Schätzung glatter monotoner Funktionen, basierend auf der Maximierung geglätteter Likelihoodfunktionen

Diplomarbeit

Institut für Mathematik
Mathematisch-Naturwissenschaftliche Fakultät II
Humboldt-Universität zu Berlin

Eingereicht von Jens Haase, geboren am 2. September 1968

Betreuer: Professor Enno Mammen

Berlin, den 5. Oktober 1995

Zusammenfassung

Ein nichtparametrisches Regressionsmodell wird untersucht, dessen Beobachtungsvariablen den Verteilungen einer einparametrigen Exponentialfamilie genügen. Zwei Schätzer der Regressionsfunktion werden hergeleitet, die ein *geglättetes Log-Likelihood-Funktional* maximieren – einer ohne Einschränkung, der andere unter allen monotonen Funktionen. Mit geeigneten Voraussetzungen werden asymptotische Konvergenzraten hergeleitet.

Mein Dank gilt meinen Eltern.

Professor Mammen danke ich für die Anregung zum Thema dieser Diplomarbeit und für zahlreiche Gespräche und Hinweise.

Allen Mitarbeitern und Mitarbeiterinnen des Institutes für Mathematik, besonders Professor Bunke, danke ich für ihre Bemühungen, mathematische Kenntnisse und Denkweisen zu vermitteln.

Matthias Küchler danke ich für die freundliche Unterstützung bei der Simulation und Berechnung der Daten für die Beispiele in Kapitel 5.

Inhaltsverzeichnis

1 Einführung

Nichtparametrische Verfahren werden in der statistischen Praxis für gewöhnlich eingesetzt, wenn wenig oder nur qualitative Merkmale des zu untersuchenden Gegenstandes bekannt sind. Im Gegensatz zu parametrischen Methoden kleiden sie die vorhandenen oder noch zu sammelnden Daten nicht von vornherein in ein mehr oder weniger starres Gerüst, in dem „nur noch“ eine gewisse Anzahl Parameter angepaßt werden. Jedoch kann es auch in nichtparametrischen Verfahren sinnvoll sein, qualitative Eigenschaften, wie z.B. Unimodalität oder Monotonie, vorauszusetzen. So ist es naheliegend anzunehmen, daß ein bestimmtes Krankheitsbild häufiger auftritt, wenn ein schädlicher Einfluß bei ansonsten gleichbleibenden Bedingungen zunimmt. Ähnlich dürfte der erwartete Gesamtschaden einer Naturkatastrophe mit deren Schwere steigen. In solchen Fällen interessieren oftmals nicht nur die klassischen Parameter, wie Korrelationskoeffizient oder ein irgendwie definierter „durchschnittlicher“ Anstieg, sondern auch konkretere qualitative oder quantitative Zusammenhänge.

Die Anfänge nichtparametrischer Statistik mit qualitativen Einschränkungen liegen in den fünfziger Jahre. So schlug Grenander 1956 einen Maximum-Likelihood-Schätzer für eine monotone Dichte vor (Grenander [3]). Einen Überblick über die Entwicklung in diesem Bereich gibt z.B. das Vorwort in Robertson, Wright und Dykstra [10].

In der vorliegenden Arbeit wird ein nichtparametrisches Regressionsmodell betrachtet, dessen Beobachtungen den Verteilungen einer einparametrigen Exponentialfamilie $\mathcal{F}$ genügen:

$$Y_i \sim F_\theta \in \mathcal{F} \tag{1.i}$$

Dabei wird die Verbindung zum zugehörigen Designpunkt im einfachen Modell mittels des Erwartungswertes hergestellt:

$$\mathsf{E}Y_i = m(x_i) \quad \text{bzw.} \quad \mathsf{E}(Y_i \,|\, X_i = t) = m(t) \tag{1.ii}$$

In einer Erweiterung dieses Modells wird der (bedingte) Erwartungswert über eine bekannte monotone Funktion G an die gesuchte Regressionsfunktion m gekoppelt.

Die Designpunkte $x_1^{(n)}, \ldots, x_n^{(n)}$ werden als Zahlen zwischen $[0,1]$ angesetzt. Wie (1.ii) erkennen läßt, werden sowohl deterministische Vektoren, genauer äquidistante Punkte zwischen 0 und 1, als auch iid-verteilte Zufallsvariablen mit stetiger Verteilung auf $[0,1]$ als Design betrachtet.

Ziel der vorliegenden Arbeit ist es, entsprechend einem definierten Optimalitätskriterium – dem sog. *geglätteten Log-Likelihood-Funktional* – einen uneingeschränkten Schätzer und einen monotonen Schätzer zu konstruieren. Es werden asymptotische Konvergenzraten der beiden Schätzer gegen die wahre Regressionsfunktion und der Differenz beider gegen Null in der Supremumsnorm hergeleitet. Ebenso wird die asymptotische Konvergenzrate der Ableitung des uneingeschränkten Schätzers gegen die Ableitung der Regressionsfunktion angegeben und bewiesen.

Das *geglättete Log-Likelihood-Funktional* ist definiert als Integral über die geglätteten Log-Likelihoods der einzelnen Beobachtungen. Die Glättung vererbt sich kanonisch auf den uneingeschränkten Schätzer, so daß die gängige Theorie zum Tragen kommt, wie sie z.B. in Green und Silverman [2] oder Hastie und Tibshirani [6] beschrieben ist.

In Kapitel 2 wird das einfache Regressionsmodell definiert und der uneingeschränkte Schätzer in beiden Designs hergeleitet und seine asymptotischen Eigenschaften unter geeigneten Bedingungen bewiesen. Es stellt sich heraus, daß unter den angenommenen Bedingungen der bekannte Nadaraya-Watson-Schätzer der hier gesuchte Schätzer ist.

Das Modell monotoner Regressionsfunktionen, im weiteren der Einfachheit wegen monotones Modell genannt, wird in Kapitel 3 beschrieben, und der monotone Schätzer wird auf der Basis des uneingeschränkten Schätzers konstruiert. Es ist interessant, daß der Umweg über die Kleinste-Quadrate-Theorie zum Ziel führt. Als Nebenergebnis fällt ab, daß unter den allgemein angesetzten Bedingungen der in Kapitel 2 hergeleitete Schätzer auch ein analog definiertes *geglättetes Kleinste-Quadrate-Funktional* maximiert.

Die oben angedeutete Erweiterung des Modells wird in Kapitel 4 behandelt. Die meisten Erkenntnisse übertragen sich analog ohne weitergehende Betrachtungen.

In Kapitel 5 schließlich sind einige Plots präsentiert, die die Ergebnisse aus den Kapiteln 2 und 3 illustrieren.

2 Das einfache Regressionsmodell

In diesem Kapitel wird das einfache Regressionsmodell definiert und betrachtet. Die Modellbeschreibungen in den folgenden Kapiteln beziehen sich weitgehend auf den Ansatz hier und werden dort nur in ihren Abweichungen vom einfachen Modell ausgeführt.

Es wird ein Optimalitätskriterium – das geglättete Log-Likelihood-Funktional – für einen Schätzer der Regressionsfunktion aufgestellt, ein Schätzer, der es erfüllt, hergeleitet, und es werden asymptotische Eigenschaften dieses Schätzers bewiesen.

Die meisten Beweise sind in einem eigenen Abschnitt am Ende des Kapitels zusammengefaßt.

2.1 Definition des Modells

Das in den Voraussetzungen 1 bis 3 beschriebene Modell gilt über das gesamte Kapitel 2. Der Voraussetzung 2 sind einige direkte Folgerungen, die keiner weiteren Erläuterungen bedürfen, angefügt.

1. Für jedes $n \in \mathbb{N}$ seien $x_1^{(n)}, \ldots, x_n^{(n)}$ reelle Zahlen im Intervall $[0, 1]$. Sie können für jedes n

   (a) äquidistant vorgegeben sein oder

   (b) Beobachtungen von unabhängig und identisch verteilten Zufallsvariablen $X_1^{(n)}, \ldots, X_n^{(n)}$ sein, deren Verteilung stetig bezüglich des Lebesgue-Maßes mit der Dichte f ist. Die Dichte soll dabei folgende Eigenschaften erfüllen:

- f ist in $[0,1]$ differenzierbar. Die Ableitung sei Lipschitz-stetig mit der Lipschitz-Konstanten $c_{f'}$.
- $c_* := 1/2 \times \min_{t\in[0,1]} f(t) > 0$ und $c^* := 4 \times \max_{t\in[0,1]} f(t) < \infty$

Die zugehörige Verteilungsfunktion sei mit F bezeichnet.

Es sei $(x_1^{(n)}, \ldots, x_n^{(n)})$ mit $x_\cdot^{(n)}$ und $(X_1^{(n)}, \ldots, X_n^{(n)})$ mit $X_\cdot^{(n)}$ bezeichnet.

Im weiteren sei der Ansatz (a) mit *äquidistantem* oder auch *determinstischem* Design und der Ansatz (b) mit *zufälligem* Design bezeichnet.

2. Für jedes $n \in \mathbb{N}$ seien $Y_1^{(n)}, \ldots, Y_n^{(n)}$ voneinander unabhängige, reellwertige ZGen auf einem gemeinsamen Wahrscheinlichkeitsraum. Im zufälligen Design seien sie zusätzlich auf demselben Raum wie $X_1^{(n)}, \ldots, X_n^{(n)}$ gegeben.

   Es sei $Y_i^{(n)} \sim F_{\theta_i^{(n)}} \in \mathcal{F}$, wobei $\mathcal{F} := \{F_\theta : \theta \in \Theta\}$ für eine konvexe Teilmenge Θ von $\mathbb{R}$ eine Familie von Verteilungen bezüglich des σ-endlichen Maßes μ über $(\mathbb{R}, \mathcal{B})$ ist, deren Dichten durch $f_\theta(y) = e^{y\theta - \beta(\theta)}$ gegeben sind. Dieser Modellansatz und ein paar Beispiele sind z.B. in McCullagh und Nelder [1], Seiten 28ff, beschrieben. In diesem Modell gilt für den Erwartungswert und die Varianz:

   $$\mathsf{E}Y = \beta'(\theta) \quad \text{und} \quad \mathsf{Var}Y = \beta''(\theta)$$

   falls $Y \sim F_\theta$. β' ist umkehrbar. Also kann die Familie $\mathcal{F}$ auch durch die Erwartungswerte ihrer Verteilungen parametrisiert werden.

   Es sei vorausgesetzt, daß

   - $\mu(\mathbb{R}/\beta'(\Theta)) = 0$ gilt.
     Da $\beta'(\Theta)$ konvex ist, impliziert diese technische Bedingung, daß jedes mögliche Mittel von Beobachtungen auch Erwartungswert einer Verteilung dieser Familie ist.

   Es sei zwischen Designpunkten $x_i^{(n)}$ und Beobachtungen $Y_i^{(n)}$ die Beziehung

   $$\mathsf{E}Y_i^{(n)} = m(x_i^{(n)})$$

   im äquidistanten Design und

   $$\mathsf{E}(Y_i^{(n)} | X_i^{(n)} = t) = m(t)$$

   im zufälligen Design angenommen.

   Dabei sei $m : [0,1] \to \beta'(\Theta)$ eine unbekannte Regressionsfunktion. Sie sei differenzierbar und ihre Ableitung Lipschitz-stetig mit der Konstanten $c_{m'}$.

   Aus technischen Gründen muß zusätzlich vorausgesetzt werden, daß eine Umgebung U von $m([0,1])$ existiert, die Teilmenge von $\beta'(\Theta)$ ist.

   Diese Funktion ist zu schätzen.

   In dem beschriebenen Modell gilt:

- β ist zweimal stetig differenzierbar und β' umkehrbar und die Umkehrung ψ von β' ist ebenfalls stetig differenzierbar.
- Es sei $\Sigma := \psi \circ m([0,1]) \subseteq \Theta$. Dieses Σ ist beschränkt und abgeschlossen, da $\psi \circ m$ stetig ist; außerdem enthält Σ keine Randpunkte der natürlichen Parametermenge Θ.

  Dann existieren ein $U_\Sigma < \infty$ und ein $u_\Sigma > 0$, so daß für alle $\theta \in \Sigma$ gilt: $\mathsf{E}e^{u_\Sigma |Y|}$ existiert und ist kleiner als U_Σ, falls $Y \sim F_\theta$.
- β' ist streng monoton wachsend, da im Inneren von Θ gilt: $\beta''(\theta) = \mathsf{Var}Y > 0$.

3. Es sei K ein Glättungskern mit dem Träger $[-1,1]$ und den folgenden Eigenschaften:

   (a) $K: \mathbb{R} \to [0,\infty)$ ist symmetrisch und monoton wachsend links von 0 und fallend rechts von 0.

   (b) $\int_{-1}^{1} K(u)du = 1$

   (c) K sei stetig differenzierbar und die Ableitung sei Lipschitz-stetig mit der Konstanten $c_{K'}$.

   Es sei $K_h(u) := K(u/h)/h$ für alle $h > 0$ und $u \in \mathbb{R}$.

Wenn keine Mehrdeutigkeiten möglich sind, wird im weiteren bei $x_i^{(n)}$, $X_i^{(n)}$ und $Y_i^{(n)}$ oftmals das ${}^{(n)}$ weggelassen.

2.2 Das geglättete Log-Likelihood-Funktional

Die Log-Likelihood-Funktion l_i der ZV Y_i hat die Gestalt

$$l_i(\theta) = Y_i\theta - \beta(\theta),$$

wobei $\theta \in \Theta$ der Parameter der Exponentialfamilie ist.

Um eine Funktion $m : [0,1] \to \beta'(\Theta)$ bei gegebenen Beobachtungen bewerten zu können, werden die $l_1(.), \ldots, l_n(.)$ in jedem $t \in [0,1]$ mit $\theta = \psi(m(t))$ ausgewertet, mit dem Glättungskern K_h gemittelt und über alle t integriert.

Man erhält die folgende Definition:

Definition 2.1 *Für ein $m : [0,1] \to \beta'(\Theta)$ wird das Funktional $\mathcal{L}^s$, das durch*

$$\mathcal{L}^s(m) := \sum_{i=1}^{n} \int_0^1 l_i(\psi(m(t)))K_h(t - x_i)dt$$

definiert ist, als geglättetes Log-Likelihood-Funktional bezeichnet.

Eine Funktion $\hat{m}_h : [0,1] \to \beta'(\Theta)$ wird als Maximum-geglätteter-Likelihood-Schätzer, kurz „MgL-Schätzer", bezeichnet, wenn sie

$$\hat{m}_h \in \underset{m:[0,1]\to\beta'(\Theta)}{\mathit{argmax}} \; \mathcal{L}^s(m)$$

erfüllt.

Anhand der Definition ist zu ersehen, daß mit einem $\hat{m}_h$ auch ein $\tilde{m}_h$ MgL-Schätzer ist, wenn beide sich nur auf einer ν-Nullmenge ($\nu(A) := \sum_{i=1}^n \int_A K_h(t-x_i)dt \quad \forall A \in \mathcal{B}_{[0,1]}$) unterscheiden; speziell, wenn sie sich nur auf einer Lebesgue-Nullmenge unterscheiden. Im folgenden wird zwischen Elementen einer solchen Äquivalenzklasse und der Klasse selbst nicht unterschieden.

2.3 Überlegungen für das zufällige Design

In diesem Abschnitt werden die Grundlagen zur Behandlung des zufälligen Designs bewiesen. Es wird gezeigt, daß es ausreicht, Aussagen über Folgen von Teilmengen aller möglichen Designvektoren zu beweisen, solange die Wahrscheinlichkeiten für diese Teilmengen gegen Eins konvergieren, wenn die Anzahl der Designpunkte groß wird.

Asymptotische Aussagen können also gegebenenfalls unter der Bedingung betrachtet werden, daß der Designvektor in einer solchen Teilmenge liegt. Die Aussage gilt dann asymptotisch auch unbedingt.

Lemma 2.1 *Es seien*

- $H_.$ *eine Folge* $\{H_n\}_{n\in\mathbb{N}}$ *von logischen Zufallsvariablen. Jedes* $H_n(.)$ *hänge von einem* n*-Vektor* $x_.^{(n)} := (x_1^{(n)}, \ldots, x_n^{(n)}) \in [0,1]^n$ *ab.*
- $M_.^a$ *und* $M_.^b$ *seien zwei Folgen* $\{M_n^a\}_{n\in\mathbb{N}}$ *bzw.* $\{M_n^b\}_{n\in\mathbb{N}}$*, wobei jedes* M_n^a*,* M_n^b *eine Menge aus* $\mathcal{B}_{[0,1]^n}^n$ *ist.*

Falls dann

$$P(X_.^{(n)} \in M_n^a) \underset{n\to\infty}{\longrightarrow} 1 \quad \text{und} \quad P(X_.^{(n)} \in M_n^b) \underset{n\to\infty}{\longrightarrow} 1$$

gilt, folgt:

$$\lim_{n\to\infty} P(H_n(X_.^{(n)})|X_.^{(n)} \in M_n^a) = \lim_{n\to\infty} P(H_n(X_.^{(n)})|X_.^{(n)} \in M_n^b)$$

Lemma 2.2 *Es seien eine Folge* $H_.$ *logischer Zufallsvariablen und eine Folge* $M_.^a$ *von Teilmengen von* $[0,1]^n$ *entsprechend Lemma 2.1 gegeben.*

Dann gilt:

$$\lim_{n\to\infty} P(H_n(X_.^{(n)})|X_.^{(n)} \in M_n^a) = \lim_{n\to\infty} P(H_n(X_.^{(n)}))$$

Lemma 2.3 *Seien* M_n^a *und* M_n^b *wie in Lemma 2.1 mit* $M_n^b \subseteq M_n^a$ *für alle* $n \in \mathbb{N}$*. Ferner sei für jedes* $\eta < \infty$ *eine Folge* $H_.^\eta$ *entsprechend Lemma 2.1 gegeben und es gelte:*

$$\forall q > 0 \;\; \exists \eta_q < \infty : \;\; P(H_n^{\eta_q}(X_.^{(n)})|X_.^{(n)} \in M_n^a) = O(n^{-q})$$

Dann gilt:

$$\forall q > 0 : \;\; P(H_n^{\eta_q}(X_.^{(n)})|X_.^{(n)} \in M_n^b) = O(n^{-q})$$

Lemma 2.4 *Es seien $M^j_\cdot$ $(j = a, b)$ zwei Mengenfolgen, die die Bedingungen des Lemmas 2.1 erfüllen.*

Dann erfüllt auch die Mengenfolge $M^c_\cdot := \{M^a_n \cap M^b_n\}_{n\in\mathbb{N}}$ diese Bedingungen.

Lemma 2.5 *Die Mengenfolgen $M^j_\cdot := \{M^j_n\}_{n\in\mathbb{N}}$ $\quad j = 1, 2, 3$ erfüllen die Bedingung $P(X^{(n)}_\cdot \in M^j_n) \underset{n\to\infty}{\longrightarrow} 1$ des Lemmas 2.1:*

1. *Es sei $h = h_n \underset{n\to\infty}{\longrightarrow} 0$ so, daß $nh^2_n \underset{n\to\infty}{\longrightarrow} \infty$.*

$$\begin{aligned} M^1_n := \{&x^{(n)}_\cdot \in [0,1]^n |\ \forall t \in [0,1] \\ &\#\{i \in \{1,\dots,n\}|\ x^{(n)}_i \in (t-h_n, t+h_n)\} \in [b_* n h_n, b^* n h_n]\}, \end{aligned}$$

wobei $b_* := c_*/2$ und $b^* := 3c^*/2$.

2. *h wie in 1.*

$$\begin{aligned} M^2_n := \{&x^{(n)}_\cdot \in [0,1]^n |\ \forall t \in [0,1] \\ &\#\{i \in \{1,\dots,n\}|\ x^{(n)}_i \in (t-\frac{h_n}{2}, t+\frac{h_n}{2})\} \in [\frac{b_* n h_n}{2}, \frac{b^* n h_n}{2}]\} \end{aligned}$$

3. *$h = h_n \underset{n\to\infty}{\longrightarrow} 0$ sei so, daß für ein $\alpha > 0$ gilt: $nh_n \geq n^\alpha$*

$$M^3_n := \{x^{(n)}_\cdot \in [0,1]^n |\ \forall t \in [0,1]\ \exists i \in \{1,\dots,n\}:\ x^{(n)}_i \in (t-\frac{h_n}{2}, t+\frac{h_n}{2})\}$$

Diese drei Mengenfolgen werden im weiteren angewendet.

Ein Designvektor aus M^1_n stellt sicher, daß für jedes $t \in [0,1]$ die Anzahl der Designpunkte, die in der h-Umgebung von t liegen, nach oben und nach unten entsprechend beschränkt ist. Speziell ist damit der Term $K_h(t-x_i)$ nur für eben diese Anzahl von Designpunkten positiv.

Ein Designvektor aus M^2_n ist analog einem aus M^1_n, jedoch wird hier die Anzahl der Designpunkte in der $h/2$-Umgebung von t festgelegt. Für diese Punkte ist der Term $K_h(t-x_i)$ größer als $K(1/2)$ und damit gleichmäßig größer als Null.

Ein Designvektor aus M^3_n stellt lediglich sicher, daß für jedes $t \in [0,1]$ ein Designpunkt in der $h/2$-Umgebung von t liegt. Dann ist $\sum_{i=1}^n K_h(t-x_i) \geq K(1/2) \quad \forall t \in [0,1]$. Wenn die Voraussetzungen von 2. erfüllt sind, gilt natürlich $M^2_n \subseteq M^3_n$, jedoch sind die Voraussetzungen von 3. etwas schwächer.

2.4 Der MgL-Schätzer der Regressionsfunktion

Entsprechend Definition 2.1 maximiert der MgL-Schätzer $\hat{m}_h$ der Regressionfunktion m das Funktional $\mathcal{L}^s$. Der folgende Satz gibt die Gestalt dieses Schätzers explizit an und trifft Aussagen über seine Existenz.

Satz 2.6 *Es sei $h = h_n \underset{n\to\infty}{\longrightarrow} 0$ so, daß $nh \underset{n\to\infty}{\longrightarrow} \infty$ im determistischen Design, bzw. so, daß für ein $\alpha > 0$ im zufälligen Design $nh \geq n^\alpha$ gilt.*

Dann ist die Funktion

$$\hat{m}_h(t) := \frac{\sum_{i=1}^n Y_i K_h(t-x_i)}{\sum_{i=1}^n K_h(t-x_i)} \quad \text{für alle } t \in [0,1]$$

in beiden Fällen der MgL-Schätzer der Regressionsfunktion m.

Im deterministischen Modell ist der MgL-Schätzer für genügend großes n wohldefiniert. Im zufälligen Modell geht die Wahrscheinlichkeit, daß er wohldefiniert ist, für $n \to \infty$ gegen 1.

Wenn der MgL-Schätzer wohldefiniert ist, so ist er bis auf Nullmengen bezüglich des Lebesgue-Maßes eindeutig.

Es stellt sich also heraus, daß der MgL-Schätzer gleich dem bekannten Nadaraya-Watson-Schätzer ist.

Im weiteren werden die folgenden Bezeichnungen gebraucht:

- $\vartheta : [0,1] \to \mathbb{R}$ mit $\vartheta(t) := m(t)f(t)$
- $\vartheta_h : [0,1] \to \mathbb{R}$ mit $\vartheta_h(t) := n^{-1}\sum_{i=1}^n m(x_i)K_h(t-x_i)$
- $\hat{\vartheta}_h : [0,1] \to \mathbb{R}$ mit $\hat{\vartheta}_h(t) := n^{-1}\sum_{i=1}^n Y_i K_h(t-x_i)$
- $\hat{f}_h : [0,1] \to \mathbb{R}$ mit $\hat{f}_h(t) := n^{-1}\sum_{i=1}^n K_h(t-x_i)$
- $m_h : [0,1] \to \mathbb{R}$ mit $m_h(t) := n^{-1}\sum_{i=1}^n m(t)K_h(t-x_i)$

Mit diesen Bezeichnungen gilt: $\hat{m}_h(t) = \hat{\vartheta}_h(t)/\hat{f}_h(t)$.

$\hat{f}_h(t)$ kann als Schätzer der Dichte f im zufälligen Design interpretiert werden. Im äquidistanten Design oder im zufälligen Design mit $f \equiv 1$ ist $\hat{\vartheta}_h$ ein etwas von $\hat{m}_h$ abweichender Schätzer von m, der aber asymptotisch ähnliche Eigenschaften hat. ϑ, ϑ_h, $\hat{\vartheta}_h$, m_h und $\hat{f}_h$ werden hier jedoch nicht weiter betrachtet, sondern nur als Hilfskonstrukte benutzt.

2.5 Punktweise asymptotische Eigenschaften des MgL-Schätzers

In diesem Abschnitt sollen punktweise asymptotische Eigenschaften des MgL-Schätzers bewiesen werden. Darauf aufbauend werden im Abschnitt 2.6 globale Eigenschaften des Schätzers hergeleitet.

Es sei $h = h_n \underset{n\to\infty}{\longrightarrow} 0$ so gewählt, daß $nh^2 \underset{n\to\infty}{\longrightarrow} 0$.

Sei $t \in (0,1)$ beliebig fest. Alle Eigenschaften werden im folgenden Sinne gleichmäßig in $t \in (0,1)$ bewiesen: Für alle $n \in \mathbb{N}$ so, daß $t \in (h_n, 1-h_n)$ gilt, sind Konstanten etc. unabhängig von t und n.

Soweit es möglich ist, werden zufälliges und deterministisches Design gemeinsam behandelt. Es ist zu beachten, daß ein Vektor $x_{\cdot}^{(n)}$ äquidistanter Designpunkte für genügend großes n in jeder der in Lemma 2.5 definierten Mengen M_n^i liegt. Dieses n hängt nur von der Wahl von $\{h_n\}_{n\in\mathbb{N}}$ ab.

Eine ausführlichere Angabe der punktweisen asymptotischen Eigenschaften des Nadaraya-Watson-Schätzers einschließlich der Konstanten im asymptotisch dominierenden Term und eine Implementation in **S** zur Berechnung des Schätzers findet sich z.B. in Härdle [5].

Lemma 2.7 *Es gilt:*

$$|\hat{\vartheta}_h(t) - \vartheta_h(t)| = O_P((nh)^{-1/2}).$$

Der Term $\hat{\vartheta}_h(t) - \vartheta_h(t)$ hat Erwartungswert Null und kann somit als „Varianzanteil" des Zählers im MgL-Schätzer angesehen werden.

In den beiden folgenden Lemmata soll der „Biasanteil" $\vartheta_h(t) - m_h(t)$ des Zählers im MgL-Schätzer betrachtet werden. Dieser Term ist nur im deterministischen Modell nicht zufällig und deshalb nur dort der Bias. Im zufälligen Design ist er der bedingte Bias unter der Bedingung eines fest gewählten Designvektors $x_\cdot^{(n)}$.

Lemma 2.8 *Im deterministischen Design gilt:*

$$|\vartheta_h(t) - m_h(t)| = O(n^{-1}) + O(h^2).$$

Lemma 2.9 *Im zufälligen Modell gilt:*

$$|\vartheta_h(t) - m_h(t)| = O_P(h^2) + O_P((h/n)^{1/2}).$$

Satz 2.10 *Für $h_n := c_h n^{-1/5}$ gilt im zufälligen und im determistischen Design:*

$$|\hat{m}_h(t) - m(t)| = O_P(n^{-2/5}).$$

$h_n = c_h n^{-1/5}$ ist die Bandbreite, bei der Standardabweichung und Erwartungswert von $|\hat{m}_h(t) - m(t)|$ mit derselben Rate gegen Null konvergieren.

In Hastie und Tibshirani [6] ist das Problem des Ausgleiches von Varianz und Bias ausführlich dargestellt. Betrachtungen zum asymptotischen Optimalitätsbegriff können in Korostelev und Tsybakov [8] nachgelesen werden.

Beweis: Es gilt:

$$|\hat{m}_h(t) - m(t)| \leq \frac{1}{\hat{f}_h(t)} \times [|\hat{\vartheta}_h(t) - \vartheta_h(t)| + |\vartheta_h(t) - m_h(t)|]$$

Entsprechend Lemma 2.7 ist der linke Summand in der eckigen Klammer ein $O_P((nh)^{-1/2})$; der rechte Summand ist im deterministischen Design wegen Lemma 2.8 ein $O(n^{-1}) + O(h^2)$ und im zufälligen Design wegen Lemma 2.9 ein $O_P((h/n)^{1/2}) + O_P(h^2)$.

Damit ist in beiden Designs die Summe in der eckigen Klammer von der Größenordnung $O_P((nh)^{-1/2}) + O_P(h^2)$. Das gleicht sich für $h := c_h n^{-1/5}$ aus. Die Konvergenzrate für die Summe in eckigen Klammern ist dann $n^{-2/5}$.

Im zufälligen Design folgt mit Lemma 2.2 und Lemma 2.5, daß

$$|\hat{\vartheta}_h(t) - \vartheta_h(t)| + |\vartheta_h(t) - m_h(t)|$$

auch unter der Bedingung $X_\cdot^{(n)} \in M_n^2$ in Wahrscheinlichkeit mit der Rate $n^{-2/5}$ konvergiert.

Unter $X_\cdot \in M_n^2$ gilt aber auch:

$$\frac{1}{\hat{f}_h(t)} \in \left[\frac{2}{b^* K(0)}, \frac{2}{b_* K(1/2)}\right],$$

wobei $K(1/2) > 0$ nach Voraussetzung. Damit folgt die Behauptung, wenn nochmals Lemma 2.2 angewendet wird.

Im deterministischen Design gilt die Behauptung analog wegen Lemma 2.8 und da $x_\cdot^{(n)} \in M_n^2$ für n genügend groß.

□

Satz 2.11 *Seien $t_0 \in (0,1)$ und $h_n := c_h n^{-1/5}$ für ein $c_h > 0$.*

Dann gilt im determinstischen Design:

$$n^{2/5}(\sigma_{t_0})^{-1}(\hat{m}_h(t_0) - \mathsf{E}\hat{m}_h(t_0)) \xrightarrow{L} \Phi,$$

wobei σ_{t_0} eine positive Konstante und Φ die Verteilungsfunktion der Standardnormalverteilung ist.

2.6 Globale asymptotische Eigenschaften des MgL-Schätzers

In diesem Abschnitt werden globale asymptotischen Eigenschaften des MgL-Schätzers der Regressionsfunktion bewiesen. Dabei wird die Bandbreitenwahl $h_n := c_h n^{-1/5}$ zugrunde gelegt, die im Satz 2.10 als günstig hergeleitet wurde.

Satz 2.15 gibt die Rate, mit der der MgL-Schätzer gleichmäßig gegen die Regressionsfunktion konvergiert, an. Weitergehende Betrachtungen zum asyptotischen Optimalitätsbegriff können in Korostelev und Tsybakov [8] nachgelesen werden.

Lemma 2.12 *Sei $\{m_n\}_{n\in\mathbb{N}}$ eine Folge in $\mathbb{N}$, so daß $m_n \geq n \ \forall n \in \mathbb{N}$, und für jedes n seien $Z_i^{(n)}$ mit $i = 1, \ldots, m_n$ reellwertige Zufallsgrößen auf einem gemeinsamen Wahrscheinlichkeitsraum. Sei $\{a_n\}_{n\in\mathbb{N}}$ eine reelle Zahlenfolge und c eine Konstante und es gelte:*

$$\exists \zeta < \infty \ \exists n_0 \in \mathbb{N} \ \forall \eta \geq \zeta \ \forall n \geq n_0 \ \forall i = 1, \ldots, m_n : \ P(|Z_i^{(n)}| \geq a_n \eta) \leq c e^{-\eta}$$

Dann folgt: $\sup_{i=1,\ldots,m_n} |Z_i^{(n)}| = O_P(a_n \ln m_n)$

Korollar 2.13 *Es gilt:*

$$\sup_{i=1,\ldots,n} |Y_i^{(n)}| = O_P(\ln n)$$

Wenn $h = h_n \underset{n\to\infty}{\longrightarrow} 0$ *so ist, daß für ein* $\alpha > 0$ *gilt* $nh \geq n^\alpha$*, dann gilt außerdem:*

$$\sup_{t\in[0,1]} |\hat{m}_h(t)| \leq O_P(\ln n) \quad \textit{und} \quad \sup_{t\in[0,1]} |\hat{m}'_h(t)| \leq O_P(h^{-1} n \ln n)$$

Lemma 2.14 *Sei* $h_n = c_h n^{-1/5}$ *für ein* $c_h > 0$ *und sei für ein* $\alpha > 0$ *die Folge* $\{m_n\}_{n\in\mathbb{N}}$ *so, daß* $m_n = n^\alpha$ *für alle* $n \in \mathbb{N}$*. Für jedes* n *seien* $t_i^{(n)}$, $i = 1, \ldots, m_n$ *Werte aus* $[h_n, 1 - h_n]$.

Dann gilt:

$$\sup_{i=1,\ldots,m_n} |\hat{m}_h(t_i) - m(t_i)| = O_P(n^{-2/5}\sqrt{\ln n})$$

Satz 2.15 *Sei* $h_n := c_h n^{-1/5}$ *für ein* $c_h > 0$.

Dann gilt:

$$\sup_{t\in[h_n,1-h_n]} |\hat{m}_h(t) - m(t)| = O_P(n^{-2/5}\sqrt{\ln n})$$

Beweis: Sei I_l ein Gitter auf $[h_n, 1 - h_n]$ mit der Gitterweite $\delta_l := l^{-1}$ und sei $l = l_n := n^2$. Dann gilt:

$$\begin{aligned}\sup_{t\in[h_n,1-h_n]} |\hat{m}_h(t) - m(t)| \;\leq\; & \sup_{y\in I_l} |\hat{m}_h(y) - m(y)| \;+ \\ & \sup_{t\in[h_n,1-h_n]} \inf_{y\in I_l} |\hat{m}_h(t) - m(t) - (\hat{m}_h(y) - m(y))|\end{aligned}$$

Nach Lemma 2.14 ist der linke Summand ein $O_P(n^{-2/5}\sqrt{\ln n})$. Der rechte Summand läßt sich durch

$$\begin{aligned} & \delta_l \times \sup_{t\in[h_n,1-h_n]} |\hat{m}'_h(t) - m'(t)| \\ \leq\; & \delta_l \times \sup_{t\in[h_n,1-h_n]} |\hat{m}'_h(t)| + \delta_l \times \sup_{t\in[h_n,1-h_n]} |m'(t)| \\ \leq\; & \delta_l O_P(h^{-1} n \ln n), \\ & \text{wegen Korollar 2.13 und da } m'(t) \text{ beschränkt ist} \\ =\; & O_P((hl)^{-1} n \ln n) = O_P(n^{-4/5} \ln n)\end{aligned}$$

abschätzen.

Damit ist die Summe von der gewünschten Ordnung $O_P(n^{-2/5}\sqrt{\ln n})$.

□

2.7 Die Ableitung des MgL-Schätzers der Regressionsfunktion

In diesem Abschnitt soll die Ableitung des Schätzers $\hat{m}_h$ nach t betrachtet werden. Auf die asymptotischen Eigenschaften der Funktion selbst aufbauend werden asymptotische Eigenschaften ihrer Ableitung bewiesen.

$|\hat{m}_h'(t) - m'(t)|$ ist kleiner als die Summe der Beträge des im folgenden definierten Termes T_n und eines weiteren Termes. Es soll zuerst eine globale Asymptotik von $|T_n|$ hergeleitet werden, bevor das globale asymptotische Verhalten von $|\hat{m}_h'(t) - m'(t)|$ betrachtet wird.

Es sei vereinbart:

$$T_n(t) := \frac{\frac{d}{dt}(m_h(t) - \hat{\vartheta}_h(t))}{\hat{f}_h(t)}$$

Lemma 2.16 *Sei $h_n = c_h n^{-1/5}$ für ein $c_h > 0$ und sei für ein $\alpha > 0$ die Folge $\{m_n\}_{n \in \mathbb{N}}$ so, daß $m_n = n^\alpha$ für alle $n \in \mathbb{N}$. Für jedes n seien $t_i^{(n)}$, $i = 1, \ldots, m_n$ Werte aus $[h_n, 1 - h_n]$.*

Dann gilt:

$$\sup_{i=1,\ldots,m_n} |T_n(t_i)| = O_P(n^{-1/5}\sqrt{\ln n})$$

Lemma 2.17 *Wenn h_n so ist, daß für ein $\alpha > 0$ gilt $nh \geq n^\alpha$, so ist die Lipschitz-Konstante c_{T_n} von T_n ein $O_P(h^{-2} n^2 \ln n)$.*

Lemma 2.18 *Sei $h_n := c_h n^{-1/5}$ für ein $c_h > 0$. Dann gilt:*

$$\sup_{t \in [h_n, 1-h_n]} |T_n(t)| = O_P(n^{-1/5}\sqrt{\ln n})$$

Lemma 2.19 *Es gilt:*

$$\sup_{u \in [0,1]} \frac{\sum_{i=1}^n K_h'(u - X_i)}{\sum_{i=1}^n K_h(u - X_i)} = O_P(1)$$

im zufälligen Design. Im äquidistanten Design ist es ein $O(1)$.

Satz 2.20 *Sei $h_n := c_h n^{-1/5}$ für ein $c_h > 0$. Dann gilt:*

$$\sup_{t \in [h_n, 1-h_n]} |\hat{m}_h'(t) - m'(t)| = O_P(n^{-1/5}\sqrt{\ln n})$$

Beweis: Es gilt:

$$\hat{m}_h'(t) - m'(t) = \frac{\sum_{i=1}^n (m(t) - \hat{m}_h(t)) K_h'(t - x_i)}{\sum_{i=1}^n K_h(t - x_i)} - T_n(t)$$

Das Supremum des Betrages des zweiten Summanden hat nach Lemma 2.18 die geforderte Ordnung.

Das Supremum des Betrages des ersten Summanden kann nach oben durch

$$\sup_{t \in [h_n, 1-h_n]} |\hat{m}_h(t) - m(t)| \quad \times \sup_{t \in [0,1]} \left| \frac{\sum_{i=1}^n K_h'(t - x_i)}{\sum_{i=1}^n K_h(t - x_i)} \right|$$

abgeschätzt werden. Der linke Faktor ist nach Satz 2.15 ein $O_P(n^{-2/5}\sqrt{\ln n})$ und der rechte Faktor ist nach Lemma 2.19 ein $O_P(1)$ im zufälligen und ein $O(1)$ im deterministischen Design.

Damit folgt die Behauptung.

□

2.8 Beweise

Es folgen nun Beweise aus Kapitel 2, die zu technisch sind oder deren Ausführung für die gesamte Arbeit von untergeordnetem Interesse sind.

Beweis von Lemma 2.1:

Es gilt:

$$\begin{aligned}
\lim_{n\to\infty} P(H_n(X_\cdot^{(n)}) | X_\cdot^{(n)} \in M_n^a) &= \frac{\lim\limits_{n\to\infty} P(H_n(X_\cdot^{(n)}), X_\cdot^{(n)} \in M_n^a)}{\lim\limits_{n\to\infty} P(X_\cdot^{(n)} \in M_n^a)} \\
&= \frac{\lim\limits_{n\to\infty} P(H_n(X_\cdot^{(n)}), X_\cdot^{(n)} \in M_n^b)}{\lim\limits_{n\to\infty} P(X_\cdot^{(n)} \in M_n^b)} \\
&= \lim_{n\to\infty} P(H_n(X_\cdot^{(n)}) | X_\cdot^{(n)} \in M_n^b),
\end{aligned}$$

da $\lim\limits_{n\to\infty} P(X_\cdot^{(n)} \in M_n^i) = 1$ für $i = a, b$ nach Voraussetzung.

□

Beweis von Lemma 2.2:

Die Behauptung folgt aus Lemma 2.1 mit $M_n^b := [0,1]^n$.

□

Beweis von Lemma 2.3:
Sei $c_n^a := P(X_\cdot^{(n)} \in M_n^a)$, $c_n^b := P(X_\cdot^{(n)} \in M_n^b) \underset{n\to\infty}{\longrightarrow} 1$.
Dann gilt:

$$\begin{aligned} P(H_n^{\eta_q}(X_\cdot^{(n)})|X_\cdot^{(n)} \in M_n^b) &= \frac{\int_{M_n^b} P(H_n^{\eta_q}(X_\cdot^{(n)})|X_\cdot^{(n)} = x_\cdot^{(n)})dF^n(x_\cdot^{(n)})}{c_n^b} \\ &\leq \frac{\int_{M_n^a} P(H_n^{\eta_q}(X_\cdot^{(n)})|X_\cdot^{(n)} = x_\cdot^{(n)})dF^n(x_\cdot^{(n)})}{c_n^a c_n^b} \\ &= \frac{P(H_n^{\eta_q}(X_\cdot^{(n)})|X_\cdot^{(n)} \in M_n^a)}{c_n^b} \\ &= (c_n^b)^{-1} O(n^{-q}) = O(n^{-q}), \end{aligned}$$

da $c_n^b \underset{n\to\infty}{\longrightarrow} 1$.

□

Beweis von Lemma 2.4:
Folgt mit der Ungleichung von Bonferroni.

□

Beweis von Lemma 2.5:

1. Es wird zuerst bewiesen, daß für jedes $t_0 \in [0,1]$ die Menge M_n^{1,t_0}, die durch

   $$\{x_\cdot^{(n)} \in [0,1]^n | \#\{i \in \{1,\dots,n\} | x_i^{(n)} \in (t_0 - h, t_0 + h)\} \in [c_* nh, c^* nh]\}$$

   definiert ist, die Eigenschaft $P(X_\cdot^{(n)} \in\!| \; M_n^{1,t_0}) \leq \tilde{c}/(nh)$ für ein von t_0 unabhängiges $\tilde{c}$ erfüllt. Damit kann dann die Aussage des Lemmas gezeigt werden.

   Es sei $p_h := P(X_1 \in (t_0 - h, t_0 + h)) = \int_{t_0-h}^{t_0+h} f(u)du$. Damit folgt:

   $$\frac{p_h}{h} = \frac{1}{h}\int_{t_0-h}^{t_0} f(u)du + \frac{1}{h}\int_{t_0}^{t_0+h} f(u)du$$

   Also wird p_h/h durch $2c_*$ nach unten und durch $c^*/2$ nach oben beschränkt.

   Nun gilt: $Z := \#\{i \in \{1,\dots,n\}| \; x_i^{(n)} \in (t_0 - h_n, t_0 + h_n)\} \sim \mathrm{Bi}(n, p_h)$ und damit $\mathsf{E}Z = np_h$ und $\mathsf{Var}Z = np_h(1-p_h)$. Mit $\varepsilon := c_* nh$ und der Tschebyscheff-Ungleichung erhält man:

   $$P(X_\cdot^{(n)} \notin M_n^{1,t_0}) \leq P(|Z - \mathsf{E}Z| \geq \varepsilon) \leq \frac{\mathsf{Var}Z}{\varepsilon^2} \leq \frac{c^*}{2c_*^2 nh}$$

   Sei $l_n := [2/h]$ und seien nun Stützstellen $t_1, \dots, t_{l_n}$ äquidistant mit dem Abstand $h/2$ über das Intervall $[0,1]$ verteilt ($[x]$ sei für eine reelle Zahl x die größte ganze Zahl, die kleiner oder gleich x ist). Dann gibt es für ein beliebiges $t \in [0,1]$ ein j, so daß $(t_j - h/2, t_j + h/2) \subseteq (t-h, t+h)$ und $(t-h, t+h) \subseteq (t_j - 3h/2, t_j + 3h/2)$ gilt. Damit sind in $(t-h, t+h)$ wenigstens $b_* nh$ und höchstens $b^* nh$ Designpunkte, falls $x_\cdot^{(n)} \in M_n^{1,t_j}$. Also ist $\bigcap_{j=1}^{l_n} M_n^{1,t_j} \subseteq M_n^1$. Nun kann mit der Bonferroni-Ungleichung $P(X_\cdot^{(n)} \in M_n^1)$ folgendermaßen abgeschätzt werden:

$$\begin{aligned} P(X_\cdot^{(n)} \in M_n^1) \;&\geq\; P(X_\cdot^{(n)} \in \bigcap_{j=1}^{l_n} M_n^{1,t_j}) \geq 1 - \sum_{j=1}^{l_n} P(X_\cdot^{(n)} \notin M_n^{1,t_j}) \\ &\geq\; 1 - l_n \max_{j=1,\ldots,l_n} P(X_\cdot^{(n)} \notin M_n^{1,t_j}) \geq 1 - \frac{\tilde{c}}{2nh^2} \end{aligned}$$

was nach Voraussetzung gegen Eins konvergiert.

2. Folgt direkt aus 1.

3. Sei $t_j := jh/4$, mit $j = 1, \ldots, l_n$ und $l_n := 4[h_n^{-1}]$ und sei $A_j := \{\omega \in \Omega |\ \forall i = 1, \ldots, n : \ X_i(\omega) \notin (t_j - h/2, t_j + h/2)\}$.
   Für $p_{h,t_j} := P(X_1 \in (t_j - h/2, t_j + h/2))$ gilt:

$$\frac{p_{h,t_j}}{h} = \frac{\int_{t_j-h/2}^{t_j+h/2} f(u)du}{h} \geq \frac{c_*}{2} > 0$$

Außerdem gilt: $\{\omega \in \Omega |\ X_\cdot(\omega) \notin M_n^3\} \subseteq \bigcup_{j=1}^{l_n} A_j$ und $P(A_j) = (1 - p_{h,t_j})^n$ für alle j.

Damit gilt wegen der Bonferroni-Ungleichung:

$$\begin{aligned} P(X_\cdot \notin M_n^3) \;&\leq\; P(\bigcup_{j=1}^{l_n} A_j) \leq \sum_{j=1}^{l_n} (1 - p_{h,t_j})^n \\ &\leq\; \sum_{j=1}^{l_n} (1 - \frac{hc_*}{2})^n = l_n (1 - \frac{hc_*}{2})^n \leq l_n \left(\left(1 - \frac{2c_*}{l_n}\right)^{\frac{l_n}{4}} \right)^{nh} \\ &\leq\; l_n q^{nh} \quad \text{für ein } q < 1 \text{ und } n \text{ groß} \\ &\leq\; 4nq^{n^\alpha} \underset{n\to\infty}{\longrightarrow} 0 \end{aligned}$$

□

Beweis von Satz 2.6:

Die Funktion $\hat{m}_h$ ist wohldefiniert, falls

$$\sum_{i=1}^{n} K_h(t - x_i) > 0 \quad \forall t \in [0,1] \tag{2.i}$$

Das ist erfüllt, falls $\forall t \in [0,1]\ \exists i \in \{1, \ldots, n\} : \ x_i \in (t - h/2, t + h/2)$.

Das wird im äquidistanten Design für n genügend groß wegen $nh \to \infty$ immer erfüllt.

Im zufälligen Design läßt sich 3. von Lemma 2.5 anwenden: Sei $H_n(x_.) := \{\exists t \in [0,1] : \sum_{i=1}^n K_h(t - x_i) = 0\}$. Dann umfaßt M_n^3 alle Vektoren $x_.$, für die H_n falsch ist. Folglich ist $P(H_n(X_.)|X_. \in M_n^3) = 0 \ \forall n \in \mathbb{N}$. Mit Lemma 2.2 folgt: $\lim\limits_{n\to\infty} P(H_n(X_.)) = \lim\limits_{n\to\infty} P(H_n(X_.)|X_. \in M_n^3) = 0$.

Um $\hat{m}_h$ zu bestimmen genügt es, den Integranden der geglätteten Log-Likelihood-Funktion, also $\sum_{i=1}^n l_i(\psi(m(t)))K_h(t-x_i)$, punktweise, d.h. für jedes $t \in [0,1]$ und jedes $n \in \mathbb{N}$, zu maximieren. Da die Abbildung ψ außerdem umkehrbar und differenzierbar ist, reicht es aus, für jedes $t \in [0,1]$ die Summe $\sum_{i=1}^n l_i(\theta_t)K_h(t - x_i)$ bezüglich θ_t zu maximieren. Das optimale $\hat{m}_h(t)$ ergibt sich dann als $\hat{m}_h(t) := \beta'(\hat{\theta}_t)$.

Leitet man die Summe nach θ_t ab, setzt die Ableitung gleich Null und stellt entsprechend um, so erhält man

$$\hat{m}_h(t) = \beta'(\hat{\theta}_t) = \frac{\sum_{i=1}^n Y_i K_h(t - x_i)}{\sum_{i=1}^n K_h(t - x_i)}$$

als einzige Lösung. Da die zweite Ableitung der Summe nach θ_t gleich

$$-\sum_{i=1}^n \beta''(\theta_t) K_h(t - x_i) = -\sum_{i=1}^n K_h(t - x_i) \mathsf{Var} Z$$

für ein $Z \sim F_{\theta_t}$ und damit negativ ist, falls nur Bedingung (2.i) erfüllt ist, ist die Summe selbst konkav in $m(t)$ und wird damit von $\hat{m}_h(t)$ und nur von $\hat{m}_h(t)$ maximiert.

Folglich maximiert $\hat{m}_h$ die geglättete Log-Likelihood-Funktion $\mathcal{L}^s$.

Ferner kann eine Funktion, die auf einer Lebesgue-Nullmenge von $\hat{m}_h$ verschieden ist, $\mathcal{L}^s$ nicht auch maximieren, falls Bedingung (2.i) erfüllt ist.

$\square$

Beweis von Lemma 2.7:

1. Zufälliges Design
   Da $\mathsf{E}(\hat{\vartheta}_h(t) - \vartheta_h(t)) = 0$, genügt es wegen der Tschebyscheff-Ungleichung zu zeigen, daß gilt: $\mathsf{Var}(\hat{\vartheta}_h(t) - \vartheta_h(t)) = O((nh)^{-1})$. Die einzelnen Summanden sind nach Voraussetzung unabhängig, also gilt:

   $$\mathsf{Var}(\hat{\vartheta}_h(t) - \vartheta_h(t)) = (nh)^{-2} \sum_{i=1}^n \mathsf{Var}[(Y_i - m(X_i))K((t - X_i)/h)]$$

   Wegen $\mathsf{Var}Z = \mathsf{E}\mathsf{Var}(Z|Y) + \mathsf{Var}\mathsf{E}(Z|Y)$ und da im vorliegenden Fall der rechte Summand Null ist, kann die gesuchte Varianz wie folgt abgeschätzt werden:

   $$\begin{aligned}
   & (nh^2)^{-1}\mathsf{E}\{\mathsf{Var}[(Y_1 - m(X_1))K((t - X_1)/h)|X_1 = T]\} \\
   \le\ & (nh^2)^{-1}\mathsf{E}\{(K((t - T)/h))^2 \mathsf{Var}[(Y_1 - m(X_1))|X_1 = T]\} \\
   =\ & (nh^2)^{-1}\mathsf{E}\{(K((t - T)/h))^2 \beta''(m(T))\} \\
   \le\ & (nh^2)^{-1} \sup_{t_0 \in [0,1]} \beta''(m(t_0))(K(0))^2 \int_{t-h}^{t+h} f(u)du \\
   =\ & O((nh)^{-1})
   \end{aligned}$$

2. Äquidistantes Design
   Im äquidistanten Design gilt:

$$\mathsf{Var}((Y_i - m(x_i))K((t - x_i)/h)) \le \beta''(m(x_i))(K(0))^2 = O(1)$$

   Da aber für jedes $t \in [0,1]$ und genügend großes n nur $2nh$ Summanden in $(\hat{\vartheta}_h(t) - \vartheta_h(t))$ ungleich Null sind, läßt sich die Varianz dieser Zufallsgröße mittels

$$(nh)^{-2} \sum_{i=1}^{n} \mathsf{Var}((Y_i - m(x_i))K((t - x_i)/h) \le (nh)^{-2} nhO(1) = O((nh)^{-1})$$

   abschätzen. Wie im zufälligen Design folgt die Behauptung mit der Tschebyscheff-Ungleichung.

□

Beweis von Lemma 2.8:

Da für genügend großes n nur $2nh$ Summanden in die Summe eingehen, und in diesen Summanden $|x_i - t| \le h$ gilt, folgt:

$$\begin{aligned}
& \left| \frac{1}{nh} \sum_{i=1}^{n} (m(x_i) - m(t)) K\left(\frac{t - x_i}{h}\right) \right| \\
\le\ & \frac{|m'(t)|}{nh} \left| \sum_{i=1}^{n} (x_i - t) K\left(\frac{t - x_i}{h}\right) \right| + 2c_{m'} K(0) h^2 \\
& (m \text{ wird in } t \text{ mit quadratischem Restglied entwickelt}) \\
\le\ & \frac{|m'(t)|}{nh} \left| \sum_{j=-[nh]}^{[nh]} \frac{j}{n} K\left(\frac{j}{nh}\right) \right| + \frac{|m'(t)|K(0)}{nh} \sum_{j=-[nh]}^{[nh]} \frac{1}{n} + 2c_{m'} K(0) h^2 \\
\le\ & 0 + \frac{2|m'(t)|K(0)}{n} + 2c_{m'} K(0) h^2 \\
& (\text{da } K \text{ symmetrisch ist}) \\
=\ & O(n^{-1}) + O(h^2)
\end{aligned}$$

□

Beweis von Lemma 2.9:

Der Betrag des Erwartungswert der ZG $(nh)^{-1} \sum_{i=1}^{n} (m(X_i) - m(t))K((t - X_i)/h)$ ist gleich $h^{-1} | \int_0^1 (m(u) - m(t))K((t - u)/h) f(u) du|$. Da für großes n die h-Umgebung von t innerhalb $(0,1)$ liegt, können $t - h$ und $t + h$ als Integrationsgrenzen angesetzt werden. Setzt man ferner $\rho(u) := (m(u) - m(t))f(u)$ und entwickelt diese Funktion bei t mit quadratischem Restglied, so ergibt sich für eine Konstante C_0, die von $c_{m'}$ und $c_{f'}$ abhängt, die folgende Abschätzung für das obige Integral:

$$\frac{1}{h}\left|\rho'(t)\int_{t-h}^{t+h}(u-t)K\left(\frac{t-u}{h}\right)du\right| + \frac{C_0K(0)}{2h}\int_{-h}^{h}u^2du$$

Da K symmetrisch ist, ist der erste Summand gleich Null. Der zweite Summand ist ein $O(h^2)$.

Die Varianz der ZG läßt sich nach oben abschätzen durch:

$$\begin{aligned}
&\frac{1}{nh^2}\ \mathsf{Var}\ (m(X_1)-m(t))K\left(\frac{X_1-t}{h}\right)\\
&\quad\leq \frac{1}{nh^2}\mathsf{E}(m(X_1)-m(t))^2K^2\left(\frac{X_1-t}{h}\right)\\
&\quad= \frac{1}{nh^2}\int_0^1(m(u)-m(t))^2K^2\left(\frac{u-t}{h}\right)f(u)du
\end{aligned}$$

Das läßt sich mit $\rho(u) := (m(u)-m(t))^2f(u)$ wegen $\rho(t) = \rho'(t) = 0$ und ansonsten zum Erwartungswert anlogen Überlegungen durch

$$\frac{C_1K^2(0)}{nh^2}\int_{-h}^{h}u^2du = O(hn^{-1})$$

nach oben abschätzen.

Wegen $\mathsf{E}|Z| \leq |\mathsf{E}Z| + \sqrt{\mathsf{Var}Z}$ für eine quadratisch integrierbare Zufallsgröße Z gilt damit:

$$\mathsf{E}\left|\frac{1}{nh}\sum_{i=1}^{n}(m(X_i)-m(t))K\left(\frac{t-X_i}{h}\right)\right| = O(h^2)+O((h/n)^{1/2})$$

Mit der Markov-Ungleichung folgt daraus die Behauptung.

□

Beweis von Satz 2.11:

Setzt man

$$\varepsilon_i^{(n)} := \frac{Y_i - m(x_i)}{\hat{f}_h(t_0)},$$

so erhält man $\mathsf{E}\varepsilon_i^{(n)} = 0$ und $\mathsf{Var}\varepsilon_i^{(n)} = \beta''(\psi(m(x_i)))(\hat{f}_h(t_0))^{-2}$.

Setzt man ferner $A_{in} = hK_h(t_0 - x_i)$, so erhält man in den Bezeichnungen von Humak [7], Seite 67:

$$\begin{aligned}
H_n^{-1} &= \frac{n^{1/5}\hat{f}_h(t_0)}{\sqrt{\sum_{i=1}^n \beta''(\psi(m(x_i)))(K_h(t_0-x_i))^2}} \quad \text{und}\\
\xi_n &= H_n^{-1}n^{4/5}(\hat{m}_h(t_0) - \mathsf{E}\hat{m}_h(t_0))
\end{aligned}$$

Für dieses H_n gilt: $n^{-2/5}H_n \underset{n\to\infty}{\longrightarrow} \sqrt{\beta''(\psi(m(t_0)))\int_{-1}^1 K^2(u)du}$.

Sind die Voraussetzungen des Zentralen Grenzwertsatzes aus Humak [7], Satz 2.4.3. erfüllt, so erhält man, daß ξ_n in Verteilung gegen Φ, also gegen die Standardnormalverteilung konvergiert. Mit

$$\sigma_{t_0} := \lim_{n\to\infty} n^{-2/5} H_n = \sqrt{\beta''(\psi(m(t_0))) \int_{-1}^{1} K^2(u)du}$$

folgt dann die Behauptung.

In der Tat sind die Voraussetzungen dieses Zentralen Grenzwertsatzes erfüllt. Es sei $\mathcal{F}_\varepsilon$ die Menge der Verteilungen der oben definierten ZGen $\varepsilon_i^{(n)}$ und es sei $\delta_i^{(n)} := (\hat{f}_h(t_0))^{-1}$.

I: $\max\limits_{i=1,\ldots,n} A_{in}(\sum_{j=1}^n A_{jn}^2)^{-1} A_{in}$

$$= \max_{i=1,\ldots,n} \frac{(hK_h(t_0 - x_i))^2}{\sum_{j=0}^n (hK_h(t_0 - x_j))^2} \leq \frac{(K(0))^2}{\sum_{j=0}^n (K((t_0 - x_j)/h))^2} \underset{n\to\infty}{\longrightarrow} 0$$

II: $\sup\limits_{F\in\mathcal{F}_\varepsilon} \int_{\{|x|>c\}} x^2 dF(x)$

$$= \sup_{\theta\in\Sigma} \int_{\{y:\, |\delta_i^{(n)}(y-m(x_i))|>c\}} (\delta_i^{(n)}(y - m(x_i)))^2 e^{y\theta-\beta(\theta)} d\mu(y)$$

$$\underset{c\to\infty}{\longrightarrow} 0, \qquad \text{da } \Sigma \text{ beschränkt ist.}$$

III: $\inf\limits_{F\in\mathcal{F}_\varepsilon} \int_{\mathbb{R}} x^2 dF(x) \geq \inf\limits_{i,n} \delta_i^{(n)} \times \inf\limits_{\theta\in\Sigma} \beta''(\theta) \geq r > 0$

$\square$

Beweis von Lemma 2.12:

Es genügt zu zeigen: $\exists q < \infty : \ P(\sup_{i=1,\ldots,m_n} |Z_i^{(n)}| \leq q a_n \ln m_n) \underset{n\to\infty}{\longrightarrow} 1$

Sei $q := \max(\zeta, 1)$ und sei $n \geq n_0$. Dann gilt:

$$\begin{aligned}
P(\sup_{i=1,\ldots,m_n} |Z_i^{(n)}| &\leq q a_n \ln m_n) \\
&\geq 1 - \sum_{i=1}^{m_n} P(|Z_i^{(n)}| \geq a_n q \ln m_n) \\
&\quad \text{(Ungleichung von Bonferroni)} \\
&\geq 1 - \sum_{i=1}^{m_n} c e^{-\rho_{m_n}} \\
&\quad \quad \text{(mit } \rho_{m_n} := q \ln m_n \geq \zeta \text{ und da } m_n \geq n \geq n_0) \\
&= 1 - m_n c e^{-q \ln m_n} = 1 - c m_n^{1-q} \underset{n\to\infty}{\longrightarrow} 1, \ \text{ da } q > 1
\end{aligned}$$

$\square$

Beweis von Korollar 2.13:

Wegen $\mathsf{E}e^{u_\Sigma|Z|} \leq U_\Sigma$ für $Z \sim F_\theta$ und alle $\theta \in \Sigma$ folgt mit der Markov-Ungleichung für alle $n \in \mathbb{N}$, alle $i = 1, \ldots, n$ und alle η:

$$P(|Y_i|u_\Sigma \geq \eta) \leq U_\Sigma e^{-\eta}$$

Also erfüllen $Y_1, \ldots, Y_n$ die Voraussetzungen des Lemmas 2.12 und damit folgt: $\sup_{i=1,\ldots,n} |Y_i| = O_P(\ln n)$.

Nach Lemma 2.2 und Lemma 2.5 genügt es im zufälligen Design, die Behauptung unter der Bedingung $X_\cdot^{(n)} \in M_n^3$ zu zeigen. Für das äquidistante Design folgt die Behauptung dann, da für großes n die Designpunkte diese Bedingung sowieso erfüllen.

Unter dieser Bedingung ist $\hat{m}_h$ auf $[0, 1]$ wohldefiniert.

Per Definition gilt dann: $|\hat{m}_h(t)| \leq \sup_{i=1,\ldots,n} |Y_i|$ für alle $t \in [0, 1]$. Folglich gilt auch $\sup_{t\in[0,1]} |\hat{m}_h(t)| \leq O_P(\ln n)$.

Die Ableitung von $\hat{m}_h$ ist:

$$\hat{m}_h'(t) = \frac{\sum_{i=1}^n (Y_i - \hat{m}_h(t))K_h'(t - x_i)}{\sum_{i=1}^n K_h(t - x_i)}$$

Also erhält man:

$$\begin{aligned}
\sup_{t\in[0,1]} |\hat{m}_h'(t)| &\leq \left(\sup_{i=1,\ldots,n} |Y_i| + \sup_{t\in[0,1]} |\hat{m}_h(t)| \right) \times \sup_{t\in[0,1]} \frac{\sum_{i=1}^n |K'((t - x_i)/h)|}{h \sum_{i=1}^n K((t - x_i)/h)} \\
&\leq O_P(\ln n) \frac{nS_K}{h} = O_P(h^{-1} n \ln n) \\
&\quad \text{für} \quad S_K := \frac{\sup_{t\in[-1,1]} |K'(t)|}{K(1/2)} < \infty
\end{aligned}$$

□

Beweis von Lemma 2.14:

Zuerst wird die Konvergenzrate $n^{-2/5}\sqrt{\ln n}$ für jedes t punktweise nachgewiesen. Konstanten etc. sind auch hier in dem Sinne wie in Abschnitt 2.5 unabhängig von n und t. Analog der Vorgehensweise dort wird das Verhalten des „Varianzterms“ und des „Biasterms“ getrennt betrachtet. Darauf aufbauend wird anschließend die Aussage des Lemmas bewiesen. Der Beweis ist an Härdle und Mammen [4] angelehnt.

Es wird erst das zufällige Design betrachtet. Sei $t \in (0, 1)$ beliebig fest.

- „Varianzterm“
  Sei $q > 0$ beliebig fest. Es wird gezeigt, daß dann ein $C_V < \infty$ existiert, so daß

  $$P(n^{-1}|\sum_{i=1}^n (Y_i - m(X_i))K_h(t - X_i)| \geq C_V n^{-2/5}\sqrt{\ln n}| X_\cdot \in M_n^1) = O(n^{-q})$$

  gilt.

Mittels der Markov-Ungleichung läßt sich die linke Seite abschätzen durch:

$$e^{-C_V \ln n}\mathsf{E}\left[\exp\left\{n^{-3/5}\sqrt{\ln n}\sum_{i:\,X_i\in(t-h,t+h)}(Y_i-m(X_i))K_h(t-X_i)\right\}\middle|\,X_{\cdot}\in M_n^1\right]+ \tag{2.ii}$$

$$e^{-C_V \ln n}\mathsf{E}\left[\exp\left\{-n^{-3/5}\sqrt{\ln n}\sum_{i:\,X_i\in(t-h,t+h)}(Y_i-m(X_i))K_h(t-X_i)\right\}\middle|\,X_{\cdot}\in M_n^1\right]$$

Der Erwartungswert im ersten Summanden läßt sich wegen der Unabhängigkeit der Beobachtungen (X_i, Y_i) in ein Produkt umformen, das wegen $X_{\cdot} \in M_n^1$ höchstens b^*nh Faktoren ungleich Eins hat:

$$\prod_{i:X_i\in(t-h,t+h)}\mathsf{E}[\exp\{n^{-2/5}\sqrt{\ln n}(Y_i-m(X_i))K((t-X_i)/h)\}|\,X_i\in(t-h,t+h)]$$

Diese Faktoren lassen sich in der bekannten Form als Reihe entwickeln:

$$\begin{aligned}&\sum_{j=0}^{\infty}\mathsf{E}[(j!)^{-1}(n^{-2/5}\sqrt{\ln n}(Y_i-m(X_i))K((t-X_i)/h))^j|\,X_i\in(t-h,t+h)]\\ &\le\ 1+\frac{(K(0))^2\ln n}{n^{4/5}}\times\\ &\qquad\sum_{j=2}^{\infty}\left(\frac{1}{j!}\left(\frac{\sqrt{\ln n}K(0)}{n^{2/5}}\right)^{j-2}\mathsf{E}[(Y_i-m(X_i))^j|\,X_i\in(t-h,t+h)]\right)\\ &\le\ 1+\frac{\tilde{C}\ln n}{n^{4/5}}\end{aligned}$$

Wegen $\mathsf{E}e^{u_\Sigma|Y_i|}\le U_\Sigma$ ist für alle $i=1\dots n$ die Summe $\sum_{j=2}^{\infty}\mathsf{E}(u_\Sigma^{j-2}|Y_i|^j)/j!$ beschränkt. Außerdem ist m beschränkt, und somit ist für n so groß, daß $\sqrt{\ln n}K(0)n^{-2/5}\le u_\Sigma$ gilt, die Reihe im vorletzten Term beschränkt.

Also ist der linke Summand in (2.ii) nach oben beschränkt durch:

$$e^{-C_V\ln n}\left(1+\frac{\tilde{C}\ln n}{n^{4/5}}\right)^{b^*n^{4/5}}\le n^{b^*\tilde{C}-C_V}=n^{-q},$$

falls C_V entsprechend gewählt wird.

Der rechte Summand in (2.ii) läßt sich analog abschätzen.

- „Biasterm“
  Sei $q>0$ beliebig fest. Es wird gezeigt, daß dann ein $C_B<\infty$ existiert, so daß

  $$P(n^{-1}|\sum_{i=1}^{n}(m(X_i)-m(t))K_h(t-X_i)|\ge C_Bn^{-2/5}\sqrt{\ln n}|X_{\cdot}\in M_n^1)=O(n^{-q})$$

  gilt.

Die linke Seite läßt sich nach oben abschätzen durch:

$$P\left(e_* + n^{-1}|\sum_{i:\, X_i\in(t-h,t+h)}((m(X_i)-m(t))K_h(t-X_i)-e_i)| \geq \frac{C_B\sqrt{\ln n}}{n^{2/5}}\middle| X_. \in M_n^1\right)$$

wobei $e_i := \mathsf{E}[(m(X_i)-m(t))K_h(t-X_i)|\, X_i \in (t-h,t+h)]$ und $e_* := n^{-1}|\sum_{i:\, X_i\in(t-h,t+h)} e_i|$. Wegen Lemma 2.2 und Lemma 2.9 ist $|\vartheta_h(t) - m_h(t)|$ unter der Bedingung $X_. \in M_n^1$ ein $O_P(n^{-2/5})$ ist. Damit ist e_i für alle $i:\ X_i \in (t-h,t+h)$ ein $O(n^{-1/5})$. Unter der Bedingung $X_. \in M_n^1$ gilt also: $e_* \leq b^*h \max_{i:\, X_i\in(t-h,t+h)} e_i$. Damit existiert ein festes c_{e_*} so, daß sich die Wahrscheinlichkeit weiter abschätzen läßt durch:

$$P\left(n^{-1}|\sum_{i:\, X_i\in(t-h,t+h)}((m(X_i)-m(t))K_h(t-X_i)-e_i)| \geq \frac{C_B\sqrt{\ln n}-c_{e_*}}{n^{2/5}}\middle| X_. \in M_n^1\right)$$

Analog zum „Varianzterm“ erhält man für diese Wahrscheinlichkeit die Abschätzung:

$$e^{-C_B \ln n + c_{e_*}\sqrt{\ln n}}\left(1+\frac{\tilde{C}\ln n}{n^{4/5}}\right)^{b^* n^{4/5}} \leq n^{b^*\tilde{C}+c_{e_*}/\sqrt{\ln n}-C_B} = O(n^{-q}),$$

falls C_B entsprechend gewählt wird.

Werden beide Terme zusammengefaßt, so erhält man, daß für jedes $q > 0$ ein $C_q < \infty$ existiert, so daß gilt:

$$P(n^{-1}|\sum_{i=1}^{n}(Y_i - m(t))K_h(t-X_i)| \geq C_q n^{-2/5}\sqrt{\ln n}|\, X_. \in M_n^1) = O(n^{-q})$$

Nun kann Lemma 2.3 angewendet werden, und es folgt:

$$P(n^{-1}|\sum_{i=1}^{n}(Y_i - m(t))K_h(t-X_i)| \geq C_q n^{-2/5}\sqrt{\ln n}|\, X_. \in M_n^1 \cap M_n^2) = O(n^{-q})$$

Analog dem Beweis von Satz 2.10 existiert ein $\tilde{C}_q < \infty$, so daß gilt:

$$P(|\hat{m}_h(t) - m(t)| \geq \tilde{C}_q n^{-2/5}\sqrt{\ln n}|\, X_. \in M_n^1 \cap M_n^2) = O(n^{-q})$$

Mit der Bonferroni-Ungleichung kann man nun schließen:

$$\begin{aligned} P(\sup_{j=1,\dots,m_n} & |\hat{m}_h(t_j) - m(t_j)| \leq C_q n^{-2/5}\sqrt{\ln n}|\, X_. \in M_n^1 \cap M_n^2) \\ \geq\ & P(X_. \in M_n^1 \cap M_n^2) \\ & - \sum_{j=1}^{m_n} P(|\hat{m}_h(t_j) - m(t_j)| \geq C_q n^{-2/5}\sqrt{\ln n}|\, X_. \in M_n^1 \cap M_n^2) \\ \geq\ & P(X_. \in M_n^1 \cap M_n^2) - m_n O(n^{-q}) \\ =\ & P(X_. \in M_n^1 \cap M_n^2) - O(n^{\alpha-q}) \underset{n\to\infty}{\longrightarrow} 1 \end{aligned}$$

Mit Lemma 2.2 folgt daraus

$$P(\sup_{j=1,\dots,m_n} |\hat{m}_h(t_j) - m(t_j)| \le C_q n^{-2/5}\sqrt{\ln n}) \underset{n\to\infty}{\longrightarrow} 1,$$

womit im zufälligen Design die Behauptung folgt.

Im äquidistanten Design wird die Konvergenz des „Varianzterms“ analog hergeleitet – die Bedingung $X_. \in M_n^1$ erübrigt sich. Der „Biasterm“ ist nicht zufällig, seine Konvergenzrate folgt unmittelbar aus Lemma 2.8. Beide Terme können analog zum zufälligen Design zusammengefaßt werden, da die Bedingung $x_. \in M_n^2$ für n genügend groß erfüllt wird.

□

Beweis von Lemma 2.16:

Es wird analog zum Beweis des Lemma 2.14 vorgegangen: Das zufällige Design wird betrachtet, das äquidistante folgt dann fast unmittelbar.

Im zufälligen Design wird zuerst das punktweise asymptotische Verhalten des Zählers, getrennt nach „Varianz-“ und „Biasterm“, hergeleitet – unter der Bedingung, daß $X_.$ in M_n^1 liegt. Wenn für beide Terme die Behauptung

$$\forall q > 0\, \exists C_* < \infty : \; P(|\text{Term}| \ge C_* n^{-1/5}\sqrt{\ln n}\,|X_. \in M_n^1) = O(n^{-q})$$

nachgewiesen ist, so können beide wie im Beweis von Lemma 2.14 zusammengefaßt und nach $X_. \in M_n^1 \cap M_n^2$ bedingt werden. Division durch den Nenner $\hat{f}_h(t)$ ändert unter dieser Bedingung nichts an der Größenordnung, d.h. dann folgt:

$$\forall q > 0\, \exists \tilde{C}_q < \infty : \; P(|T_n(t)| \ge \tilde{C}_q n^{-1/5}\sqrt{\ln n}\,|X_. \in M_n^1) = O(n^{-q})$$

Mit der Bonferroni-Ungleichung und Lemma 2.2 folgt dann die Behauptung im zufälligen Design analog zum Beweis von Lemma 2.14.

Der Zähler von $T_n(t)$ läßt sich zerlegen in eine Summe aus

- „Varianzterm“
$$n^{-1}\sum_{i=1}^{n}(m(X_i) - Y_i)K_h'(t - X_i)$$
- und „Biasterm“
$$n^{-1}\sum_{i=1}^{n}\left(m'(t)K_h(t - X_i) + (m(t) - m(X_i))K_h'(t - X_i)\right).$$

Die oben behauptete Eigenschaft beider Terme wird im folgenden getrennt bewiesen.

- „Varianzterm“
Der Nachweis von
$$\forall q > 0\, \exists C_* < \infty : \; P(|\text{Varianzterm}| \ge C_* n^{-1/5}\sqrt{\ln n}\,|X_. \in M_n^1) = O(n^{-q})$$

kann genauso wie im Beweis von Lemma 2.14 durchgeführt werden. Die Konvergenzrate verschlechtert sich um $n^{1/5}$, da statt K die Ableitung K' in den Term eingeht. Entsprechend wird $K'((t-X_i)/h)$ durch $\sup_{u\in[-1,1]} K'(u) < \infty$ für alle i mit $X_i \in (t-h, t+h)$ abgeschätzt.

- „Biasterm"
  Auch hier kann die benötigte Eigenschaft analog zum Beweis von Lemma 2.14 bewiesen werden. Es sei

  $$e_i := \mathsf{E}\left[m'(t)K_h(t-X_i) + (m(t)-m(X_i))K'_h(t-X_i) \middle| X_i \in (t-h,t+h)\right]$$

  und $e_* := n^{-1}|\sum_{i:\, X_i\in(t-h,t+h)} e_i|$.
  Es gilt: $e_i = O(1)$ für alle $i = 1,\ldots,n$, denn:

  $$\begin{aligned} e_i &\le \gamma_{t,h}\left|\int_{t-h}^{t+h}(m'(t)K_h(t-u)+(m(t)-m(u))K'_h(t-u))f(u)du\right| \\ &\qquad \text{wobei } \gamma_{t,h} := \left(\int_{t-h}^{t+h} f(v)dv\right)^{-1} \\ &\le \gamma_{t,h}\left|\int_{t-h}^{t+h}(m'(t)K_h(t-u)-m'(t)(u-t)K'_h(t-u))f(u)du\right| \\ &\quad +\gamma_{t,h}\left|\int_{t-h}^{t+h} c_{m'}(u-t)^2K'_h(t-u))f(u)du\right| \\ &= \gamma_{t,h}\left|m'(t)\int_{t-h}^{t+h}\rho'(u)f(u)du\right| + O(1) \\ &\qquad \text{wobei } \rho(u) := (u-t)K_h(t-u) \\ &\le \gamma_{t,h}|m'(t)|\left|[\rho(u)f(u)]_{t-h}^{t+h} - \int_{t-h}^{t+h}\rho(u)f'(u)du\right| + O(1) \\ &= O(1) \end{aligned}$$

  Wie im Beweis zum Lemma 2.14 gilt unter der Bedingung $X_. \in M_n^1$ die Abschätzung $e_* \le b^*h\max_{i:\, X_i\in(t-h,t+h)} e_i$. Analog folgt die Behauptung

  $$\forall q>0\,\exists C_*<\infty:\ P(|\text{Biasterm}| \ge C_* n^{-1/5}\sqrt{\ln n}\,|\,X_. \in M_n^1) = O(n^{-q})$$

Im deterministischen Design muß analog Lemma 2.14 noch die Konvergenzrate des dann nicht zufälligen „Biasterms" bewiesen werden. Es gilt die folgende Abschätzung:

$$\begin{aligned} &n^{-1}\left|\sum_{i=1}^n(m'(t)K_h(t-x_i)+(m(t)-m(x_i))K'_h(t-x_i))\right| \\ \le &\left|n^{-1}m'(t)\sum_{i=1}^n(K_h(t-x_i)+(t-x_i)K'_h(t-x_i))\right| \\ &+\ 2h^3c_{m'}\max_{u\in[-1,1]}|K'_h(u)| \end{aligned}$$

$$
\begin{aligned}
&\le \left| n^{-1} m'(t) \sum_{i=1}^{n} \rho'(x_i) \right| + O(h), \\
&\quad \text{wobei } \rho(u) := (u-t)K_h(t-u) \\
&= \left| n^{-1} m'(t) \sum_{i=1}^{n} \rho'(i/n) \right| + O(h) \\
&= \left| n^{-1} m'(t) n \int_0^1 \rho'(u)du + n^{-1} m'(t) nhO((nh^2)^{-1}) \right| + O(h) \\
&\quad \text{(Rechteckregel)} \\
&= O((nh)^{-1}) + O(h), \text{ da } \int_0^1 \rho'(u)du = 0 \text{ für } n \text{ genügend groß} \\
&= O(n^{-1/5})
\end{aligned}
$$

□

Beweis von Lemma 2.17:

Zuerst wird das zufällige Design betrachtet. Sei $X_.^{(n)} \in M_n^3$. Unter dieser Bedingung ist T_n auf $[0,1]$ wohldefiniert.

Wenn m und K zweimal differenzierbar sind, so sind die zweiten Ableitungen durch $c_{m'}$ bzw. $c_{K'}$ beschränkt, T_n ist differenzierbar, und die Ableitung ist:

$$
\frac{\sum_{i=1}^n \left[hm''(t)K\left(\frac{t-X_i}{h}\right) + 2m'(t)K'\left(\frac{t-X_i}{h}\right) + \frac{1}{h}(m(t)-Y_i)K''\left(\frac{t-X_i}{h}\right) \right]}{h \times \sum_{i=1}^n K\left(\frac{t-X_i}{h}\right)}
$$
$$
- \frac{\left(\sum_{i=1}^n \left[hm'(t)K\left(\frac{t-X_i}{h}\right) + (m(t)-Y_i)K'\left(\frac{t-X_i}{h}\right) \right]\right) \times \left(\sum_{i=1}^n K'\left(\frac{t-X_i}{h}\right)\right)}{h^2 \times \left(\sum_{i=1}^n K\left(\frac{t-X_i}{h}\right)\right)^2}
$$

Da K und m und die ersten und zweiten Ableitungen beschränkt sind, läßt sich dieser Term durch

$$
\sup_{i=1,\dots,n} |Y_i| \times \frac{\left(\sum_{i=1}^n K'\left(\frac{t-X_i}{h}\right)\right)^2}{\left(h\sum_{i=1}^n K\left(\frac{t-X_i}{h}\right)\right)^2}
$$

abschätzen.

Da in der $h/2$-Umgebung von jeden $t \in [0,1]$ wenigstens ein X_i ist, gibt es ein $\tilde{C} < \infty$, so daß die Wahrscheinlichkeit unter der Bedingung $X_.^{(n)} \in M_n^3$ dafür, daß der Betrag des rechten Faktors nach oben durch $\tilde{C}h^{-2}n^2$ abgeschätzt wird, gleich Eins ist.

Ferner gilt entsprechend Lemma 2.13: $\sup_{i=1,\dots,n} |Y_i| = O_P(\ln n)$. Mit Lemma 2.2 folgt $\sup_{t\in[0,1]} |T_n'(t)| \le O_P(h^{-2}n^2 \ln n)$.

Ist m oder K nicht zweimal differenzierbar, so kann c_{T_n} ähnlich abgeschätzt werden: Für $m''(t)$ wird $c_{m'}$ und für $K''((t-x_i)/h)$ wird $c_{K'}$ eingesetzt.

Im äquidistanten Fall folgt die Behauptung analog, da $x_.^{(n)} \in M_n^3$ für genügend großes n gilt und die Bedingung an die Designpunkte entfällt.

□

Beweis von Lemma 2.18:

Sei I_l ein Gitter auf $[h_n, 1-h_n]$ mit der Gitterweite $\delta_l := l^{-1}$ und sei $l = l_n := n^3$. Dann gilt:

$$\sup_{t\in[h_n,1-h_n]} |T_n(t)| \le \sup_{y\in I_l} |T_n(y)| \ + \sup_{t\in[h_n,1-h_n]} \inf_{y\in I_l} |T_n(t) - T_n(y)|$$

Nach Lemma 2.16 ist der linke Summand ein $O_P(n^{-1/5}\sqrt{\ln n})$. Der rechte Summand läßt sich durch

$$\delta_l \times c_{T_n}$$

abschätzen, was nach Lemma 2.17 ein $O_P(l^{-1}h^{-2}n^2 \ln n) = O_P(n^{-3/5} \ln n) \le O_P(n^{-1/5}\sqrt{\ln n})$ ist. Damit folgt die Behauptung.

□

Beweis von Lemma 2.19:

Unter $X_\cdot^{(n)} \in M_n^1 \cap M_n^2$ ist der Betrag des Zählers durch $b^* nh \sup K'$ nach oben und der Betrag der Nenners durch $c_* nhK(1/2)$ nach unten beschränkt. Beides gilt gleichmäßig in $t \in [0,1]$.

Im deterministischen Design folgt die Behauptung unmittelbar, da $x_\cdot^{(n)} \in M_n^1 \cap M_n^2$ für n genügend groß gilt.

Im zufälligen Design folgt, daß ein $\tilde{C} < \infty$ existiert, so daß gilt:

$$P\left(\sup_{u\in[0,1]} \left| \frac{\sum_{i=1}^n K_h'(u-X_i)}{\sum_{i=1}^n K_h(u-X_i)} \right| \le \tilde{C} \,\middle|\, X_\cdot \in M_n^1 \cap M_n^2 \right) = 1$$

Nach Lemma 2.2 konvergiert die unbedingte Wahrscheinlichkeit gegen Eins, womit die Behauptung folgt.

□

3 Das monotone Regressionsmodell

In diesem Kapitel soll das Modell aus Kapitel 2 unter der Bedingung betrachtet werden, daß die unbekannte Regressionsfunktion m monoton ist. Um umständliche Unterscheidungen zwischen *monoton wachsend* und *monoton fallend* zu vermeiden, wird im Wertebereich der Regressionsfunktion m mit der linearen Ordnung $\preceq$ gearbeitet, welche wahlweise $\le$ oder $\ge$ sein kann, so daß die Theorie auf Modelle sowohl mit wachsender als auch mit fallender Regressionsfunktion angewendet werden kann.

In Kapitel 1 in Robertson, Wright und Dykstra [10] wird die in diesem Kapitel vorgestellte Theorie ausführlicher, jedoch für Funktionen mit diskretem Definitionsbereich entwickelt.

In den beiden ersten Abschnitten dieses Kapitels wird das monotone Modell definiert und werden einige Eigenschaften monotoner Funktionen hergeleitet. Die Konstruktion des monotonen MgL-Schätzers wird über den Umweg der

Kleinste-Quadrate-Theorie durchgeführt. Das ist der Inhalt des 3. Abschnittes. Im 4. Abschnitt wird der monotone MgL-Schätzer definiert und es wird bewiesen, daß er gleich dem in Abschnitt 3 konstruierten Schätzer ist. Im 5. Abschnitt werden asymptotische Konvergenzraten eines leicht modifizierten monotonen Schätzers gezeigt. Die meisten Beweise sind wieder in einem eigenen Abschnitt am Ende dieses Kapitels zusammengefaßt.

3.1 Das einfache monotone Modell

Definition 3.1 *Monotone Funktion, Konvexe Funktion*
Seien S und M Teilmengen von $\mathbb{R}$.

- *Eine Funktion $h: S \to M$ heißt monoton, falls gilt:*

$$\forall u, v \in S: \quad u < v \Rightarrow h(u) \preceq h(v)$$

Falls sogar $h(u) \prec h(v)$ folgt, so wird h auch streng monoton genannt.

- *Seien S und M außerdem konvex in $\mathbb{R}$.*

Eine Funktion $h: S \to M$ heißt konvex, falls gilt:

$$\forall u, v \in S \;\; \forall \alpha \in (0,1): \quad h(\alpha u + (1-\alpha)v) \preceq \alpha h(u) + (1-\alpha)h(v)$$

Falls sogar überall

$$h(\alpha u + (1-\alpha)v) \prec \alpha h(u) + (1-\alpha)h(v)$$

gilt, wird h auch streng konvex genannt.

Setzt man in dieser Definition $\preceq := \leq$ oder $\preceq := \geq$ so erhält man die gewöhnlichen Definitionen von wachsender bzw. fallender Monotonie und Konvexität bzw. Konkavität. Während des gesamten Kapitels stehe $\preceq$ für eine dieser beiden Relationen. Werden im Text verbale Vergleiche wie „größer als“, „kleiner als“ o. ä. getroffen, so sei das bzgl. der $\preceq$ bzw. der $\prec$ Relation verstanden. Im Unterschied dazu werden Elemente des Definitionsbereiches solcher Funktionen verbal mit den Formulierungen „links von“, „rechts von“ o. ä. verglichen.

Zusätzlich zu den Modellannahmen aus Kapitel 2.1 sei also in diesem Abschnitt angenommen, daß die Regressionsfunktion m monoton bezüglich $\preceq$ ist.

3.2 Verhalten monotoner und konvexer Funktionen

In diesem Abschnitt sollen einige Eigenschaften monotoner und konvexer Funktionen hergeleitet werden.

Die beiden folgenden Lemmata beschreiben das Aussehen einer beliebigen monotonen Funktion $h : [0, 1] \to \mathbb{R}$: Der linke und der rechte Grenzwert an jeder Stelle des Definitionsbereiches existieren, der Funktionswert muß zwischen beiden liegen. Als Unstetigkeitsstellen kommen damit nur Sprungstellen in Frage. Ferner läßt sich der Definitionsbereich in eine Menge disjunkter, aber aneinander grenzender offener Intervalle zerlegen, auf denen die Funktion entweder konstant oder streng monoton ist. Die Punkte zwischen diesen Intervallen seien im weiteren als *Bruchstellen* der Funktion bezeichnet.

Lemma 3.1 *Sei* $h : [0,1] \to \mathbb{R}$ *eine monotone Funktion. Dann existiert für jedes* $t_0 \in (0,1)$ *der links- und der rechtsseitige Grenzwert* $h(t_0-)$ *und* $h(t_0+)$. *Für* $t_0 = 0$ *existiert der rechtsseitige und für* $t_0 = 1$ *der linksseitige Grenzwert.*

Lemma 3.2 *Der Definitionsbereich einer monotonen Funktion läßt sich in eine Menge disjunkter, aneinander grenzender offener Intervalle sowie einpunktiger Mengen zerlegen, auf denen die Funktion entweder konstant oder streng monoton ist.*

Das Verhalten konvexer Funktionen wird im Lemma 3.4 auf das Verhalten monotoner Funktionen zurückgeführt; dazu wird das folgende Lemma benötigt:

Lemma 3.3 *Es sei* $h : [0,1] \to \mathbb{R}$ *eine konvexe Funktion und es sei der Differenzenquotient* D_{x,h_0} *für* $x \in (0,1]$ *und* $h_0 < 0$ *(linksseitig) oder für* $x \in [0,1)$ *und* $h_0 > 0$ *(rechtsseitig) definiert durch:*

$$D_{x,h_0} := \frac{h(x+h_0) - h(x)}{h_0}$$

Dann gilt:

1. *für alle* $x \in (0,1)$ *und* $h_1 < 0$ *und* $h_2 > 0$: $D_{x,h_1} \preceq D_{x,h_2}$
2. *für alle* $x \in [0,1)$ *und* $0 < h_1 < h_2$: $D_{x,h_1} \preceq D_{x,h_2}$
3. *für alle* $x \in (0,1]$ *und* $h_2 < h_1 < 0$: $D_{x,h_2} \preceq D_{x,h_1}$
4. *für alle* $x_1 < x_2 \in [0,1]$ *und* $h_1 > 0$ *und* $h_2 < 0$ *so, daß* $x_1 + h_1 < x_2 + h_2$: $D_{x_1,h_1} \preceq D_{x_2,h_2}$.

Lemma 3.4 *Der Definitionsbereich einer konvexen Funktion* h *läßt sich in paarweise disjunkte, offene Intervalle zerlegen, auf denen* h *entweder linear oder streng konvex ist. Die restlichen Punkte sind isoliert und die links- und rechtsseitigen Ableitungen an einer solchen Stelle* t_0 *sind größer gleich* h' *links von* t_0 *und kleiner gleich* h' *rechts von* t_0.

3.3 Der monotone KQS der Regressionsfunktion

Um den MgL-Schätzer einer monotonen Regressionsfunktion zu ermitteln, erweist es sich als günstig, das Problem zunächst in der Kleinsten-Quadrate-Theorie zu betrachten.

Strenggenommen handelt es sich nicht um eine Kleinste-Quadrate-Schätzung, sondern um einen analog zum MgL-Schätzer definierten „Kleinste-geglättete-Quadrate-Schätzer". Um die Bezeichnungen jedoch nicht unnötig kompliziert zu machen, wird im folgenden die Bezeichnung KQS verwendet.

3.3.1 Der uneingeschränkte KQS und der monotone KQS

Definition 3.2 *Uneingeschränkter Kleinste-Quadrate-Schätzer*
Die Funktion $\hat{g}_h : [0,1] \to \mathbb{R}$ *wird (uneingeschränkter) Kleinste-Quadrate-Schätzer der Regressionsfunktion* m *genannt, falls sie das Funktional*

$$\mathcal{L}^q(g) := \sum_{i=1}^{n} \int_0^1 (Y_i - g(t))^2 K_h(t - x_i) dt$$

unter allen Funktionen $g : [0,1] \to \mathbb{R}$ *minimiert.*

Satz 3.5 *Falls* $x_\cdot^{(n)} \in M_n^3$*, so gilt bis auf Lebesgue-Nullmengen:*

$$\hat{g}_h(t) = \frac{\hat{\vartheta}_h(t)}{\hat{f}_h(t)} \text{ für alle } t \in [0,1]$$

Das heißt, daß der KQS der Regressionsfunktion gleich ihrem MgL-Schätzer ist. Damit gelten auch alle asymptotischen Eigenschaften des MgL-Schätzers für den KQS. Andererseits folgt daraus auch, daß $\mathcal{L}^q(\hat{m}_h)$ endlich ist.

Definition 3.3 *Monotoner KQS der Regressionsfunktion.*
Die Funktion $\hat{g}_h^* : [0,1] \to \mathbb{R}$ *wird monotoner Kleinste-Quadrate-Schätzer der Regressionsfunktion* m *genannt, falls sie das Funktional* $\mathcal{L}^q$ *unter allen monotonen Funktionen* $g : [0,1] \to \mathbb{R}$ *minimiert.*

3.3.2 Zusammenhang zwischen uneingeschränktem und monotonem KQS – Monotonisierung des uneingeschränkten KQS

Im folgenden Satz wird der Zusammenhang zwischen uneingeschränktem KQS und monotonem KQS dargestellt:

Satz 3.6 *Der monotone KQS* $\hat{g}_h^*$ *minimiert unter allen monotonen Funktionen* $g : [0,1] \to \mathbb{R}$ *das Funktional*

$$\int_0^1 (\hat{g}_h(t) - g(t))^2 w_h(t) dt,$$

wobei $w_h(t) := n\hat{f}_h(t) = \sum_{i=1}^n K_h(t-x_i)$*. Nur Funktionen, die dieses Funktional minimieren, sind monotone KQS.*

Mit Hilfe dieses Satzes lassen sich die weiteren Betrachtungen auf dieses Funktional beschränken. Es muß lediglich $w_h(t) > 0$ vorausgesetzt werden – d.h. $x_\cdot^{(n)} \in M_n^3$. Die Anzahl und Lage der Designpunkte und die Bandweite h beeinflussen die Wahl des monotonen KQS nur mittels des uneingeschränkten KQS und der Gewichtsfunktion w_h.

Speziell folgt aus diesem Satz, daß der monotone KQS gleich der mit w_h gewichteten Kleinste-Quadrate-Monotonisierung des uneingeschränkten KQS ist.

3.3.3 Existenz und Eindeutigkeit des monotonen KQS

Durch Satz 3.6 ist es möglich, das Problem des monotonen KQS bei bekanntem uneingeschränktem KQS auf die Kleinste-Quadrate-Monotonisierung desselben zu reduzieren. Darüber hinaus kann unter den Voraussetzungen von Satz 3.5 davon ausgegangen werden, daß $\hat{g}_h$ stetig und beschränkt und mithin $\int_0^1 \hat{g}_h^2(t)w_h(t)dt$ existiert und endlich ist. Außerdem ist w_h dann per Definition stetig und überall positiv.

In diesem Abschnitt nun sollen Existenz und Eindeutigkeit der K-Q-Monotonisierung des uneingeschränkten KQS nachgewiesen werden.

Die zu monotonisierende Funktion sei mit g bezeichnet, die Gewichtsfunktion mit w.

Satz 3.7 *Es sei* $w : [0,1] \to \mathbb{R}$ *eine stetige und überall positive Funktion und es sei*

$$\langle g, h \rangle := \int_0^1 g(t)h(t)w(t)dt.$$

Dann bildet die Menge H *aller Restklassen fast überall gleicher Funktionen* $h : [0,1] \to \mathbb{R}$ *mit* $\langle h, h \rangle < \infty$ *einen Hilbert-Raum; d.h.*

- $\langle .,. \rangle$ *ist ein Skalarprodukt auf* H,
- $||h|| := \sqrt{\langle h, h \rangle}$ *ist eine Metrik auf* H *und*
- H *ist vollständig bzgl.* $||.||$

Beweis: Siehe zum Beispiel Reed und Simon [9], Kapitel II.

□

Definition 3.4 *Eine Funktion* $h \in H$ *heiße* L_2*-monoton, falls eine Nullmenge* N *existiert, so daß* $h(t_1) \preceq h(t_2)$ *für beliebige* $t_1 < t_2$ *außerhalb* N. *Die Menge der* L_2*-monotonen Funktionen in* H *sei mit* C *bezeichnet.*

Lemma 3.8 *Die Menge* C *aller* L_2*-monotonen Funktionen in* H *ist abgeschlossen und konvex.*

Satz 3.9 *Zu jedem* $g \in H$ *existiert ein eindeutig bestimmtes* $g^* \in C$, *so daß*

$$g^* = \underset{f \in C}{argmin} \, ||g - f||$$

Für diese Funktion gilt außerdem:

$$\langle g - g^*, g^* \rangle = 0 \tag{3.i}$$
$$\langle g - g^*, g^* - f \rangle \geq 0 \tag{3.ii}$$
$$||g - f||^2 \geq ||g - g^*||^2 + ||g^* - f||^2 \tag{3.iii}$$

Die beiden letzten Ungleichungen gelten jeweils für alle $f \in C$.

Gilt andererseits für ein $u \in C$

$$\langle g - u, u - f \rangle \geq 0 \qquad \forall f \in C, \tag{3.iv}$$

so ist diese Funktion gleich g^*.

Satz 3.9 sichert die Existenz und Eindeutigkeit einer L_2-Restklasse g^*, die g im Kleinste-Quadrate-Sinne optimal monotonisiert. Im weiteren sei mit g^* immer eine beschränkte monotone Version dieser Restklasse bezeichnet; ihre Existenz wird im folgenden Lemma gezeigt. Damit wird es möglich, die Betrachtungen im Raum aller quadratisch integrierbaren Funktionen fortzusetzen und umständliche Restklassenbetrachtungen zu vermeiden.

Definition 3.5 *Eine Funktion $\tilde{h} : [0,1] \to \mathbb{R}$ heiße monotone Version der L_2-monotonen Restklasse h, falls $\tilde{h}$ monoton ist.*

Da der Definitionsbereich abgeschlossen ist, ist eine solche monotone Version auch immer beschränkt.

Lemma 3.10 *Zu einem stetigen und beschränkten g existiert eine monotone und mithin beschränkte Version $\tilde{g}^*$ der eindeutig bestimmten L_2-monotonen Restklasse g^*, die g im Kleinste-Quadrate-Sinne optimal monotonisiert.*

Also ist $\tilde{g}^*$ bis auf Lebesgue-Nullmengen eindeutig bestimmt.

Im weiteren stehe g^* für eine solche beschränkte und monotone Version und es seien für ein $a \in g^*([0,1])$ die Bezeichnungen

$$t_a := \inf\{t \in [0,1] : g^*(t) \succeq a\} \quad \text{und} \quad t^a := \sup\{t \in [0,1] : g^*(t) \preceq a\}$$

vereinbart.

3.3.4 Eigenschaften des monotonen KQS

Die nun folgenden Lemmata behandeln Eigenschaften der Kleinsten-Quadrate-Monotonisierung. Diese Eigenschaften werden im folgenden Abschnitt die Konstruktion des monotonen KQS erlauben.

Lemma 3.11 *Für jedes $c \in g^*([0,1])$ gilt:*

$$\int_{[g^*=c]} g(t)w(t)dt = c \int_{[g^*=c]} w(t)dt.$$

Lemma 3.12 *Sei $[a,b]$ ein Intervall, auf dem g monoton fallend ist, d.h. es gilt: $\forall t_1 < t_2 \in [a,b] : \; g(t_2) \preceq g(t_1)$.*

Dann ist g^* auf (a,b) konstant.

Lemma 3.13 *Es sei $J_c^{g^*} = (t_c, t^c) \in \mathcal{J}_1^{g^*}$ ein Intervall, auf dem g^* konstant c ist, wobei $t_c > 0$ und $t^c < 1$ sei.*
Dann gilt: $g(t_c) = c = g(t^c)$

Lemma 3.14 *Es sei $J_x^{g^*} = (t_*, t^*) \in \mathcal{J}_2^{g^*}$ ein Intervall, auf dem g^* streng monoton ist.*
Dann gilt: $c_* := g^*(t_*+) = g(t_*)$ und $c^* := g^*(t^*-) = g(t^*)$.

Lemma 3.15 *Es seien $t_* < t^*$ zwei Bruchstellen von g^*, die je ein Intervall, auf dem g^* konstant oder streng monoton ist, nach links bzw. nach rechts begrenzen, und es sei g auf (t_*, t^*) monoton.*
Dann ist $g(t) = g^*(t)$ für alle $t \in (t_*, t^*)$.

Aus den Lemmata 3.11 bis 3.15 folgt also, daß g auf Intervallen, auf denen g^* konstant ist ($\mathcal{J}_1^{g^*}$), ebenfalls konstant oder sogar teilweise fallend ist. Der Wert, den g^* auf einem solchen Intervall annimmt, ist der mit w gewichtete Durchschnitt von g auf eben diesem Intervall.

Andererseits ist g^* gleich g auf den Intervallen, auf denen g^* streng monoton ist ($\mathcal{J}_2^{g^*}$).

Um g^* eindeutig festzulegen, genügt es also, die Menge der Bruchstellen dieser Funktion zu kennen. Dazu sind das folgende Korollar und das Lemma 3.17 hilfreich.

Korollar 3.16 *Sei $a \in g^*([0,1])$ beliebig fest.*
Dann gilt:

$$\int_0^{t^a} g^*(t)w(t)dt = \int_0^{t^a} g(t)w(t)dt \quad \textit{und} \quad \int_0^{t_a} g^*(t)w(t)dt = \int_0^{t_a} g(t)w(t)dt$$

Lemma 3.17 *Für alle $t_* \in [0,1]$ gilt:*

$$\int_0^{t_*} g^*(t)w(t)dt \preceq \int_0^{t_*} g(t)w(t)dt$$

3.3.5 Konstruktion des monotonen KQS

In diesem Abschnitt soll die Funktion konstruiert werden, die g im Kleinste-Quadrate-Sinne optimal monotonisiert. Es seien $u(t)$ und $v(t)$ mit $t \in [0,1]$ zwei Funktionen, die durch

$$u(t_0) := \int_0^{t_0} w(t)dt \quad \text{und} \quad v(t_0) := \int_0^{t_0} g(t)w(t)dt$$

definiert sind. Da w und g stetig sind, sind u und v stetig differenzierbar und u ist wegen $w(t) > 0$ für alle $t \in [0,1]$ zusätzlich streng monoton bezüglich $\leq$ und damit umkehrbar und die Umkehrung ist stetig. $\mathcal{U}$ und $\mathcal{V}$ seien die Wertebereiche von u und v. Sie sind beschränkt und konvex.

Dann läßt sich eine Funktion $\tilde{v} : \mathcal{U} \to \mathcal{V}$ mit $\tilde{v}(s) := v(u^{-1}(s))$ konstruieren. Wegen $u'(t) = w(t)$, läßt sich $\tilde{v}$ auch mittels

$$\tilde{v}(s_0) = \int_0^{s_0} g(u^{-1}(s))ds$$

angeben. Da u^{-1} stetig ist, ist $\tilde{v}$ stetig differenzierbar.

Definition 3.6 *Es sei $h : \mathcal{U} \to \mathcal{V}$ eine Funktion, wobei $\mathcal{U}$ und $\mathcal{V}$ konvexe Teilmengen von $\mathbb{R}$ seien sollen. Dann heißt h^* größte konvexe Minorante von h, falls*

1. *h^* konvex ist*
2. *für alle $s \in \mathcal{U}$ gilt $h^*(s) \preceq h(s)$ und*
3. *für jede Funktion h^+, die 1. und 2. erfüllt, gilt: $h^+(s) \preceq h^*(s)$ ist.*

Satz 3.18 *Zu jeder Funktion $h : \mathcal{U} \to \mathcal{V}$ existiert eine eindeutig bestimmte größte konvexe Minorante h^*.*

Die größte konvexe Minorante von $\tilde{v}$ sei nun mit v^* bezeichnet. Dann ist $g^*(t) := v^{*\prime}(u(t))$ die gesuchte Kleinste-Quadrate-Monotonisierung von g. Das wird in den nun folgenden Lemmata bewiesen.

Lemma 3.19 *Die Funktion $g^* : [0,1] \to \mathbb{R}$, die an allen Stellen, an denen v^* differenzierbar ist, durch*

$$g^*(t) := v^{*\prime}(u(t))$$

definiert ist und an den verbleibenden isolierten Punkten gleich der linksseitigen Ableitung von v^* an der Stelle $u(t)$ gesetzt wird, erfüllt die Eigenschaften der Kleinsten-Quadrate-Monotonisierung, die in den Lemmata 3.11 bis 3.17 gezeigt wurden.

Damit ist klar, daß die in Lemma 3.19 definierte Funktion g^* die Bedingungen der Lemmata 3.11 bis 3.17 erfüllt. Es bleibt zu zeigen, daß g^* die einzige Funktion ist, die diese Bedingungen erfüllt.

Lemma 3.20 *Es sei $h^* : [0,1] \to \mathbb{R}$ eine monotone Funktion, die die Bedingungen der Lemmata 3.11 bis 3.17 erfüllt.*

Dann ist h^* gleich der in Lemma 3.19 definierten Funktion g^*.

Also existiert genau eine monotone Funktion, die die Bedingungen der Lemmata 3.11 bis 3.17 erfüllt. Damit gilt der folgende Satz:

Satz 3.21 *Die Funktion $g^* : [0,1] \to \mathbb{R}$, die in Lemma 3.19 definiert ist, ist die Kleinste-Quadrate-Monotonisierung der Funktion g.*

Setzt man $g := \hat{g}_h$ und $w := n\hat{f}_h$, so ist $\hat{g}_h^* := g^*$ der bis auf Lebesgue-Nullmnegen eindeutig bestimmte monotone KQS der Regressionsfunktion m, falls nur $x_\cdot^{(n)} \in M_n^3$.

Beweis: Die erste Teil der Behauptung folgt aus der Existenz und Eindeutigkeit einer K-Q-Monotonisierung (Lemma 3.10) und aus Lemma 3.20.

Der zweite Teil folgt dann aus Satz 3.6.

□

3.4 Der monotone MgL-Schätzer der Regressionsfunktion

In diesem Abschnitt wird hergeleitet, daß der monotone KQS des vorherigen Abschnittes auch der monotone MgL-Schätzer ist. Dazu soll dieser Begriff zuerst definiert werden.

Definition 3.7 *Monotoner MgL-Schätzer*

Als monotoner MgL-Schätzer der Regressionsfunktion m wird eine monotone Funktion $\hat{m}_h^* : [0,1] \to \beta'(\Theta)$ bezeichnet, die das Funktional

$$\mathcal{L}^{sm}(m) := \sum_{i=1}^{n} \int_0^1 l_i(\psi(m(t)))K_h(t-x_i)dt$$

unter allen monotonen Funktionen $m : [0,1] \to \beta'(\Theta)$ maximiert.

Die drei folgenden Lemmata werden zum Beweis der oben angedeuteten Behauptung benötigt.

g und w seien wie im vorigen Abschnitt, g^* sei die K-Q-Monotonisierung der Funktion g.

Lemma 3.22 *Für eine beliebige integrierbare Funktion $\Psi : g^*([0,1]) \to \mathbb{R}$ gilt:*

$$\int_0^1 (g(t) - g^*(t))\Psi(g^*(t))w(t)dt = 0$$

In Anlehnung an Abschnitt 1.5 in Robertson, Wright und Dykstra [10] seien die folgenden Festlegungen getroffen:

Es sei $\Phi : \beta'(\Theta) \to \mathbb{R}$ eine differenzierbare und bezüglich der $\leq$ Relation streng konvexe Funktion. Ihre Ableitung sei mit φ bezeichnet.

Für $y, z \in \beta'(\Theta)$ sei die Funktion Δ_Φ definiert durch:

$$\Delta_\Phi(y,z) = \Phi(y) - \Phi(z) - (y-z)\varphi(z)$$

Für alle y und z gilt dann offensichtlich: $\Delta_\Phi(y,z) \geq 0$, wobei $=$ genau dann gilt, wenn $z = y$. Außerdem folgt aus der Festlegung von Δ_Φ:

$$\Delta_\Phi(r,t) = \Delta_\Phi(r,s) + \Delta_\Phi(s,t) + (r-s)(\varphi(s) - \varphi(t)) \qquad (3.v)$$

Nun läßt sich folgendes Lemma formulieren und beweisen:

Lemma 3.23 *Wenn $h : [0,1] \to \beta'(\Theta)$ monoton ist, dann gilt:*

$$\int_0^1 \Delta_\Phi(g(t), h(t))w(t)dt \geq \int_0^1 \Delta_\Phi(g(t), g^*(t))w(t)dt + \int_0^1 \Delta_\Phi(g^*(t), h(t))w(t)dt \tag{3.vi}$$

Folglich minimiert g^* das Funktional

$$\int_0^1 \Delta_\Phi(g(t), h(t))w(t)dt$$

unter allen monotonen Funktionen $h : [0,1] \to \beta'(\Theta)$ und maximiert in dieser Menge das Funktional

$$\int_0^1 [\Phi(h(t)) + (g(t) - h(t))\varphi(h(t))]w(t)dt.$$

Die minimierende (maximierende) Funktion ist eindeutig bestimmt.

Dieses Lemma erlaubt es nun, zu beweisen, daß g^* auch der monotone MgL-Schätzer ist. Es sei festgelegt:

$$\Phi(y_0) := \int_\chi^{y_0} \psi(y)dy$$

für ein festes $\chi \in \beta'(\Theta)$ und für jedes $y_0 \in \beta'(\Theta)$.

Da β' und mithin auch ψ streng monoton wachsend bezüglich $\leq$ ist, ist Φ streng konvex bezüglich $\leq$. Für φ erhält man dann: $\varphi(y) = \psi(y)$.

Es gilt das folgende Lemma:

Lemma 3.24 *Es sei $x_\cdot^{(n)} \in M_n^3$ und es seien $w(t) := \sum_{i=1}^n K_h(t - x_i)$ und $g(t) := (\sum_{i=1}^n Y_i K_h(t - x_i))/w(t)$.*

Für eine Funktion $l : [0,1] \to \beta'(\Theta)$ gilt dann:

$$\begin{aligned}\int_0^1 [\Phi(l(t)) \;&+\; (g(t) - l(t))\varphi(l(t))]w(t)dt \\ &= \int_0^1 \sum_{i=1}^n l_i(\psi(l(t)))K_h(t - x_i)dt + \tilde{\mathcal{C}}\end{aligned}$$

für ein festes $\tilde{\mathcal{C}}$.

Damit ergibt sich der folgende Satz:

Satz 3.25 *Falls $x_\cdot^{(n)} \in M_n^3$, dann ist der monotone MgL-Schätzer der Regressionsfunktion m gleich ihrem monotonen Kleinste-Quadrate-Schätzer.*

Beweis: Gemäß Lemma 3.24 wird $\mathcal{L}^{sm}(.)$ durch dieselben Funktionen maximiert wie das zweite Funktional in Lemma 3.23. Das jedoch wird – wenn g und w wie in Lemma 3.24 festgelegt werden – durch den monotonen KQS maximiert.

□

3.5 Modifizierter monotoner MgL-Schätzer

In diesem Abschnitt sollen asymptotische Eigenschaften des monotonen MgL-Schätzers hergeleitet werden. Dabei stellt es sich als notwendig heraus, den Schätzer etwas zu modifizieren, um Randprobleme zu umgehen.

Lemma 3.26 *Für den monotonen MgL-Schätzer $\hat{m}_h^*$ und den uneingeschränkten MgL-Schätzer $\hat{m}_h$ gilt:*

$$\sup_{t\in[0,1]} |\hat{m}_h^*(t) - m(t)| \leq \sup_{t\in[0,1]} |\hat{m}_h(t) - m(t)| \tag{3.vii}$$

Der Fehler des monotonen MgL-Schätzers ist in der Supremums-Norm gemessen also nicht größer als der Fehler des uneingeschränkten MgL-Schätzers. Das gilt jedoch nur, falls das Supremum über das gesamte Intervall $[0,1]$ genommen wird. Die asymptotische Konvergenzrate des uneingeschränkten MgL-Schätzers andererseits ist nur für das Supremum über das Intervall $[h_n, 1-h_n]$ bewiesen – und gilt in der Tat auch nur über diesem Intervall. Es läßt sich also nicht folgern, daß auch $\sup_{t\in[h,1-h]} |\hat{m}_h^*(t) - m(t)|$ mit der Rate $n^{-2/5}\sqrt{\ln n}$ gegen Null konvergiert.

Um dieses Problem zu umgehen, sei der folgende, leicht modifizierte monotone MgL-Schätzer definiert.

Definition 3.8 *Modifizierter monotoner MgL-Schätzer*

Als modifizierter monotoner MgL-Schätzer der Regressionsfunktion m wird eine monotone Funktion $\hat{m}_h^{*m} : [h, 1-h] \to \beta'(\Theta)$ bezeichnet, die das Funktional

$$\mathcal{L}_h^{sm}(m) := \sum_{i=1}^{n} \int_h^{1-h} l_i(\psi(m(t)))K_h(t-x_i)dt$$

unter allen monotonen Funktionen $m : [h, 1-h] \to \beta'(\Theta)$ maximiert.

Da die Konstruktion des monotonen MgL-Schätzers aus dem uneingeschränkten MgL-Schätzer lediglich von diesem und der Funktion $w(t) := \sum_{i=1}^n K_h(t-x_i)$ abhängt, kann $\hat{m}_h^{*m}$ auf dem Intervall $[h, 1-h]$ aus $\hat{m}_h$ genau so konstruiert werden. Es ist lediglich als Definitionsbereich der Funktionen $[h, 1-h]$ anstatt $[0,1]$ zugrunde zu legen. Auch Lemma 3.26 gilt entsprechend.

Der so modifizierte Schätzer „erbt“ nicht die Randprobleme des uneingeschränkten MgL-Schätzers. Deshalb hat $\hat{m}_h^{*m}$ die folgenden Eigenschaften:

Satz 3.27 *Sei $h_n := c_h n^{-1/5}$ für ein $c_h > 0$.*

Dann gilt:

$$\sup_{t\in[h_n,1-h_n]} |\hat{m}_h^{*m}(t) - m(t)| = O_P(n^{-2/5}\sqrt{\ln n})$$

und

$$\sup_{t\in[h_n,1-h_n]} |\hat{m}_h^{*m}(t) - \hat{m}_h(t)| = O_P(n^{-2/5}\sqrt{\ln n})$$

Beweis: Die erste Gleichung gilt wegen Satz 2.15 und Lemma 3.26. Die zweite folgt dann mit

$$\sup_{t\in[h_n,1-h_n]} |\hat{m}_h^{*m}(t) - \hat{m}_h(t)| \leq \sup_{t\in[h_n,1-h_n]} |\hat{m}_h^{*m}(t) - m(t)| + \sup_{t\in[h_n,1-h_n]} |m(t) - \hat{m}_h(t)|$$

□

3.6 Beweise

Es folgen nun die Beweise aus Kapitel 3, die zu technisch sind, oder deren Ausführung für die gesamte Arbeit von untergeordnetem Interesse ist.

Beweis von Lemma 3.1:

Sei $t_0 \in (0,1)$. Da h monoton ist, ist für hinreichend kleines $\varepsilon > 0$ die Menge $h((t_0-\varepsilon, t_0))$ nach oben und die Menge $h((t_0, t_0+\varepsilon))$ nach unten beschränkt. Damit hat $\{h(t_n)\}_{n\in\mathbb{N}}$ für jede Folge $\{t_n\}_{n\in\mathbb{N}}$, die von links oder von rechts gegen t_0 konvergiert, wenigstens einen Häufungspunkt. Hätte sie mehr als einen, so müßten zwei Teilfolgen existieren, die gegen zwei verschiedene Häufungspunkte konvergieren. Dann könnte h aber nicht monoton sein. Somit müssen die links- und rechtsseitigen Grenzwerte existieren.

Für $t_0 = 0$ und $t_0 = 1$ existiert der rechts- bzw. linksseitige Grenzwert entsprechend.

□

Beweis von Lemma 3.2:

Es sei $h : [0,1] \to \mathbb{R}$ eine monotone Funktion. Dann sei definiert:

- $M_1^h := \{x \in h([0,1]) : \int_{[h=x]} dt > 0\}$, wobei $[h = x] := \{t \in [0,1] : h(t) = x\}$. Da h monoton ist, ist $[h = x]$ konvex in $\mathbb{R}$.
- $M_2^h := h([0,1])/M_1^h$

Dann hat $h^{-1}(\{x\})$ für jedes $x \in M_2^h$ genau ein Element, denn aus $t_1 < t_2 \in h^{-1}(\{x\})$ folgt $h(t) = x$ für alle $t \in [t_1, t_2]$ und damit $\int_{[h=x]} dx \geq t_2 - t_1 > 0$. Dann aber wäre $x \in M_1^h$.

Sei nun

$$\mathcal{J}_1^h := \{J_x^h := (\inf[h = x], \sup[h = x]) | \, x \in M_1^h\}$$

und

$$\mathcal{J}_2^h := \{J_x^h | \, x \in M_2^h\}/\{\emptyset\},$$

wobei für $x \in M_2^h$ definiert sei:

$$J_x^h := (\sup\{\bigcup_{y\in M_1^h;\, y\preceq x} J_y^h\}, \inf\{\bigcup_{y\in M_1^h;\, x\preceq y} J_y^h\})$$

Damit enthalten $\mathcal{J}_1^h$ und $\mathcal{J}_2^h$ nur offene Intervalle und es gelten die folgenden Eigenschaften:

1. Die Menge $M := [0,1]/(\bigcup_{x \in M_1^h \cup M_2^h} J_x^h)$ enthält nur isolierte Punkte.
2. Seien $x \in M_1^h$ und $y \in M_2^h$. Dann sind J_x^h und J_y^h disjunkt.
3. h ist auf jedem Intervall $(t_*, t^*) \in J_2^h$ streng monoton.

Beweis:

1. Angenommen, es gäbe ein Intervall $(t_*, t^*) \subseteq M$. Dann wäre dieses Intervall disjunkt mit J_x^h für jedes $x \in M_1^h \cup M_2^h$ nach Definition von M.

   Speziell wäre $(t_*, t^*) \cap J_y^h = \emptyset$ für alle $y \in M_1^h$. Daraus folgt für ein $x \in h((t_*, t^*))$:

   $$\sup\{\bigcup_{y \in M_1^h;\, y \preceq x} J_y^h\} \leq t_* \quad \text{und} \quad t^* \leq \inf\{\bigcup_{y \in M_1^h;\, x \preceq y} J_y^h\}$$

   Dann aber ist $(t_*, t^*) \subseteq J_x^h$, wobei $x \in M_2^h$. Das aber ist ein Widerspruch zur Voraussetzung.

2. Wenn $x \preceq y$, dann gilt:

   $$\sup J_x^h \leq \sup\{\bigcup_{z \in M_1^h;\, z \preceq y} J_z^h\} < t \quad \forall t \in J_y^h$$

   Also sind J_x^h und J_y^h disjunkt.

   Analog, falls $y \preceq x$:

   $$\forall t \in J_y^h: \;\; t < \inf\{\bigcup_{z \in M_1^h;\, y \preceq z} J_z^h\} \leq \sup J_x^h$$

   Auch in diesem Fall sind J_x^h und J_y^h disjunkt.

3. Angenommen, es gäbe $t_1 < t_2$ aus (t_*, t^*) so, daß $h(t_1) = h(t_2)$.
   Dann wäre $h((t_1, t_2)) = \{h(t_1)\}$ und damit auch $h(t_1) \in M_1^h$ und $(t_1, t_2) \subseteq J_{h(t_1)}^h$. Dann aber wäre $(t_*, t^*) \cap J_{h(t_1)}^h \neq \emptyset$, was im Widerspruch zu 2. stehen würde.

$\square$

Beweis von Lemma 3.3:

1. Sei $x_1 := x + h_1$ und $x_2 := x + h_2$ und sei $\alpha := h_2/(h_2 - h_1)$. Dann gilt wegen der Konvexität von f:

   $$h(\alpha x_1 + (1-\alpha)x_2) \preceq \alpha h(x_1) + (1-\alpha)h(x_2)$$

   Daraus folgt $(h_2 - h_1)h(x) \preceq h_2 h(x_1) - h_1 h(x_2)$, woraus die Behauptung folgt.

2. Angenommen, es gilt $D_{x,h_2} \prec D_{x,h_1}$.
   Dann folgt mit $\alpha := (h_2 - h_1)/h_2$:
   $$\begin{aligned} h(\alpha x + (1-\alpha)(x+h_2)) = h(x+h_1) &\succ \frac{h_1}{h_2}(h(x+h_2) - h(x)) + h(x) \\ &= \alpha h(x) + (1-\alpha)h(x+h_2), \end{aligned}$$
   was ein Widerspruch zur Konvexität von h ist.
3. Analog 2.
4. Folgt aus 2. und 3.: $D_{x_1,h_1} \preceq D_{x_1,x_2-x_1} = D_{x_2,x_1-x_2} \preceq D_{x_2,h_2}$

□

Beweis von Lemma 3.4:

Aus Lemma 3.3 folgt: Ein linksseitiger Differenzenquotient in $x_0 \in (0,1)$ wird nach oben durch einen beliebigen rechtsseitigen Differenzenquotienten in x_0 beschränkt. Außerdem ist er monoton wachsend für $h_0 \longrightarrow 0-$. Also existiert die linksseitige Ableitung bei x_0. Analog existiert die rechtsseitige Ableitung und sie ist größer oder gleich der linksseitigen.

Außerdem ist für $x_1 < x_2$ die rechtsseitige Ableitung bei x_1 kleiner oder gleich der linksseitigen Ableitung bei x_2.

Es sei nun die Funktion h^l der linksseitigen Ableitungen von h betrachtet. Offensichtlich ist sie monoton. Wendet man die Kenntnisse aus den Lemmata 3.1 und 3.2 auf h^l an, so läßt sich $[0,1]$ in paarweise disjunkte offene Intervalle zerlegen, auf denen h^l entweder streng monoton oder konstant ist. Es bleiben isolierte Bruchstellen übrig. Außerdem ist h^l bis auf Sprungstellen stetig. Wegen 4. von Lemma 3.3 muß die rechtsseitige Ableitung an allen Stellen von $(0,1)$ bis auf die Sprungstellen von h^l gleich der linksseitigen Ableitung sein.

Also ist h bis auf isolierte Punkte differenzierbar und die Ableitung h' hat dieselben Eigenschaften wie h^l.

□

Beweis von Satz 3.5:

Analog dem Beweis von Satz 2.6, wobei

$$\frac{d}{d\hat{g}_h(t)} \sum_{i=1}^{n} (Y_i - \hat{g}_h(t))^2 K_h(t - x_i) = -2 \sum_{i=1}^{n} (Y_i - \hat{g}_h(t)) K_h(t - x_i)$$

und

$$\frac{d^2}{d(\hat{g}_h(t))^2} \sum_{i=1}^{n} (Y_i - \hat{g}_h(t))^2 K_h(t - x_i) = 2 \sum_{i=1}^{n} K_h(t - x_i)$$

□

Beweis von Satz 3.6:

Nach Satz 3.5 gilt bis auf Nullmengen:

$$\sum_{i=1}^{n}(Y_i - \hat{g}_h(t))K_h(t-x_i) = \sum_{i=1}^{n} Y_i K_h(t-x_i) - \hat{g}_h(t)\sum_{i=1}^{n} K_h(t-x_i) = 0$$

Damit gilt ebenfalls:

$$\int_0^1 \sum_{i=1}^{n}(Y_i - \hat{g}_h(t))(\hat{g}_h(t) - g(t))K_h(t-x_i)dt = 0$$

für alle Funktionen g.

Also läßt sich $\mathcal{L}^q(g)$ darstellen durch:

$$\int_0^1 \sum_{i=1}^{n}(Y_i - \hat{g}_h(t))^2 K_h(t-x_i)dt + \int_0^1 (\hat{g}_h(t) - g(t))^2 w_h(t)dt$$

Da der linke Summand von g nicht abhängt, ist die Funktion $\hat{g}_h^*$ so zu wählen, daß sie den rechten Summanden minimiert.

□

Beweis von Lemma 3.8:

- Abgeschlossen
  Es sei $\{g_n\}_{n\in\mathbb{N}}$ eine Folge aus C, die gegen $g^* \in H$ konvergiert.
  z.Z.: $g^* \in C$
  Seien $N_n \subseteq [0,1]$ solche Nullmengen, daß jedes g_n außerhalb N_n monoton ist, und $N_* \subseteq [0,1]$ sei eine solche Nullmenge, daß $\forall t \in [0,1]/N_*$: $g_n(t) \underset{n\to\infty}{\longrightarrow} g^*(t)$ gilt – wäre N_* keine Nullmenge, so könnte $||g^* - g_n||^2$ nicht gegen Null konvergieren.

  Dann ist $N := N_* \cup (\bigcup_{n=1}^{\infty} N_n)$ eine Nullmenge. Für zwei beliebige $t_1 < t_2$ aus $[0,1]/N$ gilt aber:

  $$g_n(t_1) \preceq g_n(t_2) \;\; \forall n \in \mathbb{N} \;\text{ und }\; g_n(t_i) \underset{n\to\infty}{\longrightarrow} g^*(t_i) \;\text{ für }\; i = 1,2$$

  Daraus folgt: $g^*(t_1) \preceq g^*(t_2)$.

- Konvex
  Seien g_1 und g_2 aus C; N sei die Nullmenge, außerhalb derer beide Funktionen monoton sind. $\alpha \in [0,1]$ sei beliebig fest.
  Dann gilt für beliebige $t_1 < t_2$ aus $[0,1]/N$:

  $$\alpha g_1(t_1) + (1-\alpha)g_2(t_1) \preceq \alpha g_1(t_2) + (1-\alpha)g_2(t_2)$$

  Damit ist auch $\alpha g_1 + (1-\alpha)g_2$ L_2-monoton.

□

Beweis von Satz 3.9:

Es sei $\{g_n\}_{n\in\mathbb{N}} \subseteq C$ so gewählt, daß $||g_n - g|| \underset{n\to\infty}{\longrightarrow} \gamma := \inf_{f\in C} ||f-g||$.

Für ein beliebiges $\varepsilon > 0$ gilt nun:

$$\exists N_\varepsilon \;\; \forall n,m \geq N_\varepsilon : \;\; \gamma \leq ||g-g_n||, ||g-g_m|| \leq \gamma+\varepsilon$$

Außerdem ist $h := (g_n+g_m)/2 \in C$, da C konvex ist und deshalb gilt auch: $\gamma \leq ||g-h||$. Damit folgt:

$$\begin{aligned}
2\gamma^2 + ||g_n-g_m||^2/2 &= 2\gamma^2 + ||h-g_n||^2 + ||h-g_m||^2 \\
&\leq 2||g-h||^2 + ||h-g_n||^2 + ||h-g_m||^2 \\
&\quad +2\langle g-h, 2h-g_n-g_m\rangle, \quad \text{da } 2h-g_n-g_m = 0 \\
&= ||g-h+h-g_n||^2 + ||g-h+h-g_m||^2 \\
&\leq 2(\gamma+\varepsilon)^2
\end{aligned}$$

Also gilt die Ungleichung $||g_n - g_m|| \leq \sqrt{8\varepsilon\gamma + 4\varepsilon^2} \underset{\varepsilon\to 0}{\longrightarrow} 0$. Damit ist $\{g_n\}_{n\in\mathbb{N}}$ eine Cauchy-Folge in C, konvergiert also gegen ein $g^* \in C$, da C abgeschlossen und H vollständig ist.

$||g-f||$ wird offensichtlich durch g^* unter allen $f \in C$ minimiert. Sei nun $f \in C$ beliebig fest. Dann ist $(1-\alpha)g^* + \alpha f \in C$ für alle $\alpha \in [0,1]$ und $||g - ((1-\alpha)g^* + \alpha f)||^2$ nimmt bezüglich α bei $\alpha = 0$ den minimalen Wert γ an. Daraus folgt, daß $\frac{d}{d\alpha}||g - [(1-\alpha)g^* + \alpha f]||^2 \,|_{\alpha=0+} \geq 0$, da sonst $||g - .||^2$ für ein $\alpha > 0$ kleiner wäre als γ, was aber nicht möglich ist. Andererseits ist diese Ableitung das Doppelte von (3.ii). (3.iii) folgt unmittelbar aus (3.ii). In (3.ii) kann man $f := cg^*$ einmal für ein $c > 1$ und einmal für ein $c \in (0,1)$ setzen und erhält damit (3.i).

Angenommen, es existieren zwei Funktionen u_1 und u_2 in C, die den Abstand zu g minimieren. Dann gilt nach (3.ii):

$$\langle g-u_1, u_1-u_2\rangle \geq 0 \;\text{ und }\; \langle u_2-g, u_1-u_2\rangle \geq 0$$

Addiert man beide Ungleichungen, so erhält man:

$$||u_1-u_2||^2 \leq 0$$

Und daraus folgt $u_1 = u_2$.

Sei nun $u \in C$ wie in (3.iv) gegeben. Dann gilt für jedes $f \in C$:

$$||g-f||^2 = ||g-u||^2 + ||u-f||^2 + 2\langle g-u, u-f\rangle \geq ||g-u||^2$$

Also minimiert auch u den Abstand zu g und muß deshalb gleich g^* sein.

□

Beweis von Lemma 3.10:

Unter dem essentiellen Supremum und dem essentiellen Infimum einer Funktion $h : [0,1] \to \mathbb{R}$ seien die Terme

$$\inf_{\substack{N \subset [0,1] \\ N \text{ ist Nullmenge}}} \; \sup_{t\in[0,1]/N} h(t) \quad \text{bzw.} \quad \sup_{\substack{N \subset [0,1] \\ N \text{ ist Nullmenge}}} \; \inf_{t\in[0,1]/N} h(t)$$

verstanden.

Da g beschränkt ist, sind das Supremum und das Infimum von g endlich. Sie sind auch die Grenzen für das essentielle Supremum und das essentielle Infimum von g^*. Ansonsten würde ein h^*, das überall dort, wo diese Grenzen über- bzw. unterschritten werden, diese Grenzen annimmt und sonst gleich g^* ist, L_2-monoton sein, aber g besser monotonisieren. Also sind auch das essentielle Supremum und das essentielle Infimum von g^* endlich.

Nun existiert eine Nullmenge N, so daß $g^*(t_1) \preceq g^*(t_2)$ für beliebige $t_1 < t_2$ außerhalb N gilt und g^* außerhalb von N beschränkt ist. Für ein beliebiges $t \in N$ ist dann die Menge $M_t := \{\alpha \in \mathbb{R} : \ \forall s \in [0,t]/N \ \forall u \in [t,1]/N \ g^*(s) \preceq \alpha \preceq g^*(u)\}$ nichtleer. Außerdem ist $\bigcup_{t \in N \cap (0,1)} M_t$ durch $\inf_{t \in [0,1]/N} g^*(t)$ und $\sup_{t \in [0,1]/N} g^*(t)$ beschränkt.

Man setze $\tilde{g}^*(t) := g^*(t)$ für $t \in [0,1]/N$ und wähle für jedes $t \in N$ einen Funktionswert aus M_t. Dann ist $\tilde{g}^*$ monoton und beschränkt.

□

Beweis von Lemma 3.11:

Falls $\int_{[g^*=c]} w(t)dt = 0$, so ist auch $\int_{[g^*=c]} g(t)w(t)dt = 0$. Sei also c so, daß $\int_{[g^*=c]} w(t)dt > 0$. Dann gilt:

$$\int_0^1 (g(t) - g^*(t))^2 w(t)dt = \int_{\overline{[g^*=c]}} (g(t) - g^*(t))^2 w(t)dt + \int_{[g^*=c]} (g(t) - c)^2 w(t)dt$$

Der Term $\int_{[g^*=c]} (g(t) - \xi)^2 w(t)dt$ wird minimal, falls

$$\xi := \frac{\int_{[g^*=c]} g(t)w(t)dt}{\int_{[g^*=c]} w(t)dt}$$

Sei $\varepsilon := c - \xi$

1. Angenommen, $\varepsilon > 0$ und $\preceq := \leq$.
   Dann sei h^*_α für jedes $\alpha \in [0,1]$ definiert durch

$$h^*_\alpha(t) := \begin{cases} g^*(t), & \text{falls } t \notin [t_{c-\alpha\varepsilon}, t^c] \\ c - \alpha\varepsilon, & \text{sonst} \end{cases}$$

   Dieses h^*_α ist monoton. Jedoch gilt:

$$\begin{aligned} ||g - h^*_\alpha||^2 &= \int_{\overline{[t_{c-\alpha\varepsilon}, t^c]}} (g(t) - g^*(t))^2 w(t)dt + \int_{t_{c-\alpha\varepsilon}}^{t^c} (g(t) - c + \alpha\varepsilon)^2 w(t)dt \\ &= \int_{\overline{[t_{c-\alpha\varepsilon}, t^c]}} (g(t) - g^*(t))^2 w(t)dt + \int_{t_c}^{t^c} (g(t) - c + \alpha\varepsilon)^2 w(t)dt \\ &\quad + \int_{t_{c-\alpha\varepsilon}}^{t_c} (g(t) - g^*(t))^2 w(t)dt + \int_{t_{c-\alpha\varepsilon}}^{t_c} (g^*(t) - c + \alpha\varepsilon)^2 w(t)dt \\ &\quad + 2\int_{t_{c-\alpha\varepsilon}}^{t_c} (g(t) - g^*(t))(g^*(t) - c + \alpha\varepsilon) w(t)dt \end{aligned}$$

$$\begin{aligned}
&\leq \int_{\overline{[t_{c-\alpha\varepsilon},t^c]}} (g(t) - g^*(t))^2 w(t)dt + \int_{t_c}^{t^c} (g(t) - g^*(t) + \alpha\varepsilon)^2 w(t)dt \\
&\quad + \int_{t_{c-\alpha\varepsilon}}^{t_c} (g(t) - g^*(t))^2 w(t)dt + (\alpha\varepsilon)^2 \int_{t_{c-\alpha\varepsilon}}^{t_c} w(t)dt \\
&\quad +2\alpha\varepsilon \int_{t_{c-\alpha\varepsilon}}^{t_c} |(g(t) - g^*(t))| w(t)dt \\
&= \int_{\overline{[t_c,t^c]}} (g(t) - g^*(t))^2 w(t)dt + \int_{t_c}^{t^c} (g(t) - g^*(t))^2 w(t)dt \\
&\quad +(\alpha\varepsilon)^2 \int_{t_{c-\alpha\varepsilon}}^{t^c} w(t)dt + 2\alpha\varepsilon \int_{t_c}^{t^c} (g(t) - c) w(t)dt \\
&\quad +2\alpha\varepsilon \int_{t_{c-\alpha\varepsilon}}^{t_c} |g(t) - g^*(t)| w(t)dt \\
&\leq \int_0^1 (g(t) - g^*(t))^2 w(t)dt \\
&+ \ (\alpha\varepsilon)^2 \left(\int_{t_{c-\alpha\varepsilon}}^{t^c} w(t)dt + 2 \sup_{[t_{c-\alpha\varepsilon},t_c]} |g(t) - g^*(t)| w(t) \right) \\
&- \ 2\alpha\varepsilon \int_{t_c}^{t^c} \varepsilon w(t)dt + 2\alpha\varepsilon \int_{t_c}^{t^c} (g(t) - \xi) w(t)dt,
\end{aligned}$$

wobei der letzte Summand nach Definition von ξ Null ist.

$$= ||g - g^*||^2 + O(\alpha^2) - 2\alpha\varepsilon^2 \int_{t_c}^{t^c} w(t)dt$$

Ferner ist $\int_{t_c}^{t^c} w(t)dt > 0$ nach Voraussetzung. Damit existiert ein $\alpha > 0$, so daß $||g - h_\alpha^*||^2 < ||g - g^*||^2$. Das aber ist ein Widerspruch zur Optimalität von g^*.

2. Die anderen Fälle mit $\varepsilon \neq 0$ führen analog zum Widerspruch.

Also folgt: $c = \xi$ und damit die Behauptung.

□

Beweis von Lemma 3.12:

Angenommen, g^* ist auf (a, b) nicht konstant.
Dann existieren $t_* < t^* \in (a, b)$ so, daß $g^*(t_*) \prec g^*(t^*)$.

1. Sei $g(t_*) \preceq g^*(t_*)$.
   Man setze
   $$h^*(t) := \begin{cases} g^*(t_*), & \text{falls}\, t \in (t_*, b) \\ g^*(t) & \text{sonst} \end{cases}$$
   Dieses h^* ist monoton und auf dem Intervall $[t_*, t^*]$ gilt
   $$g(t) \preceq g(t_*) \preceq g^*(t_*) = h^*(t) \preceq g^*(t),$$
   und auf dem Intervall (t^*, b) gilt:
   $$g(t) \preceq g(t_*) \preceq g^*(t_*) = h^*(t) \prec g^*(t).$$

Damit folgt $\int_{t_*}^b (h^*(t) - g(t))^2 w(t)dt < \int_{t_*}^b (g^*(t) - g(t))^2 w(t)dt$, während außerhalb von $[t_*, b]$ die Funktionen h^* und g^* gleich sind. Das aber ist ein Widerspruch zur Optimalität von g^*.

2. Sei $g^*(t^*) \preceq g(t^*)$.
   Mit
   $$h^*(t) := \begin{cases} g^*(t^*), & \text{falls}\, t \in (a, t^*) \\ g^*(t) & \text{sonst} \end{cases}$$
   ergibt sich analog zu 1. ein Widerspruch.

3. Alle anderen Fälle, d.h.
   $$g^*(t_*) \prec g(t_*) \;\text{ und }\; g(t^*) \prec g^*(t^*)$$
   Es sei definiert:
   $$s_* := \inf_{s \in (a,b)} \{g(s) \preceq g^*(s)\} \;\text{ und }\; s^* := \sup_{s \in (a,b)} \{g^*(s) \preceq g(s)\}$$
   Aus dieser Definition folgt $g(s) = g^*(s)$ für alle $s \in (s_*, s^*)$ und $g(s) \neq g^*(s)$ für alle $s \notin [s_*, s^*]$.

   Da ferner g^* wachsend und g fallend sind, kann $g((s_*, s^*)) = g^*((s_*, s^*))$ höchstens ein Element enthalten. Diese Menge ist leer, genau dann, wenn $s_* = s^*$.

   Aus den Annahmen dieses Falles folgen weiterhin die Ungleichungen $t_* \leq s_* \leq s^* \leq t^*$.

   Es sei $\alpha := g(s^*)$. Wegen der Stetigkeit von g gilt dann auch $\alpha = g(s_*)$.

   Dann ist die Funktion
   $$h^*(t) := \begin{cases} \alpha, & \text{falls }\; t \in (a, b) \\ g^*(t) & \text{sonst} \end{cases}$$
   monoton und es gelten die Ungleichungen
   $$\forall t \in (a, s_*) : \; g^*(t) \preceq g^*(s_*) \preceq \alpha = g(s_*) \preceq g(t)$$
   und
   $$\forall t \in (s^*, b) : \; g(t) \preceq g(s^*) = \alpha \preceq g^*(s^*) \preceq g^*(t)$$
   und mithin die beiden Abschätzungen
   $$\begin{aligned} \int_a^{s_*} (h^*(t) - g(t))^2 w(t)dt &= \int_a^{s_*} (\alpha - g(t))^2 w(t)dt \\ &\leq \int_a^{s_*} (g^*(t) - g(t))^2 w(t)dt \end{aligned}$$
   und
   $$\begin{aligned} \int_{s^*}^b (h^*(t) - g(t))^2 w(t)dt &= \int_{s^*}^b (\alpha - g(t))^2 w(t)dt \\ &\leq \int_{s^*}^b (g^*(t) - g(t))^2 w(t)dt, \end{aligned}$$

wobei in einer der beiden Ungleichungen $<$ gilt, da sonst $g^*(t_*) = \alpha = g^*(t^*)$ gelten müßte, was im Widerspruch zur Annahme steht.

Somit hat auch dieses h^* einen kleineren Abstand zu g als g^*, was ein Widerspruch ist.

Alle möglichen Fälle sind zum Widerspruch geführt.

□

Beweis von Lemma 3.13:
Entsprechend der Definition von t_c und t^c muß $g^*(s) \prec c \prec g^*(u)$ für alle $s < t_c$ und für alle $u > t^c$ gelten.

1. Angenommen, $c \neq g(t_c)$.

   (a) Angenommen, $c \prec g(t^c)$.
   Dann existiert wegen der Stetigkeit von g ein $\varepsilon > 0$, so daß $c \prec g(t)$ für alle $t \in (t_c - \varepsilon, t_c)$. Man setze

   $$h^*(t) := \begin{cases} g^*(t), & \text{falls}\, t \notin (t_c - \varepsilon, t_c) \\ c & \text{sonst.} \end{cases}$$

   Dieses h^* ist monoton und hat auf $(t_c - \varepsilon, t_c)$ einen kleineren Abstand zu g als g^*. Damit ist auch die Monotonisierung von g besser. Das aber ist ein Widerspruch zur Optimalität von g^*.

   (b) Angenommen, $g(t_c) \prec c$.
   Dann existiert wegen der Stetigkeit von g ein $\varepsilon > 0$, so daß $g(t) \prec (c + g(t_c))/2$ für alle $t \in (t_c - \varepsilon, t_c + \varepsilon)$. Man setze

   $$h^*(t) := \begin{cases} g^*(t), & \text{falls}\, t \notin (t_c - \varepsilon, t_c + \varepsilon) \\ d & \text{sonst,} \end{cases}$$

   wobei $d := \max\{g^*(t_c - \varepsilon), (c + g(t_c))/2\} \prec c$. Dann ist dieses h^* monoton und hat auf $(t_c - \varepsilon, t_c + \varepsilon)$ einen kleineren Abstand zu g als g^*. Also ist auch hier die Monotonisierung von g durch h^* besser als durch g^*. Das aber ist ein Widerspruch zur Optimalität von g^*.

2. Angenommen, $c \neq g(t^c)$.
   Dieser Fall kann analog zum Widerspruch geführt werden.

□

Beweis von Lemma 3.14:
Wegen der Monotonie von g^* muß $g^*(s) \preceq c_*$ und $c^* \preceq g^*(u)$ für alle $s < t_*$ und für alle $u > t^*$ gelten.

1. Angenommen, $c_* \neq g(t_*)$.

(a) Angenommen, $c_* \prec g(t_*)$.
Weil g stetig ist und weil $g^*(t) \underset{t \to t_*+}{\longrightarrow} c_*$, existiert ein $\varepsilon > 0$ so, daß für alle $t \in (t_*, t_* + \varepsilon)$ gilt:

$$\delta := (c_* + g(t_*))/2 \prec g(t) \;\text{ und }\; g^*(t) \prec \delta$$

Man setze:

$$h^*(t) := \begin{cases} g^*(t), & \text{falls}\, t \notin (t_*, t_* + \varepsilon) \\ g^*((t_* + \varepsilon)-) & \text{sonst.} \end{cases}$$

Dann ist h^* monoton und hat auf $(t_*, t_* + \varepsilon)$ einen kleineren Abstand zu g als g^*. Damit ist auch die Monotonisierung von g besser. Das aber ist ein Widerspruch zur Optimalität von g^*.

(b) Angenommen, $g(t_*) \prec c_*$.
Dann existiert wegen der Stetigkeit von g ein $\varepsilon > 0$ so, daß für alle $t \in (t_*, t_* + \varepsilon)$ gilt: $g(t) \prec c_*$. Setzt man nun

$$h^*(t) := \begin{cases} g^*(t), & \text{falls}\, t \notin (t_*, t_* + \varepsilon) \\ c_* & \text{sonst,} \end{cases}$$

Dieses h^* ist monoton und hat auf $(t_*, t_* + \varepsilon)$ einen kleineren Abstand zu g als g^* zu g. Also ist auch die Monotonisierung von g durch h^* besser als durch g^*. Das aber ist ein Widerspruch zur Optimalität von g^*.

2. Angenommen, $c^* \neq g(t^*)$.
Dieser Fall kann analog zum Widerspruch geführt werden.

□

Beweis von Lemma 3.15:

Sei $t_* > 0$ und $t^* < 1$. Dann folgt aus den Lemmata 3.13 und 3.14, daß $g(t_*) = g^*(t_*+)$ und $g(t^*) = g^*(t^*-)$ gilt.

Sollte die Behauptung nicht gelten, dann ist

$$h^*(t) := \begin{cases} g^*(t), & \text{falls}\, t \notin (t_*, t^*) \\ g(t) & \text{sonst,} \end{cases}$$

eine monotone Funktion und hat einen kleineren Abstand zu g als g^*, was ein Widerspruch ist.

Ist $t_* = 0$ oder $t^* = 1$, so ist h^* ebenfalls monoton, falls man zusätzlich noch $h^*(0) := g(0)$ bzw. $h^*(1) := g(1)$ setzt. Wiederum ist der Abstand zwischen h^* und g kleiner als der zwischen g^* und g.

□

Beweis von Korollar 3.16:

Die Menge aller Bruchstellen der Funktion g^* sei mit M bezeichnet. Ferner sei für ein $\varsigma \in M$ definiert: $\varsigma_+ := \inf\{\tau \in M : \tau > \varsigma\}$. Sollte $\{\tau \in M : \tau > \varsigma\}$ leer sein, so sei $\varsigma_+ := 1$.

Offensichtlich gilt wegen Lemma 3.11 und Lemma 3.15:

$$\int_{\varsigma}^{\varsigma_+} g^*(t)w(t)dt = \int_{\varsigma}^{\varsigma_+} g(t)w(t)dt$$

für alle $\varsigma \in M$. Denn falls nicht $\varsigma = \varsigma_+$ gilt, so liegt zwischen beiden Integralgrenzen entweder ein konstanter Abschnitt von g^* oder ein streng monotoner.

Setzt man $\varsigma^0 := \sup\{\varsigma \in M | \, \varsigma_+ \leq t^a\}$, so ist $\varsigma_+^0 < t^a$ höchstens dann, wenn g^* auf $(\varsigma_+^0, \varsigma_{++}^0)$ wachsend und damit gleich g ist. Folglich gilt:

$$\begin{aligned}
\int_0^{t^a} g^*(t)w(t)dt &= \sum_{\varsigma \in M:\, \varsigma_+ \leq t^a} \int_{\varsigma}^{\varsigma_+} g^*(t)w(t)dt + \int_{\varsigma_+}^{t^a} g^*(t)w(t)dt \\
&= \sum_{\varsigma \in M:\, \varsigma_+ \leq t^a} \int_{\varsigma}^{\varsigma_+} g(t)w(t)dt + \int_{\varsigma_+}^{t^a} g(t)w(t)dt \\
&= \int_0^{t^a} g(t)w(t)dt
\end{aligned}$$

Wegen $\int_{t_a}^{t^a} g^*(t)w(t)dt = \int_{t_a}^{t^a} g(t)w(t)dt$ in jedem Fall, gilt auch die zweite Gleichung.

□

Beweis von Lemma 3.17:

Die Behauptung wird nur für $\preceq := \leq$ gezeigt. Für $\preceq := \geq$ kann der Beweis analog geführt werden.

Angenommen, es existiert ein $t_* \in [0, 1]$ so, daß

$$\int_0^{t_*} g(t)w(t)dt + \varepsilon < \int_0^{t_*} g^*(t)w(t)dt$$

für ein $\varepsilon > 0$ gilt. Sei $c := g^*(t_*)$.

Aus Korollar 3.16 folgt: $\int_0^{t_c} g(t)w(t)dt = \int_0^{t_c} g^*(t)w(t)dt$. Also ist

$$\varepsilon < \int_{t_c}^{t_*} (g^*(t) - g(t))w(t)dt \tag{3.viii}$$

und deshalb auch $t_c < t_*$.

Es sei nun die Funktion h_α^* für ein $\alpha > 0$ betrachtet:

$$h_\alpha^*(t) := \begin{cases} g^*(t), & \text{falls}\, t \notin [t_{c-\alpha\varepsilon}, t_*] \\ c - \alpha\varepsilon & \text{sonst} \end{cases}$$

Dieses h_α^* ist monoton. Jedoch gilt:

$$
\begin{aligned}
||h_\alpha^* - g||^2 &= \int_{\overline{[t_{c-\alpha\varepsilon}, t_*]}} (g^*(t) - g(t))^2 w(t)dt + \int_{t_{c-\alpha\varepsilon}}^{t_*} (c - \alpha\varepsilon - g(t))^2 w(t)dt \\
&= \int_{\overline{[t_{c-\alpha\varepsilon}, t_*]}} (g^*(t) - g(t))^2 w(t)dt \\
&\quad + \int_{t_{c-\alpha\varepsilon}}^{t_c} (c - \alpha\varepsilon - g^*(t))^2 w(t)dt + \int_{t_{c-\alpha\varepsilon}}^{t_c} (g^*(t) - g(t))^2 w(t)dt \\
&\quad +2 \int_{t_{c-\alpha\varepsilon}}^{t_c} (c - \alpha\varepsilon - g^*(t))(g^*(t) - g(t))w(t)dt \\
&\quad + \int_{t_c}^{t_*} (c - \alpha\varepsilon - g^*(t))^2 w(t)dt + \int_{t_c}^{t_*} (g^*(t) - g(t))^2 w(t)dt \\
&\quad +2 \int_{t_c}^{t_*} (c - \alpha\varepsilon - g^*(t))(g^*(t) - g(t))w(t)dt \\
&\leq ||g^* - g||^2 + \int_{t_{c-\alpha\varepsilon}}^{t_c} (\alpha\varepsilon)^2 w(t)dt \\
&\quad +2 \int_{t_{c-\alpha\varepsilon}}^{t_c} \alpha\varepsilon |g^*(t) - g(t)| w(t)dt \\
&\quad + \int_{t_c}^{t_*} (\alpha\varepsilon)^2 w(t)dt - 2 \int_{t_c}^{t_*} \alpha\varepsilon (g^*(t) - g(t))w(t)dt \\
&= ||g^* - g||^2 + O((\alpha\varepsilon)^2) - 2\alpha\varepsilon \int_{t_c}^{t_*} (g^*(t) - g(t))w(t)dt
\end{aligned}
$$

Wegen (3.viii) ist die letzte Zeile für ein α nahe genug bei Null kleiner als $||g^* - g||^2$, was ein Widerspruch zur Optimalität von g^* ist.

□

Beweis von Satz 3.18:

Die Eindeutigkeit ist klar: Gäbe es zwei voneinander verschiedene größte konvexe Minoranten, so so erfüllen beide 1. und 2. der Definition, aber mindestens eine von beiden könnte 3. nicht erfüllen.

Zur Existenz: An der Stelle $s_0 \in \mathcal{U}$ sei definiert:

$$h^*(s_0) := \sup_{g \in \Gamma} g(s_0),$$

wobei Γ die Menge aller Funtionen von $\mathcal{U}$ nach $\mathcal{V}$ sei, die 1. und 2. der Definition erfüllen.

Offensichtlich erfüllt diese Funktion 2. der Definition und, falls sie konvex seien sollte, auch 3. Es bleibt also zu zeigen, daß h^* konvex ist.

Angenommen h^* ist nicht konvex. Dann existieren s_1 und s_2 aus $\mathcal{U}$ und ein $\alpha \in (0, 1)$ so, daß

$$\alpha h^*(s_1) + (1 - \alpha)h^*(s_2) \prec h^*(\alpha s_1 + (1 - \alpha)s_2)$$

Dann aber existiert eine Funktion $h^+ \in \Gamma$, so daß h^+ an der Stelle $\alpha s_1 + (1-\alpha)s_2$ beliebig nah an h^* herankommt. Also gilt für ein $h^+ \in \Gamma$:

$$\alpha h^*(s_1) + (1-\alpha)h^*(s_2) \prec h^+(\alpha s_1 + (1-\alpha)s_2)$$

Andererseits ist $h^+(s_i) \preceq h^*(s_i)$ für $i = 1, 2$, womit auch

$$\alpha h^+(s_1) + (1-\alpha)h^+(s_2) \prec h^+(\alpha s_1 + (1-\alpha)s_2)$$

folgt. Dann aber ist h^+ nicht konvex.

Dieser Widerspruch beweist die Konvexität von h^*.

□

Beweis von Lemma 3.19:

Es ist zu beachten, daß die Transformation $u : [0,1] \to \mathcal{U}$ nichts an Monotonie, strenger Monotonie oder Konstanz der Funktion g^* im Vergleich zu $v^{*\prime}$ verändert.

Entsprechend Abschnitt 3.2 kann nun festgestellt werden, daß sich das Intervall $\mathcal{U}$ in paarweise disjunkte, offene Intervalle zerlegen läßt, auf denen v^* entweder linear oder streng konvex ist.

An den Endpunkten t_e dieser Intervalle gilt $v^*(t_e+) = \tilde{v}(t_e)$, wenn t_e ein linker Endpunkt ist bzw. $v^*(t_e-) = \tilde{v}(t_e)$, wenn t_e ein rechter Endpunkt ist.

Zum Beweis sei ein solcher Abschnitt betrachtet. Rechts des rechten Endpunktes muß ein Intervall folgen, auf dem v^* einen größeren Anstieg hat als am rechten Rand des Intervalles selbst. Wenn dann $v^*(t_e) \neq \tilde{v}(t_e)$ und, da v^* eine Minorante von $\tilde{v}$ ist, mithin $v^*(t_e) \prec \tilde{v}(t_e)$ gilt, könnte eine Funktion h^* konstruiert werden, die $\tilde{v}$ ebenfalls nicht übersteigt, aber auf einem Intervall $(t_e - \varepsilon, t_e + \varepsilon)$ etwas größer als v^* und trotzdem konvex ist. Dann aber wäre v^* nicht mehr die größte konvexe Minorante von $\tilde{v}$. Ist $t_e = 1$, so kann h^* auf $(1-\varepsilon, 1)$ entsprechend konstruiert werden.

Ebenso kann argumentiert werden, wenn t_e linker Endpunkt eines Intervalles ist.

Außerdem läßt sich feststellen, daß auf Intervallen, auf denen v^* streng konvex ist, $v^* \equiv \tilde{v}$ gilt.

Angenommen, es gäbe ein Intervall, auf dem das nicht gilt. Da beide Funktinen stetig sind, müßte es ein Teilintervall I geben, auf dem $v^*(t) \succ \tilde{v}(t)\, \forall t \in I$ gilt. Damit kann eine Funktion h^* konstruiert werden, die auf I größer ist als v^*, aber kleiner oder gleich $\tilde{v}$ und konvex ist. Wiederum wäre v^* nicht die größte konvexe Minorante von $\tilde{v}$.

- Lemma 3.11
  Sei $c \in g^*([0,1])$. Dann ist $u([g^* = c])$ ein Intervall, auf dem v^* linear ist. An den Endpunkten dieses Intervalls, $u(t_c)$ und $u(t^c)$, sind v^* und $\tilde{v}$ gleich groß. Also folgt:

  $$v^*(u(t^c)) - v^*(u(t_c)) = \tilde{v}(u(t^c)) - \tilde{v}(u(t_c))$$

Und damit:

$$c\int_{t_c}^{t^c} w(t)dt = c\int_{u(t_c)}^{u(t^c)} ds = \int_{u(t_c)}^{u(t^c)} g(u^{-1}(s))ds = \int_{t_c}^{t^c} g(t)w(t)dt$$

- Lemma 3.12
  Es sei $[a,b]$ ein Intervall, auf dem g fallend ist. Dann ist $\tilde{v}$ auf dem Intervall $[u(a),u(b)]$ konkav, d.h. es gilt für alle $u_1,u_2 \in [u(a),u(b)]$ und für alle $\alpha \in (0,1)$: $\alpha\tilde{v}(u_1)+(1-\alpha)\tilde{v}(u_2) \preceq \tilde{v}(\alpha u_1+(1-\alpha)u_2)$. Das gilt speziell für $u_1 := u(a)$ und $u_2 := u(b)$.

  Angenommen, v^* ist nicht linear auf $[u_1,u_2]$. Dann ist die Funktion h^*, die durch

  $$h^*(s) := \begin{cases} v^*(s), & \text{falls}\, s \notin (u_1,u_2) \\ \alpha v^*(u_1)+(1-\alpha)v^*(u_2) & \text{sonst, mit}\, \alpha := \frac{u_2-s}{u_2-u_1} \in (0,1) \end{cases}$$

  definiert wird, einerseits eine Minorante von $\tilde{v}$, andererseits ist sie wegen der Konvexität von v^* auf (u_1,u_2) größer als v^*.

  Es bleibt zu zeigen, daß h^* konvex ist.
  Seien $s_1 < s_2 \in \mathcal{U}$ und $\alpha \in (0,1)$ beliebig fest.
  z.Z.: $h^*(\alpha s_1+(1-\alpha)s_2) \preceq \alpha h^*(s_1)+(1-\alpha)h^*(s_2)$

  1. Sei $s_1 < u_1$ und $s_2 \in (u_1,u_2)$.
     Es sei $s := \alpha s_1+(1-\alpha)s_2$. Wegen der Konvexität von g^* gilt: $v^*(s_2) \preceq h^*(s_2)$. Damit folgt die Behauptung sofort, falls $s \in (s_1,u_1)$. Andererseits kann man feststellen, daß die Sekante von h^* zwischen s_1 und s_2 bei u_1 größer oder gleich $h^*(u_1) = v^*(u_1)$ und bei s_2 gleich $h^*(s_2)$ ist. Da aber h^* zwischen u_1 und s_2 selbst linear ist, muß die Sekante auch auf (u_1,s_2) größer gleich h^* sein.
  2. Sei $s_1 \in (u_1,u_2)$ und $u_2 < s_2$.
     Dieser Fall kann analog Fall 1 behandelt werden.
  3. Sei $s_1 < u_1$ und $u_2 < s_2$.
     Die Behauptung folgt aus 1. und 2.
  4. Alle anderen Fälle spielen sich entweder ganz innerhalb des linearen Abschnittes zwischen u_1 und u_2 ab oder vollständig links oder rechts von ihm. In diesen Fällen ist die Konvexität nach Voraussetzung gegeben.

  Also ist v^* linear auf $[u(a),u(b)]$, und folglich ist g^* auf $[a,b]$ konstant.

- Lemma 3.13
  Wenn g^* auf (t_c,t^c) konstant gleich c ist, so ist v^* auf $(u(t_c),u(t^c))$ linear mit dem Anstieg c und es gilt, wie bereits festgestellt wurde:

  $$v^*(u(t_c)) = \tilde{v}(u(t_c)) \quad \text{und} \quad v^*(u(t^c)) = \tilde{v}(u(t^c))$$

  Zwischen $u(t_c)$ und $u(t^c)$ ist v^* kleiner oder gleich $\tilde{v}$.

Das heißt, daß der lineare Abschnitt von v^* zwischen $u(t_c)$ und $u(t^c)$ eine Tangente an $\tilde{v}(u(t_c))$ und an $\tilde{v}(u(t^c))$ ist. Damit aber sind die links- bzw. rechtsseitigen Ableitungen beider Funktionen an diesen Stellen gleich. Also gilt auch $c = \tilde{v}'(u(t_c)) = g(t_c)$ und $c = \tilde{v}'(u(t^c)) = g(t^c)$.

- Lemma 3.14
  Wenn g^* auf (t_*, t^*) streng monoton ist, so ist v^* auf $(u(t_*), u(t^*))$ streng konvex. Also ist v^* auf diesem Intervall gleich $\tilde{v}$. Folglich müssen auch die rechtsseitigen und die linksseitigen Ableitungen beider Funktionen an $u(t_*)$ bzw. $u(t^*)$ gleich sein . Wie im vorigen Punkt folgt die Behauptung.

- Lemma 3.15
  Wenn g auf (t_*, t^*) monoton ist, so ist $\tilde{v}$ auf $(u(t_*), u(t^*))$ konvex. Also ist v^* auf diesem Intervall gleich $\tilde{v}$. Damit müssen auch die Ableitungen beider Funktionen für alle $s \in (u(t_*), u(t^*))$ gleich sein.

- Korollar 3.16
  Da dieses Korollar aus den vorhergehenden Lemmata folgt, gilt es auch hier.

- Lemma 3.17
  Dieses Lemma gilt für g^*, da v^* eine Minorante von $\tilde{v}$ ist.

□

Beweis von Lemma 3.20:

Es sei definiert:

$$\rho^*(s_0) := \int_0^{u^{-1}(s_0)} h^*(t)w(t)dt$$

Da h^* und g^* auf streng monotonen Abschnitten gleich g sind, muß $s_* := \inf\{s \in \mathcal{U} : \rho^*(s) \neq v^*(s)\}$ am linken Rand eines linearen Intervalls von ρ^* oder v^* liegen, sofern überhaupt $s_* < u(1)$ gilt.

Sollte $s_* < u(1)$ sein und es den linken Rand eines linearen Abschnittes von ρ^* bilden, dann ist der Anstieg von ρ^* auf diesem linearen Abschnitt kleiner als der von v^*. Dann aber ist ρ^* auf der ganzen Länge dieses Abschnittes echt kleiner als v^*, denn v^* ist konvex. Am rechten Ende s^* dieses linearen Abschnittes kann ρ^* in diesem Fall $\tilde{v}$ nicht tangential berühren. Dann aber verletzt h^* Korollar 3.16 an der Stelle $u^{-1}(s^*)$.

Sollte andererseits s_* der linken Rand eines linearen Abschnittes von v^* sein, dann ist der Anstieg von v^* auf diesem linearen Abschnitt kleiner als der von ρ^*. Da v^* am rechten Rand s^* dieses linearen Abschnittes $\tilde{v}$ berührt, gilt dann $\tilde{v}(s^*) \prec \rho^*(s^*)$. Das aber widerspricht Lemma 3.17.

Also ist g^* die einzige monotone Funktion, die die Bedingungen der Lemmata 3.11 bis 3.17 erfüllt.

□

Beweis von Lemma 3.22:

Seien m und ς_+ festgelegt wie im Beweis von Korollar 3.16. Dann gilt für jedes $\varsigma \in M$:

$$\int_{\varsigma}^{\varsigma_+} g^*(t)\Psi(g^*(t))w(t)dt = \int_{\varsigma}^{\varsigma_+} g(t)\Psi(g^*(t))w(t)dt$$

Falls $[\varsigma, \varsigma_+]$ keine Nullmenge ist, so ist g^* dort konstant oder streng monoton. Im ersten Fall ist $\Psi(g^*(t))$ auf (ς, ς_+) unabhängig von t und die Gleichung folgt aus Lemma 3.11. Im zweiten Fall sind g und g^* auf (ς, ς_+) identisch und folglich sind die Integranden auf beiden Seiten punktweise gleich. Mit

$$\int_0^1 g^*(t)\Psi(g^*(t))w(t)dt = \sum_{\varsigma \in M} \int_{\varsigma}^{\varsigma_+} g^*(t)\Psi(g^*(t))w(t)dt$$
$$= \sum_{\varsigma \in M} \int_{\varsigma}^{\varsigma_+} g(t)\Psi(g^*(t))w(t)dt = \int_0^1 g(t)\Psi(g^*(t))w(t)dt$$

folgt die Behauptung.

□

Beweis von Lemma 3.23:

Mit (3.v) erhält man, daß die Differenz der linken Seite und der rechten Seite von (3.vi) gleich

$$\int_0^1 (g(t) - g^*(t))(\varphi(g^*(t)) - \varphi(h(t)))w(t)dt$$

ist. Das wiederum ist gleich Null für $h = g^*$ und sonst wegen Lemma 3.22 gleich

$$-\int_0^1 (g(t) - g^*(t))\varphi(h(t))w(t)dt$$

Da Φ bezüglich $\leq$ konvex ist, ist φ bezüglich $\leq$ monoton, und mithin ist $\varphi \circ h$ monoton bezüglich $\preceq$. Also können (3.i) und (3.ii) in Satz 3.9 auf $\varphi \circ h$ angewendet werden. Damit ist der letzte Term ≥ 0. Daraus folgt (3.vi). Wenn $h = g^*$, so gilt in (3.vi) das Gleichheitszeichen.

Wegen $\Delta_\Phi \geq 0$ und $\Delta_\Phi(y, z) = 0 \Leftrightarrow y = z$ ist $\int_0^1 \Delta_\Phi(g^*(t), h(t))w(t)dt \geq 0$ und $=$ genau dann, wenn $h \equiv g^*$ bis auf Nullmengen. Folglich minimiert g^* und nur g^* das erste Funktional. Setzt man die Definition von Δ_Φ in dieses Funktional ein, multipliziert mit -1 und beachtet, das $\Phi(g(t))$ nicht von h abhängt, folgt, daß g^* und nur g^* das zweite Funktional maximiert.

□

Beweis von Lemma 3.24:

Wählt man $\mathcal{C}$ und $\mathcal{C}_*$ entsprechend, so gilt für ein beliebiges $y_0 \in \beta'(\Theta)$:

$$\begin{aligned}
\beta(\psi(y_0)) &= \int_{\psi(\chi)}^{\psi(y_0)} \beta'(\theta)d\theta + \mathcal{C} = \int_{\psi(\chi)}^{\psi(y_0)} \psi^{-1}(\theta)d\theta + \mathcal{C} \\
&= \int_{\chi}^{y_0} y\psi'(y)dy + \mathcal{C} \\
&= [y\psi(y)]_{y=\chi}^{y_0} - \int_{\chi}^{y_0} \psi(y)dy + \mathcal{C} \\
&= y_0\varphi(y_0) - \Phi(y_0) + \mathcal{C}_*
\end{aligned}$$

Setzt man das mit $y_0 = l(t)$ in den Integranden der linken Seite der zu beweisenden Gleichung ein, so kann er zu folgendem Term umgeformt werden:

$$\sum_{i=1}^{n} [Y_i\psi(l(t)) - \beta(\psi(l(t))) + \mathcal{C}_*]K_h(t - x_i)$$

Zieht man $\mathcal{C}_*$ aus der Summe heraus, multipliziert mit $w(t)$, integriert von 0 bis 1 und setzt $\tilde{\mathcal{C}} = \mathcal{C}_* \int_0^1 w(t)dt$, so erhält man die Behauptung.

□

Beweis von Lemma 3.26:

Angenommen, (3.vii) gilt nicht. Dann existiert wegen der Stetigkeit von $\hat{m}_h^*$ und von m ein t_0, so daß gilt:

$$|\hat{m}_h^*(t_0) - m(t_0)| > \sup_{t \in [0,1]} |\hat{m}_h(t) - m(t)|$$

Für dieses t_0 gilt offensichtlich: $\hat{m}_h^*(t_0) \neq \hat{m}_h(t_0)$. Also existiert entsprechend der Konstruktion von $\hat{m}_h^*$ aus $\hat{m}_h$ ein Intervall (t_*, t^*), das t_0 enthält und auf dem $\hat{m}_h^*$ konstant ist, während $\hat{m}_h$ am linken Rand größer und am rechten Rand kleiner als $\hat{m}_h^*$ ist.

Wenn $\hat{m}_h^*(t_0) \prec \hat{m}_h(t_0)$, dann exitiert ein $t_1 \in (t_0, t^*)$, so daß $\hat{m}_h(t_1) \prec \hat{m}_h^*(t_1)$ gilt. Andererseits ist m wachsend, d.h es gilt: $m(t_0) \prec m(t_1)$. Also folgt: $|\hat{m}_h^*(t_0) - m(t_0)| < |\hat{m}_h(t_1) - m(t_1)|$. Das aber ist ein Widerspruch zur Annahme.

Analog kann der Fall $\hat{m}_h(t_0) \prec \hat{m}_h^*(t_0)$ behandelt werden. Es existiert dann ein $t_1 \in (t_*, t_0)$, so daß $\hat{m}_h^*(t_0) \prec \hat{m}_h(t_0)$ gilt. Weil dann $m(t_1) \prec m(t_0)$ gilt, folgt wiederum ein Widerspruch zur Annahme.

Also folgt (3.vii).

□

4 Das erweiterte Regressionsmodell

In diesem Kapitel werden die Modelle aus den Kapiteln 2 und 3 leicht erweitert. Es wird nicht mehr davon ausgegangen, daß die unbekannte Regressionsfunktion m den Erwartungswert der Beobachtungen liefert, sondern daß sie das Argument einer weiteren Funktion G ist, welche ihrerseits die Erwartungswertfunktion ist. Dabei sei G monoton und erfülle gewisse Glattheitsbedingungen.

Die meisten Eigenschaften des einfachen und des monotonen Modells übertragen sich direkt auf den hier betrachteten Ansatz.

4.1 Modell und geglättetes Log-Likelihood-Funktional

Das in diesem Abschnitt behandelte Modell ist eine Erweiterung des Modells aus Kapitel 2. Alle in 2.1 genannten Voraussetzungen und vereinbarten Bezeichnungen sollen auch hier gelten, soweit es nicht im folgenden explizit anderes angegeben ist.

Es gelte $\mathsf{E}Y_i^{(n)} = G(m(x_i^{(n)}))$ im äquidistanten Design, bzw. $\mathsf{E}(Y_i^{(n)}|X_i^{(n)} = t) = G(m(t))$ im zufälligen Design, jeweils für alle $n \in \mathbb{N}$ und für alle $i = 1, \ldots, n$. Dabei sei $G : \mathrm{J} \to \beta'(\Theta)$ eine bekannte Funktion mit:

(a) J ist ein kompaktes Intervall in $\mathbb{R}$.

(b) G ist streng monoton und differenzierbar. Die Ableitung ist Lipschitz-stetig.

(c) Die Umkehrfunktion H von G ist ebenfalls differenzierbar und die Ableitung ist Lipschitz-stetig. c_H und $c_{H'}$ sind die Lipschitz-Konstanten von H bzw. H'.

$m : [0,1] \to \mathrm{J}$ sei eine unbekannte Regressionsfunktion. Sie ist differenzierbar und die Ableitung ist Lipschitz-stetig. m ist zu schätzen.

Es ist zu beachten, daß mit diesen Voraussetzungen die Verknüpfung $G \circ m$ die Eigenschaften erfüllt, die die Regressionsfunktion aus Kapitel 2 erfüllen soll.

Die Log-Likelihood-Funktion l_i der Zufallsvariable Y_i in Abhängigkeit vom Parameter $\theta \in \Theta$ hat die gleiche Gestalt wie in Kapitel 2. Jedoch ergibt sich für das zur Schätzung von m zu maximierende Funktional die folgende Definition:

Definition 4.1 *Für ein $m : [0,1] \to \mathrm{J}$ wird das Funktional $\mathcal{L}^{sG}$, das durch*

$$\mathcal{L}^{sG}(m) := \sum_{i=1}^{n} \int_0^1 l_i(\psi(G(m(t))))K_h(t - x_i)dt$$

definiert ist, als geglättetes Log-Likelihood-Funktional bezeichnet.

Eine Funktion $\hat{m}_h : [0,1] \to \mathrm{J}$ wird als Maximum-geglättete-Likelihood-Schätzer bezeichnet, wenn sie

$$\hat{m}_h \in \underset{m:[0,1]\to \mathrm{J}}{argmax}\ \mathcal{L}^{sG}(m)$$

erfüllt.

Zur Eindeutigkeit gilt das gleiche wie für das einfache Regressionsmodell.

4.2 Der MgL-Schätzer der Regressionsfunktion

Der folgende Satz gibt die Gestalt des Schätzers explizit an und trifft Aussagen über seine Existenz.

Satz 4.1 *Es sei $h = h_n \underset{n\to\infty}{\longrightarrow} 0$ so, daß $nh \underset{n\to\infty}{\longrightarrow} \infty$ im determistischen Design, bzw. so, daß im zufälligen Design für ein $\alpha > 0$ gilt: $nh \geq n^\alpha$.*

Dann ist die Funktion

$$\hat{m}_h(t) := H\left(\frac{\sum_{i=1}^n Y_i K_h(t-x_i)}{\sum_{i=1}^n K_h(t-x_i)}\right) \quad \textit{für alle } t \in [0,1]$$

in beiden Fällen der MgL-Schätzer der Regressionsfunktion m.

Im deterministischen Modell ist der MgL-Schätzer für genügend großes n wohldefiniert. Im zufälligen Modell geht die Wahrscheinlichkeit, daß er wohldefiniert ist, für $n \to \infty$ gegen 1.

Wenn der MgL-Schätzer wohldefiniert ist, so ist er bis auf Nullmengen bezüglich des Lebesgue-Maßes eindeutig.

Beweis: Der Beweis kann analog dem Beweis von Satz 2.6 geführt werden, wobei zu beachten ist, daß G' entweder strikt positiv oder strikt negativ ist.

□

Da $G \circ m$ alle Voraussetzungen der im Kapitel 2 definierten Regressionsfunktion erfüllt, kann $G \circ m$ auch entsprechend geschätzt werden. Sei $\hat{g}_h$ der MgL-Schätzer von $G \circ m$ im Sinne des einfachen Modells. Dann gelten auch alle bisher bewiesenen Eigenschaften dieses Schätzers. Das wird im folgenden zum Beweis von Eigenschaften des MgL-Schätzers im erweiterten Modell genutzt.

Satz 4.2 *In beiden Designs gilt, falls $h_n := c_h n^{-1/5}$ für ein $c_h > 0$:*

$$\sup_{t\in[h_n,1-h_n]} |\hat{m}_h(t) - m(t)| = O_P(n^{-2/5}\sqrt{\ln n})$$

und

$$\sup_{t\in[h_n,1-h_n]} |\hat{m}'_h(t) - m'(t)| = O_P(n^{-1/5}\sqrt{\ln n})$$

Beweis: Es gilt:

$$\begin{aligned}\sup_{t\in[h_n,1-h_n]} |\hat{m}_h(t) - m(t)| &= \sup_{t\in[h_n,1-h_n]} |H(\hat{g}_h(t)) - H(G(m(t)))| \\ &\leq c_H \sup_{t\in[h_n,1-h_n]} |\hat{g}_h(t) - G(m(t))| \\ &= O_P(n^{-2/5}\sqrt{\ln n})\end{aligned}$$

entsprechend Satz 2.15. Ferner gilt:

$$\begin{aligned}|\hat{m}'_h(t) - m'(t)| &= \left|H'(\hat{g}_h(t))\hat{g}'_h(t) - H'(G(m(t)))\frac{dG(m(t))}{dt}\right| \\ &\leq |H'(G(m(t)))||\hat{g}'_h(t) - \frac{dG(m(t))}{dt}| + c_{H'}|\hat{g}_h(t) - G(m(t))| \\ &\leq C_1|\hat{g}'_h(t) - \frac{dG(m(t))}{dt}| + c_{H'}\hat{g}'_h(t)|\hat{g}_h(t) - G(m(t))|\end{aligned}$$

Damit folgt aus Satz 2.15 und Satz 2.20:

$$\sup_{t\in[h_n,1-h_n]} |\hat{m}'_h(t) - m'(t)| = O_P(n^{-1/5}\sqrt{\ln n})$$

□

4.3 Das erweiterte monotone Modell

Analog der Erweiterung des einfachen Regressionsmodells wird hier das monotone Modell erweitert.

Es sollen alle Voraussetzungen des erweiterten Regressionsmodells in Abschnitt 4.1 gelten, und es sei $m : [0,1] \to \mathrm{J}$ eine bezüglich $\preceq$ monotone Funktion.

Wiederum erfüllt $G \circ m$ alle Voraussetzungen der Regressionsfunktion im einfachen monotonen Modell, wobei je nach Monotonie von G die Monotonie von $G \circ m$ gleich der von m oder der von m entgegengesetzt ist.

Definition 4.2 *Für ein* $m : [0,1] \to \mathrm{J}$ *wird das Funktional* $\mathcal{L}^{smG}$*, das durch*

$$\mathcal{L}^{smG}(m) := \sum_{i=1}^{n} \int_0^1 l_i(\psi(G(m(t))))K_h(t - x_i)dt$$

definiert ist, als geglättetes Log-Likelihood-Funktional bezeichnet.

Als MgL-Schätzung wird eine monotone Funktion $\hat{m}_h^*$ *bezeichnet, die* $\mathcal{L}^{smG}$ *unter allen monotonen Funktionen aus* $\mathrm{J}^{[0,1]}$ *maximiert.*

Es gilt der Satz:

Satz 4.3 *Es gilt:*

$$\hat{m}_h^* = H \circ \widehat{(G \circ m)}_h^*,$$

wobei $\widehat{(G \circ m)}_h^*$ *der MgL-Schätzer gemäß des einfachen monotonen Modells ist.*

Beweis: Da G umkehrbar ist und da eine Funktion $m : [0,1] \to \mathrm{J}$ genau dann monoton ist, wenn $G \circ m : [0,1] \to \beta'(\Theta)$ monoton ist – in die gleiche oder entgegengesetzte Richtung wie m, je nach Monotonie von G – besteht eine eineindeutige Zuordnung zwischen den Funktionen, aus denen $\hat{m}_h^*$ gewählt wird und denen, unter denen $\widehat{(G \circ m)}_h^*$ gemäß des einfachen monotonen Modells gesucht wird. Also maximiert $H \circ \widehat{(G \circ m)}_h^*$ das Funktional $\mathcal{L}^{smG}$ genau dann, wenn $\widehat{(G \circ m)}_h^*$ das Funktional $\mathcal{L}^{sm}$ maximiert.

□

Auch die Modifizierung läßt sich analog zu Abschnitt 3.5 durchführen. Für die Modifizierung des monotonen Schätzer gilt der Satz:

Satz 4.4 *Sei* $h_n := c_h n^{-1/5}$ *für ein* $c_h > 0$.
Dann gilt:

$$\sup_{t\in[h_n,1-h_n]} |\hat{m}_h^{*m}(t) - m(t)| = O_P(n^{-2/5}\sqrt{\ln n})$$

und

$$\sup_{t\in[h_n,1-h_n]} |\hat{m}_h^{*m}(t) - \hat{m}_h(t)| = O_P(n^{-2/5}\sqrt{\ln n})$$

Beweis: Lemma 3.26 gilt im erweiterten Modellansatz analog. Es sagt für den modifizierten monotonen MgL-Schätzer aus:

$$\sup_{t\in[h_n,1-h_n]} |\hat{m}_h^{*m}(t) - m(t)| \leq \sup_{t\in[h_n,1-h_n]} |\hat{m}_h(t) - m(t)|$$

Damit und mit Satz 4.2 folgt die erste Gleichung. Die zweite folgt dann wie im Beweis von Satz 3.27.

□

5 Plots

5.1 Einführung und Erläuterung

In diesem Kapitel wird eine Reihe Plots präsentiert, die einige Aspekte der in dieser Arbeit betrachteten Schätzer illustrieren soll.

Die Plots beruhen auf simulierten Daten. Die Simulation und alle folgenden Rechnungen wurden mit Hilfe eines Programmes durchgeführt, das Matthias Küchler im Rahmen eines von Professor Mammen betreuten Forschungsprojektes erarbeitete.

Bei der Auswahl der Regressionsfunktion habe ich versucht sicherzustellen, daß der wesentliche Effekt der Monotonisierung hinreichend gut illustriert wird. Die allen Plots zugrundeliegende Regressionsfunktion ist streng monoton wachsend, hat einen horizontalen Wendepunkt bei 1/2, während ihr Anstieg zu den Rändern hin zunimmt.

Zwangsläufig sind die Daten und der zu ihrer Erzeugung verwendete Algorithmus diskreter Natur. Auch wurde nicht der MgL-Schätzer errechnet, sondern ein leicht abgewandelter Kernschätzer, der jedoch im äquidistanten Design praktisch das gleiche Ergebnis liefert. Ferner wurden jeweils links und rechts des Intervalles $[0, 1]$ soviele weitere äquidistante Designpunkte in die Rechnungen einbezogen, wie notwendig waren, um Randprobleme zu vermeiden.

Das von Matthias Küchler erarbeitete Programm stellte die simulierten Zufallszahlen, die Funktionswerte des Kernschätzers an den Designpunkten und deren Kleinste-Quadrate-Monotonisierung zur Verfügung. Die Plots wurden aus diesen Daten mit Excel 5.0[†] erstellt, auch Interpolationen zwischen den Funktionswerten basieren auf dem Graphikprogramm.

[†]Warenzeichen der Microsoft Corporation

Abbildung 5.1: „$h = 0.05$“ *Plots zweier simulierter Datensätze (Punkte), der wahren Regressionsfunktion (gepunktete Linie), einer Kernschätzung der Regressionsfunktion und ihrer Montonisierung (durchgezogene Linien).*

Alle Unterschiede in der Erstellung der Plots zu der in dieser Arbeit vorgestellten Theorie beeinträchtigen nicht die Aussagekraft der Plots hinsichtlich der Theorie. Die durch sie verursachten Abweichungen sind deutlich zu gering, um auf den Plots wahrgenommen zu werden. Unterschiede gibt es bezüglich des Kernschätzers – es ist nicht ganz der Nadaraya-Watson-Schätzer – und der Monotonisierung – sie wird zwangsläufig diskret berechnet. Näheres dazu wird im folgenden Abschnitt angegeben.

5.2 Details der Simulation

Die der Simulation zugrundeliegende Regressionsfunktion ist gegeben durch:

$$m(t) := \begin{cases} \sin(\pi t), & \text{falls } t < 1/2 \\ 2 - \sin(\pi t), & \text{falls } 1/2 \leq t \end{cases}$$

Dieses m ist differenzierbar und die Ableitung ist L-stetig.

Abbildung 5.2: „$h = 0.1$“ *Plots zweier simulierter Datensätze (Punkte), der wahren Regressionsfunktion (gepunktete Linie), einer Kernschätzung der Regressionsfunktion und ihrer Montonisierung (durchgezogene Linien).*

An den äquidistanten Designpunkten $x_i = i/100$, für $i = -nh, \ldots, 100 + nh$ wurde dazu eine Realisierung einer normalverteilten ZG mit Erwartungswert Null und Varianz 0.25 addiert, d.h.

$$Y_i = m(i/100) + Z_i/2,$$

wobei die Z_i unabhängige, standardnormalverteilte ZG sind. Dieser Modellansatz entspricht also dem in dieser Arbeit behandelten deterministischen Design.

Für $i = 0, \ldots, 99$ wurden diese Daten geglättet:

$$\hat{Y}_i := (nh)^{-1} \sum_{j=-nh}^{n+nh} Y_i K((i-j)/nh)$$

Das ist nicht ganz der Nadaraya-Watson-Schätzer, im äquidistanten Design ist der Unterschied jedoch gering. Der Kern K ist dabei definiert durch:

$$K(x) := \frac{L(x)}{\int_{-1}^{1} L(v)dv} \quad \text{und} \quad L(x) := \begin{cases} (1-x^2)^2, & \text{falls } |x| < 1 \\ 0 & \text{sonst} \end{cases}$$

Der so definierte Kern ist differenzierbar und die Ableitung ist L-stetig.

Diese geglätteten Daten wurden monotonisiert, d.h. es wurden $\hat{Y}_i^*$ für $i = 0, \ldots, 99$ ausgerechnet, so daß gilt: $\hat{Y}_{i-1}^* \leq \hat{Y}_i^*$ und $\sum_{i=0}^{99} (\hat{Y}_i - \hat{Y}_i^*)^2$ ist minimal.

5.3 Die Plots im einzelnen

Es wurden drei Serien mit den $h = 0.05$, 0.1 und 0.2 simuliert. Jeweils zwei Plots aus diesen Serien werden im folgenden vorgestellt und kurz kommentiert.

Abbildung 5.3: „$h = 0.2$“ *Plots zweier simulierter Datensätze (Punkte), der wahren Regressionsfunktion (gepunktete Linie), einer Kernschätzung der Regressionsfunktion und ihrer Montonisierung (durchgezogene Linien).*

In allen Plots ist die wahre Regressionsfunktion als gepunktete Linie, die Schätzung und ihre Monotonisierung sind als durchgehende Linien dargestellt.

Abbildung 5.1, Serie „$h = 0.05$“

Diese Serie zeigt deutlich, daß die Bandbreite zu klein gewählt wurde. In beiden Plots schwankt die Schätzung stark, über den gesamten Definitionsbereich hinweg hat sie lokale Extrema. Es ist gut zu erkennen, wie die Monotonisierung aussieht und daß mehrere lokale Extrema der uneingeschränkten Schätzung durch einen konstanten Abschnitt monotonisiert werden können (linker Plot, Mitte).

Offensichtlich schätzt der monotone Schätzer die wahre Regressionsfunktion erheblich besser als der uneingeschränkte Schätzer.

Abbildung 5.2, Serie „$h = 0.1$“

Der linke Plot sieht geradezu „klassisch“ aus. Beide Schätzungen sind insgesamt recht gut. Die uneingeschränkte Schätzung hat in der Mitte einen deutlichen und links und rechts davon zwei kleine Buckel, welche von der Monotonisierung in der erwarteten Weise durch konstante Abschnitte ersetzt werden.

Im rechten Plot wird die Regressionsfunktion fast auf dem gesamten Intervall $[0, 1]$ unterschätzt. Aber auch hier sind lokale Extrema der uneingeschränkten Schätzung in der Mitte konzentriert.

Abbildung 5.3, Serie „$h = 0.2$“

Im linken Plot unterscheiden sich beide Schätzer nur marginal, im rechten gar nicht. In beiden Fällen wird die wahre Regressionsfunktion sehr gut geschätzt.

Es stellt sich allerdings die Frage, ob hier nicht bereits zu sehr geglättet wird, was durch die sehr glatte Regressionsfunktion vertuscht werden könnte.

Bezeichnungen

$[x]$	$= \max\{n \in \mathbb{N} : n \leq x\}$ für alle $x \in \mathbb{R}$
$\#M$	Anzahl der Elemente der Menge M
$\mathcal{B}^n$	Klasse der Borelmengen in $\mathbb{R}^n$
$\mathcal{B}^n_A$	$= \{B \in \mathcal{B}^n \| B \subseteq A\}$ für alle $A \in \mathcal{B}^n$
$\left.\begin{matrix}\mathsf{E}(X\|Y=y)\\ \mathsf{Var}(X\|Y=y)\end{matrix}\right\}$	bedingter Erwartungswert bzw. bedingte Varianz der ZG X unter der Bedingung „$Y = y$“
$\left.\begin{matrix}\mathsf{E}(X\|Y)\\ \mathsf{Var}(X\|Y)\end{matrix}\right\}$	Funktion der ZG Y, realisierungsweise festgelegt mittels $\mathsf{E}(X\|Y = y)$ bzw. $\mathsf{Var}(X\|Y = y)$
$X \sim F$	X hat die Verteilung F
$O(.)$	$a_n = O(b_n)$, falls a_n/b_n für alle n beschränkt ist
$O_P(.)$	$X_n = O_P(b_n)$, falls gilt: $\forall q > 0\ \exists C_q < \infty\ \exists n_0 \in \mathbb{N}\ \forall n \geq n_0 : P(\|X_n/b_n\| \geq C_q) \leq q$
$X_n \xrightarrow{L} F$	Konvergenz in Verteilung: Die Verteilungsfunktionen F_n der ZGen X_n konvergieren gegen die Verteilungsfunktion F in allen Stetigkeitspunkten von F.

Literatur

[1] P. McCullagh, J. A. Nelder: Generalized Linear Models. Chapman and Hall, London, New York 1989.

[2] P. J. Green, B. W. Silverman: Nonparametric regression and generalized linear models. Chapman and Hall, London 1994.

[3] U. Grenander: On the theory of mortality measurement, Part II. Skand. Akt. 39, p. 125 – 153.

[4] W. Härdle, E. Mammen: Comparing Nonparametric versus Parametric Regression Fits. The Annals of Statistics, Vol. 21, 1993, p. 1926 – 1947.

[5] W. Härdle: Smoothing Techniques, *With Implementations in* **S**. Springer-Verlag, New York 1991.

[6] T. J. Hastie, R. J. Tibshirani: Generalized Additive Models. Chapman and Hall, London 1990.

[7] K. M. S. Humak: Statistische Methoden der Modellbildung, Band 1. Akademie-Verlag, Berlin 1977.

[8] A. P. Korostelev, A. B. Tsybakov: Minimax Theory of Image Reconstruction. Springer-Verlag, New York 1993.

[9] M. Reed, B. Simon: Methods of Modern Mathematical Physics I: Functional Analysis. Academic Press, San Diego.

[10] T. Robertson, F. T. Wright, R. L. Dykstra: Order Restricted Statistical Inference. John Wiley & Sons 1988.

Thesen

i. Im einfachen Regressionsmodell mit äquidistantem Design existiert für geeignete Wahl der Bandbreite ein bis auf Borel-Nullmengen eindeutig bestimmter MgL-Schätzer.

Für geeignete Wahl der asymptotischen Abhängigkeit der Bandweite von der Anzahl der Designpunkte konvergiert im einfachen Regressionsmodell mit zufälligem Design die Wahrscheinlichkeit, daß der MgL-Schätzer existiert und bis auf Borelmengen eindeutig bestimmt ist, mit größer werdender Anzahl der Designpunkte gegen Eins.

ii. Im einfachen Regressionsmodell konvergiert der MgL-Schätzer punktweise mit der Rate $n^{-2/5}$ gegen die wahre Regressionsfunktion, wenn die Bandweite so gewählt wird, daß Bias und Standardabweichung des MgL-Schätzers mit derselben Rate gegen Null konvergieren.

Mit der Rate $n^{-2/5}\sqrt{\ln n}$ konvergiert unter dieser Bedingung die Differenz von MgL-Schätzer und wahrer Regressionsfunktion in der Supremumsnorm gegen Null. Die Ableitung dieser Differenz konvergiert in der Supremumsnorm mit der Rate $n^{-1/5}\sqrt{\ln n}$ gegen Null.

iii. Der monotone MgL-Schätzer ist gleich der geeignet gewichteten Kleinste-Quadrate-Monotonisierung des MgL-Schätzers. Existenz und Eindeutigkeit folgen denselben Regeln wie beim MgL-Schätzer.

Es wird ein explizites Verfahren zur Konstruktion des monotonen MgL-Schätzers angegeben.

iv. Im monotonen Regressionsmodell konvergiert ein an den Rändern des Definitionsbereiches leicht modifizierter monotoner MgL-Schätzer in der Supremumsnorm mit derselben Rate gegen die wahre Regressionsfunktion wie der MgL-Schätzer. Folglich konvergiert die Differenz beider Schätzer mit dieser Rate gegen Null.

v. Der MgL-Schätzer und der monotone MgL-Schätzer existieren in einem geeignet erweiterten Modell analog. Die Rate der Konvergenz beider Schätzer gegen die wahre Regressionsfunktion und die Rate der Konvergenz der Differenz beider gegen Null ist die gleiche wie in einfachen Modell.

Erklärung

Hiermit erkläre ich, daß ich diese Diplomarbeit selbständig und nur unter Verwendung der angegebenen Literatur und Hilfen angefertigt habe.

Berlin, den 5. Oktober 1995

Jens Haase