}(t) \right| &= O\left(\frac{\log h_r^{-1}}{nh_r}\right)^{1/2} + O(h_d) \text{ f.s.} \\ \sup_{t \in J} \left| (\hat{\phi}_{h_d})'(t) - (m^{-1})'(t) \right| &= O\left(\frac{\log h_r^{-1}}{nh_r h_d^2}\right)^{1/2} + o(1) \text{ f.s.} \end{aligned}$$

Konvergenzraten des inversen isotonen Schätzers bei viermaliger Differenzierbarkeit der Regressionsfunktion

Es seien nun m und f viermal stetig differenzierbar, K_r und K_d seien zweimal stetig differenzierbare Kerne der Ordnung 2 und $K_d(\pm 1) = K'_d(\pm 1) = 0$. Die Bandbreiten erfüllen die Bedingungen $h_r, h_d \rightarrow 0$, $nh_r, nh_d \rightarrow \infty$, $h_r = O(n^{-1/5})$, $h_d^2 \log h_r^{-1} / h_r^{5/2} \rightarrow 0$ und $h_r^{1/2} (\log h_r^{-1})^2 / nh_d^4 = O(1)$.

Lemma A.2 *Unter den oben genannten Bedingungen gilt*

$$\begin{aligned} \sup_t |(\hat{\phi}_{h_d})^{(s)}(t) - (m^{-1})^{(s)}(t)| &= O\left(\frac{\log h_r^{-1}}{nh_r^{2s+1}}\right)^{1/2} + O(h_d^2) \text{ für } s = 0, 1, 2 \\ \sup_t |(\hat{\phi}_{h_d})^{(3)}(t) - (m^{-1})^{(3)}(t)| &= O\left(\frac{\log h_r^{-1}}{nh_r^7}\right)^{1/2} + o(h_d). \end{aligned}$$

Konvergenzraten des inversen isotonen Ableitungsschätzers

In Kapitel 4 wird die gleichmäßig fast sichere Konvergenz der Verteilungsfunktion des Ableitungsschätzers benötigt. Die Konvergenzraten sind in folgendem Lemma aufgeführt.

Lemma A.3 *Es seien m und f dreimal stetig differenzierbar, K_r und K_d seien zweimal stetig differenzierbare Kerne der Ordnung 2 mit $K_d(\pm 1) = K_r(\pm 1) = 0$. Die Bandbreiten erfüllen die Bedingungen $h_r, h_d \rightarrow 0$, $nh_r, nh_d \rightarrow \infty$, $h_r = O(n^{-1/7})$, $h_d/h_r^{3/2} \rightarrow 0$ und $(\log h_r^{-1})^{3/2} / nh_r^5 h_d = o(1)$. Dann gilt*

$$\begin{aligned} \sup_t |\hat{\psi}_{h_d}(t) - m'^{-1}(t)| &= O\left(\frac{\log h_r^{-1}}{nh_r^3}\right)^{1/2} + O(h_d^2) \\ \sup_t |\hat{\psi}'_{h_d}(t) - (m'^{-1})'(t)| &= O\left(\frac{\log h_r^{-1}}{nh_r^5}\right)^{1/2} + o(h_d) \end{aligned}$$

Beweis von Lemma A.1 und A.2 Beide Beweise verwenden dieselbe Zerlegung des Supremums in

$$\sup_{t \in J} \left| \hat{\phi}_{h_d}^{(s)}(t) - (m^{-1})^{(s)}(t) \right| \leq T_1^{(s)} + T_2^{(s)},$$

mit

$$\begin{aligned} T_1^{(s)} &:= \sup_{t \in J} \left| \hat{\phi}_{h_d}^{(s)}(t) - \phi_{h_d}^{(s)}(t) \right| \\ T_2^{(s)} &:= \sup_{t \in J} \left| \phi_{h_d}^{(s)}(t) - (m^{-1})^{(s)}(t) \right|, \end{aligned}$$

$s = 0, 1$ in Lemma A.1 und $s = 0, \dots, 3$ in Lemma A.2.

Mit der Zerlegung aus (3.10) erhält man für $T_1^{(s)}$, $s = 0, \dots, 3$

$$T_1^{(s)} \leq \Delta_n^{(1),(s)} + \frac{1}{2} \Delta_n^{(2),(s)}$$

mit

$$\begin{aligned} \Delta_n^{(1),(s)} &= \sup_{t \in J} \left| (\Delta_n^{(1)})^{(s)}(t) \right| \\ \Delta_n^{(2),(s)} &= \sup_{t \in J} \left| (\Delta_n^{(2)})^{(s)}(t) \right| \end{aligned}$$

und $\Delta_n^{(k)}(t)$, $k = 1, 2$ sind definiert wie in (3.12) und (3.13).

(i) In der Situation von Lemma A.1 erhält man eine weitere Abschätzung durch

$$\begin{aligned} \Delta_n^{(k),(s)} &\leq \frac{1}{h_d^{k+s-1}} \int_{-1}^1 \left| K_d^{(k+s-1)}(v) \right| \sup_{t \in J} \left| (m^{-1})'(t + h_d v) \right| \\ &\quad \times \left(\sup_{t \in J} \left| (\hat{m} - m) \circ m^{-1}(t + h_d v) \right| \right)^k dv \\ &\leq \frac{1}{h_d^{k+s-1}} \sup_{z \in [0,1]} \left| (m^{-1})'(z) \right| \left(\sup_{z \in [0,1]} |\hat{m}(z) - m(z)| \right)^k \int_{-1}^1 \left| K_d^{(k+s-1)}(v) \right| dv. \end{aligned}$$

Als stetige Funktion auf einem kompakten Intervall ist $(m^{-1})'$ beschränkt. Für den Nadaraya-Watson-Schätzer gilt wie bei Mack und Silverman (1982), hier jedoch für einmalige Differenzierbarkeit von m und f unter der Annahme $h_r = o(\log h_r^{-1}/nh_r)$

$$\sup_{z \in J} |\hat{m}(z) - m(z)| \stackrel{f.s.}{=} O\left(\frac{\log h_r^{-1}}{nh_r}\right)^{1/2}.$$

Damit erhält man die Ordnung

$$\Delta_n^{(k),(s)} = O\left(\frac{\log h_r^{-1}}{nh_r h_d^{2(k+s-1)/k}}\right)^{k/2} \quad \text{f.s. für } k = 1, 2, \quad s = 0, 1.$$

ist $k = 1$, so ergibt dies nach Voraussetzung (3.9)

$$\Delta_n^{(1),(s)} = O\left(\frac{\log h_r^{-1}}{nh_r h_d^{2s}}\right)^{1/2}$$

und für $k = 2$

$$O\left(\frac{\log h_r^{-1}}{nh_r h_d^{1+s}}\right) = O\left(\frac{\log h_r^{-1}}{nh_r h_d^{2s}}\right)^{1/2}.$$

Wie in (3.14) erhält man für den deterministischen Teil $T_2^{(0)}$

$$T_2^{(0)} \leq h_d \int_{-1}^1 |u| |K_d(u)| du \sup_{z \in J} \left| (m^{-1})'(z) \right| = O(h_d), \quad (\text{A.1})$$

$T_2^{(1)}$ kann abgeschätzt werden durch

$$\begin{aligned} T_2^{(1)} &= \sup_{t \in J} \left| \frac{\partial}{\partial t} \left(\frac{1}{h_d} \int_0^1 \int_{-\infty}^t K_d \left(\frac{m(v) - u}{h_d} \right) dudv \right) - (m^{-1})'(t) \right| \\ &= \sup_{t \in J} \left| \frac{1}{h_d} \int_0^1 K_d \left(\frac{m(v) - t}{h_d} \right) dv - (m^{-1})'(t) \right| \\ &= \sup_{t \in J} \left| \int_{-1}^1 K_d(v) (m^{-1})'(t + h_d v) dv - (m^{-1})'(t) \right| \\ &= \sup_{t \in J} \left| \frac{1}{h_d} K_d(v) m^{-1}(t + h_d v) \Big|_{-1}^1 - \frac{1}{h_d} \int_{-1}^1 K_d'(v) m^{-1}(t + h_d v) dv - (m^{-1})'(t) \right| \\ &= \sup_{t \in J} \left| -\frac{1}{h_d} \int_{-1}^1 K_d'(v) dv m^{-1}(t) - \int_{-1}^1 v K_d'(v) \left\{ (m^{-1})'(t + h_d \lambda v) - (m^{-1})'(t) \right\} dv \right| \\ &\leq \int_{-1}^1 |v| |K_d'(v)| \sup_{t \in J} \left| (m^{-1})'(t + h_d v) - (m^{-1})'(t) \right| dv \end{aligned} \quad (\text{A.2})$$

mit $\lambda \in [0, 1]$. Wegen der gleichmäßigen Stetigkeit von $(m^{-1})'$ konvergiert die Funktion $(m^{-1})'(t + h_d \lambda v)$ in v gleichmäßig gegen $(m^{-1})'(t)$ und es ist $T_2^{(1)} = o(1)$.

(ii) Um die Aussage von Lemma A.2 zu beweisen, wird ausgenutzt, dass in dieser Situation auch der Kern K_r und somit der Schätzer \hat{m} dreimal stetig differenzierbar ist. Damit kann für $\Delta_n^{(2),(s)}$, $s = 0, \dots, 3$ eine partielle Integration durchgeführt und so die Konvergenzrate verbessert werden. Es ist wie bisher

$$\begin{aligned} \Delta_n^{(1)}(t) &= \frac{1}{h_d} \int_0^1 K_d \left(\frac{m(v) - t}{h_d} \right) (\hat{m}(v) - m(v)) dv \\ &= \int_{-1}^1 K_d(v) (\hat{m} - m)(m^{-1}(t + h_d v)) (m^{-1})'(t + h_d v) dv = O_P \left(\frac{\log h_r^{-1}}{n h_r} \right)^{1/2}, \end{aligned}$$

und für $1 \leq s \leq 3$ mit der Bezeichnung $d(x) = \hat{m}(x) - m(x)$

$$\begin{aligned} \frac{\partial^s}{\partial t^s} \Delta_n^{(1)}(t) &= \int_{-1}^1 K_d(v) \left\{ \sum_{j=0}^s \binom{s}{j} \frac{\partial^j}{\partial t^j} [d \circ m^{-1}](t + h_d v) \frac{\partial^{s-j}}{\partial t^{s-j}} (m^{-1})'(t + h_d v) \right\} dv \\ &= O_P \left(\frac{\log h_r^{-1}}{n h_r^{2s+1}} \right)^{1/2} \end{aligned}$$

denn die einzelnen Summanden sind mit ähnlichen Abschätzungen wie in Mack und Silverman (1982) fast sicher von der Ordnung

$$\begin{aligned} \frac{\partial^j}{\partial t^j} [d \circ m^{-1}](t + h_d v) \frac{\partial^{s-j}}{\partial t^{s-j}} (m^{-1})'(t + h_d v) &= O\left(\frac{\log h_r^{-1}}{nh_r^{2j+1}}\right)^{1/2} \text{ für } j = 0, 1, \\ \frac{\partial^2}{\partial t^2} [d \circ m^{-1}](t + h_d v) (m^{-1})^{(s-1)}(t + h_d v) &= d^{(1)}(x_n) z_1(t + h_d v) + d^{(2)}(x_n) z_2(t + h_d v) \\ &= O\left(\frac{\log h_r^{-1}}{nh_r^5}\right)^{1/2} \text{ für } j = 2 \text{ und } s = 2, 3 \end{aligned}$$

und

$$\begin{aligned} \frac{\partial^3}{\partial t^3} [d \circ m^{-1}](t + h_d v) (m^{-1})'(t + h_d v) &= d^{(3)}(x_n) z_3(t + h_d v) + 2d^{(2)}(x_n) z_4(t + h_d v) \\ &\quad + d^{(1)}(x_n) z_5(t + h_d v) \\ &= O\left(\frac{\log h_r^{-1}}{nh_r^7}\right)^{1/2} \end{aligned}$$

mit gleichmäßig beschränkten Funktionen $z_i(x)$, $i = 1, \dots, 5$ und $x_n = m^{-1}(t + h_d v)$. Somit gilt auch

$$\Delta_n^{(1),(s)} = O_P\left(\frac{\log h_r^{-1}}{nh_r^{2s+1}}\right)^{1/2}.$$

Weiter ergibt $\Delta_n^{(2)}(t)$ mit partieller Integration und den Bezeichnungen aus obiger Abschätzung

$$\begin{aligned} \Delta_n^{(2)}(t) &= \frac{1}{h_d} \int_{-1}^1 K_d'(v) (\hat{m} - m)^2 (m^{-1}(t + h_d v)) (m^{-1})'(t + h_d v) dv (1 + o_P(1)) \quad (\text{A.3}) \\ &= - \int_{-1}^1 K_d(v) \{2d(x_n) d^{(1)}(x_n) (m^{-1})'^2(t + h_d v) + d^2(x_n) (m^{-1})^{(2)}(t + h_d v)\} dv \quad (\text{A.4}) \\ &= O_P\left(\frac{\log h_r^{-1}}{nh_r^2}\right) = o_P\left(\frac{\log h_r^{-1}}{nh_r}\right)^{1/2}. \end{aligned}$$

Die Ordnungen der ersten und zweiten Ableitung lassen sich durch Ableiten von (A.4) bestimmen. Mit den Bezeichnungen aus der Abschätzung von $\Delta_n^{(1),(s)}(t)$ ist zum Beispiel

$$\begin{aligned} \frac{\partial}{\partial t} \Delta_n^{(2)}(t) &= - \int_{-1}^1 K_d(v) \{d^{(1)2}(x_n) y_1(t + h_d v) + d(x_n) d^{(2)}(x_n) y_1(t + h_d v) \\ &\quad + d(x_n) d^{(1)}(x_n) y_3(t + h_d v) + d^2(x_n) y_4(t + h_d v)\} dv \\ &= O_P\left(\frac{\log h_r^{-1}}{nh_r^3}\right) = o_P\left(\frac{\log h_r^{-1}}{nh_r^3}\right)^{1/2}. \end{aligned}$$

Eine weitere Ableitung liefert

$$\frac{\partial^2}{\partial t^2} \Delta_n^{(2)}(t) = O_P\left(\frac{\log h_r^{-1}}{nh_r^4}\right) = o_P\left(\frac{\log h_r^{-1}}{nh_r^5}\right)^{1/2}.$$

Die Regressionsfunktion ist nach Annahme nur viermal stetig differenzierbar. Daher kann die Größenordnung von $\Delta_n^{(2),(s)}(t)$ für $s = 3$ nicht durch weiteres Ableiten von (A.4) bestimmt werden. Statt dessen wird nun in der Darstellung (A.3) dreimal abgeleitet. Man erhält

$$\frac{\partial^3}{\partial t^3} \Delta_n^{(2)}(t) = O_P\left(\frac{\log h_r^{-1}}{nh_r^4 h_d}\right) = o_P\left(\frac{\log h_r^{-1}}{nh_r^7}\right)^{1/2}$$

und zusammenfassend gilt

$$\Delta_n^{(2),(s)} = o_P\left(\frac{\log h_r^{-1}}{nh_r^{2s+1}}\right)^{1/2}.$$

Bei der Betrachtung des deterministischen Teils ist zu beachten, dass hier eine Taylorentwicklung höherer Ordnung möglich ist und es ist

$$\begin{aligned} T_2^{(0)} &= \sup_{t \in J} \left| \frac{1}{h_d} \int_0^1 \int_{-\infty}^t K_d\left(\frac{m(v) - u}{h_d}\right) dudv - m^{-1}(t) \right| \\ &= \sup_{t \in J} \left| m^{-1}(t - h_d) + h_d \int_{-1}^1 (m^{-1})'(t + h_d z) \int_z^1 K_d(v) dv dz - m^{-1}(t) \right| \\ &= h_d^2 \sup_{t \in J} |(m^{-1})^{(2)}(t)| \kappa_2(K_d) + o(h_d^2) = O(h_d^2) \end{aligned}$$

im Fall $s = 0$. Ist hingegen $s = 1, 2$, so ergibt sich

$$\begin{aligned} T_2^{(s)} &= \sup_{t \in J} \left| \frac{\partial^s}{\partial t^s} \left(\int_{-1}^1 K_d(v) (m^{-1})'(t + h_d v) dv - (m^{-1})'(t) \right) \right| \\ &= \sup_{t \in J} \left| \int_{-1}^1 K_d(v) (m^{-1})^{(s+1)}(t + h_d v) dv - (m^{-1})^{(s+1)}(t) \right| \\ &= h_d^2 \int_{-1}^1 v^2 K_d(v) dv \sup_{t \in J} |(m^{-1})^{(s+2)}(t)| + o(h_d^2) = O(h_d^2). \end{aligned}$$

Im Fall $s = 3$ ist $(m^{-1})^{(3)}$ für obige Taylorentwicklung nicht mehr hinreichend oft differenzierbar. Daher zeigt eine ähnliche Rechnung wie in (A.1)

$$T_2^{(3)} = h_d \sup_{t \in J} \left| \int_{-1}^1 v K_d(v) (m^{-1})^{(4)}(t + h_d v) dv \right| = o(h_d)$$

□

Beweis von Lemma A.3 Aus der Zerlegung (4.23) von $\hat{\psi}_{h_d} - m'^{-1}$ und den Abschätzungen (4.25), (4.26), (4.35), (4.36), (4.39) und (4.40) folgt

$$\begin{aligned} \sup_t |\hat{\psi}_{h_d}(t) - m'^{-1}(t)| &= O\left(\frac{\log h_r^{-1}}{nh_r^3}\right)^{1/2} + O\left(\frac{\log h_r^{-1}}{nh_r^4}\right) + O(h_d^2) \\ &= O\left(\frac{\log h_r^{-1}}{nh_r^3}\right)^{1/2} + O(h_d^2), \\ \sup_t |\hat{\psi}'_{h_d}(t) - (m'^{-1})'(t)| &= O\left(\frac{\log h_r^{-1}}{nh_r^5}\right)^{1/2} + O\left(\frac{\log h_r^{-1}}{nh_r^4 h_d}\right) + o(h_d) \\ &= O\left(\frac{\log h_r^{-1}}{nh_r^5}\right)^{1/2} + o(h_d). \end{aligned}$$

□

A.2 Beweis von Satz 5.5

Um die Rechnungen in diesem Abschnitt übersichtlicher zu gestalten wird im Folgenden der Index r bei Kern und Bandbreite weggelassen. Der Beweis folgt der Struktur der Beweise von Theorem 1 und 4 in Hall (1984). Im ersten Schritt wird das asymptotische Verhalten des Integrals über $S_k^2(x)\hat{f}^2(x)$ mit stochastisch konvergenter Gewichtsfunktion $v_n(x)$ bewiesen. Dieses Resultat wird dann auf das Integral über $S_k^2(x)$ mit deterministischer Gewichtsfunktion $w(x)$ übertragen, indem man $v_n(x) = w(x)/\hat{f}^2(x)$ setzt.

Das im ersten Schritt betrachtete Integral hat die Zerlegung

$$\int_A S_k^2(x)\hat{f}^2(x)v_n(x)dx = I_{n1} + 2I_{n2} + 2I_{n3} + I_{n4}$$

mit

$$I_{n1} = \frac{1}{n^2h^{2k+2}} \sum_{i=1}^n (Y_i - m(X_i))^2 \int_A K^{(k)2}\left(\frac{x - X_i}{h}\right)v_n(x)dx \quad (\text{A.5})$$

$$I_{n2} = \frac{1}{n^2h^{2k+2}} \sum_{i < j} (Y_i - m(X_i))(Y_j - m(X_j)) \int_A K^{(k)}\left(\frac{x - X_i}{h}\right)K^{(k)}\left(\frac{x - X_j}{h}\right)v_n(x)dx \quad (\text{A.6})$$

$$I_{n3} = \frac{1}{nh^{k+1}} \sum_{i=1}^n (Y_i - m(X_i)) \int_A K^{(k)}\left(\frac{x - X_i}{h}\right)\hat{g}^{(k)}(x)v_n(x)dx \quad (\text{A.7})$$

$$I_{n4} = \int_A (\hat{g}^{(k)}(x))^2 v_n(x)dx. \quad (\text{A.8})$$

Mit dieser Zerlegung lässt sich das folgende asymptotische Verhalten zeigen. Der Erwartungswert und die Varianz einer Zufallsvariablen Z bedingt auf die Stichprobe X_1, \dots, X_n seien dabei bezeichnet mit $E_{\mathcal{X}}[Z]$ beziehungsweise $\text{Var}_{\mathcal{X}}(Z)$.

Satz A.4 *Unter den Bedingungen aus Satz 5.5 mit einer stochastisch gleichmäßig konvergenten Gewichtsfunktion v_n deren Grenzwert v sei gilt für $k = 0$*

$$\int_A S_0^2(x)\hat{f}^2(x)v_n(x)dx = E_{\mathcal{X}}[I_{n1}] + I_{n4} + (n^{-2}h^{-1}\alpha_{1,0} + h^4n^{-1}\alpha_{2,0})^{1/2}Z_n$$

und für $k = 1$ und 2

$$\int_A S_k^2(x)\hat{f}^2(x)v_n(x)dx = E_{\mathcal{X}}[I_{n1}] + I_{n4} + (n^{-2}h^{-1-4k}\alpha_{1,k})^{1/2}Z_n$$

mit einer asymptotisch standardnormalverteilten Zufallsvariablen Z_n und den in Satz 5.5 definierten Konstanten $\alpha_{1,k}$ und $\alpha_{2,0}$.

Beweis Der Beweis gliedert sich in drei Teile. Im ersten Teil wird gezeigt, dass sich I_{n1} als Summe seines bedingten Erwartungswertes und eines asymptotisch vernachlässigbaren Rests schreiben lässt. Teil zwei und drei widmen sich der asymptotischen Normalität von I_{n2} und I_{n3} .

(i) Es ist

$$\begin{aligned}
& n^4 h^{4k+4} \mathbb{E}_{\mathcal{X}}[(I_{n1} - \mathbb{E}_{\mathcal{X}}[I_{n1}])^2] \\
&= \mathbb{E}_{\mathcal{X}} \left[\left(\sum_{i=1}^n \{(Y_i - m(X_i))^2 - \sigma^2(X_i)\} \int_A K^{(k)2} \left(\frac{x - X_i}{h} \right) v_n(x) dx \right)^2 \right] \\
&= \sum_{i=1}^n \mathbb{E}_{\mathcal{X}}[\{(Y_i - m(X_i))^2 - \sigma^2(X_i)\}^2] \left(\int_A K^{(k)2} \left(\frac{x - X_i}{h} \right) v_n(x) dx \right)^2.
\end{aligned} \tag{A.9}$$

Falls X_i aus der Menge $A^{\lambda h} = \{x \in \mathbb{R} \mid \inf_{a \in A} |x - a| < \lambda h\}$ ist (der Kern K besitzt hier den kompakten Träger $[-\lambda, \lambda]$), kann das Integral abgeschätzt werden durch

$$\int_A K^{(k)2} \left(\frac{x - X_i}{h} \right) v_n(x) dx = O_P(1) h \int_A K^{(k)2}(x) dx = O_P(h)$$

Da das vierte Moment $\mathbb{E}_{\mathcal{X}}[\{(Y_i - m(X_i))^2 - \sigma^2(X_i)\}^2]$ für alle i wohldefiniert ist, erhält man durch Einsetzen der vorhergehenden Abschätzung in (A.9)

$$n^4 h^{4k+4} \mathbb{E}_{\mathcal{X}}[(I_{n1} - \mathbb{E}_{\mathcal{X}}[I_{n1}])^2] = O_P(nh^2)$$

beziehungsweise

$$\mathbb{E}_{\mathcal{X}}[(I_{n1} - \mathbb{E}_{\mathcal{X}}[I_{n1}])^2] = O_P\left(\frac{1}{n^3 h^{4k+2}}\right).$$

Mit der bedingten Markov-Ungleichung gilt für eine Folge $\{\lambda_n\}_{n \in \mathbb{N}}$ mit $\lambda_n \rightarrow \infty$ für $n \rightarrow \infty$ die stochastische Konvergenz

$$\mathbb{P}\left(|I_{n1} - \mathbb{E}_{\mathcal{X}}[I_{n1}]| > \frac{\lambda_n}{n^{3/2} h^{2k+1}} \mid X_1, \dots, X_n\right) \leq \frac{1}{\lambda_n} O_P(1) \xrightarrow{n \rightarrow \infty} 0.$$

Somit konvergiert auch die unbedingte Wahrscheinlichkeit gegen null und es ist

$$I_{n1} = \mathbb{E}_{\mathcal{X}}[I_{n1}] + o_P\left(\frac{1}{n^{3/2} h^{2k+1}}\right) = \mathbb{E}_{\mathcal{X}}[I_{n1}] + o_P\left(\frac{1}{n h^{2k+1/2}}\right).$$

(ii) In diesem Abschnitt wird der Beitrag von I_{n2} zur asymptotischen Verteilung bestimmt. Dazu seien W_{nij} und \tilde{W}_{nij} definiert als

$$\begin{aligned}
W_{nij} &= \int_A K^{(k)} \left(\frac{x - X_i}{h} \right) K^{(k)} \left(\frac{x - X_j}{h} \right) v_n(x) dx \\
\tilde{W}_{nij} &= \int_A K^{(k)} \left(\frac{x - X_i}{h} \right) K^{(k)} \left(\frac{x - X_j}{h} \right) v(x) dx.
\end{aligned}$$

Beide Zufallsvariablen sind $\sigma(X_1, \dots, X_n)$ -messbar. Weiter sei die Zufallsvariable Y_{nj} definiert als

$$Y_{nj} = (Y_j - m(X_j)) \sum_{i=1}^{j-1} (Y_i - m(X_i)) W_{nij}.$$

Mit diesen Definitionen erhält man

$$\begin{aligned} n^2 h^{2k+2} I_{n2} &= \sum_{1 \leq i < j \leq n} (Y_i - m(X_i))(Y_j - m(X_j)) W_{nij} \\ &= \sum_{j=2}^n (Y_j - m(X_j)) \sum_{i=1}^{j-1} (Y_i - m(X_i)) W_{nij} \\ &= \sum_{j=2}^n Y_{nj}. \end{aligned}$$

Für einen festen Index j ist die Zufallsvariable Y_{nj} messbar bezüglich der von X_1, \dots, X_n und Y_1, \dots, Y_j erzeugten σ -Algebra $\mathcal{A}_{n,j}$ wobei $\mathcal{A}_{n,0} = \sigma(X_1, \dots, X_n)$ sei. Dann ist wegen der Unabhängigkeit der Zufallsgrößen $(X_1, Y_1), \dots, (X_{j-1}, Y_{j-1}), X_j, \dots, X_n$

$$E[Y_j | \mathcal{A}_{n,j-1}] = E[Y_j | X_j] = m(X_j)$$

und der bedingte Erwartungswert von Y_{nj} ergibt

$$E[Y_{nj} | \mathcal{A}_{n,j-1}] = (E[Y_j | \mathcal{A}_{n,j-1}] - m(X_j)) \sum_{i=1}^{j-1} (Y_i - m(X_i)) W_{nij} = 0.$$

Diese Eigenschaft charakterisiert die Folge $\{(S_{ni} = \sum_{j=2}^i Y_{nj}, \mathcal{A}_{ni}), 2 \leq i \leq n < \infty\}$ als Martingal-Differenzschema. Es ist das Ziel, auf das eben definierte Schema einen zentralen Grenzwertsatz (vgl. Hall, Heyde (1980), Corollary 3.2) anzuwenden, um den Einfluss von I_{n2} auf die asymptotische Verteilung der betrachteten Statistik zu untersuchen. Das weitere Vorgehen teilt sich, bestimmt durch die nachzuweisenden Bedingungen in zwei Schritte auf. Im ersten Schritt wird gezeigt, dass die bedingte Varianz eines Diagonalelements S_{nn} dieser Folge stochastisch konvergiert, während im zweiten Schritt eine Lindeberg-Bedingung nachgewiesen wird.

Die bedingte Varianz von S_{nn} ist gegeben durch

$$\begin{aligned} V_n^2 &= \sum_{j=2}^n E[Y_{nj}^2 | \mathcal{A}_{n,j-1}] = \sum_{j=2}^n E[(Y_j - m(X_j))^2 | \mathcal{A}_{n,j-1}] \left(\sum_{i=1}^{j-1} (Y_i - m(X_i)) W_{nij} \right)^2 \\ &= \sum_{j=2}^n \sigma^2(X_j) \sum_{i=1}^{j-1} (Y_i - m(X_i))^2 W_{nij}^2 \end{aligned}$$

$$\begin{aligned}
& + 2 \sum_{j=2}^n \sigma^2(X_j) \sum_{1 \leq i < k \leq j-1} (Y_i - m(X_i))(Y_k - m(X_k)) W_{nij} W_{nkj} \\
& = V_{n1} + V_{n2},
\end{aligned}$$

wobei sich die Bezeichnungen V_{n1} und V_{n2} aus der Zerlegung in zwei Summen im vorhergehenden Schritt ergeben.

Lemma A.5 *Mit den oben genannten Bedingungen und Definitionen gilt*

$$n^{-2}h^{-3}V_{n1} \xrightarrow{P} \frac{1}{4}\alpha_{1,k} \quad (\text{A.10})$$

und

$$n^{-2}h^{-3}V_{n2} \xrightarrow{P} 0 \quad (\text{A.11})$$

für $n \rightarrow \infty$.

Beweis von Lemma A.5 Da $v_n(x)$ stochastisch gleichmäßig gegen $v(x)$ konvergiert, reicht es für die stochastische Konvergenz von V_{n1} den Ausdruck

$$\tilde{V}_{n1} = \sum_{j=2}^n \sigma^2(X_j) \sum_{i=1}^{j-1} (Y_i - m(X_i))^2 \tilde{W}_{nij}^2$$

zu betrachten, in dem die stochastisch konvergente Funktion $v_n(x)$ durch ihren Grenzwert $v(x)$ ersetzt ist. Es ist dann $V_{n1} = \tilde{V}_{n1}(1 + o_P(1))$. Für den Term \tilde{V}_{n1} wird nun nachgewiesen, dass er stochastisch gegen seinen bedingten Erwartungswert konvergiert und für diesen dann ein Resultat über U-Statistiken angewandt um die behauptete Konvergenz zu zeigen.

$$\begin{aligned}
\mathbb{E}_{\mathcal{X}}[(\tilde{V}_{n1} - \mathbb{E}_{\mathcal{X}}[\tilde{V}_{n1}])^2] &= \mathbb{E}_{\mathcal{X}} \left[\left(\sum_{j=2}^n \sigma(X_j) \sum_{i=1}^{j-1} \{(Y_i - m(X_i))^2 - \sigma^2(X_i)\} \tilde{W}_{nij} \right)^2 \right] \\
&= \mathbb{E}_{\mathcal{X}} \left[\left(\sum_{i=1}^{n-1} \{(Y_i - m(X_i))^2 - \sigma^2(X_i)\} \sum_{j=i+1}^n \sigma^2(X_j) \tilde{W}_{nij} \right)^2 \right] \\
&= \sum_{i=1}^{n-1} \mathbb{E}_{\mathcal{X}}[\{(Y_i - m(X_i))^2 - \sigma^2(X_i)\}^2] \left(\sum_{j=i+1}^n \sigma^2(X_j) \tilde{W}_{nij} \right)^2 \\
&= \sum_{i=1}^{n-1} \mu_4(X_i) \left(\sum_{j=i+1}^n \sigma^2(X_j) \tilde{W}_{nij} \right)^2. \quad (\text{A.12})
\end{aligned}$$

Die Varianz $\sigma^2(X_i)$ ist nach Voraussetzung beschränkt. Desweiteren kann der Betrag von \tilde{W}_{nij} abgeschätzt werden durch

$$\begin{aligned}
|\tilde{W}_{nij}| &= \left| \int_A K^{(k)}\left(\frac{x-X_i}{h}\right) K^{(k)}\left(\frac{x-X_j}{h}\right) v(x) dx \right| \\
&\leq \int_A \left| K^{(k)}\left(\frac{x-X_i}{h}\right) \right| \left| K^{(k)}\left(\frac{x-X_j}{h}\right) \right| \\
&= h \int_{-\lambda}^{\lambda} |K^{(k)}(x)| \left| K^{(k)}\left(x + \frac{X_i-X_j}{h}\right) \right| v(x) dx \\
&= hC,
\end{aligned} \tag{A.13}$$

mit einer Konstanten C , da der abgeleitete Kern nur dann ungleich null ist, wenn X_l in dem Intervall $A^{h\lambda}$ liegt. Eingesetzt in die Abschätzung (A.12) ergibt dies

$$\mathbb{E}_{\mathcal{X}}[(\tilde{V}_{n1} - \mathbb{E}_{\mathcal{X}}[\tilde{V}_{n1}])^2] \leq C^3 n h^3 \sum_{\substack{i \leq n-1 \\ X_i \in A^\varepsilon}} \mu_4(X_i) \sum_{\substack{i+1 \leq j \\ X_j \in A^\varepsilon}} |\tilde{W}_{nij}|. \tag{A.14}$$

Beide Summen in (A.14) enthalten höchstens $n-1$ Summanden. Daher kann der Erwartungswert abgeschätzt werden durch

$$\begin{aligned}
&\mathbb{E}[\mathbb{E}_{\mathcal{X}}[(\tilde{V}_{n1} - \mathbb{E}_{\mathcal{X}}[\tilde{V}_{n1}])^2]] \tag{A.15} \\
&\leq C_1 n^3 h^3 \int_{A^\varepsilon} \int_{A^\varepsilon} \mu_4(y) \int_A |K^{(k)}\left(\frac{x-y}{h}\right)| \left| K^{(k)}\left(\frac{x-z}{h}\right) \right| dx f(y) f(z) dy dz \\
&\leq C_2 n^3 h^5 \int_{A^\varepsilon} \mu_4(y) f(y) dy \int_{-\lambda}^{\lambda} \int_{-\lambda}^{\lambda} |K^{(k)}(x)| \left| K^{(k)}(x+z) \right| dx dz \\
&= O(n^3 h^5) \tag{A.16}
\end{aligned}$$

mit Konstanten $C_1 = C^3 \sup_{x \in A^\varepsilon} |v(x)|$ und $C_2 = C_1 \sup_{u \in A^\varepsilon} |f(u)|$. Die letzte Identität gilt wegen der Voraussetzung $\int_{A^\varepsilon} \mu_4(y) f(y) dy < \infty$. Mit der Eigenschaft der iterierten Erwartung erhält man als Darstellung von \tilde{V}_{n1}

$$\tilde{V}_{n1} = \mathbb{E}_{\mathcal{X}}[\tilde{V}_{n1}] + O_P(n^{3/2} h^{5/2}).$$

Nun wird der bedingte Erwartungswert mit Hilfe des unbedingten Erwartungswerts ausgedrückt. Dazu bezeichne $A_n = \mathbb{E}_{\mathcal{X}}[\tilde{V}_{n1}]$ und $a_n = \mathbb{E}[A_n] = \mathbb{E}[\tilde{V}_{n1}]$. Die Statistik

$$U_n = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} H_n(X_i, X_j)$$

ist eine U-Statistik mit symmetrischem Kern

$$H_n(X_i, X_j) = \sigma^2(X_i) \sigma^2(X_j) \tilde{W}_{nij}^2 - \mathbb{E}[\sigma^2(X_i) \sigma^2(X_j) \tilde{W}_{nij}^2]$$

der Ordnung 2. Definiert man $h_1(X_j) = \mathbb{E}[H_n(X_i, X_j)|X_j]$, so ist die Projektion von U_n gegeben durch

$$U_n^* = \frac{2}{n} \sum_{i=1}^n h_1(X_i).$$

Mit dieser U-Statistik und ihrer Projektion lässt sich die Differenz $A_n - a_n$ schreiben als

$$D_n = A_n - a_n = \frac{n(n-1)}{2} U_n$$

und die Projektion ist

$$D_n^* = \frac{n(n-1)}{2} U_n^* = (n-1) \sum_{i=1}^n h_1(X_i).$$

Die L_2 -Konvergenz von D_n kann nun auf die von D_n^* und $D_n - D_n^*$ zurückgeführt werden,

$$\begin{aligned} \mathbb{E}[D_n^2] &\leq 2(\mathbb{E}[D_n^{*2}] + \mathbb{E}[(D_n - D_n^*)^2]) && \text{(A.17)} \\ &= 2\left((n-1)^2 \mathbb{E}\left[\left(\sum_{i=1}^n h_1(X_i)\right)^2\right] + \mathbb{E}\left[\left(\sum_{1 \leq i < j \leq n} H(X_i, X_j) - (n-1) \sum_{i=1}^n h_1(X_i)\right)^2\right]\right) \\ &= 2\left((n-1)^2 \mathbb{E}\left[\left(\sum_{i=1}^n h_1(X_i)\right)^2\right] + \mathbb{E}\left[\left(\sum_{1 \leq i < j \leq n} H(X_i, X_j) - \sum_{i=1}^n \sum_{j \neq i} h_1(X_i)\right)^2\right]\right) \\ &= 2\left((n-1)^2 \mathbb{E}\left[\left(\sum_{i=1}^n h_1(X_i)\right)^2\right] \right. \\ &\quad \left. + \mathbb{E}\left[\left(\sum_{1 \leq i < j \leq n} H(X_i, X_j) - \sum_{1 \leq i < j \leq n} h_1(X_i) - \sum_{1 \leq j < i \leq n} h_1(X_i)\right)^2\right]\right) \\ &= 2\left((n-1)^2 \mathbb{E}\left[\left(\sum_{i=1}^n h_1(X_i)\right)^2\right] + \mathbb{E}\left[\left(\sum_{1 \leq i < j \leq n} (H(X_i, X_j) - h_1(X_i) - h_1(X_j))\right)^2\right]\right) \\ &= 2\left((n-1)^2 \sum_{i=1}^n \mathbb{E}[h_1^2(X_i)] + \sum_{i \neq j} \mathbb{E}[h_1(X_i)] \mathbb{E}[h_1(X_j)] \right. \\ &\quad \left. + \sum_{1 \leq i < j \leq n} \mathbb{E}[(H(X_i, X_j) - h_1(X_i) - h_1(X_j))^2] \right. \\ &\quad \left. + \sum_{\substack{i < j, k < l \\ i \neq k, j \neq l}} \mathbb{E}[(H(X_i, X_j) - h_1(X_i) - h_1(X_j))(H(X_k, X_l) - h_1(X_k) - h_1(X_l))]\right) \\ &\leq 2\left(n^3 \mathbb{E}[h_1^2(X_1)] + \sum_{i \neq j} \mathbb{E}[h_1(X_i)] \mathbb{E}[h_1(X_j)] + n^2 \mathbb{E}[(H(X_1, X_2) - h_1(X_1) - h_1(X_2))^2] \right. \\ &\quad \left. + \sum_{\substack{i < j, k < l \\ i \neq k, j \neq l}} \mathbb{E}[(H(X_i, X_j) - h_1(X_i) - h_1(X_j))(H(X_k, X_l) - h_1(X_k) - h_1(X_l))]\right). \end{aligned}$$

Die letzte Abschätzung gilt, da die Zufallsvariablen X_i für $i = 1, \dots, n$ unabhängig und identisch verteilt sind. Die erste noch verbleibende Summe ergibt null, denn

$$\mathbb{E}[h_1(X_1)] = \mathbb{E}[\mathbb{E}[H(X_i, X_j)|X_i]] = \mathbb{E}[H(X_i, X_j)] = 0$$

nach Definition von $H(X_i, X_j)$, $h_1(X_i)$ und der Eigenschaft der iterierten Erwartung. Die zweite Summe enthält Summanden, in denen zwei Indizes gleich sind. Aus diesem Grund ist es hier nicht direkt ersichtlich, dass auch diese Summe null ergibt. Ohne Einschränkung seien die Indizes i und k gleich. Die anderen Fälle ergeben sich durch ein Symmetrieargument. Es ist

$$\begin{aligned} & \mathbb{E}[(H(X_i, X_j) - h_1(X_i) - h_1(X_j))(H(X_i, X_l) - h_1(X_i) - h_1(X_l))] \\ &= \mathbb{E}[\mathbb{E}[(H(X_i, X_j) - h_1(X_i) - h_1(X_j))(H(X_i, X_l) - h_1(X_i) - h_1(X_l))|X_i]] \\ &= \mathbb{E}[\mathbb{E}[(H(X_i, X_j) - h_1(X_i) - h_1(X_j))|X_i] \mathbb{E}[(H(X_i, X_l) - h_1(X_i) - h_1(X_l))|X_i]] \\ &= \mathbb{E}[(\mathbb{E}[H(X_i, X_j)|X_i] - h_1(X_i) - \mathbb{E}[h_1(X_j)])(\mathbb{E}[H(X_i, X_l)|X_i] - h_1(X_i) - \mathbb{E}[h_1(X_l)])] \\ &= \mathbb{E}[h_1(X_j)] \mathbb{E}[h_1(X_l)] = 0. \end{aligned}$$

Somit gilt nach (A.17) und obiger Diskussion der Summen

$$\begin{aligned} \mathbb{E}[(A_n - a_n)^2] &\leq n^3 \mathbb{E}[h_1^2(X_1)] + n^2 \mathbb{E}[(H(X_1, X_2) - h_1(X_1) - h_1(X_2))^2] \\ &\leq C_1 n^3 \mathbb{E}[\sigma^4(X_1)\sigma^4(X_2)\tilde{W}_{n12}^4] \\ &\leq C_2 n^3 \mathbb{E}[\tilde{W}_{n12}^4] \end{aligned} \tag{A.18}$$

wobei die zweite Abschätzung einerseits durch

$$\begin{aligned} \mathbb{E}[h_1^2(X_1)] &= \mathbb{E}[\mathbb{E}^2[H(X_1, X_2)|X_1]] \\ &\leq \mathbb{E}[(\mathbb{E}_{\mathcal{X}}[\sigma^2(X_1)\sigma^2(X_2)\tilde{W}_{n12}^2|X_1])^2] \\ &\leq \mathbb{E}[\sigma^4(X_1)\sigma^4(X_2)\tilde{W}_{n12}^4] \end{aligned}$$

und andererseits durch

$$\begin{aligned} \mathbb{E}[(H(X_1, X_2) - h_1(X_1) - h_1(X_2))^2] &\leq C_4(\mathbb{E}[H^2(X_1, X_2)] + \mathbb{E}[h_1^2(X_1)] + \mathbb{E}[h_1^2(X_2)]) \\ &\leq 3C_4 \mathbb{E}[\sigma^4(X_1)\sigma^4(X_2)\tilde{W}_{n12}^4] \end{aligned}$$

begründet ist. In der letzten Abschätzung von (A.18) wird die gleichmäßige Beschränktheit von σ ausgenutzt. Nun wird der Erwartungswert $\mathbb{E}[\tilde{W}_{n12}^4]$ abgeschätzt,

$$\begin{aligned} \mathbb{E}[\tilde{W}_{n12}^4] &\leq \mathbb{E} \left[\left(\int \left| K^{(k)} \left(\frac{x - X_1}{h} \right) \right| \left| K^{(k)} \left(\frac{x - X_2}{h} \right) \right| v(x) dx \right)^4 \right] \\ &= h^4 \int \int \left(\left| K^{(k)}(x) \right| \left| K^{(k)} \left(x + \frac{y - z}{h} \right) \right| v(x) dx \right)^4 f(y) f(z) dy dz \\ &= h^5 \int \int \left(\left| K^{(k)}(x) \right| \left| K^{(k)}(x + z) \right| v(x) dx \right)^4 f(y) f(y - hz) dy dz \\ &= O(h^5). \end{aligned}$$

Mit (A.18) ist die Differenz $E[(A_n - a_n)^2]$ von der Größenordnung $O(n^3h^5)$. Insgesamt bedeutet das für \tilde{V}_{n1}

$$\tilde{V}_{n1} = a_n + O_P(n^{3/2}h^{5/2}).$$

Zum Abschluss der Betrachtung von V_{n1} wird der Grenzwert von a_n bestimmt. Es ist

$$\begin{aligned} a_n &= E\left[\sum_{i=2}^n \sigma^2(X_i) \sum_{j=1}^{i-1} \sigma^2(X_j) \tilde{W}_{nij}^2\right] \\ &= \frac{n(n-1)}{2} E[\sigma^2(X_1)\sigma^2(X_2)\tilde{W}_{n12}^2] \\ &= \frac{n(n-1)}{2} \int_{A^\varepsilon} \int_{A^\varepsilon} \sigma^2(y)\sigma^2(z)f(y)f(z) \left(\int_A K^{(k)}\left(\frac{x-y}{h}\right)K^{(k)}\left(\frac{x-z}{h}\right)v(x)dx\right)^2 dydz \\ &= \frac{1}{2}n(n-1)h^3 \int_{-\lambda}^{\lambda} \int_{-\lambda}^{\lambda} \sigma^2(y)\sigma^2(y-hz)f(y)f(y-hz) \\ &\quad \times \left(\int_{-\lambda}^{\lambda} K^{(k)}(x)K^{(k)}(x+z)v(y+hx)dx\right)^2 dydz \\ &= \frac{1}{2}n^2h^3 \left(\int_{-\lambda}^{\lambda} \sigma^4(y)f^2(y)v^2(y)dy\right) \left(\int_{-\lambda}^{\lambda} \left(\int_{-\lambda}^{\lambda} K^{(k)}(x)K^{(k)}(x+z)dx\right)^2 dz\right) (1+o(1)). \end{aligned}$$

Somit gilt die Konvergenz

$$n^{-2}h^{-3}\tilde{V}_{n1} \xrightarrow{P} \frac{1}{2} \left(\int_{-\lambda}^{\lambda} \sigma^4(y)f^2(y)v^2(y)dy\right) \left(\int_{-\lambda}^{\lambda} \left(\int_{-\lambda}^{\lambda} K^{(k)}(x)K^{(k)}(x+z)dx\right)^2 dz\right)$$

und die erste Aussage von Lemma A.5 ist bewiesen.

Um die stochastische Konvergenz von V_{n2} gegen null zu beweisen wird wiederum die Eigenschaft der iterierten Erwartung ausgenutzt. Es ist

$$\begin{aligned} E_{\mathcal{X}}[V_{n2}^2] &= E_{\mathcal{X}} \left[\left(2 \sum_{j=2}^n \sigma^2(X_j) \sum_{1 \leq i < k \leq j-1} (Y_i - m(X_i))(Y_k - m(X_k))W_{nij}W_{nkj} \right)^2 \right] \\ &= 4 E_{\mathcal{X}} \left[\left(\sum_{1 \leq i < k \leq n-1} (Y_i - m(X_i))(Y_k - m(X_k)) \sum_{j=k+1}^n \sigma^2(X_j)W_{nij}W_{nkj} \right)^2 \right] \\ &= 4 \sum_{1 \leq i < k \leq n-1} \sigma^2(X_i)\sigma^2(X_k) \left(\sum_{j=k+1}^n \sigma^2(X_j)W_{nij}W_{nkj} \right)^2 \\ &= O_P(1) \sum_{1 \leq i < k \leq n-1} \left(\sum_{j=k+1}^n \tilde{W}_{nij}\tilde{W}_{nkj} \right)^2 \end{aligned} \tag{A.19}$$

wegen der gleichmäßigen Beschränktheit von σ und der stochastischen Konvergenz von $v_n(x)$ gegen $v(x)$. Ist $1 \leq j < k \leq n-1$, so gilt für den Erwartungswert der Summanden

$$\begin{aligned} \mathbb{E} \left[\left(\sum_{j=k+1}^n \tilde{W}_{nij} \tilde{W}_{nkj} \right)^2 \right] &\leq \mathbb{E} \left[\left(\sum_{j=k+1}^n |\tilde{W}_{nij}| |\tilde{W}_{nkj}| \right)^2 \right] \\ &= (n-k) \mathbb{E}[|\tilde{W}_{n13}|^2 |\tilde{W}_{n23}|^2] + (n-k)(n-k-1) \mathbb{E}[|\tilde{W}_{n13}| |\tilde{W}_{n23}| |\tilde{W}_{n14}| |\tilde{W}_{n24}|]. \end{aligned} \quad (\text{A.20})$$

Im Folgenden bezeichne \int das Integral über den Bereich $[-\lambda, \lambda]$ und die Konstante C sei definiert als $C = 1/(\sup |v(x)|)^4$. Der erste Erwartungswert in obiger Formel lässt sich abschätzen durch

$$\begin{aligned} &C \mathbb{E}[|\tilde{W}_{n13}|^2 |\tilde{W}_{n23}|^2] \\ &\leq \mathbb{E} \left[\left(\int_A \left| K^{(k)} \left(\frac{x - X_1}{h} \right) \right| \left| K^{(k)} \left(\frac{x - X_3}{h} \right) \right| dx \right)^2 \right. \\ &\quad \left. \times \left(\int_A \left| K^{(k)} \left(\frac{y - X_2}{h} \right) \right| \left| K^{(k)} \left(\frac{y - X_3}{h} \right) \right| dy \right)^2 \right] \\ &= h^4 \int_{A^\varepsilon} \int_{A^\varepsilon} \int_{A^\varepsilon} \left(\int \left| K^{(k)}(x) \right| \left| K^{(k)} \left(x + \frac{x_1 - x_3}{h} \right) \right| dx \right)^2 \\ &\quad \times \left(\int \left| K^{(k)}(y) \right| \left| K^{(k)} \left(\frac{x_2 - x_3}{h} \right) \right| dx \right)^2 f(x_1) f(x_2) f(x_3) dx_1 dx_2 dx_3 \\ &= h^6 \int_{A^\varepsilon} \int \int \left(\int \left| K^{(k)}(x) \right| \left| K^{(k)}(x + x_1) \right| dx \right)^2 \\ &\quad \times \left(\int \left| K^{(k)}(y) \right| \left| K^{(k)}(y + x_2) \right| dy \right)^2 f(x + hx_1) f(y + hx_2) f(x_3) dx_1 dx_2 dx_3 \\ &= O(h^6), \end{aligned}$$

der zweite durch

$$\begin{aligned} &C \mathbb{E}[|\tilde{W}_{n13}| |\tilde{W}_{n23}| |\tilde{W}_{n14}| |\tilde{W}_{n24}|] \\ &\leq \int_A \int_A \int_A \int_A \mathbb{E} \left[\left| K^{(k)} \left(\frac{x - X_1}{h} \right) \right| \left| K^{(k)} \left(\frac{z - X_1}{h} \right) \right| \right] \\ &\quad \times \mathbb{E} \left[\left| K^{(k)} \left(\frac{y - X_2}{h} \right) \right| \left| K^{(k)} \left(\frac{u - X_2}{h} \right) \right| \right] \mathbb{E} \left[\left| K^{(k)} \left(\frac{x - X_3}{h} \right) \right| \left| K^{(k)} \left(\frac{y - X_3}{h} \right) \right| \right] \\ &\quad \times \mathbb{E} \left[\left| K^{(k)} \left(\frac{z - X_4}{h} \right) \right| \left| K^{(k)} \left(\frac{u - X_4}{h} \right) \right| \right] dx dy dz du \\ &= h^4 \int_A \int_A \int_A \int_A \left(\int \left| K^{(k)}(x_1) \right| \left| K^{(k)} \left(x_1 + \frac{z - x}{h} \right) \right| f(x - hx_1) dx_1 \right) \\ &\quad \times \left(\int \left| K^{(k)}(x_2) \right| \left| K^{(k)} \left(x_2 + \frac{u - y}{h} \right) \right| f(y - hx_2) dx_2 \right) \\ &\quad \times \left(\int \left| K^{(k)}(x_3) \right| \left| K^{(k)} \left(x_3 + \frac{y - x}{h} \right) \right| f(x - hx_3) dx_3 \right) \end{aligned}$$

$$\times \left(\int |K^{(k)}(x_4)| \left| K^{(k)}\left(x_4 + \frac{u-z}{h}\right) \right| f(z - hx_4) dx_4 \right) dx dy dz du.$$

Mit der Substitution $z = x + ha$, $y = x + hb$ und $u = x + hb + hc$ ergibt dies

$$\begin{aligned} & C \mathbb{E}[|\tilde{W}_{n13}| |\tilde{W}_{n23}| |\tilde{W}_{n14}| |\tilde{W}_{n24}|] \\ & \leq h^7 \int \int \int \int_A \left(\int |K^{(k)}(x_1)| |K^{(k)}(x_1 + a)| f(x - hx_1) dx_1 \right) \\ & \quad \times \left(\int |K^{(k)}(x_2)| |K^{(k)}(x_2 + c)| f(x + hb - hx_2) dx_2 \right) \\ & \quad \times \left(\int |K^{(k)}(x_3)| |K^{(k)}(x_3 + b)| f(x - hx_3) dx_3 \right) \\ & \quad \times \left(\int |K^{(k)}(x_4)| |K^{(k)}(x_4 - a + b + c)| f(x + ha - hx_4) dx_4 \right) dx da db dc. \\ & = O(h^7). \end{aligned}$$

Unter Verwendung von (A.19) und (A.20) bedeutet das

$$\mathbb{E}X[V_{n2}^2] = O_P(n^3 h^6 + n^4 h^7)$$

beziehungsweise

$$\mathbb{E}[(n^{-2} n^{-3} V_{n2})^2] = \mathbb{E}[\mathbb{E}X[(n^{-2} h^{-3} V_{n2})^2]] = O(n^{-1} + h) = o(1)$$

und die Aussage von Lemma A.5 ist bewiesen. \square

Es fehlt noch der Nachweis einer Lindeberg-Bedingung für das Dreiecksschema von Martingaldifferenzen, welcher in folgendem Lemma geführt wird.

Lemma A.6 *Für alle $\varepsilon > 0$ gilt*

$$n^{-2} h^{-3} \sum_{i=2}^n \mathbb{E}X[Y_{ni}^2 I\{|Y_{ni}| > \varepsilon n h^{3/2}\}] \xrightarrow{P} 0.$$

Beweis von Lemma A.6 Der Beweis folgt im Wesentlichen dem von Hall (1984) unter Beachtung, dass die Zufallsvariablen W_{nij} und \tilde{W}_{nij} nicht notwendig positiv sind. Es sei

$$Z_{ni} = \sum_{j=1}^{i-1} (Y_j - m(X_j)) W_{nij},$$

dann lässt sich Y_{ni} schreiben als

$$Y_{ni} = (Y_i - m(X_i)) Z_{ni}$$

und die Zufallsvariablen Z_{ni} sind $\mathcal{A}_{n,i-1}$ -messbar. Ist $l > 0$ konstant und $|Y_{ni}| > \varepsilon nh^{3/2}$, so ist entweder $|Y_i - m(X_i)| > l$ oder es folgt im Fall $|Y_i - m(X_i)| \leq l$ die Ungleichung $l|Z_{ni}| > \varepsilon nh^{3/2}$. Das bedeutet für den bedingten Erwartungswert

$$\begin{aligned} & \mathbb{E}[Y_{ni}^2 I\{|Y_{ni}| > \varepsilon nh^{3/2}\} | \mathcal{A}_{n,i-1}] \\ & \leq \mathbb{E}[(Y_i - m(X_i))^2 Z_{ni}^2 I\{|Y_i - m(X_i)| > l\} + I\{|Z_{ni}| > \varepsilon nh^{3/2}/l\} | \mathcal{A}_{n,i-1}] \\ & = Z_{ni}^2 \left\{ \mathbb{E}[(Y_i - m(X_i))^2 I\{|Y_i - m(X_i)| > l\} | \mathcal{A}_{n,i-1}] + \sigma^2(X_i) I\{|Z_{ni}| > \varepsilon nh^{3/2}/l\} \right\}. \end{aligned}$$

Beachtet man, dass die von X_1, \dots, X_n erzeugte σ -Algebra Teilmenge der σ -Algebra $\mathcal{A}_{n,i-1}$ ist, so lassen sich die Summanden in der Lindeberg-Bedingung abschätzen durch

$$\begin{aligned} \mathbb{E}_{\mathcal{X}}[Y_{ni}^2 I\{|Y_{ni}| > \varepsilon nh^{3/2}\}] &= \mathbb{E}_{\mathcal{X}}[\mathbb{E}[Y_{ni}^2 I\{|Y_{ni}| > \varepsilon nh^{3/2}\} | \mathcal{A}_{n,i-1}]] \\ &\leq \mathbb{E}_{\mathcal{X}}[Z_{ni}^2] \mathbb{E}_{\mathcal{X}}[(Y_i - m(X_i))^2 I\{|Y_i - m(X_i)| > l\}] \\ &\quad + \sigma^2(X_i) \mathbb{E}_{\mathcal{X}}[Z_{ni}^2 I\{|Z_{ni}| > \varepsilon nh^{3/2}/l\}] \\ &\leq C_1 \left(\sum_{j=1}^{i-1} W_{nij}^2 \right) \mathbb{E}_{\mathcal{X}}[(Y_i - m(X_i))^2 I\{|Y_i - m(X_i)| > l\}] \\ &\quad + C_1 \mathbb{E}_{\mathcal{X}}[Z_{ni}^2 I\{|Z_{ni}| > \varepsilon nh^{3/2}/l\}], \end{aligned}$$

wobei die Konstante C_1 aus der gleichmäßigen Abschätzung der Varianz $\sigma^2(X_i)$ stammt. Mit der Ungleichung (A.13) im Beweis von Lemma A.5 lässt sich zeigen, dass gleichmäßig in i gilt

$$\sum_{j=1}^{i-1} W_{nij} = O_P(1) \sum_{j=1}^{i-1} \tilde{W}_{nij} \leq O_P(h) \sum_{j=1}^{i-1} |\tilde{W}_{nij}|.$$

Aus diesem Grund und den vorhergehenden Abschätzungen reicht es für den Beweis von Lemma A.6 die Konvergenzen

$$n^{-2} h^{-3} \sum_{i=2}^n \mathbb{E}_{\mathcal{X}}[Z_{ni}^2 I\{|Z_{ni}| > \varepsilon nh^{3/2}/l\}] \xrightarrow{P} 0 \quad (\text{A.21})$$

für $n \rightarrow \infty$ und jedes $\varepsilon > 0$ und

$$\limsup_{n \rightarrow \infty} n^{-2} h^{-2} \sum_{i=2}^n \mathbb{E} \left[\sum_{j=1}^{i-1} |\tilde{W}_{nij}| \mathbb{E}_{\mathcal{X}}[(Y_i - m(X_i))^2 I\{|Y_i - m(X_i)| > l\}] \right] \xrightarrow{l \rightarrow \infty} 0 \quad (\text{A.22})$$

nachzuweisen. Für den Beweis von (A.21) wird die Ljapunov-Bedingung

$$n^{-4} h^{-6} \sum_{i=2}^n \mathbb{E}_{\mathcal{X}}[Z_{ni}^4] \xrightarrow{P} 0.$$

gezeigt. Ein Summand der obigen Summe lässt sich abschätzen durch

$$\begin{aligned}
\mathbb{E}_{\mathcal{X}}[Z_{ni}^4] &= \sum_{j=1}^{i-1} \mathbb{E}_{\mathcal{X}}[(Y_j - m(X_j))^4] (|W_{nij}|)^4 \\
&\quad + 6 \sum_{1 \leq j < k \leq i-1} \mathbb{E}_{\mathcal{X}}[(Y_j - m(X_j))^2 (Y_k - m(X_k))^2] (|W_{nij}|)^2 (|W_{nik}|)^2 \\
&= \sum_{j=1}^{i-1} \mu_4(X_j) (|W_{nij}|)^4 + 6 \sum_{1 \leq j < k \leq i-1} \sigma^2(X_j) \sigma^2(X_k) (|W_{nij}|)^2 (|W_{nik}|)^2 \\
&= O_P(h^3) \sum_{j=1}^{i-1} \mu_4(X_j) |\tilde{W}_{nij}| + O_P(nh^3) \sum_{i=2}^n \sum_{j=1}^{i-1} |W_{nij}|.
\end{aligned}$$

Insgesamt erhält man nun

$$\mathbb{E} \left[\sum_{i=2}^n \mathbb{E}_{\mathcal{X}}[Z_{ni}^4] \right] = O_P(h^3) \sum_{i=2}^n \sum_{j=1}^{i-1} \mathbb{E}[\mu_4(X_j) |\tilde{W}_{nij}|] + O_P(nh^3) \sum_{i=2}^n \sum_{j=1}^{i-1} \mathbb{E}[|W_{nij}|]$$

und mit einer Abschätzung wie in (A.15) ergibt sich

$$\begin{aligned}
\sum_{i=2}^n \sum_{j=1}^{i-1} \mathbb{E}[\mu_4(X_j) |\tilde{W}_{nij}|] &= O(n^2 h^2), \\
\sum_{i=2}^n \sum_{j=1}^{i-1} \mathbb{E}[|W_{nij}|] &= O(n^2 h^2).
\end{aligned}$$

Das bedeutet

$$\sum_{i=2}^n \mathbb{E}_{\mathcal{X}}[Z_{ni}^4] = O_P(n^2 h^5) + O_P(n^3 h^5) = n^{-1} h^{-1} O_P(n^4 h^6) = o_P(n^4 h^6)$$

und beweist die Gültigkeit von (A.21).

Auch die Aussage (A.22) ist erfüllt, was durch die folgende Abschätzung deutlich wird. Die bedingten Versionen der Cauchy-Schwarz- und Markov-Ungleichung liefern

$$\begin{aligned}
&\mathbb{E}_{\mathcal{X}}[(Y_i - m(X_i))^2 I\{|Y_i - m(X_i)| > l\}] \\
&\leq (\mathbb{E}_{\mathcal{X}}[(Y_i - m(X_i))^2] \mathbb{P}(|Y_i - m(X_i)| > l | X_1, \dots, X_n))^{1/2} \leq \frac{\sigma^2(X_i)}{l}.
\end{aligned}$$

Diese Abschätzung wird in die Summe in (A.22) eingesetzt

$$n^{-2} h^{-2} \sum_{i=2}^n \sum_{j=1}^{i-1} \mathbb{E}[|W_{nij}| \mathbb{E}_{\mathcal{X}}[(Y_i - m(X_i))^2 I\{|Y_i - m(X_i)| > l\}]]$$

$$\begin{aligned}
&\leq \frac{1}{l} \frac{1}{n^2 h^2} \sum_{i=2}^n \sum_{j=1}^{i-1} \mathbb{E}[|W_{n12}| \sigma^2(X_1)] \\
&\leq \frac{1}{l} C \int \int \int |K^{(k)}(x)| |K^{(k)}(x+z)| dx f(x_1 + hz) f(x_1) dx_1 dz \xrightarrow{l \rightarrow \infty} 0.
\end{aligned}$$

und die Aussage von Lemma A.6 ist bewiesen. \square

Lemma A.5 und A.6 garantieren, dass die Voraussetzungen von Korollar 3.2 aus Hall und Heyde (1980) erfüllt sind und somit die Folge $\{n^{-1}h^{-3/2}S_{nn}\}_{n \in \mathbb{N}}$ in Verteilung gegen eine Normalverteilung mit Erwartungswert 0 und Varianz $\alpha_{1,k}/4$ konvergiert. Für die Zerlegung des integrierten quadrierten Fehlers bedeutet es nach Konstruktion der Folge $\{S_{nn}\}$, dass $2I_{n2}$ dasselbe asymptotische Verhalten aufweist wie $n^{-1}h^{-2k-1/2}\alpha_{1,k}^{1/2}Z_n$ mit einer asymptotisch standardnormalverteilten Zufallsvariablen Z_n .

Darüber hinaus erhält man mit denselben Methoden wie in Hall (1984) die stärkere Aussage

$$|P(A_n \cap \{2n^{-1}h^{2k+1/2}\alpha_1^{-1/2}I_{n2} \leq x\}) - P(A_n)\Phi(x)| \rightarrow 0$$

für alle Ereignisse $A_n \in \mathcal{A}_{n,0}$ und die Verteilungsfunktion Φ der Standardnormalverteilung. In Worten bedeutet dies, dass die Zufallsvariable I_{n2} asymptotisch unabhängig ist von der σ -Algebra $\mathcal{A}_{n,0}$.

(iii) Im dritten Schritt wird gezeigt, dass I_{n3} nur für $k = 0$ einen Beitrag zur asymptotischen Verteilung des integrierten quadrierten Fehlers liefert. Der Term I_{n3} wird zerlegt in einen Teil J_{n1} , in dem $\hat{g}^{(k)}(x)$ durch seinen Erwartungswert $\gamma_{n,k}(x) = \mathbb{E}[\hat{g}^{(k)}(x)]$ ersetzt wird und einen Teil J_{n2} , in dem die Differenz zwischen $g^{(k)}(x)$ und $\gamma_{n,k}(x)$ betrachtet wird,

$$I_{n3} = J_{n1} + J_{n2}$$

mit

$$\begin{aligned}
J_{n1} &= \frac{1}{nh^{k+1}} \sum_{i=1}^n (Y_i - m(X_i)) \int_A K^{(k)}\left(\frac{x - X_i}{h}\right) \gamma_{n,k}(x) v_n(x) dx \\
J_{n2} &= \frac{1}{nh^{k+1}} \sum_{i=1}^n (Y_i - m(X_i)) \int_A K^{(k)}\left(\frac{x - X_i}{h}\right) (\hat{g}^{(k)}(x) - \gamma_{n,k}(x)) v_n(x) dx.
\end{aligned}$$

Jeder der beiden Terme wird im Folgenden einzeln behandelt. Zunächst seien nun die Hilfsvariablen

$$Y_{ni} = (Y_i - m(X_i)) \tilde{Z}_{ni}$$

und

$$\tilde{Z}_{ni} = \int_A K^{(k)}\left(\frac{x - X_i}{h}\right) \gamma_{n,k}(x) v_n(x) dx$$

definiert, mit denen sich J_{n1} darstellen lässt als

$$J_{n1} = \frac{1}{nh^{k+1}} \sum_{i=1}^n Y_{ni}.$$

Lemma A.7 (i) *Es sei $k = 0$. Mit obigen Bezeichnungen gilt in diesem Fall*

$$n^{-1}h^{-6} \text{Var}_{\mathcal{X}} \left(\sum_{i=1}^n Y_{ni} \right) \xrightarrow{P} \frac{1}{4} \alpha_{2,0} = \int_A \sigma^2(x) v^2(x) \gamma_0^2(x) f(x) dx$$

und die bedingte Lindebergbedingung

$$n^{-1}h^{-6} \sum_{i=1}^n \text{E}_{\mathcal{X}} [Y_{ni}^2 I\{|Y_{ni}| > \varepsilon n^{1/2} h^3\}] \xrightarrow{P} 0$$

ist für alle $\varepsilon > 0$ erfüllt.

(ii) *Für $k > 0$ gilt*

$$\sum_{i=1}^n Y_{ni} = o_P(n^{1/2} h^3)$$

Beweis von Lemma A.7 Die Zufallsvariablen Y_{ni} sind bedingt auf X_1, \dots, X_n unabhängig identisch verteilt mit Erwartungswert 0. Der Erwartungswert $\gamma_{n,k}(x)$ von $\hat{g}^{(k)}(x)$ besitzt unter Berücksichtigung der Randbedingungen $K^{(j)}(\pm\lambda) = 0$, $j = 0, 1$ eine Entwicklung der Form

$$\begin{aligned} \gamma_{n,k}(x) &= \frac{1}{h^{k+1}} \text{E}[Z(x, X_1)] \\ &= \frac{1}{h^{k+1}} \int_{A^\varepsilon} K^{(k-1)}\left(\frac{x-y}{h}\right) \{(m(y) - m(x))f'(y) + m'(y)f(y)\} dy \\ &\quad + \sum_{j=1}^{k-1} \frac{1}{h^j} \binom{k}{j} m^{k-j}(x) \int_{A^\varepsilon} K^{(j-1)}\left(\frac{x-y}{h}\right) f^{(1)}(y) dy \\ &\quad + \frac{1}{h} m^{(k)}(x) \int_{A^\varepsilon} K\left(\frac{x-y}{h}\right) f(y) dy \\ &= \frac{1}{h^{k+1}} \int_{A^\varepsilon} K\left(\frac{x-y}{h}\right) \sum_{j=0}^k \binom{k}{j} (m^{(k-j)}(y) - m^{(k-j)}(x)) f^{(j)}(y) dy \\ &= \int_{-\lambda}^{\lambda} K(y) \sum_{j=0}^k \binom{k}{j} (m^{(k-j)}(x - hy) - m^{(k-j)}(x)) f^{(j)}(x - hy) dy \end{aligned}$$

$$\begin{aligned}
&= \int_{-\lambda}^{\lambda} K(y) \sum_{j=0}^k \binom{k}{j} \left[-hym^{(k-j+1)}(x) + \frac{h^2 y^2}{2} m^{(k-j+2)}(x) + o(h^2) \right] \\
&\quad \times [f^{(j)}(x) - hyf^{(j+1)}(x) + o(h)] dy \\
&= h^2 \kappa_2(K) \sum_{j=0}^k \binom{k}{j} [2m^{(k-j+1)}(x) f^{(j+1)}(x) + m^{(k-j+2)}(x) f^{(j)}(x)] + o(h^2) \\
&= h^2 \kappa_2(K) \left[m^{(k+2)}(x) f(x) + \sum_{j=0}^{k-1} \binom{k}{j+1} \frac{k+j+2}{k-j} m^{(k-j+1)}(x) f^{(j+1)}(x) \right. \\
&\quad \left. + 2m^{(1)}(x) f^{(k+1)}(x) \right] + o(h^2) \\
&= h^2 \gamma_k(x) + o(h^2) \tag{A.23}
\end{aligned}$$

wobei γ_k in Satz 5.5 und $Z(x, X_1)$ als

$$Z(x, X_1) = K^{(k)}\left(\frac{x - X_1}{h}\right) (m(X_1) - m(x)) - \sum_{j=0}^{k-1} h^{k-j} \binom{k}{j} m^{(k-j)}(x) K^{(j)}\left(\frac{x - X_1}{h}\right) \tag{A.24}$$

definiert ist. Diese Abschätzung gilt darüberhinaus gleichmäßig in $x \in A$. Desweiteren sei an dieser Stelle noch einmal darauf hingewiesen, dass ebenfalls $v_n(x) = v(x) + o_P(1)$ gleichmäßig in x gilt. Eingesetzt in \tilde{Z}_{ni} ergibt das für die bedingte Varianz

$$\begin{aligned}
\mathbb{E}_{\mathcal{X}} \left[\left(\sum_{i=1}^n Y_{ni} \right)^2 \right] &= \sum_{i=1}^n \sigma^2(X_i) \tilde{Z}_{ni} \\
&= h^4 \sum_{i=1}^n \sigma^2(X_i) \left(\int_A K^{(k)}\left(\frac{x - X_i}{h}\right) \gamma_k(x) v(x) dx \right)^2 + o_P(nh^6) \\
&= h^4 M_n + o_P(nh^6).
\end{aligned}$$

Nun ist zu entscheiden, welche Größenordnung M_n in den Fällen $k = 0, 1$ und 2 hat. Mit der Substitution $(y, z) = (x - hu, x + h(v - u))$ gilt für den Erwartungswert von M_n

$$\begin{aligned}
\mathbb{E}[M_n] &= nh^2 \int_A \int_{-\lambda}^{\lambda} \sigma^2(x - hu) f(x - hu) K^{(k)}(u) \\
&\quad \times \int_{-\lambda}^{\lambda} K^{(k)}(v) \gamma_k(x + h(v - u)) v(x + h(v - u)) dv du dx.
\end{aligned}$$

Für das asymptotische Verhalten dieses Erwartungswerts bedeutet dies im Fall $k = 0$

$$\mathbb{E}[M_n] \sim nh^2 \int_A \sigma^2(x) f(x) \gamma_0^2(x) v(x) dx,$$

und $E[M_n] = o(nh^2)$ im Fall $k = 1$ beziehungsweise $k = 2$. Desweiteren lässt sich die Varianz von M_n abschätzen durch

$$\begin{aligned} \text{Var}(M_n) &= n \text{Var}(\sigma^2(X_1) \left(K^{(k)} \left(\frac{x - X_1}{h} \right) \gamma_k(x) v(x) dx \right)^2) \\ &\leq n E \left[\sigma^4(X_1) \left(K^{(k)} \left(\frac{x - X_1}{h} \right) \gamma_k(x) v(x) dx \right)^4 \right] \\ &\leq n C_1 \int_{A^\varepsilon} \left(\int_A K^{(k)} \left(\frac{x - y}{h} \right) dx \right)^4 f(y) dy \\ &= nh^4 C \int_{A^\varepsilon} \left(\int_{-\lambda}^{\lambda} K^{(k)}(x) dx \right)^4 dy = O(nh^4) = o(n^2 h^4). \end{aligned}$$

Es gilt daher

$$\begin{aligned} E_{\mathcal{X}} \left[\left(\sum_{i=1}^n Y_{ni} \right)^2 \right] &= h^4 E[M_n] + o_P(nh^6) \\ &= \begin{cases} nh^6 \int_A \sigma^2(x) f(x) \gamma_0^2(x) v(x) dx + o_P(nh^6) & \text{für } k = 0 \\ o_P(nh^6) & \text{für } k = 1, 2 \end{cases} \end{aligned}$$

Zum Abschluss des Beweises ist die Lindeberg Bedingung in Teil (i) von Lemma A.7 nachzuweisen. Statt dessen kann die einfacher zu handhabende Ljapunov Bedingung

$$\sum_{i=1}^n E_{\mathcal{X}}[Y_{ni}^4] = \sum_{i=1}^n E_{\mathcal{X}}[(Y_i - m(X_i))^4] \tilde{Z}_{ni}^4 \xrightarrow{P} 0 \quad (\text{A.25})$$

direkt nachgerechnet werden. Die Zufallsvariablen \tilde{Z}_{ni} sind stochastisch beschränkt und es gilt für das Supremum

$$\begin{aligned} \sup_{1 \leq i \leq n} |\tilde{Z}_{ni}| &\leq \sup_{1 \leq i \leq n} \int_A \left| K \left(\frac{x - X_i}{h} \right) \right| |\gamma_{n,0}(x)| |v_n(x)| dx \\ &= O_P(h^2) n \int_A \sup_{1 \leq i \leq n} \left| K \left(\frac{x - X_i}{h} \right) \right| dx \end{aligned}$$

nachdem im letzten Schritt die Definition von $\gamma_{n,0}(x)$ eingesetzt und die gleichmäßige Beschränktheit von $v(x)$ ausgenutzt wurde. Die Kernfunktion ist auf dem Intervall $[-\lambda, \lambda]$ beschränkt und man erhält durch Substitution die Größenordnung

$$\sup_{1 \leq i \leq n} |\tilde{Z}_{ni}| = O_P(h^3).$$

Eingesetzt in (A.25) bedeutet dies

$$\sum_{i=1}^n E_{\mathcal{X}}[Y_{ni}^4] = O_P(h^{12}) \sum_{\substack{i \leq n \\ X_i \in A^\varepsilon}} \mu_4(X_i).$$

Die Summe der $\mu_4(X_i)$ wird ähnlich behandelt wie in der Rechnung nach (A.14). Es ist

$$\sum_{\substack{i \leq n \\ X_i \in A^\varepsilon}} \mathbb{E}[\mu_4(X_i)] = n \int_{A^\varepsilon} \mu_4(x) f(x) dx = O(n)$$

und damit

$$\sum_{\substack{i \leq n \\ X_i \in A^\varepsilon}} \mu_4(X_i) = O_P(n).$$

Daraus folgt

$$\sum_{i=1}^n \mathbb{E} \mathcal{X}[Y_{ni}^4] = O_P(nh^{12}) = o_P(n^2h^{12}),$$

was den Beweis von Lemma A.7 abschließt. \square

Überträgt man die Aussage von Lemma A.7 nun auf die Zufallsvariable J_{n1} , so erhält man

$$\frac{2n^{1/2}}{h^2} \alpha_{2,0}^{-1/2} J_{n1} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$$

für $k = 0$ und für $k = 1$ und 2

$$J_{n1} = o_P\left(\frac{h^{2-k}}{n^{1/2}}\right) = o_P\left(\frac{1}{nh^{2k+1/2}}\right).$$

Darüberhinaus lässt sich für $k = 0$ zeigen, dass $n^{1/2}h^{-2}J_{n1}$ asymptotisch unabhängig ist von jeder Folge von Ereignissen $A_n \in \mathcal{A}_{n,0}$. Da dasselbe für I_{n2} gilt, sind I_{n2} und J_{n1} asymptotisch unabhängig und für die Zerlegung des integrierten quadrierten Fehlers gilt

$$\begin{aligned} (n^{-2}h^{-1}\alpha_{1,0} + n^{-1}h^4\alpha_{2,0})^{-1/2}(2I_{n2} + 2J_{n1}) &\stackrel{\mathcal{D}}{=} nh^{1/2}\alpha_{1,0}^{-1/2}(2I_{n2}) + n^{1/2}h^{-2}\alpha_{2,0}^{-1/2}(2J_{n1}) \\ &\xrightarrow{\mathcal{D}} \mathcal{N}(0, 1). \end{aligned}$$

Lemma A.8 $J_{n2} = o_P(n^{-1}h^{-2k-1/2})$

Beweis Es wird der bedingte L_2 -Abstand von J_{n2} zur 0 abgeschätzt, dann erhält man mit der iterierten Erwartung und der Tatsache, dass aus L_2 -Konvergenz die stochastische Konvergenz folgt, die Aussage von Lemma A.8. Es ist

$$\begin{aligned} \mathbb{E} \mathcal{X}[J_{n2}^2] &= \frac{1}{n^2h^{2k+2}} \sum_{i=1}^n \sigma^2(X_i) \left(\int_A K^{(k)}\left(\frac{x - X_i}{h}\right) (\hat{g}^{(k)}(x) - \gamma_{n,k}(x)) v_n(x) dx \right)^2 \\ &\leq \frac{C}{n^2h^{2k+2}} \sum_{i=1}^n \left(\int_A \left| K^{(k)}\left(\frac{x - X_i}{h}\right) \right|^{1/2} \left| K^{(k)}\left(\frac{x - X_i}{h}\right) \right|^{1/2} |\hat{g}^{(k)}(x) - \gamma_{n,k}(x)| |v_n(x)| dx \right)^2 \end{aligned}$$

$$\begin{aligned}
&\leq \frac{C_1}{n^2 h^{2k+2}} \sum_{i=1}^n \left(\int_A \left| K^{(k)} \left(\frac{x - X_i}{h} \right) \right| dx \right) \left(\int_A \left| K^{(k)} \left(\frac{x - X_i}{h} \right) \right| (\hat{g}^{(k)}(x) - \gamma_{n,k}(x))^2 dx \right) \\
&= \frac{C_1}{n^2 h^{2k+2}} \left(\int \left| K^{(k)} \left(\frac{x}{h} \right) \right| dx \right) A_n
\end{aligned} \tag{A.26}$$

mit

$$A_n = \sum_{i=1}^n \int_A \left| K^{(k)} \left(\frac{x - X_i}{h} \right) \right| (\hat{g}^{(k)}(x) - \gamma_{n,k}(x))^2 dx.$$

Die erste Zeile rührt von der Abschätzung der Varianz durch eine Konstante C her. In die zweite Zeile fließen sowohl die gleichmäßige Abschätzung der Gewichte $v_n(x)$ als auch die Cauchy-Schwarz-Ungleichung ein. Es reicht nun nachzuweisen, dass A_n von der Ordnung $O_P(h^{2-2k})$ ist. Dazu wird zunächst die Differenz $(\hat{g}^{(k)}(x) - \gamma_{n,k}(x))^2$ in einen Ausdruck, der die Zufallsvariable X_i enthält und einen, der nur von X_i unabhängige Zufallsvariablen enthält zerlegt. Es ist wie in (A.23)

$$\hat{g}^{(k)}(x) = Z(x, X_i)$$

mit

$$Z(x, X_i) = K^{(k)} \left(\frac{x - X_i}{h} \right) (m(X_i) - m(x)) - \sum_{j=0}^{k-1} h^{k-j} \binom{k}{j} m^{(k-j)}(x) K^{(j)} \left(\frac{x - X_i}{h} \right).$$

Die Differenz zwischen $\hat{g}^{(k)}$ und seinem Erwartungswert kann dargestellt werden als

$$\begin{aligned}
\hat{g}^{(k)}(x) - \gamma_{n,k}(x) &= \frac{1}{nh^{k+1}} \sum_{l=1}^n \{Z(x, X_l) - \mathbb{E}[Z(x, X_l)]\} \\
&= \hat{g}_i^{(k)}(x) - \mathbb{E}[\hat{g}_i^{(k)}(x)] + \frac{1}{nh^{k+1}} \{Z(x, X_i) - \mathbb{E}[Z(x, X_i)]\}
\end{aligned}$$

mit

$$\hat{g}_i^{(k)}(x) = \frac{1}{nh^{k+1}} \sum_{l \neq i} Z(x, X_l).$$

Dann lässt sich die quadrierte Differenz abschätzen durch

$$(\hat{g}^{(k)}(x) - \gamma_{n,k}(x))^2 \leq 2(\hat{g}_i^{(k)}(x) - \mathbb{E}[\hat{g}_i^{(k)}(x)])^2 + 4 \frac{Z^2(x, X_i)}{n^2 h^{2k+2}} + 4 \frac{\mathbb{E}[Z(x, X_i)]}{n^2 h^{2k+2}}$$

und es ist

$$\begin{aligned}
\mathbb{E} \left[\left| K^{(k)} \left(\frac{x - X_i}{h} \right) \right| (\hat{g}^{(k)}(x) - \gamma_{n,k}(x))^2 \right] &\leq 2 \mathbb{E} \left[\left| K^{(k)} \left(\frac{x - X_i}{h} \right) \right| \right] \mathbb{E}[(\hat{g}_i^{(k)}(x) - \mathbb{E}[\hat{g}_i^{(k)}(x)])^2] \\
&+ \frac{4}{n^2 h^{2k+2}} \mathbb{E} \left[\left| K^{(k)} \left(\frac{x - X_i}{h} \right) \right| Z^2(x, X_i) \right] + \frac{4}{n^2 h^{2k+2}} \mathbb{E} \left[\left| K^{(k)} \left(\frac{x - X_i}{h} \right) \right| \right] \mathbb{E}[Z(x, X_i)].
\end{aligned}$$

Die Ordnung von $E[|K^{(k)}((x - X_i)/h)|]$ ist leicht zu bestimmen,

$$E \left[\left| K^{(k)} \left(\frac{x - X_i}{h} \right) \right| \right] = h \int_{-\lambda}^{\lambda} |K^{(k)}(x)| dx = O(h). \quad (\text{A.27})$$

Weiter erhält man die Abschätzung

$$\begin{aligned} E[(\hat{g}_i^{(k)}(x) - E[\hat{g}_i^{(k)}(x)])^2] &= \frac{1}{n^2 h^{2k+2}} \sum_{l \neq i} E[(Z(x, X_l) - E[Z(x, X_l)])^2] \\ &\leq \frac{1}{nh^{2k+2}} E[Z^2(x, X_l)] \\ &= \frac{1}{nh^{2k+1}} \int_{-\lambda}^{\lambda} \left\{ K^{(k)}(y)(m(x - hy) - m(x)) \right. \\ &\quad \left. - \sum_{j=0}^{k-1} h^{k-j} \binom{k}{j} m^{(k-j)}(x) K^{(j)}(y) \right\}^2 f(x - hy) dy \\ &= \frac{1}{nh^{2k-1}} \int_{-\lambda}^{\lambda} \left\{ K^{(k)}(y)(-ym'(x) + \frac{hy^2}{2}m''(x) + o(h)) \right. \\ &\quad \left. - \sum_{j=0}^{k-1} h^{k-j-1} \binom{k}{j} m^{(k-j)}(x) K^{(j)}(y) \right\}^2 f(x - hy) dy \\ &= O\left(\frac{1}{nh^{2k-1}}\right) \end{aligned} \quad (\text{A.28})$$

und

$$\frac{1}{n^2 h^{2k+2}} E \left[\left| K^{(k)} \left(\frac{x - X_i}{h} \right) \right| Z^2(x, X_i) \right] = O\left(\frac{1}{nh^{2k-1}}\right). \quad (\text{A.29})$$

Wie in (A.23) ist der Erwartungswert von $Z(x, X_i)$

$$\begin{aligned} E[Z(x, X_l)] &= \int_{A^\varepsilon} K^{(k-1)} \left(\frac{x - y}{h} \right) \{ (m(y) - m(x))f'(y) + m'(y)f(y) \} dy \\ &= h^{k+3} \gamma_k(x) + o(h^{k+3}) = O(h^{k+3}). \end{aligned} \quad (\text{A.30})$$

Eingesetzt ergibt dies

$$\frac{1}{n^2 h^{2k+2}} E^2[Z(x, X_i)] = O\left(\frac{h^4}{n^2}\right) = o(1). \quad (\text{A.31})$$

Die Abschätzungen (A.27) - (A.31) zusammengefasst ergeben

$$E \left[\left| K^{(k)} \left(\frac{x - X_i}{h} \right) \right| (\hat{g}^{(k)}(x) - \gamma_{n,k}(x))^2 \right]$$

$$\begin{aligned}
&= O(h)O\left(\frac{1}{nh^{2k-1}}\right) + O\left(\frac{1}{n^2h^{2k-1}}\right) + O(h)O\left(\frac{h^4}{n^2}\right) \\
&= \frac{1}{nh^{2k-2}}\left(O(1) + O\left(\frac{1}{nh}\right) + O\left(\frac{h^{2k+3}}{n}\right)\right) \\
&= O\left(\frac{1}{nh^{2k-2}}\right). \tag{A.32}
\end{aligned}$$

Wegen der gleichmäßigen Stetigkeit von m auf A gilt diese Abschätzung gleichmäßig für alle $x \in A$ und die Ordnung bleibt erhalten, wenn über die linke Seite von (A.32) integriert wird. Daraus ergibt sich

$$\mathbb{E}[A_n] = \sum_{i=1}^n \mathbb{E} \left[\left| K^{(k)}\left(\frac{x - X_i}{h}\right) \left(\hat{g}^{(k)}(x) - \gamma_{n,k}(x) \right)^2 \right| \right] = O(h^{2-2k})$$

und wegen der Positivität von A_n auch $A_n = O_P(h^{2-2k})$. Diese Abschätzung wird nun in (A.26) eingesetzt, man erhält

$$\mathbb{E}_{\mathcal{X}}[J_{n2}^2] = O_P\left(\frac{h^{2-2k}}{n^2h^{2k+1}}\right) = O_P\left(\frac{1}{n^2h^{4k-1}}\right) = o_P\left(\frac{1}{n^2h^{4k+1}}\right)$$

und somit die Aussage von Lemma A.8. \square

Mit dem Beweis dieses Lemmas ist gleichzeitig der Beweis von Satz A.4 abgeschlossen. \square

Bevor aus den bisher getroffenen Aussagen Satz 5.5 folgt, muss noch eine entsprechende Darstellung für die Ausdrücke $\mathbb{E}_{\mathcal{X}}[I_{n1}]$ und I_{n4} gefunden werden. Das geschieht in dem folgenden Lemma.

Lemma A.9 *Es gelten die Voraussetzungen aus Satz A.4 und die Gewichtsfunktion sei definiert als $v_n(x) = w(x)/(\hat{f}(x))^2$. Dann gilt*

$$\begin{aligned}
\mathbb{E}_{\mathcal{X}}[I_{n1}] &= \frac{1}{n^2h^{2k+2}} \sum_{i=1}^n \int_A \mathbb{E}[\sigma^2(X_i) K^{(k)^2}\left(\frac{x - X_i}{h}\right)] (\mathbb{E}[\hat{f}(x)])^{-2} w(x) dx \\
&\quad + o_P(n^{-1}h^{-2k-1/2})
\end{aligned}$$

und

$$I_{n4} = \int_A (\mathbb{E}[\hat{g}^{(k)}(x)])^2 (f(x))^{-2} w(x) dx + o_P(n^{-1}h^{-2k-1/2}).$$

Beweis Im Beweis werden folgende Taylorentwicklungen verwendet,

$$\begin{aligned}
(\hat{f}(x))^{-2} &= (\mathbb{E}[\hat{f}(x)])^{-2} - 2 \frac{(\hat{f}(x) - \mathbb{E}[\hat{f}(x)])}{(\mathbb{E}[\hat{f}(x)])^3} + o(|\hat{f}(x) - \mathbb{E}[\hat{f}(x)]|) \\
&= (\mathbb{E}[\hat{f}(x)])^{-2} + O_P(1)|\hat{f}(x) - \mathbb{E}[\hat{f}(x)]| \tag{A.33}
\end{aligned}$$

$$\begin{aligned}
(\hat{f}(x))^{-2} &= (\mathbb{E}[\hat{f}(x)])^{-2} - 2 \frac{(\hat{f}(x) - \mathbb{E}[\hat{f}(x)])}{(\mathbb{E}[\hat{f}(x)])^3} + 6 \frac{(\hat{f}(x) - \mathbb{E}[\hat{f}(x)])^2}{(\mathbb{E}[\hat{f}(x)])^4} \\
&\quad + o(|\hat{f}(x) - \mathbb{E}[\hat{f}(x)]|^2) \\
&= (\mathbb{E}[\hat{f}(x)])^{-2} - 2 \frac{(\hat{f}(x) - \mathbb{E}[\hat{f}(x)])}{(\mathbb{E}[\hat{f}(x)])^3} + O_P(1)(\hat{f}(x) - \mathbb{E}[\hat{f}(x)])^2 \quad (\text{A.34})
\end{aligned}$$

$$(\mathbb{E}[\hat{f}(x)])^{-2} = (f(x))^{-2} + O_P(1)|\mathbb{E}[\hat{f}(x)] - f(x)|. \quad (\text{A.35})$$

Mit (A.34) ist

$$\begin{aligned}
\mathbb{E}\mathcal{X}[I_{n1}] &= \frac{1}{n^2 h^{2k+2}} \sum_{i=1}^n \sigma^2(X_i) \int_A K^{(k)2} \left(\frac{x - X_i}{h} \right) (\mathbb{E}[\hat{f}(x)])^{-2} w(x) dx \\
&\quad - \frac{2}{n^2 h^{2k+2}} \sum_{i=1}^n \sigma^2(X_i) \int_A K^{(k)2} \left(\frac{x - X_i}{h} \right) (\mathbb{E}[\hat{f}(x)])^{-3} (\hat{f}(x) - \mathbb{E}[\hat{f}(x)]) w(x) dx \\
&\quad + O_P\left(\frac{1}{n^2 h^{2k+2}}\right) \sum_{i=1}^n \sigma^2(X_i) \int_A K^{(k)2} \left(\frac{x - X_i}{h} \right) (\hat{f}(x) - \mathbb{E}[\hat{f}(x)])^2 w(x) dx \\
&= B_1 - 2B_2 + B_3.
\end{aligned}$$

Für die Varianz von B_1 gilt

$$\begin{aligned}
\text{Var}(B_1) &= \frac{1}{n^4 h^{4k+4}} \sum_{i=1}^n \left\{ \mathbb{E} \left[\sigma^4(X_i) \left(\int_A K^{(k)2} \left(\frac{x - X_i}{h} \right) (\mathbb{E}[\hat{f}(x)])^{-2} w(x) dx \right)^2 \right] \right. \\
&\quad \left. - \mathbb{E}^2 \left[\sigma^2(X_i) \int_A K^{(k)2} \left(\frac{x - X_i}{h} \right) (\mathbb{E}[\hat{f}(x)])^{-2} w(x) dx \right] \right\} \\
&= \frac{1}{n^3 h^{4k+2}} \left\{ \int_{A^\varepsilon} \sigma^4(y) f(y) \left(\int_{-\lambda}^{\lambda} K^{(k)2}(x) (\mathbb{E}[\hat{f}(y - hx)])^{-2} w(y - hx) dy \right) \right. \\
&\quad \left. - \left(\int_{A^\varepsilon} \sigma^2(y) f(y) \int_{-\lambda}^{\lambda} K^{(k)2}(x) (\mathbb{E}[\hat{f}(y - hx)])^{-2} w(y - hx) dx dy \right)^2 \right\} \\
&= O\left(\frac{1}{n^3 h^{4k+2}}\right) = o\left(\frac{1}{n^2 h^{4k+1}}\right).
\end{aligned}$$

Daher gilt mit (A.35)

$$\begin{aligned}
B_1 &= \mathbb{E}[B_1] + o_P(n^{-1} h^{-2k-1/2}) \\
&= \frac{1}{n^2 h^{2k+2}} \sum_{i=1}^n \int_A \mathbb{E} \left[\sigma^2(X_i) K^{(k)} \left(\frac{x - X_i}{h} \right) \right] (f(x))^{-2} w(x) dx \\
&\quad + O_P\left(\frac{1}{n^2 h^{2k+2}}\right) \sum_{i=1}^n \int_A \mathbb{E} \left[\sigma^2(X_i) K^{(k)} \left(\frac{x - X_i}{h} \right) \right] |\mathbb{E}[\hat{f}(x)] - f(x)| w(x) dx
\end{aligned}$$

$$\begin{aligned}
& +o_P(n^{-1}h^{-2k-1/2}) \\
= & \frac{1}{n^2h^{2k+2}} \sum_{i=1}^n \int_A \mathbb{E} \left[\sigma^2(X_i) K^{(k)} \left(\frac{x - X_i}{h} \right) \right] (f(x))^{-2} w(x) dx \\
& +o_P \left(\frac{1}{nh^{2k+1}} \right) \int_A \int_{-\lambda}^{\lambda} \sigma^2(x - hy) f(x - hy) K^{(k)}(y) dy O(h^2) w(x) dx \\
& +o_P(n^{-1}h^{-2k-1/2}) \\
= & \frac{1}{n^2h^{2k+2}} \sum_{i=1}^n \int_A \mathbb{E} \left[\sigma^2(X_i) K^{(k)} \left(\frac{x - X_i}{h} \right) \right] (f(x))^{-2} w(x) dx + o_P(n^{-1}h^{-2k-1/2}).
\end{aligned}$$

Dies ist der dominierende Term in dem bedingten Erwartungswert $\mathbb{E}_{\mathcal{X}}[I_{n1}]$, die anderen beiden sind von kleinerer Ordnung. Um diese Behauptung nachzuweisen, wird zunächst der Ausdruck B_2 zerlegt in

$$\begin{aligned}
B_2 &= \frac{1}{n^3h^{2k+3}} \sum_{i,j} \int_A \mathbb{E} \left[\sigma^2(X_i) K^{(k)2} \left(\frac{x - X_i}{h} \right) \right] \xi_n(X_j, x) w_n(x) dx \\
&+ \frac{1}{n^3h^{2k+3}} \sum_{i=1}^n \int_A \zeta_n(X_i, x) \xi_n(X_i, x) w_n(x) dx \\
&+ \frac{1}{n^3h^{2k+3}} \sum_{i \neq j} \int_A \zeta_n(X_i, x) \xi_n(X_j, x) w_n(x) dx \\
&= L_{n1} + L_{n2} + L_{n3}
\end{aligned}$$

mit

$$\begin{aligned}
\xi_n(X_i, x) &= K \left(\frac{x - X_i}{h} \right) - \mathbb{E} \left[K \left(\frac{x - X_i}{h} \right) \right] \\
\zeta_n(X_i, x) &= \sigma^2(X_i) K^{(k)2} \left(\frac{x - X_i}{h} \right) - \mathbb{E} \left[\sigma^2(X_i) K^{(k)2} \left(\frac{x - X_i}{h} \right) \right] \\
w_n(x) &= w(x) (\mathbb{E}[\hat{f}(x)])^{-3}.
\end{aligned}$$

Der Erwartungswert von L_{n1} und L_{n3} ist null. Weiter gilt

$$\mathbb{E} \left[\sigma^2(X_i) K^{(k)2} \left(\frac{x - X_i}{h} \right) \right] = h \int_{-\lambda}^{\lambda} \sigma^2(x - hy) K^{(k)2}(y) f(x - hy) dy = O(h),$$

so dass man

$$\begin{aligned}
\text{Var}(L_{n1}) &= O \left(\frac{1}{n^4h^{4k+4}} \right) \sum_{j=1}^n \int_A \int_A \mathbb{E}[\xi_n(X_i, x) \xi_n(X_i, y)] w_n(x) w_n(y) dx dy \\
&= O \left(\frac{1}{n^3h^{4k+4}} \right) \int_A \int_A \int_{A^\varepsilon} K \left(\frac{x - z}{h} \right) K \left(\frac{y - z}{h} \right) f(z) dz dx dy (1 + o(1)) \\
&= O \left(\frac{1}{n^3h^{4k+2}} \right) = o \left(\frac{1}{n^2h^{4k+1}} \right)
\end{aligned}$$

und

$$\begin{aligned}
\text{Var}(L_{n3}) &= \frac{1}{n^6 h^{4k+6}} \sum_{i \neq j} \int_A \int_A \mathbb{E}[\zeta_n(X_i, x) \zeta_n(X_i, y)] \mathbb{E}[\xi_n(X_j, x) \xi_n(X_j, y)] w_n(x) w_n(y) dx dy \\
&\quad + \frac{1}{n^6 h^{4k+6}} \sum_{i \neq j} \int_A \int_A \mathbb{E}[\zeta_n(X_i, x) \xi_n(X_i, y)] \mathbb{E}[\zeta_n(X_j, x) \xi_n(X_j, y)] w_n(x) w_n(y) dx dy \\
&= O\left(\frac{1}{n^4 h^{4k+6}}\right) \int_A \int_A \left(\int_{A^\varepsilon} \sigma^4(z) K^{(k)^2}\left(\frac{x-z}{h}\right) K^{(k)^2}\left(\frac{y-z}{h}\right) f(z) dz \right) \\
&\quad \times \left(\int_{A^\varepsilon} K\left(\frac{x-z}{h}\right) K\left(\frac{y-z}{h}\right) f(z) dz \right) w_n(x) w_n(y) dx dy \\
&\quad + O\left(\frac{1}{n^4 h^{4k+6}}\right) \int_A \int_A \left(\int_{A^\varepsilon} \sigma^2(z) K^{(k)^2}\left(\frac{x-z}{h}\right) K\left(\frac{y-z}{h}\right) f(z) dz \right)^2 w_n(x) w_n(y) dx dy \\
&= O\left(\frac{1}{n^4 h^{4k+3}}\right) \int_A \int_{-\lambda}^{\lambda} \left(\int_{-\lambda}^{\lambda} \sigma^4(x-hz) K^{(k)^2}(z) K^{(k)^2}(z+y) f(x-hz) dz \right) \\
&\quad \times \left(\int_{-\lambda}^{\lambda} K(z) K(z+y) f(x-hz) dz \right) w_n(x) w_n(x+hy) dy dx \\
&\quad + O\left(\frac{1}{n^4 h^{4k+3}}\right) \int_A \int_{-\lambda}^{\lambda} \left(\int_{-\lambda}^{\lambda} \sigma^2(x-hz) K^{(k)^2}(z) K(z+y) f(x-hz) dz \right)^2 \\
&= O\left(\frac{1}{n^4 h^{4k+3}}\right) = o\left(\frac{1}{n^2 h^{4k+1}}\right)
\end{aligned}$$

erhält. Damit sind sowohl L_{n1} als auch L_{n3} von der Ordnung $o_P(n^{-1}h^{-2k-1/2})$. Der Erwartungswert von L_{n2} hat die Ordnung

$$\begin{aligned}
\mathbb{E}[L_{n2}] &= \frac{1}{n^3 h^{2k+3}} \sum_{i=1}^n \int_A \mathbb{E}[\zeta_n(X_i, x) \xi_n(X_i, x)] w_n(x) dx \\
&= \frac{1}{n^2 h^{2k+2}} \int_{A^\varepsilon} \int_{-\lambda}^{\lambda} \sigma^2(y) K^{(k)^2}(x) K(x) f(y) w_n(y-hx) dx dy + o\left(\frac{1}{n^2 h^4}\right) \\
&= O\left(\frac{1}{n^2 h^{2k+2}}\right) = o\left(\frac{1}{n h^{2k+1/2}}\right),
\end{aligned}$$

für die Varianz ergibt sich die Abschätzung

$$\begin{aligned}
\text{Var}(L_{n2}) &\leq \frac{1}{n^6 h^{4k+6}} \sum_{i=1}^n \int_A \int_A \mathbb{E}[\zeta_n(X_i, x) \zeta_n(X_i, y) \xi_n(X_i, x) \xi_n(X_i, y)] w_n(x) w_n(y) dx dy \\
&= O\left(\frac{1}{n^5 h^{4k+6}}\right) \int_A \int_A \int_{A^\varepsilon} \sigma^4(z) K^{(k)^2}\left(\frac{x-z}{h}\right) K^{(k)^2}\left(\frac{y-z}{h}\right) \\
&\quad \times K\left(\frac{x-z}{h}\right) K\left(\frac{y-z}{h}\right) f(z) dz w_n(x) w_n(y) dx dy \\
&= O\left(\frac{1}{n^5 h^{4k+4}}\right) = o\left(\frac{1}{n^2 h^{4k+1}}\right),
\end{aligned}$$

so dass L_{n2} geschrieben werden kann als

$$L_{n2} = \mathbb{E}[L_{n2}] + o_P\left(\frac{1}{nh^{2k+1/2}}\right) = O\left(\frac{1}{n^2h^{2k+1/2}}\right) + o_P\left(\frac{1}{nh^{2k+1/2}}\right) = o_P\left(\frac{1}{nh^{2k+1/2}}\right).$$

Zusammengefasst ist B_2 von der Ordnung $o_P(n^{-1}h^{-2k-1/2})$.

Der letzte Term B_3 lässt sich einfach mit der gleichmäßig fast sicheren Konvergenzrate für \hat{f} abschätzen. Es ist

$$B_3 = O_P\left(\frac{\log h^{-1}}{n^2h^{2k+2}}\right) = O_P\left(\frac{1}{nh^{2k+1/2}} \frac{\log h^{-1}}{nh^{3/2}}\right) = o_P\left(\frac{1}{nh^{2k+1/2}}\right)$$

und damit ist der erste Teil von Lemma (A.9) bewiesen.

Es folgt abschließend der Nachweis der stochastischen Entwicklung für die Zufallsvariable I_{n4} . Hierzu wird die Taylorentwicklung (A.33) in die Definition von I_{n4} eingesetzt,

$$I_{n4} = \int_A (\hat{g}^{(k)}(x))^2 (\mathbb{E}[\hat{f}(x)])^{-2} + O_P(1) \int_A (\hat{g}^{(k)}(x))^2 |\hat{f}(x) - \mathbb{E}[\hat{f}(x)]| = I_{n4}^{[1]} + I_{n4}^{[2]}.$$

Die Varianz von $I_{n4}^{[1]}$ ist

$$\text{Var}(I_{n4}^{[1]}) = \text{Var}(L_{n1} + L_{n2}) \leq \text{Var}(L_{n1}) + \text{Var}(L_{n2})$$

mit

$$\begin{aligned} L_{n1} &= \frac{1}{n^2h^{2k+2}} \sum_{i=1}^n \int_A \xi_n^{[2]}(X_i, x) (\mathbb{E}[\hat{f}(x)])^{-2} w(x) dx, \\ L_{n2} &= \frac{1}{n^2h^{2k+2}} \sum_{i \neq j} \xi_n^{[1]}(X_i, x) \xi_n^{[1]}(X_j, x) (\mathbb{E}[\hat{f}(x)])^{-2} w(x) dx, \\ \xi_n^{[l]}(X_i, x) &= Z^l(x, X_i) - \mathbb{E}[Z^l(x, X_i)], \end{aligned}$$

wobei die Zufallsvariable $Z(x, X_i)$ wie in (A.24) definiert ist. Man berechnet für die Varianz von L_{n1}

$$\begin{aligned} \text{Var}(L_{n1}) &= \frac{1}{n^4h^{4k+4}} \sum_{i=1}^n \mathbb{E} \left[\left(\int_A \xi_n^{[2]}(X_1, x) (\mathbb{E}[\hat{f}(x)])^{-2} w(x) dx \right)^2 \right] \\ &= \frac{1}{n^3h^{4k+4}} \int_A \int_A \mathbb{E}[\xi_n^{[2]}(X_1, x) \xi_n^{[2]}(X_1, y)] (\mathbb{E}[\hat{f}(x)] \mathbb{E}[\hat{f}(y)])^{-2} w(x) w(y) dx dy \\ &= \frac{1}{n^3h^{4k+4}} \int_A \int_A \mathbb{E}[Z^2(x, X_1) Z^2(y, X_1)] (\mathbb{E}[\hat{f}(x)] \mathbb{E}[\hat{f}(y)])^{-2} w(x) w(y) dx dy \\ &\quad - \frac{1}{n^3h^{4k+4}} \int_A \int_A \mathbb{E}[Z^2(x, X_1)] \mathbb{E}[Z^2(y, X_1)] (\mathbb{E}[\hat{f}(x)] \mathbb{E}[\hat{f}(y)])^{-2} w(x) w(y) dx dy \\ &= O\left(\frac{1}{n^3h^{4k-1}}\right) + O\left(\frac{1}{n^3h^{4k-2}}\right) = O\left(\frac{1}{n^3h^{4k-1}}\right) = o\left(\frac{1}{n^2h^{4k+1}}\right) \end{aligned}$$

mit einer ähnlichen Rechnung wie in (A.36). Die Varianz von L_{n2} ist

$$\begin{aligned}\text{Var}(L_{n2}) &= O\left(\frac{1}{n^2 h^{4k+4}}\right) \int_A \int_A \mathbb{E}^2[Z(x, X_i)Z(y, X_i)](\mathbb{E}[\hat{f}(x)])^{-2} w(x)w(y) dx dy \\ &= O\left(\frac{1}{n^2 h^{4k-4}}\right) = o\left(\frac{1}{n^2 h^{4k+1}}\right).\end{aligned}$$

Daher gilt

$$I_{n4}^{[1]} = \mathbb{E}[I_{n4}^{[1]}] + o_P\left(\frac{1}{nh^{2k+1/2}}\right) = \int_A \mathbb{E}[(\hat{g}^{(k)}(x))^2](\mathbb{E}[\hat{f}(x)])^{-2} w(x) dx + o_P\left(\frac{1}{nh^{2k+1/2}}\right)$$

und für den Erwartungswert erhält man

$$\mathbb{E}[(\hat{g}^{(k)}(x))^2] = \mathbb{E}[G_n^{[1]}(x)] + \frac{n-1}{n} (\mathbb{E}[G_n^{[2]}(x)])^2$$

mit

$$\begin{aligned}G_n^{[1]}(x) &= \frac{1}{nh^{2k+2}} Z^2(x, X_1) \\ G_n^{[2]}(x) &= \frac{1}{h^{k+1}} Z(x, X_1),\end{aligned}$$

wobei $Z(x, X_1)$ definiert ist wie in (A.24). Der erste Term ist vernachlässigbar, denn mit einer ähnlichen Rechnung wie in (A.28)

$$\begin{aligned}\mathbb{E}[G_n^{[1]}(x)] &= \frac{1}{nh^{2k+1}} \mathbb{E}[Z^2(x, X_1)] = O\left(\frac{1}{nh^{2k-1}}\right) \\ &= o\left(\frac{1}{nh^{2k+1/2}}\right)\end{aligned}\tag{A.36}$$

und nach (A.30) gilt

$$\begin{aligned}\mathbb{E}[G_n^{[2]}(x)] &= \frac{1}{h^{k+1}} \mathbb{E}[Z(x, X_1)] = \frac{1}{h^{k+1}} (h^{k+3} \gamma_k(x) + o(h^{k+3})) \\ &= O(h^2).\end{aligned}\tag{A.37}$$

Unter der Voraussetzung $nh^{2k+5} = O(1)$ ist der Term mit $\mathbb{E}[G_n^{[2]}(x)] = \mathbb{E}[\hat{g}^{(k)}(x)]$ asymptotisch nicht vernachlässigbar und mit obigen Abschätzungen und der Taylorentwicklung (A.35) erhält man

$$\begin{aligned}I_{n4}^{[1]} &= \int_A (\mathbb{E}[\hat{g}^{(k)}(x)])^2 (f(x))^{-2} w(x) dx + O\left(\frac{h^4}{n}\right) \\ &\quad + O(h^4) \int_A |\mathbb{E}[\hat{f}(x)] - f(x)| w(x) dx + o_P\left(\frac{1}{nh^{2k+1/2}}\right) \\ &= \int_A (\mathbb{E}[\hat{g}^{(k)}(x)])^2 (f(x))^{-2} w(x) dx + O\left(\frac{h^4}{n}\right) + O(h^6) + o_P\left(\frac{1}{nh^{2k+1/2}}\right) \\ &= \int_A (\mathbb{E}[\hat{g}^{(k)}(x)])^2 (f(x))^{-2} w(x) dx + o_P\left(\frac{1}{nh^{2k+1/2}}\right).\end{aligned}\tag{A.38}$$

Mit der Cauchy-Schwarz Ungleichung wird $I_{n^4}^{[2]}$ abgeschätzt durch

$$I_{n^4}^{[2]} \leq O_P(1) \left(\int_A (\hat{g}^{(k)}(x))^4 dx \right)^{1/2} \left(\int_A |\hat{f}(x) - \mathbb{E}[\hat{f}(x)]|^2 dx \right)^{1/2}.$$

Der zweite Faktor quadriert ist von der Ordnung

$$\int_A (\hat{f}(x) - \mathbb{E}[\hat{f}(x)])^2 dx = \int_A \text{MSE}(\hat{f}(x)) dx = O\left(\frac{1}{nh}\right),$$

für das Quadrat des ersten gilt

$$\begin{aligned} \int_A \mathbb{E}[(\hat{g}^{(k)}(x))^4] dx &= \frac{1}{n^4 h^{4k+4}} \sum_{i,j,k,l} \int_A Z(x, X_i) Z(x, X_j) Z(x, X_k) Z(x, X_l) dx \\ &= O\left(\frac{1}{n^3 h^{4k+4}}\right) \int_A \mathbb{E}[Z^4(x, X_1)] dx \\ &\quad + O\left(\frac{1}{n^2 h^{4k+4}}\right) \left(\int_A \mathbb{E}^2[Z^2(x, X_1)] dx + \int_A \mathbb{E}[Z^3(x, X_1)] \mathbb{E}[Z(x, X_1)] \right) \\ &\quad + O\left(\frac{1}{nh^{4k+4}}\right) \int_A \mathbb{E}[Z^2(x, X_1)] \mathbb{E}^2[Z(x, X_1)] dx + O\left(\frac{1}{h^{4k+4}}\right) \int_A \mathbb{E}^4[Z(x, X_1)] dx \\ &= O\left(\frac{1}{n^3 h^{4k-1}}\right) + O\left(\frac{1}{n^2 h^{4k-2}}\right) + O\left(\frac{1}{nh^{2k-5}}\right) + O(h^8). \end{aligned}$$

Eingesetzt in die Zerlegung von $I_{n^4}^{[2]}$ bedeutet dies

$$\begin{aligned} (I_{n^4}^{[2]})^2 &= \left(O\left(\frac{1}{n^3 h^{4k-1}}\right) + O\left(\frac{1}{n^2 h^{4k-2}}\right) + O\left(\frac{1}{nh^{2k-5}}\right) + O(h^8) \right) O\left(\frac{1}{nh}\right) \\ &= o\left(\frac{1}{n^2 h^{4k+1}}\right). \end{aligned}$$

Diese Abschätzung zusammen mit (A.38) schließt den Beweis von Lemma A.9 ab. \square

Eine Kombination von Satz A.4 mit der Gewichtsfunktion $v_n(x) = w(x)/(\hat{f}(x))^2$ und Lemma A.9 ergibt den Beweis von Satz 5.5.

Symbolverzeichnis

Symbol	Bedeutung	Seite
X_i	i -ter Designpunkt	11
Y_i	i -te Response-Variable	11
ε_i	i -ter Fehler	11
m	Regressionsfunktion	11
σ	Funktion der Standardabweichung von Y_i	11
n	Stichprobengröße	11
K_r, K_d	Kernfunktionen	12, 22
h_r, h_d	Bandbreiten zur Kernschätzung	12, 22
\hat{m}	Nadaraya-Watson Schätzer der Regressionsfunktion	12
\hat{m}_p	Lokaler Polynomschätzer der Ordnung p	13
$\kappa_q(K_r)$	q -tes Moment des Kerns K_r	15
b_{K_r}	Integral des quadrierten Kerns K_r	15
$\phi(g)(t)$	Verteilungsfunktion der Funktion g	19
$\phi(g)^{-1}(t)$	isotone Umordnung der Funktion g	19
$\phi_{h_d}(g)(t)$	geglättete Verteilungsfunktion der Funktion g	22
$\phi_{h_d}^{-1}(g)(t)$	geglättete isotone Umordnung der Funktion g	22
$\varphi^{-1}(g)(t)$	antitone Umordnung der Funktion g	25
$\varphi_{h_d}^{-1}(g)(t)$	geglättete antitone Umordnung der Funktion g	25
$\rho(m)(x, a)$	konvexe Umordnung der Funktion m	26
$\rho_{h_d}(m)(x, a)$	geglättete konvexe Umordnung der Funktion m	26
$\rho(m)(x), \rho_{h_d}(m)(x)$	L^2 -optimale (geglättete) konvexe Umordnung der Funktion m	26
$\varrho(m)(x, a), \varrho_{h_d}(m)(x, a)$	(geglättete) konkave Umordnung der Funktion m	28
$\varrho(m)(x), \varrho_{h_d}(m)(x)$	L^2 -optimale (geglättete) konkave Umordnung der Funktion m	28
$\hat{\phi}(t), \hat{\phi}_{h_d}(t)$	(geglättete) Verteilungsfunktion von \hat{m}	31
$\hat{\phi}^{-1}(t), \hat{\phi}_{h_d}^{-1}(t)$	(geglätteter) isotoner Schätzer	31

Symbol	Bedeutung	Seite
$\Delta_n^{(j)}(t)$	Term der Zerlegung von $\hat{\phi}_{h_d}^{-1}(t)$ im Beweis von Satz 3.2	
$s^2(t), \tilde{s}^2(t), \tau^2(t)$	Varianzen des (inversen) isotonen Schätzers	32, 33, 38
$a_{K_d, K_r}(t), b_{K_d}(t)$	Bias-Terme des (inversen) isotonen Schätzers	33, 38
\hat{m}_I	isotoner Kleinste-Quadrate Schätzer	40
$\hat{\rho}_{h_d}(x, a)$	konvexer Schätzer der Regressionsfunktion	44
$\hat{\rho}_{h_d}(x)$	L^2 -optimaler konvexer Schätzer der Regressionsfunktion	44
$\gamma^2(x)$	asymptotische Varianz des konvexen Schätzers	44
$\hat{\varrho}_{h_d}(x, a)$	konkaver Schätzer der Regressionsfunktion	46
$\hat{\varrho}_{h_d}(x)$	L^2 -optimaler konkaver Schätzer der Regressionsfunktion	46
\hat{m}_{LS}	konvexer Kleinste-Quadrate Schätzer	52
$\tilde{\Delta}_n^{(j)}, \Delta_n^{(j)}(t)$	Terme der Zerlegung von $\hat{\phi}_{h_d}^{-1}(t)$ im Beweis von Lemma 4.3 und 4.4	58, 64
\hat{f}_n	Kern-Dichte Schätzer	71
\tilde{T}_n, T_n	Teststatistiken zum Testen auf Monotonie	72
\tilde{T}	deterministische Version von \tilde{T}_n	74
A_i°	Inneres der Menge A_i	74
$\mu_4(X_i)$	auf X_i bedingtes zentriertes viertes Moment von Y_i	77
$B_n^{[1]}, B_n^{[2]}, \tilde{B}_n^{[1]}$	Bias-Terme der Teststatistik T_n	77, 78
V	asymptotische Varianz der Teststatistik T_n	77
$\hat{m}^{(k)}$	Schätzer der k -ten Ableitung von m	79
$T_n^{(k)}$	integrierter quadrierter Fehler des Schätzers der k -ten Ableitung von m	79
$S_k(x)$	Term in der Zerlegung von $\hat{m}^{(k)} - m^{(k)}$	79
$B_{n,k}^{[1]}, B_{n,k}^{[2]}$	Bias-Terme von S_k	80
$\alpha_{1,k}, \alpha_{2,0}$	Varianz-Terme von S_k	80
$\hat{\varepsilon}_i$	Residuen (geschätzte Fehlervariablen)	88
Y_i^*	Bootstrap-Variablen	88
$T_n^{*,j}$	j -te Version der Bootstrap-Statistik	
$h_{r,b}, h_{d,b}$	Bandbreiten im Bootstrap	88
I_{nj}	Terme im Beweis von Satz 5.5	100
$E_{\mathcal{X}}[Z]$	Erwartungswert von Z bedingt auf X_1, \dots, X_n	100
$\text{Var}_{\mathcal{X}}(Z)$	Varianz von Z bedingt auf X_1, \dots, X_n	100
$\mathcal{A}_{n,j}$	Von X_1, \dots, X_n und Y_1, \dots, Y_j erzeugte σ -Algebra	102

Literaturverzeichnis

- Y. AÏT-SAHALIA und J. DUARTE (1998). Nonparametric option pricing under shape restrictions. *J. Econom.* 116, 9-47.
- D. ANEVSKI und O. HÖSSJER (2006). A general asymptotic scheme for inference under order restrictions. *Ann. Statist.* 34, 1874-1930.
- Y. BARAUD, S. HUET und B. LAURENT (2003). Adaptive tests of qualitative hypotheses. *ESAIM: Probab. Statist.* 7, 147-159.
- C. BENNETT und R. SHARPLEY (1988). *Interpolation of Operators*. Academic Press, New York
- F. BLACK und M. SCHOLES (1973). The pricing of options and corporate liabilities, *Journal of Political Economy* 81, 637-654.
- A.W. BOWMAN, M.C. JONES und I. GIJBELS (1998). Testing Monotonicity of Regression. *J. Comput. Graph. Statist.* 7, 489-500.
- D. BREEDEN und R. LITZENBERGER (1978). Prices of state-contingent claims implicit in option prices. *Journal of Business* 51, 621-651 .
- H.D. BRUNK (1955). Maximum likelihood estimates of monotone parameters. *Ann. Math. Stat.* 26, 607-616 .
- H.D. BRUNK (1958). On the estimation of parameters restricted by inequalities. *Ann. Math. Stat.* 29, 437-454.
- H.D. BRUNK (1970). Estimation of isotonic regression. *Nonparametric Techniques in Statistical Inference* . 177-197, Cambridge Univ. Press, London
- H. DETTE, N. NEUMEYER und K. PILZ (2006). A simple nonparametric estimator of a strictly monotone regression function. *Bernoulli* 12, 469-490.
- J. DOMÍNGUEZ-MENCHERO, G. GONZÁLEZ-RODRÍGUEZ, M. LÓPEZ-PALOMO (2005). An L_2 point of view in testing monotone regression. *J. Nonparametric Stat.* 17, 135-153.

- M.L. DUDZINSKI und R. MYKYTOWYCZ (1961). The eye lens as an indicator of age in the wild rabbit in Australia. *CSIRO Wildlife Research* 6, 156-159.
- C. DUROT (2003). A Kolmogorov-type test for monotonicity of regression. *Stat. Probab. Lett.* 63, 425-433.
- R.L. EFROMOVICH (1999). *Nonparametric curve estimation: methods, theory, and applications*. Springer, New York.
- R.L. EUBANK (1988). *Spline smoothing and nonparametric regression*. Marcel Dekker, New York.
- J. FAN und I. GIJBELS (1997). *Local polynomial modelling and its applications*. Chapman & Hall, London.
- J. FRIEDMAN und R. TIBSHIRANI (1984). The monotone smoothing of scatterplots. *Technometrics* 26, 243-250.
- T. GASSER und H.-G. MÜLLER (1979). Kernel estimation of regression functions. *Smoothing techniques for curve estimation, Proc. Workshop, Heidelberg 1979, Lect. Notes Math.* 757, 23-68.
- S. GHOSAL, A. SEN und A.W. VAN DER VAART (2000). Testing monotonicity of regression. *Ann. Statist.* 28, 1054-1082 .
- I. GIJBELS, P. HALL, M.C. JONES und I. KOCH (2000). Tests for monotonicity of a regression mean with guaranteed level. *Biometrika* 87, 663-673.
- P. GROENEBOOM (1985). Estimating a monotone density. *Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer, Vol. II* 539-555.
- P. GROENEBOOM (1989). Brownian motion with a parabolic drift and Airy functions. *Probab. Theory Related Fields* 81, 79-109.
- P. GROENEBOOM, G. JONGBLOED und J.A. WELLNER (2001). Estimation of a convex function: characterizations and asymptotic theory. *Ann. Statist.* 19, 1653-1698.
- W. HÄRDLE (1995). *Applied nonparametric regression*. Cambridge Univ. Press.
- P. HALL (1984). Integrated square error properties of kernel estimators of regression functions. *Ann. Statist.* 12, 241-260.
- P. HALL und C.C. HEYDE (1980). *Martingale limit theory and its application*. Probability and Mathematical Statistics, Academic Press, New York.

- P. HALL und N.E. HECKMAN (2000). Testing for monotonicity of a regression mean by calibrating for linear functions. *Ann. Statist.* 28, 20-39.
- P. HALL und L.-S. HUANG (2001). Nonparametric kernel regression subject to monotonicity constraints. *Ann. Statist.* 29, 624-647.
- D.L. HANSON und G. PLEDGER (1976). Consistency in concave regression. *Ann. Statist.* 4, 1038-1050.
- C. HILDRETH (1954). Point estimates of ordinates of concave functions. *J. Amer. Statist. Assoc.* 49, 598-619.
- Y.P. MACK und B.W. SILVERMAN (1982). Weak and strong uniform consistency of kernel regression estimates. *Z. Wahrsch. Verw. Gebiete* 61, 405-415.
- E. MAMMEN (1991a). Estimating a smooth monotone regression function. *Ann. Statist.* 19, 724-740.
- E. MAMMEN (1991b). Nonparametric regression under qualitative smoothness assumptions. *Ann. Statist.* 19, 741-759.
- E. MAMMEN (1993). Bootstrap and wild bootstrap for high dimensional linear models. *Ann. Statist.* 21, 255-285.
- E. MAMMEN, J.S. MARRON, B.A. TURLACH und M.P. WAND (2001). A general projection framework for constrained smoothing. *Statist. Sci.* 16, 232-248.
- E. MAMMEN und C. THOMAS-AGNAN (1999). Smoothing splines and shape restrictions. *Scand. J. Statist.* 26, 239-252.
- H. MUCKERJEE (1988). Monotone nonparametric regression. *Ann. Statist.* 16, 741-750.
- H.-G. MÜLLER (1987). Weighted local regression and kernel methods for nonparametric curve fitting. *J. Amer. Statist. Assoc.* 82, 231-238.
- E.A. NADARAYA (1964). On estimating regression. *Theory Probab. Appl.* 9, 141-142.
- B.L.S. PRAKASA RAO (1969). Estimating a unimodal density. *Sankhyā Ser. A* 31, 23-36.
- D.A. RATKOWSKY (1983). *Nonlinear Regression Modeling*. Marcel Dekker Inc., New York.
- J. O. RAMSAY (1998). Estimating smooth monotone functions. *J. R. Statist. Soc. B* 60, 365-375.

- J. RICE (1984). Bandwidth choice for nonparametric regression. *Ann. Statist.* 12, 1215-1230.
- W. SCHLEE (1982). Nonparametric tests of the monotony and convexity of regression. In: *Nonparametric Statistical Inference, 2*, Ed. B.V. Gnedenko, M.L. Puri and I. Vincze, pp. 823-836. Amsterdam: North-Holland.
- S.R. SEARLE (1971). *Linear models*. Wiley, New York.
- G.A.F. SEBER (1977). *Linear regression analysis*. Wiley, New York.
- C.J. STONE (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* 10, 1040-1053.
- B.W. SILVERMAN (1981). Using kernel density estimates to investigate multi-modality. *J. R. Statist. Soc. B* 43, 97-99.
- F.T. WRIGHT (1981). The asymptotic behavior of monotone regression estimates. *Ann. Statist.* 9, 443-448.
- G.S. WATSON (1964). Smooth regression analysis. *Sankhyā Ser. A* 26, 359-372.
- C.F.J. WU (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Ann. Statist.* 14, 1261-1295.
- J.-T. ZHANG (2004). A simple and efficient monotone smoother using smoothing splines. *J. Nonparametric Stat.* 16, 779-796.

Danksagung

An dieser Stelle bedanke ich mich ganz herzlich bei Herrn Prof. Dr. Holger Dette für die ausgezeichnete fachliche Betreuung. Mein Dank gilt auch allen Mitarbeiterinnen und Mitarbeitern des Lehrstuhls für ihre Hilfsbereitschaft und gute Zusammenarbeit. Darüberhinaus danke ich Dr. Kay Pilz für die Bereitstellung von Daten der Sal. Oppenheim Bank, die in dieser Arbeit benötigt wurden und Prof. Dr. Geurt Jongbloed dafür, dass er mir ein R-Programm für Simulationen zu dem konvexen Kleinst-Quadrate Schätzer zur Verfügung gestellt hat. Nicht zuletzt möchte ich ganz besonders meinen Eltern Sigrid und Reinhard Birke, Simon Konzelmann sowie allen danken, die mich während der letzten Jahre unterstützt und zu dieser Arbeit ermutigt haben.

Dissertation eingereicht am 20. November 2006

Gutachter: Prof. Dr. H. Dette (Ruhr-Universität Bochum)
Prof. Dr. H. Dehling (Ruhr-Universität Bochum)
Prof. Dr. M. Denker (Georg-August-Universität zu Göttingen)

Mündliche Prüfung am 06. Februar 2007

Lebenslauf

Persönliche Daten

Name	Melanie Birke
Geburtsdatum	25. 03. 1980
Geburtsort	Dortmund

Ausbildung

1986 - 1990	Gemeinschaftsgrundschule Hörder Straße, Bochum
1990 - 1999	Lessing-Schule (Gymnasium), Bochum
Juni 1999	Abitur
1999 - 2003	Diplomstudiengang Mathematik an der Ruhr-Universität Bochum
16.12.2003	Diplom in Mathematik

Beschäftigung

April 2001 - Dez. 2003	Studentische Hilfskraft an der Fakultät für Mathematik, Ruhr-Universität Bochum
seit Januar 2004	Wissenschaftliche Mitarbeiterin an der Fakultät für Mathematik, Ruhr-Universität Bochum. Mitarbeit im SFB 475, Komplexitätsreduktion in multivariaten Datenstrukturen