Robustheitseigenschaften von Dekonvolutionsdichteschätzern bezüglich Missspezifikation der Fehlerdichte

Von der Fakultät Mathematik und Physik der Universität Stuttgart zur Erlangung der Würde eines Doktors der Naturwissenschaften (Dr. rer. nat.) genehmigte Abhandlung

Vorgelegt von

Alexander Meister

aus Filderstadt

Hauptberichter:	Prof. Dr. C.H. Hesse
Mitberichter:	Prof. Dr. M.H. Neumann (TU Braunschweig) Prof. Dr. H. Walk
Tag der mündlichen Prüfung:	23. Juli 2003

Institut für Stochastik und Anwendungen
Fakultät Mathematik und Physik
Universität Stuttgart

2003

Vorwort

Diese Dissertation entstand am Institut für Stochastik und Anwendungen der Fakultät für Mathematik und Physik der Universität Stuttgart. Bedanken möchte ich mich insbesondere bei Herrn Prof. Dr. Christian H. Hesse für die Betreuung meiner Promotion und bei den Mitberichtern Prof. Dr. Neumann und Prof. Dr. Walk.

im April 2003

Alexander Meister

Inhaltsverzeichnis

Kapitel 1

Einführung

Das Grundproblem der Dekonvolutionsdichteschätzung besteht darin, die Dichte einer Zufallsvariablen aufgrund endlich vieler verrauschter bzw. fehlerbehafteter Beobachtungen zu schätzen.
Diese Beobachtungen - ihre Anzahl sei n - werden als reelle Zufallsvariable $Y_1, \ldots, Y_n$ modelliert. Dabei setzt sich jede dieser Zufallsveränderlichen Y_j als Summe zweier reeller Zufallsvariablen X_j und ϵ_j zusammen

$$Y_j = X_j + \epsilon_j \quad , j \in \{1, \ldots, n\}.$$

Dabei seien $X_1, \epsilon_1, \ldots, X_n, \epsilon_n$ unabhängig und $X_1, \ldots, X_n$ sowie $\epsilon_1, \ldots, \epsilon_n$ seien identisch verteilt.
Die Zufallsgrößen ϵ_j stellen den Fehler der Beobachtungen dar, die Dichte ihrer Verteilung werde mit g bezeichnet.
Geschätzt werden soll nun die Dichte f der Verteilung der Zufallsvariablen X_j mit Hilfe der Beobachtungen $Y_1, \ldots, Y_n$. Ein direkter empirischer Zugang zu den Beobachtungen $X_1, \ldots, X_n$ ist also unmöglich. Gewöhnlicherweise sind aber sowohl für die Fehlerdichte g als auch für die zu schätzende Dichte f gewisse deterministische nichtparametrische Voraussetzungen bekannt wie zum Beispiel Aussagen über das Abklingverhalten der Fouriertransformierten, die gleichgradige Hölder-Stetigkeit oder die Sobolev-Norm der Dichten. Im mathematischen Modell bedeutet dies, dass wir eine Menge aller in Frage kommenden zu schätzenden Dichten f einführen können, die mit $\mathcal{F}$ bezeichnet werde, sowie eine Menge aller in Frage kommenden Fehlerdichten g, die $\mathcal{G}$ genannt werde. Die deterministischen Voraussetzungen lauten dann

$$f \in \mathcal{F}\,,\ g \in \mathcal{G}.$$

$\mathcal{F}$ und $\mathcal{G}$ sind somit Teilmengen der Klasse aller (Zufalls-)dichten $\mathcal{D}$.

$$\mathcal{D} := \{f \in L_1(\mathbb{R}) \mid f(x) \geq 0\,, \forall x \in \mathbb{R} \quad \text{und} \quad \|f\|_{L_1(\mathbb{R})} = 1\}. \tag{1.1}$$

$L_1(\mathbb{R})$ bezeichne dabei wie üblich den Banachraum aller über $\mathbb{R}$ absolut Lebesgue-integrablen Funktionen $\mathbb{R} \longrightarrow \mathbb{C}$. Dagegen liegt eine Dichte nicht notwendigerweise im Hilbertraum $L_2(\mathbb{R})$ aller über $\mathbb{R}$ absolut-quadratisch Lebesgue-integrablen Funktionen $\mathbb{R} \longrightarrow \mathbb{C}$. Daher führen wir die Menge aller Dichten, die in $L_2(\mathbb{R})$ liegen, das Symbol

$$\mathcal{D}_2 := \mathcal{D} \cap L_2(\mathbb{R}) \tag{1.2}$$

ein. Eine Dichte liegt unter milden Zusatzvoraussetzungen in $\mathcal{D}_2$. Die Forderung der Beschränktheit einer Dichte genügt beispielsweise, um ihre Mitgliedschaft in $\mathcal{D}_2$ zu garantieren (siehe *Meister* (2001)).
Die Dichte h der Zufallsvariablen Y_j ergibt sich wegen der Unabhängigkeit von X_j und ϵ_j nach einem elementaren Ergebnis der Wahrscheinlichkeitstheorie als Faltung der Dichten f von X_j und g von ϵ_j

$$h = f * g.$$

Entsprechend kann man folgende Faltungsklassen definieren:

$$\begin{aligned} f * \mathcal{G} &:= \{f * g \mid g \in \mathcal{G}\} \\ g * \mathcal{F} &:= \{f * g \mid f \in \mathcal{F}\} \\ \mathcal{H} &:= \mathcal{F} * \mathcal{G} := \{f * g \mid f \in \mathcal{F}\,, g \in \mathcal{G}\} \end{aligned} \tag{1.3}$$

Dabei ist zu beachten, dass die Faltung von Dichten das Kommutativgesetz erfüllt. Daher könnte man statt $f * \mathcal{G}$ auch $\mathcal{G} * f$ schreiben.
ψ_f bezeichne die Fouriertransformierte einer Funktion f,

$$\psi_f(x) := \int \exp(itx)\, f(t)\, dt.$$

Dabei ist zu unterscheiden, ob $f \in L_1(\mathbb{R})$ oder $f \in L_2(\mathbb{R})$ gilt. Im ersten Fall ist $\psi_\bullet$ als Fourier-Dirichlet-Transformation zu verstehen, im zweiten

als Fourier-Plancherel-Transformation. Die Unterscheidung der beiden Fälle wird im Funktionalanalysis-Buch von *D. Werner* genau dargelegt. Falls $f \in L_1(\mathbb{R}) \cap L_2(\mathbb{R})$ gilt, fallen beide Begriffe der Fouriertransformation zusammen, dies rechtfertigt, für beide Sachverhalte nur ein Symbol ψ zu verwenden.
Die 'Dekonvolutionsformel'

$$\psi_{f*\tilde{f}} = \psi_f \cdot \psi_{\tilde{f}} \tag{1.4}$$

kann in zwei Fällen angewandt werden

1. $f \in L_1(\mathbb{R})$ und $\tilde{f} \in L_1(\mathbb{R})$

   oder

2. $f \in L_2(\mathbb{R})$ und $\tilde{f} \in L_1(\mathbb{R}) \cap L_2(\mathbb{R})$.

Kein hinreichendes Kriterium für die Entfaltung (1.4) ist jedoch $f \in L_1(\mathbb{R})$ und $f \in L_2(\mathbb{R})$. In diesem Fall ist nicht einmal die Faltung zwischen f und $\tilde{f}$ wohldefiniert. Der erste Fall erlaubt die Anwendung von (1.4) auf alle Dichten f und $\tilde{f}$.

Um die Inklusion einer Faltungsklasse in $\mathcal{D}_2$ überprüfen zu können, nützt folgendes Lemma

Lemma 1.1 *Sei $f \in \mathcal{D}$ und $\tilde{f} \in \mathcal{D}_2$. Dann gilt*

$$f * \tilde{f} \in \mathcal{D}_2$$

Beweis: Dass die Faltung zweier Dichten erneut eine Dichte ergibt, ist ein gängiges Resultat aus der Wahrscheinlichkeitstheorie. Es bleibt also noch zu zeigen, dass $f * \tilde{f}$ in $L_2(\mathbb{R})$ liegt:
Da $f \in \mathcal{D} \subset L_1(\mathbb{R})$ und $\tilde{f} \in \mathcal{D} \subset L_1(\mathbb{R})$ gilt, kann (1.4) angewendet werden. Da zudem für f wie für jede Dichte $|\psi_f(t)| \leq 1$ gilt, folgt:

$$|\psi_{f*\tilde{f}}| = |\psi_f| \cdot |\psi_{\tilde{f}}| \leq |\psi_{\tilde{f}}|$$

Da $\tilde{f}$ und somit nach der Parseval-Identität auch $\psi_{\tilde{f}}$ in $L_2(\mathbb{R})$ liegen, ist nach obiger Ungleichung $|\psi_{\tilde{f}}|^2$ eine integrable Majorante für die als Fouriertransformierte zweier Dichten stetige Funktion $|\psi_{f*\tilde{f}}|^2$. Somit ist $|\psi_{f*\tilde{f}}|^2$ über $\mathbb{R}$ integrabel, und gemäß der Parseval-Identität liegt $f * \tilde{f}$ daher in $L_2(\mathbb{R})$.

■

Durch Lemma 1.1 ist erwiesen, dass eine Faltungsklasse $f * \mathcal{G}$ oder $g * \mathcal{F}$ komplett in $L_2(\mathbb{R})$ liegt, wenn $f \in L_2(\mathbb{R})$ bzw. $g \in L_2(\mathbb{R})$ gilt. Ebenso ist $\mathcal{H}$ Teilmenge von $L_2(\mathbb{R})$, wenn $\mathcal{F}$ oder $\mathcal{G}$ es ist.

In den meisten Artikeln über Dekonvolutionsschätzer wird vorausgesetzt, dass die Fehlerdichte g exakt bekannt ist, was in der hier verwendeten Modellierung zur Einelementigkeit der Menge $\mathcal{G}$ ($\mathcal{G} = \{g\}$) äquivalent ist (zum Beispiel *Stefanski und Carroll* (1990), *Fan* (1991a), *Fan* (1993), *Devroye* (1989), *Meister* (2001)). Unter bestimmten deterministischen Begebenheiten an g und $\mathcal{F}$ wird zum Beispiel in *Stefanski und Carroll* (1990) die gleichmäßige L_2-Konsistenz für den Dekonvolutionskernschätzer bewiesen. Ebenso wird gezeigt, dass dieser Schätzer die optimalen Konvergenzraten erreicht (siehe *Fan* (1993)).
Man kann jedoch nicht in allen Anwendungen der Dekonvolutionsschätzung davon ausgehen, dass die Fehlerdichte stets bekannt ist. Die Situation einer nicht exakt bekannten Fehlerdichte g wird zum Beispiel in den Artikeln von *Neumann* (1997) und *Efromovich* (1997) beleuchtet. In beiden Veröffentlichungen wird angenommen, dass zusätzliche unabhängige identisch verteilte Beobachtungen, die direkt aus der zur Fehlerdichte gehörigen Verteilung stammen und unabhängig von den Beobachtungen der verunreinigten Dichte sind, gegeben sind. Mit deren Hilfe wird ein Schätzer für die Fouriertransformierte der Fehlerdichte ermittelt.

In dieser Dissertation wird erstmals die Situation betrachtet, dass die Fehlerdichte g weder exakt bekannt ist noch zusätzliche Beobachtungen aus g vorhanden sind. Der Schätzer für die Dichte f darf somit nur auf den Zufallsvariablen $Y_1, \ldots, Y_n$ sowie auf Kenntnis der deterministischen nichtparametrischen Bedingungen an f und g basieren. Letztere werden durch Definition der Dichteklassen $\mathcal{F}$ und $\mathcal{G}$ angegeben. Vorerst werden diese Voraussetzungen aber noch variabel gehalten.
Um einen solchen Dekonvolutionsschätzer bei nicht a-priori bekannter Fehlerdichte, d.h. mehrelementiger Menge $\mathcal{G}$ ermitteln zu können, müssen selbstverständlich zuerst die Qualitätsmerkmale festgelegt werden, denen der Schätzer genügen soll. *Wir gehen davon aus, dass keine a-priori Gewichtung der Fehlerdichten gegeben ist, sondern alle $g \in \mathcal{G}$ vor Bekanntgabe der empirischen Daten gleichberechtigt sind.* Bei der Dekonvolutionsschätzung mit bekanntem g treten als asymptotische Konvergenzmerkmale einer Schätzerse-

quenz $(\hat{f}_n)_n$ die (individuelle) d^k-Konsistenz

$$E_{f,g}\, d(\hat{f}_n(Y_1,\ldots,Y_n),f)^k \quad \underset{n\to\infty}{\longrightarrow} 0 \qquad \forall f\in\mathcal{F}$$

und die gleichmäßige d^k-Konsistenz

$$\sup_{f\in\mathcal{F}} E_{f,g}\, d(\hat{f}_n(Y_1,\ldots,Y_n),f)^k \quad \underset{n\to\infty}{\longrightarrow} 0$$

auf. Dabei sei $d(\cdot,\cdot)$ eine beliebige Metrik, wobei noch gefordert werden muss, dass sowohl die Schätzfunktion als auch alle Dichten aus $\mathcal{F}$ in dem zu d gehörigen metrischen Raum liegen. k sei eine reelle Zahl größer 0. Als Abstand d kommen die von den L_p-Normen $(p\geq 1)$, der Supremum-Norm oder auch den Sobolev-Normen abgeleiteten Metriken in Frage.
Für das Schätzproblem einer nicht exakt bekannten Fehlerdichte lassen sich analoge Konsistenzbegriffe definieren. Dazu führen wir den Begriff des **robust d^k-konsistenten** Schätzers $(\hat{f}_n)_n$ ein, der

$$E_{f,g}\, d(\hat{f}_n(Y_1,\ldots,Y_n),f)^k \quad \underset{n\to\infty}{\longrightarrow} 0 \qquad \forall f\in\mathcal{F}, \forall g\in\mathcal{G} \tag{1.5}$$

erfüllt. Ebenso sprechen wir von einem **gleichmäßig robust d^k-konsistenten** Schätzer genau dann, wenn er

$$\sup_{g\in\mathcal{G}}\sup_{f\in\mathcal{F}} E_{f,g}\, d(\hat{f}_n(Y_1,\ldots,Y_n),f)^k \quad \underset{n\to\infty}{\longrightarrow} 0 \tag{1.6}$$

erfüllt. Wie man leicht erkennen kann, ist die Aussage der gleichmäßigen robusten d^k-Konsistenz stärker als die der robusten d^k-Konsistenz, d.h. (1.6) impliziert (1.5), die Umkehrung gilt ohne weitere Voraussetzungen nicht. Der Begriff der Robustheit ist ein in der mathematischen Statistik häufig verwendeter Ausdruck. In dem Kontext dieser Dissertation ist die Robustheit des Schätzverfahrens gegenüber Missspezifikation oder Unkenntnis der Fehlerdichte gemeint; die Begriffe der robusten Konsistenz berücksichtigen eben auch die Mehrelementigkeit der Menge $\mathcal{G}$.

Kapitel 2

Missspezifikation der Fehlerdichte

In *Meister* (2001) wird der Dekonvolutionsschätzer

$$\hat{f}_{\omega_n} := (B_g^\dagger B_g)^{-1} B_g^t R_{\omega_n} \hat{h}_{\omega_n} \tag{2.1}$$

eingeführt. Dabei bezeichnet B_g den Faltungsoperator mit der Fehlerdichte g, $B_g^\dagger$ den dazu adjungierten Operator und R_{ω_n} die orthogonale Projektion in den dort definierten Hilbertraum $L_2^{\omega_n}(\mathbb{R})$. $\hat{h}_{\omega_n}$ ist der Kernschätzer, der den sync-Kern K mit $\psi_K = \chi_{[-1;1]}$ verwendet. Dabei bezeichne $\chi_{[-1,1]}$ die Indikatorfunktion des Intervalls $[-1,1]$. Folglich besitzt der Kernschätzer die Form $\hat{h}_{\omega_n} = \frac{1}{n}\sum\limits_{j=0}^{n} \omega_n^{-1} K\left(\frac{\cdot - Y_j}{\omega_n}\right)$. $(\omega_n)_{n\in\mathbb{N}}$ ist die noch zu bestimmende Bandbreitenfolge.
Weiterhin sind die deterministischen Forderungen an f gestellt, dass

$$\mathcal{F} = \{f \;\; L_2(\mathbb{R})\text{-Dichte} \mid |\psi_f(t)| \leq C|t|^{-\beta} \quad \forall t \in \mathbb{R}\} \qquad (\beta > 1) \tag{2.2}$$

gilt. Diese Forderung kann modifiziert werden zu

$$\mathcal{F} = \{f \;\; L_2(\mathbb{R})\text{-Dichte} \mid \int\limits_{\omega}^{\infty} |\psi_f(t)|^2 \, dt \;\leq\; C\omega^{1-2\beta} \quad \forall \omega \geq \omega_0\}. \tag{2.3}$$

mit $\beta > 1, \omega_0 > 0, C > 0$. Ebenso werden Forderungen an die Fehlerdichte g gestellt:

$$|\psi_g(t)| \neq 0 \quad \forall t \in \mathbb{R} \text{ und } |\psi_g(t)| \geq D|t|^{-\eta} \quad \forall t \text{ mit } |t| \geq T \tag{2.4}$$

für fest gewähltes $T > 0$ und $\eta > 1/2$. Fehlerdichten, die dieser Forderung an das Abklingverhalten der Fouriertransformierten genügen, nennt man glatte Fehlerdichten. Betrachtet werden aber auch superglatte Fehlerdichten:

$$|\psi_g(t)| \neq 0 \quad \forall t \in \mathbb{R} \text{ und } |\psi_g(t)| \geq B|t|^\eta \exp(-\delta|t|^\xi) \quad \forall t \text{ mit } |t| \geq T. \quad (2.5)$$

Zusätzlich wird noch $g \in \mathcal{D}_2$ gefordert. In *Meister* (2001) werden iterative Verfahren vorgestellt, die gegen den Schätzer (2.1) konvergieren, der optimale Konvergenzraten bezüglich des Supremum des MISE (mean integrated squared error) über die Funktionenklasse (2.3) aufweist, wenn die Anzahl n der Beobachtungen gegen unendlich strebt. Dabei müssen jedoch sowohl für glatte als auch superglatte Fehlerdichten noch Schranken für die Fouriertransformierte nach oben gefordert werden.

2.1 Auswirkung auf das Supremum des MISE

In diesem Abschnitt soll das asymptotische Verhalten des Supremums des MISE für die Schätzerfolge $\hat{f}_{\omega_n}$ in (2.1) untersucht werden, wenn die Fehlerdichte g missspezifiziert wird, d.h. *anstatt der tatsächlichen Fehlerdichte* g *eine verfälschte Fehlerdichte* $\tilde{g}$ *zur Berechnung des Dekonvolutionsschätzers* $\hat{f}_{\omega_n}$ *verwendet wird.* In *Hesse* (1999) wird in Aussicht gestellt, dass für diese Situation weitere Forschung nötig ist.

Sowohl g als auch $\tilde{g}$ müssen in $\mathcal{G}$ liegen. Durch Kenntnis der Dichteklasse $\mathcal{G}$, die zur Konstruktion des Schätzers gebraucht werden darf, kann die fälschliche Wahl einer Fehlerdichte, die sogar außerhalb der Menge $\mathcal{G}$ liegt, zumindest theoretisch vermieden werden. Eine solche Fehlentscheidung kann nur aus numerischen Rundungsfehlern und Instabilitäten resultieren - dazu später mehr (siehe Abschnitt 2.3).
Die Frage nach dem asymptotischen Verhalten des Supremum des MISE soll nun beantwortet werden:

Satz 2.1 *Sei* $\tilde{g}$ *eine Dichte aus* $\mathcal{G} \subseteq \mathcal{D}_2$, *die zur Konstruktion des Dekonvolutionsschätzers verwendet wird,* g *sei die tatsächliche Fehlerdichte:*

$$\hat{f}_{\omega_n} := (B_{\tilde{g}}^{\dagger} B_{\tilde{g}})^{-1} B_{\tilde{g}}^{t} R_{\omega_n} \hat{h}_{\omega_n}. \tag{2.6}$$

Für die Wahl der Bandbreitenfolge (ω_n) *wird nur vorausgesetzt, dass die Dekonvolutionsschätzersequenz für richtige Spezifikation der Fehlerdichte* ($\tilde{g} = g$) *konsistent bzgl. des Supremum des MISE über die Dichtenklasse* $\mathcal{F}$ *ist. Alle Dichten* $\mathcal{G}$ *sollen entweder (2.4) oder (2.5) erfüllen; an* $\mathcal{F}$ *wird ausschließlich die Forderung der Inklusion in* $\mathcal{D}_2$ *gestellt.*
Dann gilt

$$\sup_{f \in \mathcal{F}} E_{f,g} \, \| \hat{f}_{\omega_n} - f \|^2_{L_2(\mathbb{R})} \overset{n \to \infty}{\longrightarrow} d_{\mathcal{F}}(g, \tilde{g}),$$

wobei der Abstand $d_{\mathcal{F}}(\cdot, \cdot)$ *definiert wird als*

$$d_{\mathcal{F}}(g, \tilde{g}) := \frac{1}{2\pi} \sup_{f \in \mathcal{F}} \int_{\mathbb{R}} \left| \frac{\psi_g(t)}{\psi_{\tilde{g}}(t)} - 1 \right|^2 |\psi_f(t)|^2 dt. \tag{2.7}$$

Anmerkungen:

- Die Wahl der Bandbreitenfolge ist laut Satz 2.1 freigestellt bis auf die Erhaltung der gleichmäßigen Konsistenz für richtige Spezifizierung. Durch passende Wahl der Folge lassen sich auch die optimalen Konvergenzraten erzielen. Die Wahl der Bandbreitenfolgen richtet sich gewöhnlich nach der Fehlerdichte und hier eben nach der angenommenen Dichte $\tilde{g}$ und nicht nach der wahren Dichte g. Wenn man die Bandbreitenfolge für eine gegebene Dichte $\tilde{g}$ entgegen der kanonischen Wahl, die zu optimalen Konvergenzraten führt, verändern möchte, bewirkt man damit nach obigem Satz keine Änderung des Grenzwertes der Schätzerfolge. Auch wenn man eine Verschlechterung der Konvergenzrate für richtig spezifizierte Dichten hinnehmen wollte, die Konsistenz an sich aber erhalten möchte, kann eine Abwandlung der Bandbreitenfolge keine Strategie sein, um etwa Konsistenz trotz Missspezifikation zu erreichen.

- Für den Abstand $d_{\mathcal{F}}$ lassen wir auch unendlich als möglichen Wert zu. Diesen Wert nimmt der Abstand immer dann an, wenn das Integral über $\mathbb{R}$ nicht existiert. Dies ist dadurch begründet, dass der Integrand nirgends negativ ist. Die Konvergenzaussage (2.7) ist so zu verstehen, dass die Folge der gleichmäßigen MISE gegen $d_{\mathcal{F}}(g,\tilde{g})$ konvergiert, wenn dieser Abstand reell ist und dass die Folge gegen $+\infty$ divergiert, wenn der Abstand unendlich ist. Genauere Überlegungen zu diesem Abstandsbegriff folgen im Anschluss an den Beweis des Satzes in Abschnitt 2.2.

Beweis:

$$\sup_{f\in\mathcal{F}} E_{f,g}\,\|(B_{\tilde{g}}^{\dagger}B_{\tilde{g}})^{-1}B_{\tilde{g}}^{\dagger}R_{\omega n}\hat{h}_{\omega_n} - f\|^2_{L_2(\mathbb{R})}$$

(Parseval-Identität)

$$= \tfrac{1}{2\pi}\sup_{f\in\mathcal{F}} E_{f,g}\int \left|\tfrac{1}{\psi_{\tilde{g}}(t)}\chi_{[-\omega_n,\omega_n]}(t)\tfrac{1}{n}\sum_{j=1}^{n}e^{itY_j} - \psi_f(t)\right|^2 dt$$

$$= \tfrac{1}{2\pi}\sup_{f\in\mathcal{F}} E_{f,g}\left(\int_{-\omega_n}^{\omega_n}\left|\tfrac{1}{n}\tfrac{1}{\psi_{\tilde{g}}(t)}\sum_{j=1}^{n}e^{itY_j} - \psi_f(t)\right|^2 dt + \int_{|t|>\omega_n}|\psi_f(t)|^2 dt\right)$$

(Satz von Fubini)

$$= \tfrac{1}{2\pi}\sup_{f\in\mathcal{F}}\left(\int_{-\omega_n}^{\omega_n} E_{f,g}\left|\tfrac{1}{n}\tfrac{1}{\psi_{\tilde{g}}(t)}\sum_{j=1}^{n}e^{itY_j} - \psi_f(t)\right|^2 dt + \int_{|t|>\omega_n}|\psi_f(t)|^2 dt\right)$$

(Beachte: $E_{f,g}e^{itY_j} = \psi_h(t)$ mit $h = f * g$.)

$$= \tfrac{1}{2\pi}\sup_{f\in\mathcal{F}}\Bigg(\int_{-\omega_n}^{\omega_n}\left(var\left(\tfrac{1}{n}\tfrac{1}{\psi_{\tilde{g}}(t)}\sum_{j=1}^{n}e^{itY_j}\right) + \left|\tfrac{\psi_h(t)}{\psi_{\tilde{g}}(t)} - \psi_f(t)\right|^2\right)dt$$
$$+ \int_{|t|>\omega_n}|\psi_f(t)|^2 dt\Bigg)$$

(Da $Y_1,\ldots,Y_n$ uiv sind, folgt für die Varianz:)

$$= \tfrac{1}{2\pi}\sup_{f\in\mathcal{F}}\left(\int_{-\omega_n}^{\omega_n}\tfrac{1-|\psi_h(t)|^2}{n|\psi_{\tilde{g}}(t)|^2}dt + \int_{-\omega_n}^{\omega_n}\left|\tfrac{\psi_g(t)}{\psi_{\tilde{g}}(t)} - 1\right|^2|\psi_f(t)|^2 dt + \int_{|t|>\omega_n}|\psi_f(t)|^2 dt\right)$$

(Da $f, h, g, \tilde{g}$ Dichten sind, ist der Betrag der Fouriertransformierten eine gerade Funktion.)

$$= \frac{1}{\pi} \sup_{f \in \mathcal{F}} \left(\underbrace{\int_0^{\omega_n} \frac{1 - |\psi_h(t)|^2}{n|\psi_{\tilde{g}}(t)|^2} dt}_{=:V(f)} + \underbrace{\int_0^{\omega_n} \left| \frac{\psi_g(t)}{\psi_{\tilde{g}}(t)} - 1 \right|^2 |\psi_f(t)|^2 dt}_{=:M(f)} + \underbrace{\int_{\omega_n}^{\infty} |\psi_f(t)|^2 dt}_{=:B(f)} \right).$$

◄ Die Schätzersequenzen seien so gewählt, dass der MISE gleichmäßig über $\mathcal{F}$ gegen 0 konvergiert, wenn man $\tilde{g} = g$ setzt. Daraus folgt, dass in obiger Gleichung $M(f) = 0$ ist und weiterhin gilt

$$\frac{1}{\pi} \sup_{f \in \mathcal{F}} \left(\int_0^{\omega_n} \underbrace{\frac{1 - |\psi_h(t)|^2}{n|\psi_{\tilde{g}}(t)|^2}}_{\geq 0, \text{da } h \text{ Dichte und somit } |\psi_h| \leq 1} dt + \int_{\omega_n}^{\infty} |\psi_f(t)|^2 dt \right) \stackrel{n \to \infty}{\longrightarrow} 0$$

Da $V(f)$ und $B(f)$ nichtnegativ sind, ist dies äquivalent zu

$$\frac{1}{\pi} \sup_{f \in \mathcal{F}} \left(\int_0^{\omega_n} \frac{1 - |\psi_h(t)|^2}{n|\psi_{\tilde{g}}(t)|^2} dt \right) \stackrel{n \to \infty}{\longrightarrow} 0$$

$$\underline{\text{und}} \qquad \frac{1}{\pi} \sup_{f \in \mathcal{F}} \left(\int_{\omega_n}^{\infty} |\psi_f(t)|^2 dt \right) \stackrel{n \to \infty}{\longrightarrow} 0 \tag{2.8}$$

$\Leftrightarrow$

$$\frac{1}{\pi} \sup_{f \in \mathcal{F}} \left(\int_0^{\omega_n} \left(\frac{1}{n} |\psi_{\tilde{g}}(t)|^{-2} - \frac{1}{n} |\psi_f(t)|^2 \right) dt \right) \stackrel{n \to \infty}{\longrightarrow} 0$$

$$\underline{\text{und}} \qquad \frac{1}{\pi} \sup_{f \in \mathcal{F}} \left(\int_{\omega_n}^{\infty} |\psi_f(t)|^2 dt \right) \stackrel{n \to \infty}{\longrightarrow} 0$$

$\Leftrightarrow$

$$\tfrac{1}{\pi}\int_0^{\omega_n} \tfrac{1}{n}|\psi_{\tilde g}(t)|^{-2}dt \;-\; \tfrac{1}{n\pi}\inf_{f\in\mathcal{F}}\int_0^{\omega_n}|\psi_f(t)|^2dt \;\overset{n\to\infty}{\longrightarrow}\; 0$$

$$\underline{\text{und}} \qquad \tfrac{1}{\pi}\sup_{f\in\mathcal{F}}\left(\int_{\omega_n}^{\infty}|\psi_f(t)|^2dt\right) \overset{n\to\infty}{\longrightarrow} 0$$

Die Bedingung $\tfrac{1}{\pi}\sup_{f\in\mathcal{F}}\left(\int_{\omega_n}^{\infty}|\psi_f(t)|^2dt\right) \overset{n\to\infty}{\longrightarrow} 0$ impliziert, dass $\sup_{f\in\mathcal{F}}\|f\|_{L_2(\mathbb{R})} < \infty$ gilt, denn für ein festes ausreichend großes m gilt dann

$$\begin{aligned}
\|f\|^2_{L_2(\mathbb{R})} &= \tfrac{1}{2\pi}\|\psi_f\|^2_{L_2(\mathbb{R})} \\
&= \tfrac{1}{2\pi}\int_{|t|\le\omega_m}\underbrace{|\psi_f(t)|^2}_{\le 1}dt + \tfrac{1}{2\pi}\int_{|t|>\omega_m}|\psi_f(t)|^2dt \;\le\; \tfrac{\omega_m}{\pi} + \tfrac{1}{2\pi}\sup_{\tilde f\in\mathcal{F}}\int_{|t|>\omega_m}|\psi_{\tilde f}(t)|^2dt \\
&\le \tfrac{\omega_m}{\pi} + 1.
\end{aligned}$$

Also ist die L_2 -Norm aller $f\in\mathcal{F}$ gleichgradig beschränkt. Daraus wiederum folgt

$$0 \;\le \tfrac{1}{\pi n}\inf_{f\in\mathcal{F}}\int_0^{\omega_n}|\psi_f(t)|^2dt \;\le\; n^{-1}\sup_{f\in\mathcal{F}}\|f\|^2_{L_2(\mathbb{R})} \;\overset{n\to\infty}{\longrightarrow} 0$$

Folglich gilt $\tfrac{1}{\pi n}\inf_{f\in\mathcal{F}}\int_0^{\omega_n}|\psi_f(t)|^2dt \overset{n\to\infty}{\longrightarrow} 0$.

Daher ist (2.8) äquivalent zu

$$\frac{1}{n\pi}\int_0^{\omega_n}|\psi_{\tilde g}(t)|^{-2}dt \overset{n\to\infty}{\longrightarrow} 0 \quad \text{und} \quad \frac{1}{\pi}\sup_{f\in\mathcal{F}}\int_{\omega_n}^{\infty}|\psi_f(t)|^2dt \overset{n\to\infty}{\longrightarrow} 0 \tag{2.9}$$

Die Bedingung (2.9) sagt aus, dass $\sup_{f\in\mathcal{F}}B(f)$ und $\sup_{f\in\mathcal{F}}V(f)$ gegen 0 auch für $\tilde g \neq g$ konvergieren, zumal $0 \le 1-|\psi_h(t)|^2 \le 1$ gilt. ▶

Fahren wir nun mit der Betrachtung des Supremum des MISE für $\tilde{g} \neq g$ fort. Dieser beträgt nach obigen Berechnungen

$$\frac{1}{\pi} \sup_{f \in \mathcal{F}} \left(V(f) + B(f) + M(f) \right). \tag{2.10}$$

Da für alle $f \in \mathcal{F}$ V, B, M nichtnegativ sind, besitzt (2.10) mit der Aussage (2.9) denselben Grenzwert wie

$$\frac{1}{\pi} \sup_{f \in \mathcal{F}} \underbrace{\int_0^{\omega_n} \left| \frac{\psi_g(t)}{\psi_{\tilde{g}(t)}} - 1 \right|^2 |\psi_f(t)|^2 dt}_{=M(f)} \tag{2.11}$$

im Falle der Konvergenz. Ebenso divergiert (2.11) aber auch genau dann gegen ∞, wenn (2.10) dies tut.
Es stellt sich nun die Frage, wie sich die Folge (ω_n) für $n \to \infty$ verhält. Von dieser Folge kennt man nur die Bedingungen (2.9). Wir werden in diesem Zusammenhang zwei Fälle unterscheiden:

1. Fall: $\exists T > 0 \, \forall f \in \mathcal{F} \, \forall t \geq T \, : \, \psi_f(t) = 0$
Daraus folgt, dass die Menge $\{T \geq 0 \mid \psi_f(t) = 0 \quad \forall f \in \mathcal{F} \, \forall t \geq T\}$ nicht leer ist und auch positive Elemente enthält. Im Folgenden wird gezeigt, dass diese Menge in $\mathbb{R}$ abgeschlossen ist:
Sei $(T_n)_n$ eine Folge in $\{T \geq 0 \mid \psi_f(t) = 0 \quad \forall f \in \mathcal{F} \, \forall t \geq T\}$, die gegen ein $T \geq 0$ konvergiert. Dann gilt $\psi_f(T_n) = 0 \quad \forall f \in \mathcal{F}, \forall n \in \mathbb{N}$. Für jedes festgehaltene $f \in \mathcal{F}$ folgt dann $\psi_f(T) = 0$, da ψ_f als Fouriertransformierte einer Dichte stetig ist. Für ein beliebiges $t \in \mathbb{R}$ mit $t > T$ existiert ein $n \in \mathbb{N}$, so dass für alle $m \geq n$ gilt: $t \geq T_m$, da $T_n \overset{n \to \infty}{\longrightarrow} T$. Da jedes T_m in der Menge $\{T \geq 0 \mid \psi_f(t) = 0 \quad \forall f \in \mathcal{F} \, \forall t \geq T\}$ liegt, gilt $\forall f \in \mathcal{F} \, : \, \psi_f(t) = 0$.
Insgesamt folgt, dass $\forall f \in \mathcal{F}, \forall t$ mit $t \geq T \, : \, \psi_f(t) = 0$ und somit dass T Element der Menge $\{T \geq 0 \mid \psi_f(t) = 0 \quad \forall f \in \mathcal{F} \, \forall t \geq T\}$ ist. Also ist $\{T \geq 0 \mid \psi_f(t) = 0 \quad \forall f \in \mathcal{F} \, \forall t \geq T\}$ in $\mathbb{R}_0^+$ abgeschlossen.
Aus der Abgeschlossenheit und Beschränktheit nach unten folgt, dass das Minimum der nicht leeren Menge $\{T \geq 0 \mid \psi_f(t) = 0 \quad \forall f \in \mathcal{F} \, \forall t \geq T\}$ in $\mathbb{R}_0^+$ existiert. Zudem ist leicht zu erkennen, dass 0 nicht in $\{T \geq 0 \mid \psi_f(t) = 0 \quad \forall f \in \mathcal{F} \, \forall t \geq T\}$ liegen kann, denn sonst wäre die Fouriertransformierte aller Funktionen aus $\mathcal{F}$ konstant 0 und die Elemente von $\mathcal{F}$ wären keine Dichten $(\psi_f(0) = 0 \neq 1)$. Daher gilt $T := \min\{T \geq 0 \mid \psi_f(t) = 0 \quad \forall f \in \mathcal{F} \, \forall t \geq T\} > 0$.

Weiterhin folgt die Aussage:

$$\forall \epsilon > 0 \, \exists N \, \forall n \geq N \, : \, \omega_n \geq T - \epsilon.$$

Zum Beweis dieser Behauptung nehmen wir an, es gelte $\exists \epsilon > 0 \, \forall N \, \exists n \geq N$: $\omega_n < T - \epsilon$. Daher gibt es ein $\epsilon > 0$, so dass eine Teilfolge $(\omega_{\sigma(n)})_n$ existiert mit $\omega_{\sigma(n)} < T - \epsilon$, $\forall n \in \mathbb{N}$. $\epsilon > 0$ kann dabei als beliebig klein angenommen werden, solange es positiv bleibt, schließlich schwächt eine Verkleinerung von ϵ die Aussage ab. Auf diese Weise kann $T - \epsilon > 0$ garantiert werden. Für diese Folge gilt

$$0 \overset{n \to \infty}{\longleftarrow} \sup_{f \in \mathcal{F}} \int\limits_{\omega_{\sigma(n)}}^{\infty} |\psi_f(t)|^2 dt \geq \sup_{f \in \mathcal{F}} \int\limits_{T-\epsilon}^{\infty} |\psi_f(t)|^2 dt > 0.$$

Damit haben wir einen Widerspruch gefunden. Der letzte Schritt der Ungleichungskette ist durch die Definition von T als Minimum obiger Menge und der Stetigkeit von ψ_f begründet. Wegen der Fallbedingung gilt:

$$\begin{aligned} \infty > \sup_{f \in \mathcal{F}} \int\limits_0^T \left| \frac{\psi_g(t)}{\psi_{\tilde{g}}(t)} - 1 \right|^2 |\psi_f(t)|^2 dt \\ = \sup_{f \in \mathcal{F}} \int\limits_0^\infty \left| \frac{\psi_g(t)}{\psi_{\tilde{g}}(t)} - 1 \right|^2 |\psi_f(t)|^2 dt \end{aligned}$$

Weiter gilt für jedes $\epsilon > 0$ und für alle $n \geq N$ ab einem gewissen N

$$\begin{aligned} &\sup_{f \in \mathcal{F}} \int\limits_0^{\omega_n} \left| \tfrac{\psi_g(t)}{\psi_{\tilde{g}}(t)} - 1 \right|^2 |\psi_f(t)|^2 dt + \sup_{f \in \mathcal{F}} \int\limits_{T-\epsilon}^{T} \left| \tfrac{\psi_g(t)}{\psi_{\tilde{g}}(t)} - 1 \right|^2 |\psi_f(t)|^2 dt \\ &\geq \sup_{f \in \mathcal{F}} \int\limits_0^{\infty} \left| \tfrac{\psi_g(t)}{\psi_{\tilde{g}}(t)} - 1 \right|^2 |\psi_f(t)|^2 dt \\ &\geq \sup_{f \in \mathcal{F}} \int\limits_0^{\omega_n} \left| \tfrac{\psi_g(t)}{\psi_{\tilde{g}}(t)} - 1 \right|^2 |\psi_f(t)|^2 dt. \end{aligned}$$

Wählt man $\epsilon > 0$ ausreichend klein, so kann man für jede positive Zahl erreichen, dass $\sup_{f \in \mathcal{F}} \int\limits_{T-\epsilon}^{T} \left| \tfrac{\psi_g(t)}{\psi_{\tilde{g}}(t)} - 1 \right|^2 |\psi_f(t)|^2 dt \leq \int\limits_{T-\epsilon}^{T} \left| \tfrac{\psi_g(t)}{\psi_{\tilde{g}}(t)} - 1 \right|^2 dt$ kleiner als diese wird, denn die stetige Funktion $\left|\frac{\psi_g}{\psi_{\tilde{g}}} - 1\right|^2$ ist über jedes Kompaktum integrabel, die Konvergenz gegen 0 liefert dann der Satz von der majorisierten Konvergenz. Dies impliziert

$$\sup_{f\in\mathcal{F}} \int_0^{\omega_n} \left|\frac{\psi_g(t)}{\psi_{\tilde{g}}(t)} - 1\right|^2 |\psi_f(t)|^2 dt \overset{n\to\infty}{\longrightarrow} \sup_{f\in\mathcal{F}} \int_0^{\infty} \left|\frac{\psi_g(t)}{\psi_{\tilde{g}}(t)} - 1\right|^2 |\psi_f(t)|^2 dt.$$

Damit ist der Satz im 1. Fall bewiesen. Schreiten wir nun zum

2. Fall: $\forall T > 0\, \exists f \in \mathcal{F}\, \exists t > T \, : \, \psi_f(t) \neq 0$.

Nehmen wir an, $(\omega_n)_n$ divergiere nicht gegen unendlich. Dann existiert eine nach oben durch ein $S > 0$ beschränkte Teilfolge $(\omega_{\sigma(n)})_n$. Nach der Annahme existiert ein $f' \in \mathcal{F}$ und ein $t > S$, so dass $\psi_{f'}(t) \neq 0$ gilt. Das impliziert unter Verwendung von (2.9)

$$0 \overset{n\to\infty}{\longleftarrow} \sup_f \int_{\omega_{\sigma(n)}}^{\infty} |\psi_f(t)|^2 dt \geq \sup_f \int_S^{\infty} |\psi_f(t)|^2 dt \geq \int_S^{\infty} |\psi_{f'}(t)|^2 dt > 0.$$

Damit ist ein Widerspruch erreicht und $(\omega_n)_n$ divergiert gegen ∞. $\sup_{f\in\mathcal{F}} \int_0^{\omega_n} \left|\frac{\psi_g(t)}{\psi_{\tilde{g}}(t)} - 1\right|^2 |\psi_f(t)|^2 dt$ konvergiert somit gegen $\sup_{f\in\mathcal{F}} \int_0^{\infty} \left|\frac{\psi_g(t)}{\psi_{\tilde{g}}(t)} - 1\right|^2 |\psi_f(t)|^2 dt$, wenn dieser Ausdruck als reelle Zahl existiert, anderenfalls divergiert $\sup_{f\in\mathcal{F}} \int_0^{\omega_n} \left|\frac{\psi_g(t)}{\psi_{\tilde{g}}(t)} - 1\right|^2 |\psi_f(t)|^2 dt$ gegen unendlich, diese intuitive Folgerung wird in Lemma 2.3 in allgemeinerem Kontext bewiesen. Daher verzichte ich an dieser Stelle auf den exakten Beweis des Spezialfalls von Lemma 2.3.

Unter Ausnützung der Symmetrie des Betrages der Fouriertransformierten von Dichten folgt insgesamt die Behauptung des Satzes.
■

2.2 Der Abstand $d_{\mathcal{F}}$

Durch Satz 2.1 wird die zentrale Bedeutung des dort definierten Abstandes zweier Fehlerdichten $d_{\mathcal{F}}(\cdot,\cdot)$ bei der Untersuchung des Verhaltens des Supremum des MISE hinsichtlich der Missspezifikation der Fehlerdichte deutlich. In diesem Abschnitt möchte ich daher diesen Abstandsbegriff etwas beleuchten.

$$\begin{aligned} d_{\mathcal{F}} \; : \; \mathcal{G} \times \mathcal{G} &\longrightarrow \mathbb{R} \cup \{+\infty\} \\ (g, \tilde{g}) &\longmapsto \tfrac{1}{2\pi} \sup_{f \in \mathcal{F}} \int \left| \tfrac{\psi_g(t)}{\psi_{\tilde{g}}(t)} - 1 \right|^2 |\psi_f(t)|^2 \, dt. \end{aligned} \tag{2.12}$$

(Unbestimmte Integrale seien als Integrale über ganz $\mathbb{R}$ zu verstehen.) Auffällig ist zum einen, dass der Abstand $d_{\mathcal{F}}$ zweier Fehlerdichten aus $\mathcal{G}$ auch unendlich groß sein kann, und zum anderen, dass $d_{\mathcal{F}}$ nicht symmetrisch ist und g und $\tilde{g}$ also nicht vertauscht werden dürfen. Daher wird durch $d_{\mathcal{F}}$ keine Metrik auf $\mathcal{G}$ definiert! Dennoch lassen sich bekannte Eigenschaften des Abstandes herleiten:

- $d_{\mathcal{F}}$ ist positiv semidefinit, d.h.

$$\begin{aligned} d_{\mathcal{F}}(g, \tilde{g}) \geq 0 \qquad & \forall g, \tilde{g} \in \mathcal{G} \\ d_{\mathcal{F}}(g, g) = 0 \qquad & \forall g \in \mathcal{G} \end{aligned}$$

  Die erste Bedingung folgt aus der Nichtnegativität des Integranden, die zweite ist elementar.

- Daran schließt sich unmittelbar die Frage nach der positiven Definitheit von $d_{\mathcal{F}}$ an. Da der Integrand nichtnegativ ist, gilt

$$d_{\mathcal{F}}(g, \tilde{g}) = 0 \;\Leftrightarrow\; \left| \tfrac{\psi_g(t)}{\psi_{\tilde{g}}(t)} - 1 \right|^2 |\psi_f(t)|^2 = 0 \text{ für LB-f.a. } t \in \mathbb{R}, \forall f \in \mathcal{F}$$

Da der Integrand stetig ist, folgt:

$$\begin{aligned} &\Leftrightarrow\; \left| \tfrac{\psi_g(t)}{\psi_{\tilde{g}}(t)} - 1 \right|^2 |\psi_f(t)|^2 = 0 \quad \forall t \in \mathbb{R}, \forall f \in \mathcal{F} \\ &\Leftrightarrow\; \psi_g(t) = \psi_{\tilde{g}}(t) \vee \psi_f(t) = 0 \quad \forall t \in \mathbb{R}, \forall f \in \mathcal{F} \\ &\Leftrightarrow\; g * f = \tilde{g} * f \,, \forall f \in \mathcal{F} \;\Rightarrow\; g = \tilde{g} \end{aligned}$$

  $d_{\mathcal{F}}$ ist folglich genau dann positiv definit, wenn gilt

$$g * f = \tilde{g} * f \,, \forall f \in \mathcal{F} \;\Rightarrow\; g = \tilde{g}. \tag{2.13}$$

Dies ist eine sehr milde Bedingung. Hinreichend dafür ist nach obigem bereits, dass $\mathcal{F}$ eine Dichte mit nirgends verschwindender Fouriertransformierten enthält (z.B. Normalverteilungsdichten, Laplace-Dichten).

- $\mathcal{G}$ besitze die Eigenschaft, dass für beliebig, aber fest ausgewählte $g, \tilde{g} \in \mathcal{G}$ auch $\lambda g + (1-\lambda)\tilde{g}$ für alle $\lambda \in [0,1]$ in $\mathcal{G}$ liege. Insbesondere konvexe Mengen $\mathcal{G}$ erfüllen diese Eigenschaft sogar für beliebige $g, \tilde{g}$. Dann gilt

$$\begin{aligned}
&d_{\mathcal{F}}(\lambda g + (1-\lambda)\tilde{g}, \tilde{g}) \\
&= \tfrac{1}{2\pi} \sup_{f \in \mathcal{F}} \int \left| \tfrac{\lambda\psi_g(t) + (1-\lambda)\psi_{\tilde{g}}}{\psi_{\tilde{g}}(t)} - 1 \right|^2 |\psi_f(t)|^2 \, dt \\
&= \tfrac{1}{2\pi} \sup_{f \in \mathcal{F}} \int \left| \lambda \tfrac{\psi_g(t)}{\psi_{\tilde{g}}(t)} + 1 - \lambda - 1 \right|^2 |\psi_f(t)|^2 \, dt \\
&= \tfrac{1}{2\pi} \sup_{f \in \mathcal{F}} \int \left| \lambda \tfrac{\psi_g(t)}{\psi_{\tilde{g}}(t)} - \lambda \right|^2 |\psi_f(t)|^2 \, dt \\
&= \lambda^2 \cdot d_{\mathcal{F}}(g, \tilde{g}).
\end{aligned} \tag{2.14}$$

- Für $g, \tilde{g} \in \mathcal{G}$ gelte : $|\psi_g(t)| \geq |\psi_{\tilde{g}}(t)| \quad \forall t \in \mathbb{R}$, so folgt

$$d_{\mathcal{F}}(g, \tilde{g}) \geq d_{\mathcal{F}}(\tilde{g}, g) \tag{2.15}$$

Beweis:

$$\begin{aligned}
d_{\mathcal{F}}(g, \tilde{g}) &= \tfrac{1}{2\pi} \sup_{f \in \mathcal{F}} \int \left| \tfrac{\psi_g(t)}{\psi_{\tilde{g}}(t)} - 1 \right|^2 |\psi_f(t)|^2 \, dt \\
&= \tfrac{1}{2\pi} \sup_{f \in \mathcal{F}} \int \left| 1 - \tfrac{\psi_{\tilde{g}}(t)}{\psi_g(t)} \right|^2 \underbrace{\left| \frac{\psi_g(t)}{\psi_{\tilde{g}}(t)} \right|^2}_{\geq 1, \forall t} |\psi_f(t)|^2 \, dt \\
&\geq \tfrac{1}{2\pi} \sup_{f \in \mathcal{F}} \int \left| 1 - \tfrac{\psi_{\tilde{g}}(t)}{\psi_g(t)} \right|^2 |\psi_f(t)|^2 \, dt \\
&= d_{\mathcal{F}}(\tilde{g}, g).
\end{aligned}$$

Die Ungleichung (2.15) ist bei der Wahl der Fehlerdichte von Bedeutung, darauf werde ich im vierten Abschnitt dieses Kapitels näher eingehen. Zunächst möchte ich aber noch die Frage nach Verbesserungsmöglichkeiten der Schätzersequenz hinsichtlich ihres asymptotischen Verhaltens aufgreifen.

2.3 Alternative Kernschätzer

Der Dekonvolutionsdichteschätzer (2.1) entsteht durch Entfaltung eines Kernschätzers mit der Fehlerdichte g. Der dabei verwendete Kern ist der so genannte sync-Kern $\left(K(x) := \frac{\sin(x)}{\pi x}\right)$. Im vorigen Abschnitt wurde lediglich die Bandbreitenfolge $(\omega_n)_n$ variiert. In *Meister* (2001) wird jedoch auch der Fall betrachtet, dass die Fehlerdichte $\tilde{g}$ nicht in $L_2(\mathbb{R})$ liegt. Dabei wird als Kern die Funktion $K(x) = \frac{2}{\pi x^2}\left(\cos(\frac{1}{2}x) - \cos(x)\right)$ verwendet und der Projektionsoperator leicht verändert. Es stellt sich also die Frage, ob eine andere Wahl des Kernes eine Veränderung der asymptotischen Qualität der Schätzersequenz bewirkt. Hierbei bietet es sich an, den allgemeinen Kernschätzer

$$\hat{f}_n(x) := \frac{1}{2\pi} \int \exp(-itx) L_n(t) \frac{1}{n} \sum_{j=1}^{n} e^{itY_j} / \psi_g(t)\, dt \qquad (2.16)$$

mit der Folge L_n zu betrachten, wobei nicht zwingend $L_n = \psi_K(\cdot/\omega_n)$ gilt. Allerdings muss $L_n/\psi_{\tilde{g}} \in L_2(\mathbb{R})$ sein, um die Existenz eines Schätzers $\hat{f}_n$ in $L_2(\mathbb{R})$ nach (2.16) zu gewährleisten . $Y_1, \ldots, Y_n$ seien die verrauschten Beobachtungen. Dieser Schätzer tritt unter anderem bei *Neumann* (1997) auf. Für den sync-Kern erhält man $L_n = \chi_{[-\omega_n, \omega_n]}$ und der Kern $K(x) = \frac{2}{\pi x^2}\left(\cos(\frac{1}{2}x) - \cos(x)\right)$ lässt sich durch $L_n = \psi^2_{\omega_n K(\omega_n \cdot)}$ darstellen. Als Forderung an die Schätzerfolge soll nach wie vor bestehen bleiben, dass für richtige Wahl der Fehlerdichte gleichmäßige Konsistenz bestehen soll.
Auch können die Voraussetzungen an $\mathcal{G}$ abgeschwächt werden. Anstatt die Inklusion in der Klasse der glatten oder superglatten Dichten zu fordern, genügt es, $\psi_g(t) \neq 0, \forall t \in \mathbb{R}, \forall g \in \mathcal{G}$ vorauszusetzen. Eine weitere Definitionsausweitung sollte an die Wahl einer Fehlerdichte getätigt werden. Im ersten Abschnitt wird eingeräumt, dass die Fehlerdichte g zwar falsch gewählt, aber dennoch eine Dichte in $\mathcal{G}$ ist. Davon kann man bei numerischen Anwendungen von Dekonvolutionsdichteschätzern nicht ausgehen, zumal die Menge aller Dichten $\mathcal{D}$ Teilmenge des Randes der Einheitssphäre des $L_1(\mathbb{R})$ ist und somit nirgends dicht in $L_1(\mathbb{R})$ liegt. Eine beliebig kleine positive Störung der

Dichte bezüglich der L_1 -Norm, wie sie bei numerischen Berechnungen mit Rundungen immer auftritt, kann die verwendete Funktion $\tilde{g}$ aus $\mathcal{D}$ und somit auch aus $\mathcal{G}$ hinausdrängen. Daher verallgemeinern wir, dass an die Stelle von $\psi_{\tilde{g}}$ nun eine Funktion $\xi : \mathbb{R} \to \mathbb{C}$ mit $\inf_{t\in[-R,R]} |\xi(t)| > 0\,,\ \forall R > 0$ tritt. Letztere Forderung an ξ muss gestellt werden, um die Existenz von (2.16) zu sichern.
Die Dekonvolutionsschätzerfolge, die nun in zweifacher Hinsicht verallgemeinert wurde, lautet also

$$\hat{f}_n(x) := \frac{1}{2\pi} \int \exp(-itx) L_n(t) \frac{1}{n} \sum_{j=1}^{n} e^{itY_j} / \xi(t)\, dt. \tag{2.17}$$

Lemma 2.1 *Sei der Schätzer $\hat{f}_n$ durch (2.17) definiert und für ξ gelte $\inf_{|t|\le T} |\xi(t)| > 0$ für jedes $T > 0$ sowie $L_n/\xi \in L_2(\mathbb{R})$. Dann ergibt sich das Supremum des MISE der Schätzerfolge $(\hat{f}_n)_{n\in\mathbb{N}}$ aus (2.17) als*

$$\begin{aligned}
&\sup_{f\in\mathcal{F}} E_{f,g} \, \|\hat{f}_n - f\|^2_{L_2(\mathbb{R})} \\
&= \frac{1}{2\pi} \sup_{f\in\mathcal{F}} \int \left(\left| \frac{L_n(t)}{\xi(t)} \right|^2 \cdot \frac{1 - |\psi_{f*g}(t)|^2}{n} + \left| L_n(t) \frac{\psi_f(t)\psi_g(t)}{\xi(t)} - \psi_f(t) \right|^2 \right) dt
\end{aligned}$$

Beweis: Betrachten wir das Risiko

$$\sup_{f\in\mathcal{F}} E_{f,g} \, \|\hat{f}_n - f\|^2_{L_2(\mathbb{R})}$$

$$= \sup_{f\in\mathcal{F}} E_{f,g} \, \| \tfrac{1}{2\pi} \int \exp(-it\cdot) L_n(t) \tfrac{1}{n} \sum_{j=1}^{n} e^{itY_j} / \xi(t)\, dt - f \|^2_{L_2(\mathbb{R})}$$

(Parseval-Identität)

$$= \tfrac{1}{2\pi} \sup_{f\in\mathcal{F}} E_{f,g} \int \left| L_n(t) \tfrac{1}{n} \sum_{j=1}^{n} e^{itY_j} / \xi(t)\, dt - \psi_f(t) \right|^2 dt$$

(Satz von Fubini)

$$= \frac{1}{2\pi}\sup_{f\in\mathcal{F}} \int E_{f,g}\left|L_n(t)\frac{1}{n}\sum_{j=1}^{n} e^{itY_j}/\xi(t)\,dt - \psi_f(t)\right|^2 dt$$

$$= \frac{1}{2\pi}\sup_{f\in\mathcal{F}} \int \Bigg(var_{f,g}\left(L_n(t)\frac{1}{n}\sum_{j=1}^{n} e^{itY_j}/\xi(t)\right) + \left|E_{f,g}L_n(t)\frac{1}{n}\sum_{j=1}^{n} e^{itY_j}/\xi(t) - \psi_f(t)\right|^2\Bigg)dt$$

$$= \frac{1}{2\pi}\sup_{f\in\mathcal{F}} \int \left(\left|\frac{L_n(t)}{\xi(t)}\right|^2 \cdot \frac{1-|\psi_{f*g}(t)|^2}{n} + \left|L_n(t)\frac{\psi_f(t)\psi_g(t)}{\xi(t)} - \psi_f(t)\right|^2\right) dt$$

■

Dieser Term ist nach unten durch $\frac{1}{2\pi}\sup_{f\in\mathcal{F}} \int |\psi_f(t)|^2 \left|L_n(t)\frac{\psi_g(t)}{\xi(t)} - 1\right|^2 dt$ und durch $\frac{1}{2\pi}\sup_{f\in\mathcal{F}} \int \left|\frac{L_n(t)}{\xi(t)}\right|^2 \cdot \frac{1-|\psi_{f*g}(t)|^2}{n} dt$ sowie nach oben durch

$$\frac{1}{2\pi}\sup_{f\in\mathcal{F}} \int \left|\frac{L_n(t)}{\xi(t)}\right|^2 \cdot \frac{1-|\psi_{f*g}(t)|^2}{n} dt + \frac{1}{2\pi}\sup_{f\in\mathcal{F}} \int \left|L_n(t)\frac{\psi_f(t)\psi_g(t)}{\xi(t)} - \psi_f(t)\right|^2 dt$$

abschätzbar.
Wenn es sich bei der Funktion ξ um die Fouriertransformierte einer Dichte handelt, genügt es zu fordern, dass ξ nirgends verschwindet.

Lemma 2.2 *Die gewählte Funktion ξ sei die Fouriertransformierte einer Dichte aus $\mathcal{G}$. Die Fouriertransformierten aller Dichten aus $\mathcal{G}$ mögen nirgends verschwinden. Für richtige Spezifizierung der Fehlerdichte, d.h. die Fouriertransformierte der tatsächlichen Fehlerdichte sei ξ, konvergiere das Supremum des MISE gegen 0. Dann gilt*

$$\sup_{f\in\mathcal{F}} \int |\psi_f(t)|^2 |L_n(t) - 1|^2\, dt \overset{n\to\infty}{\longrightarrow} 0 \tag{2.18}$$

und

$$\sup_{f\in\mathcal{F}} \int_{-R}^{R} \left| L_n(t)\frac{\psi_g(t)}{\xi(t)} - \frac{\psi_g(t)}{\xi(t)} \right|^2 |\psi_f(t)|^2 \, dt \overset{n\to\infty}{\longrightarrow} 0 \tag{2.19}$$

für jedes $R > 0$ *und für alle* $g \in \mathcal{G}$ *und*

$$\frac{1}{n} \int \left| \frac{L_n(t)}{\xi(t)} \right|^2 dt \overset{n\to\infty}{\longrightarrow} 0. \tag{2.20}$$

Beweis: Nach Lemma 2.1 folgt aus der gleichmäßigen L_2-Konsistenz zum einen

$$\Rightarrow \quad \tfrac{1}{2\pi}\sup_{f\in\mathcal{F}} \int |\psi_f(t)|^2 \, |L_n(t) - 1|^2 \; dt \quad \overset{n\to\infty}{\longrightarrow} 0 \text{ , also (2.18)}$$

$(R > 0$ beliebig$)$

$$\Rightarrow \quad \tfrac{1}{2\pi}\sup_{f\in\mathcal{F}} \int_{-R}^{R} |\psi_f(t)|^2 \, |L_n(t) - 1|^2 \; dt \quad \overset{n\to\infty}{\longrightarrow} 0$$

Da wegen der Voraussetzungen (ξ ist Dichte aus $\mathcal{G}$). $\xi(t) \neq 0, \forall t \in \mathbb{R}$ und der Stetigkeit von ξ $\inf\limits_{|t|\leq R} |\xi(t)|^2 > 0$ gilt, folgt weiter

$$\sup_{|t|\leq R} \left| \frac{\psi_g(t)}{\xi(t)} \right|^2 \leq \frac{1}{\inf\limits_{|t|\leq R} |\xi(t)|^2} < \infty.$$

Damit gilt

$$\tfrac{1}{2\pi}\sup_{f\in\mathcal{F}} \int_{-R}^{R} \sup_{|\tau|\leq R} \left| \tfrac{\psi_g(\tau)}{\xi(\tau)} \right|^2 |\psi_f(t)|^2 \, |L_n(t) - 1|^2 \; dt \quad \overset{n\to\infty}{\longrightarrow} 0$$

$$\geq \tfrac{1}{2\pi}\sup_{f\in\mathcal{F}} \int_{-R}^{R} \left| L_n(t)\tfrac{\psi_g(t)}{\xi(t)} - \tfrac{\psi_g(t)}{\xi(t)} \right|^2 |\psi_f(t)|^2 \, dt$$

$$\geq 0$$

Also gilt (2.19)

$$\frac{1}{2\pi}\sup_{f\in\mathcal{F}}\int\limits_{-R}^{R}\left|L_n(t)\frac{\psi_g(t)}{\xi(t)}-\frac{\psi_g(t)}{\xi(t)}\right|^2|\psi_f(t)|^2\,dt \stackrel{n\to\infty}{\longrightarrow} 0$$

für alle $R>0$.

Als zweite Konsistenzbedingung erhält man

$$\begin{aligned}
&\tfrac{1}{2\pi}\sup_{f\in\mathcal{F}}\int\left|\tfrac{L_n(t)}{\xi(t)}\right|^2\cdot\tfrac{1-|\psi_{f*g}(t)|^2}{n}dt \qquad \stackrel{n\to\infty}{\longrightarrow} 0\\
&\geq \tfrac{1}{2\pi}\int\limits_{|t|\geq T}\left|\tfrac{L_n(t)}{\xi(t)}\right|^2\cdot\tfrac{1-|\psi_{f*g}(t)|^2}{n}dt\\
&\geq 0
\end{aligned}$$

für beliebiges $T>0$ und eine beliebige Dichte $f\in\mathcal{F}$. Da $f*g$ eine Dichte ist, gilt nach Ergebnissen der Wahrscheinlichkeitstheorie $\psi_{f*g}(t)=1\Leftrightarrow t=0$ und $|\psi_{f*g}(t)|\leq 1, \forall t\in\mathbb{R}$. Aus der Stetigkeit von ψ_{f*g} folgt schließlich $\inf\limits_{|t|\geq T}(1-|\psi_{f*g}(t)|^2)>0$. Daraus folgt weiter

$$\begin{aligned}
&\int\limits_{|t|\geq T}\left|\tfrac{L_n(t)}{\xi(t)}\right|^2\cdot\tfrac{1-|\psi_{f*g}(t)|^2}{n}dt \qquad\qquad \stackrel{n\to\infty}{\longrightarrow} 0\\
&\geq \tfrac{1}{n}\underbrace{\inf_{|t|\geq T}(1-|\psi_{f*g}(t)|^2)}_{\text{unabhängig von } n}\int\limits_{|t|\geq T}\left|\tfrac{L_n(t)}{\xi(t)}\right|^2dt\\
&\geq 0
\end{aligned}$$

$$\Rightarrow \tfrac{1}{n}\int\limits_{|t|\geq T}\left|\tfrac{L_n(t)}{\xi(t)}\right|^2dt \stackrel{n\to\infty}{\longrightarrow} 0 \text{ für jedes } T>0.$$

Wir wählen ein $f\in\mathcal{F}$ beliebig aus und bestimmen das noch frei wählbare $T>0$ so, dass $|\psi_f(t)|>\frac{1}{2}$ für alle t mit $|t|\leq T$. Dies ist möglich, weil $\psi_f(0)=1$ und ψ_f stetig ist.

$$\sup_{f\in\mathcal{F}} \int |\psi_f(t)|^2 \left|L_n(t)-1\right|^2 dt \overset{n\to\infty}{\longrightarrow} 0$$

$$\geq \int_{-T}^{T} |\psi_f(t)|^2 \left|L_n(t)-1\right|^2 dt$$

$$\geq \tfrac{1}{2}\int_{-T}^{T} \left|L_n(t)-1\right|^2 dt$$

Das bedeutet also, dass die Funktionenfolge $(L_n - 1)_n$ im Raum $L_2([-T,T])$ konvergiert und somit beschränkt ist. Da 1 als Funktion dieses Raumes aufgefasst die (endliche) Norm $\sqrt{2T}$ besitzt, ist die Folge $(L_n)_n$ in der $L_2([-T,T])$ -Norm beschränkt.
Wenn man jetzt noch bedenkt, dass die Forderung $\inf\limits_{|t|\leq T}|\xi(t)| > 0$ nach der Voraussetzung, dass die Fouriertransformierten der Dichten aus $\mathcal{G}$ nirgends verschwinden, erfüllt ist, folgt (2.20)

$$\frac{1}{n}\int_{-T}^{T} \left|\frac{L_n(t)}{\xi(t)}\right|^2 dt \overset{n\to\infty}{\longrightarrow} 0$$

und somit

$$\begin{aligned} &\tfrac{1}{n}\int \left|\tfrac{L_n(t)}{\xi(t)}\right|^2 dt \\ &= \tfrac{1}{n}\int_{-T}^{T} \left|\tfrac{L_n(t)}{\xi(t)}\right|^2 dt + \tfrac{1}{n}\int_{|t|\geq T} \left|\tfrac{L_n(t)}{\xi(t)}\right|^2 dt \overset{n\to\infty}{\longrightarrow} 0. \end{aligned}$$

■

Kehren wir jetzt zur Betrachtung des MISE im Falle der Missspezifikation zurück. Wenn ξ nicht Fouriertransformierte einer Dichte aus $\mathcal{G}$ ist, gibt es keine richtige Spezifizierung der Fehlerdichte. Da wir weiterhin den allgemeinen Fall betrachten wollen, werde ich - wenn nötig - die Bedingungen (2.19) oder (2.20) fordern. Bedingung (2.19) folgt aus der Eigenschaft $\frac{1}{2\pi}\sup\limits_{f\in\mathcal{F}} \int |\psi_f(t)|^2 \left|L_n(t)-1\right|^2 dt \overset{n\to\infty}{\longrightarrow} 0$, wie im Beweis dargelegt. Diese Konsistenzbedingung hängt also nicht von der Wahl der Fehlerdichte ab, sondern stellt eine notwendige Eigenschaft der Kernfunktion dar. An (2.19) soll daher nicht gerüttelt werden. Um einen zentralen Satz über Dekonvolutionskernschätzer beweisen zu können, benötigt man folgendes Lemma

Lemma 2.3 *Es gilt*

$$\frac{1}{2\pi}\sup_{f\in\mathcal{F}}\int_{-R}^{R}|\psi_f(t)|^2\left|\frac{\psi_g(t)}{\xi(t)}-1\right|^2 dt \overset{R\to\infty}{\longrightarrow} \frac{1}{2\pi}\sup_{f\in\mathcal{F}}\int|\psi_f(t)|^2\left|\frac{\psi_g(t)}{\xi(t)}-1\right|^2 dt.$$

Dabei ist die Aussage als Konvergenz zu verstehen, wenn $\frac{1}{2\pi}\sup_{f\in\mathcal{F}}\int|\psi_f(t)|^2\left|\frac{\psi_g(t)}{\xi(t)}-1\right|^2 dt < +\infty$ *ist, und anderenfalls als bestimmte Divergenz gegen* $+\infty$. *Wir lassen also wie bei der Definition des Abstandes* $d_{\mathcal{F}}$ *auch* $+\infty$ *als möglichen Wert zu.*

Beweis: Es wird eine Fallunterscheidung benötigt:

<u>1. Fall</u>: $\frac{1}{2\pi}\sup_{f\in\mathcal{F}}\int|\psi_f(t)|^2\left|\frac{\psi_g(t)}{\xi(t)}-1\right|^2 dt < +\infty$

$\sup_{f\in\mathcal{F}}\int_{-R}^{R}|\psi_f(t)|^2\left|\frac{\psi_g(t)}{\xi(t)}-1\right|^2 dt$ ist bzgl. R monoton steigend und durch den proponierten Grenzwert $\frac{1}{2\pi}\sup_{f\in\mathcal{F}}\int|\psi_f(t)|^2\left|\frac{\psi_g(t)}{\xi(t)}-1\right|^2 dt$ nach oben beschränkt. Andererseits muss $\frac{1}{2\pi}\sup_{f\in\mathcal{F}}\int|\psi_f(t)|^2\left|\frac{\psi_g(t)}{\xi(t)}-1\right|^2 dt$ kleinste obere Schranke sein, da es für jedes $\epsilon > 0$ ein $\tilde{f}(\epsilon)\in\mathcal{F}$ gibt, so dass

$$\begin{aligned}
\sup_{f\in\mathcal{F}}\int|\psi_f(t)|^2\left|\frac{\psi_g(t)}{\xi(t)}-1\right|^2 dt \quad &\le \int|\psi_{\tilde{f}}(t)|^2\left|\frac{\psi_g(t)}{\xi(t)}-1\right|^2 dt + \epsilon/2\\
(\exists R(\tilde{f}(\epsilon),\epsilon)>0 :) \quad &\le \int_{-R(\tilde{f}(\epsilon),\epsilon)}^{R(\tilde{f}(\epsilon),\epsilon)}|\psi_{\tilde{f}}(t)|^2\left|\frac{\psi_g(t)}{\psi_{\xi(t)}}-1\right|^2 dt + \epsilon/2 + \epsilon/2\\
&\le \sup_{f\in\mathcal{F}}\int_{-R(\tilde{f}(\epsilon),\epsilon)}^{R(\tilde{f}(\epsilon),\epsilon)}|\psi_f(t)|^2\left|\frac{\psi_g(t)}{\xi(t)}-1\right|^2 dt + \epsilon
\end{aligned}$$

<u>2. Fall</u>: $\frac{1}{2\pi}\sup_{f\in\mathcal{F}}\int|\psi_f(t)|^2\left|\frac{\psi_g(t)}{\xi(t)}-1\right|^2 dt = \infty$

Es existiert also eine Dichtenfolge $(f_n)_n$, so dass $\frac{1}{2\pi}\int|\psi_{f_n}(t)|^2\left|\frac{\psi_g(t)}{\xi(t)}-1\right|^2 dt$

für $n \to \infty$ gegen unendlich divergiert. Für jedes f_n lässt sich dann eine positive Zahl R_n finden, so dass auch $\frac{1}{2\pi}\int\limits_{-R_n}^{R_n} |\psi_{f_n}(t)|^2 \left|\frac{\psi_g(t)}{\xi(t)} - 1\right|^2 dt$ und folglich $\frac{1}{2\pi}\sup\limits_{f\in\mathcal{F}}\int\limits_{-R_n}^{R_n} |\psi_f(t)|^2 \left|\frac{\psi_g(t)}{\xi(t)} - 1\right|^2 dt$ für $n \to \infty$ gegen unendlich streben. Wegen der steigenden Monotonie bzgl. R folgt daraus

$$\sup_{f\in\mathcal{F}}\int\limits_{-R}^{R} |\psi_f(t)|^2 \left|\frac{\psi_g(t)}{\xi(t)} - 1\right|^2 dt \overset{R\to+\infty}{\longrightarrow} \infty.$$

■

Alle Hoffnungen, durch geschickte Wahl von ξ oder der Kernfunktion eine Verbesserung der asymptotischen Qualität der Schätzersequenz herbeizuführen, begräbt folgender Satz

Satz 2.2 *Sei der Schätzer $\hat{f}_n$ durch (2.17) definiert und für ξ gelte $\inf\limits_{|t|\leq T}|\xi(t)| > 0$ für jedes $T > 0$ und $L_n/\xi \in L_2(\mathbb{R})$, $\forall n \in \mathbb{N}$. Die Konsistenzbedingung (2.19) gelte. Dann gilt*
(a) Die Schätzersequenz $(\sup\limits_{f\in\mathcal{F}} E_{f,g}\, \|\hat{f}_n - f\|^2_{L_2(\mathbb{R})})_{n\in\mathbb{N}}$ besitzt keinen Häufungspunkt, der kleiner als $\frac{1}{2\pi}\sup\limits_{f\in\mathcal{F}}\int |\psi_f(t)|^2 \left|\frac{\psi_g(t)}{\xi(t)} - 1\right|^2 dt$ ist.
(b) Wenn zusätzlich die Bedingungen (2.18), (2.20) sowie $|L_n(t)| \leq 1$, $\forall t \in \mathbb{R}$, $\forall n \in \mathbb{N}$ gelten, dann konvergiert (bzw. divergiert) $(\sup\limits_{f\in\mathcal{F}} E_{f,g}\, \|\hat{f}_n - f\|^2_{L_2(\mathbb{R})})_{n\in\mathbb{N}}$ gegen $\frac{1}{2\pi}\sup\limits_{f\in\mathcal{F}}\int |\psi_f(t)|^2 \left|\frac{\psi_g(t)}{\xi(t)} - 1\right|^2 dt$.

Anmerkungen:

- Wenn es sich bei ξ um die Fouriertransformierte einer $\mathcal{G}$-Dichte $\tilde{g}$ handelt, entspricht $\frac{1}{2\pi}\sup\limits_{f\in\mathcal{F}}\int |\psi_f(t)|^2 \left|\frac{\psi_g(t)}{\xi(t)} - 1\right|^2 dt$ dem aus dem vorigen Abschnitt bekannten Abstand $d_{\mathcal{F}}(g, \tilde{g})$.

- Die nur für Teil (b) benötigte hinreichende Forderung $|L_n(t)| \leq 1$, $\forall t \in \mathbb{R}$, $\forall n \in \mathbb{N}$ erscheint realistisch und ist erfüllt, wenn $L_n = \psi_K(t/\omega_n)$ mit einer Dichte K und einer beliebigen Bandbreitenfolge $(\omega_n)_n$ konstruiert wird.

Beweis: ad (a): $\sup\limits_{f\in\mathcal{F}} E_{f,g} \|\hat{f}_n - f\|^2_{L_2(\mathbb{R})}$ kann nach Lemma 2.1 durch $\frac{1}{2\pi}\sup\limits_{f\in\mathcal{F}} \int |\psi_f(t)|^2 \left|L_n(t)\frac{\psi_g(t)}{\xi(t)} - 1\right|^2 dt$ nach unten abgeschätzt werden. Betrachten wir für diese untere Schranke die folgende Ungleichungskette

$$\frac{1}{2\pi}\sup_{f\in\mathcal{F}} \int |\psi_f(t)|^2 \left|L_n(t)\frac{\psi_g(t)}{\xi(t)} - 1\right|^2 dt$$

$$\geq \frac{1}{2\pi}\sup_{f\in\mathcal{F}} \int_{-R}^{R} |\psi_f(t)|^2 \left|L_n(t)\frac{\psi_g(t)}{\xi(t)} - 1\right|^2 dt$$

$$= \frac{1}{2\pi}\sup_{f\in\mathcal{F}} \int_{-R}^{R} |\psi_f(t)|^2 \left|L_n(t)\frac{\psi_g(t)}{\xi(t)} - \frac{\psi_g(t)}{\xi(t)} + \frac{\psi_g(t)}{\xi(t)} - 1\right|^2 dt$$

$$\begin{aligned}
= \frac{1}{2\pi}\sup_{f\in\mathcal{F}} \{ &\int_{-R}^{R} |\psi_f(t)|^2 \left|L_n(t)\frac{\psi_g(t)}{\xi(t)} - \frac{\psi_g(t)}{\xi(t)}\right|^2 dt \\
&+ 2\cdot Re \int_{-R}^{R} |\psi_f(t)|^2 \left(L_n(t)\frac{\psi_g(t)}{\xi(t)} - \frac{\psi_g(t)}{\xi(t)}\right) \overline{\left(\frac{\psi_g(t)}{\xi(t)} - 1\right)} dt \\
&+ \int_{-R}^{R} \left|\frac{\psi_g(t)}{\xi(t)} - 1\right|^2 |\psi_f(t)|^2 dt\}
\end{aligned}$$

(Cauchy-Schwarz'sche Ungleichung)

$$\begin{aligned}
\geq \frac{1}{2\pi}\sup_{f\in\mathcal{F}} \{ &\int_{-R}^{R} |\psi_f(t)|^2 \left|L_n(t)\frac{\psi_g(t)}{\xi(t)} - \frac{\psi_g(t)}{\xi(t)}\right|^2 dt \\
&-2\cdot \left(\int_{-R}^{R} |\psi_f(t)|^2 \left|L_n(t)\frac{\psi_g(t)}{\xi(t)} - \frac{\psi_g(t)}{\xi(t)}\right|^2 dt\right)^{1/2} \cdot \left(\int_{-R}^{R} \left|\frac{\psi_g(t)}{\xi(t)} - 1\right|^2 |\psi_f(t)|^2 dt\right)^{1/2} \\
&+ \int_{-R}^{R} \left|\frac{\psi_g(t)}{\xi(t)} - 1\right|^2 |\psi_f(t)|^2 dt\}
\end{aligned}$$

$$\begin{aligned}
= \frac{1}{2\pi}\sup_{f\in\mathcal{F}} \{ &\left(\int_{-R}^{R} |\psi_f(t)|^2 \left|L_n(t)\frac{\psi_g(t)}{\xi(t)} - \frac{\psi_g(t)}{\xi(t)}\right|^2 dt\right)^{1/2} \\
&- \left(\int_{-R}^{R} |\psi_f(t)|^2 \left|\frac{\psi_g(t)}{\xi(t)} - 1\right|^2 dt\right)^{1/2} \}^2
\end{aligned}$$

$$\geq \frac{1}{2\pi}\Bigg\{\left(\sup_{f\in\mathcal{F}}\int_{-R}^{R}|\psi_f(t)|^2\left|L_n(t)\frac{\psi_g(t)}{\xi(t)}-\frac{\psi_g(t)}{\xi(t)}\right|^2 dt\right)^{1/2}$$
$$-\left(\sup_{f\in\mathcal{F}}\int_{-R}^{R}|\psi_f(t)|^2\left|\frac{\psi_g(t)}{\xi(t)}-1\right|^2 dt\right)^{1/2}\Bigg\}^2$$

Mit (2.19) folgt, dass $(\sup\limits_{f\in\mathcal{F}} E_{f,g}\,\|\hat{f}_n - f\|^2_{L_2(\mathbb{R})})_{n\in\mathbb{N}}$ keinen Häufungspunkt besitzen kann, der kleiner als $\frac{1}{2\pi}\sup\limits_{f\in\mathcal{F}}\int_{-R}^{R}|\psi_f(t)|^2\left|\frac{\psi_g(t)}{\xi(t)}-1\right|^2 dt$ ist für jedes $R>0$. Wegen Lemma 2.3 kann es auch keinen Häufungspunkt geben, der kleiner als $\frac{1}{2\pi}\sup\limits_{f\in\mathcal{F}}\int|\psi_f(t)|^2\left|\frac{\psi_g(t)}{\xi(t)}-1\right|^2 dt$. Damit ist (a) bewiesen.

ad (b): Hier können wir uns auf den Fall $\frac{1}{2\pi}\sup\limits_{f\in\mathcal{F}}\int|\psi_f(t)|^2\left|\frac{\psi_g(t)}{\xi(t)}-1\right|^2 dt<\infty$ beschränken, im anderen Fall ist bereits durch (a) alles bewiesen, denn eine nach unten beschränkte reelle Zahlenfolge ohne endliche Häufungspunkte divergiert gegen $+\infty$ (Satz von Bolzano-Weierstraß). Das Risiko $(\sup\limits_{f\in\mathcal{F}} E_{f,g}\,\|\hat{f}_n - f\|^2_{L_2(\mathbb{R})})_{n\in\mathbb{N}}$ ist gemäß Lemma 2.1 durch $\frac{1}{2\pi}\sup\limits_{f\in\mathcal{F}}\int\left|\frac{L_n(t)}{\xi(t)}\right|^2\cdot\underbrace{\frac{1-|\psi_{f*g}(t)|^2}{n}}_{\leq 1/n}dt+$ $\frac{1}{2\pi}\sup\limits_{f\in\mathcal{F}}\int\left|L_n(t)\frac{\psi_f(t)\psi_g(t)}{\xi(t)}-\psi_f(t)\right|^2 dt$ nach oben abschätzbar. Aus Bedingung (2.20), die ja nun zusätzlich zur Verfügung steht, folgt, dass der erste Summand in obigem Term gegen 0 konvergiert. Der zweite Summand wird im Beweis von Teil (a) nach unten abgeschätzt und lässt sich auch nach oben begrenzen

$$
\begin{aligned}
&\frac{1}{2\pi}\sup_{f\in\mathcal{F}}\int |\psi_f(t)|^2 \left| L_n(t)\frac{\psi_g(t)}{\xi(t)} - 1\right|^2 dt \\
&= \frac{1}{2\pi}\sup_{f\in\mathcal{F}}\int |\psi_f(t)|^2 \left| L_n(t)\frac{\psi_g(t)}{\xi(t)} - L_n(t) + L_n(t) - 1\right|^2 dt \\
&= \frac{1}{2\pi}\sup_{f\in\mathcal{F}}\int |\psi_f(t)|^2 \Big(\left| L_n(t)\frac{\psi_g(t)}{\xi(t)} - L_n(t)\right|^2 + 2\,Re\Big(\Big(L_n(t)\frac{\psi_g(t)}{\xi(t)} - L_n(t)\Big) \\
&\qquad \cdot \big(L_n(t)-1\big)\Big) + |L_n(t)-1|^2\Big) dt \\
&\leq \frac{1}{2\pi}\sup_{f\in\mathcal{F}}\int |\psi_f(t)|^2 \Big(\left| L_n(t)\frac{\psi_g(t)}{\xi(t)} - L_n(t)\right|^2 + 2\left| L_n(t)\frac{\psi_g(t)}{\xi(t)} - L_n(t)\right| \\
&\qquad \cdot |L_n(t)-1| + |L_n(t)-1|^2\Big) dt
\end{aligned}
$$

(Cauchy-Schwarz'sche Ungleichung)

$$
\begin{aligned}
&\leq \frac{1}{2\pi}\sup_{f\in\mathcal{F}}\Big\{ \int |\psi_f(t)|^2 \left| L_n(t)\frac{\psi_g(t)}{\xi(t)} - L_n(t)\right|^2 dt \\
&\quad + 2\Big(\int |\psi_f(t)|^2 \left| L_n(t)\frac{\psi_g(t)}{\xi(t)} - L_n(t)\right|^2 dt\Big)^{1/2} \Big(\int |\psi_f(t)|^2 |L_n(t)-1|^2 dt\Big)^{1/2} \\
&\quad + \int |\psi_f(t)|^2 |L_n(t)-1|^2 dt \Big\} \\
&= \frac{1}{2\pi}\sup_{f\in\mathcal{F}}\Big\{\Big(\int |\psi_f(t)|^2 \left| L_n(t)\frac{\psi_g(t)}{\xi(t)} - L_n(t)\right|^2 dt\Big)^{1/2} \\
&\qquad + \Big(\int |\psi_f(t)|^2 |L_n(t)-1|^2 dt\Big)^{1/2}\Big\}^2 \\
&\leq \frac{1}{2\pi}\Big\{\sup_{f\in\mathcal{F}}\Big(\int |\psi_f(t)|^2 \left| L_n(t)\frac{\psi_g(t)}{\xi(t)} - L_n(t)\right|^2 dt\Big)^{1/2} \\
&\qquad + \sup_{f\in\mathcal{F}}\Big(\int |\psi_f(t)|^2 |L_n(t)-1|^2 dt\Big)^{1/2}\Big\}^2
\end{aligned}
$$

(Wegen $|L_n(t)| \leq 1$ folgt weiter)

$$\leq \frac{1}{2\pi}\left\{\sup_{f\in\mathcal{F}}\left(\int|\psi_f(t)|^2\left|\frac{\psi_g(t)}{\xi(t)}-1\right|^2 dt\right)^{1/2} + \sup_{f\in\mathcal{F}}\left(\int|\psi_f(t)|^2|L_n(t)-1|^2 dt\right)^{1/2}\right\}^2$$

Aufgrund (2.18) konvergiert

$$\sup_{f\in\mathcal{F}}\left(\int|\psi_f(t)|^2|L_n(t)-1|^2 dt\right)^{1/2}$$

$$= \left(\sup_{f\in\mathcal{F}}\int|\psi_f(t)|^2|L_n(t)-1|^2 dt\right)^{1/2} \stackrel{n\to\infty}{\longrightarrow} 0.$$

Daraus folgt die Beschränktheit der Folge $(\sup_{f\in\mathcal{F}} E_{f,g}\,\|\hat{f}_n - f\|^2_{L_2(\mathbb{R})})_{n\in\mathbb{N}}$ sowie die Tatsache, dass diese Folge keine Häufungspunkte besitzen kann, die größer als $\sup_{f\in\mathcal{F}}\int|\psi_f(t)|^2\left|\frac{\psi_g(t)}{\xi(t)}-1\right|^2 dt$ sind. Nach (a) kann die Folge aber auch keine Häufungspunkte besitzen, die kleiner als dieser Ausdruck sind. Es existiert also genau ein Häufungspunkt der beschränkten reellen Zahlenfolge, und daher konvergiert nach Erkenntnissen der elementaren Analysis (Bolzano-Weierstraß) die Folge gegen jenen Term.
■

Der in *Meister* (2001) für Fehlerdichten außerhalb $L_2(\mathbb{R})$ verwendete Dekonvolutionskernschätzer mit dem Kern $K(x) = \frac{2}{\pi x^2}\left(\cos(\frac{1}{2}x) - \cos(x)\right)$ erfüllt beziehungsweise erzielt bei geeigneter Wahl der Bandbreitenfolge in Abstimmung auf die Fehlerdichte Konsistenz und strebt mit Lemma 2.2 und Satz 2.2 gegen $\sup_{f\in\mathcal{F}}\int|\psi_f(t)|^2\left|\frac{\psi_g(t)}{\xi(t)}-1\right|^2 dt$.
Ebenso sagt der Satz aus, dass die größere Freiheit des Modells, dass ξ nicht Fouriertransformierte einer Dichte aus $\mathcal{G}$ sein muss, keinen Vorteil hinsichtlich der asymptotischen Verzerrung beschert, allerdings die Struktur der asymptotischen Verzerrung auch beibehält.

2.4 Wahl der Fehlerdichte

Um eine Dekonvolutionsschätzung durchführen zu können, muss eine Dichte aus $\mathcal{G}$ ausgewählt werden. Wenn $\mathcal{G}$ mehrelementig ist, kann eine Missspezifizierung der Fehlerdichte nicht ausgeschlossen werden. Eine a-priori-Gewichtung der Fehlerdichten existiert nicht, wie im ersten Kapitel postuliert. Es stellt sich nun die Frage, wie $\tilde{g} \in \mathcal{G}$ zu wählen ist, um die asymptotische Verzerrung des Supremum des MISE, die nach den Sätzen 2.1 und 2.2 mit $d_{\mathcal{F}}(g, \tilde{g})$ identisch ist, minimal zu halten.
(2.15) könnte hier helfen. Bei jeweils zwei konkurrierenden Fehlerdichten ist diejenige die bessere Wahl, deren Fouriertransformierte betragsmäßig punktweise überall größer als die der anderen Dichte ist. Leider sind zwei Fehlerdichten durch diese Relation im Allgemeinen nicht total geordnet, mehrere oder unendlich viele Fehlerdichten erst recht nicht. Es lassen sich also längst nicht alle Dichten auf diese Weise miteinander vergleichen.
Betrachten wir nun ein sehr wichtiges Beispiel:
Nehmen wir an, $\mathcal{G}$ sei zweielemtig und bestehe aus der Laplace-Dichte g_L mit $g_L(x) = \frac{1}{2}\exp(-|x|)$ und $\psi_{g_L}(t) = \frac{1}{1+t^2}$ und der Standardnormalverteilungsdichte g_N mit $g_N(x) = \frac{1}{\sqrt{2\pi}}\exp(-x^2/2)$ und $\psi_{g_N}(t) = \exp(-x^2/2)$. Dies sind die Standardbeispiele für glatte bzw. superglatte Fehlerdichten. Von der Menge $\mathcal{F}$ wird nur gefordert, dass sie die Laplace-Dichte g_L enthalte und dass alle Elemente aus $\mathcal{F}$ in ihrer $L_2(\mathbb{R})$ -Norm gleichgradig beschränkt sind. Diese Forderung ist bei geeigneter Wahl der Parameter in den Modellräumen, wie sie zum Beispiel in *Neumann* (1997) auftauchen, erfüllt.

- Nehmen wir an, die Fehlerdichte sei in Wirklichkeit g_L, doch zur Dekonvolutionsschätzung wird g_N verwendet. Für die asymptotische Verzerrung ergibt sich

$$
\begin{aligned}
d_{\mathcal{F}}(g_L, g_N) &= \tfrac{1}{2\pi} \sup_{f\in\mathcal{F}} \int \left| \tfrac{\psi_{g_L}(t)}{\psi_{g_N}(t)} - 1 \right|^2 |\psi_f(t)|^2 \, dt \\
&\geq \tfrac{1}{2\pi} \int \left| \tfrac{(1+t^2)^{-1}}{\exp(-t^2/2)} - 1 \right|^2 |\psi_{g_L}(t)|^2 \, dt \\
&= \tfrac{1}{2\pi} \int \left| \tfrac{\exp(t^2/2)}{1+t^2} - 1 \right|^2 \left| \tfrac{1}{1+t^2} \right|^2 dt \\
&= \tfrac{1}{2\pi} \int \left| \tfrac{\exp(t^2/2)-1-t^2}{(1+t^2)^2} \right|^2 dt
\end{aligned}
$$

Der Integrand besteht aus einem Bruch, der im Zähler mit exponentieller Geschwindigkeit anwächst und im Nenner in der achten Potenz. Daher divergiert der Integrand für $t \to +\infty$ und für $t \to -\infty$ bestimmt gegen $+\infty$. Daher existiert das Integral als reelle Zahl nicht und es gilt

$$d_{\mathcal{F}}(g, \tilde{g}) = +\infty.$$

$\Rightarrow$ **Katastrophe:** Das Supremum des MISE des Dekonvolutionskernschätzer ist nicht nur keine Nullfolge, sondern wächst sogar über alle positiven Grenzen. Selbst ein Schätzer, der auf Ignorierung der Verrauschung der Beobachtungen basiert und in Folge dessen nicht konsistent ist wie etwa der Kernschätzer der fehlerbehafteten direkt beobachteten Dichte, liefert hier ein geringeres asymptotisches Risiko als der Dekonvolutionsschätzer mit missspezifizierter Fehlerdichte. Der Widersinn besteht also darin, dass das Risiko des Schätzers im Trend um so größer wird, je mehr Beobachtungen verwendet werden. Bei nichtparametrischer Dichteschätzung wird stets mit sehr großen Stichprobenumfängen gearbeitet, daher kann diese bestimmte Divergenz des MISE bei Anwendungen eine totale Fehlschätzung verursachen.

- Gehen wir jetzt davon aus, dass umgekehrt g_N die tatsächliche Dichte sei und g_L missspezifiziert werde. In diesem Falle gilt für die asymptotische Verzerrung

$$\begin{aligned} d_{\mathcal{F}}(g_N, g_L) &= \tfrac{1}{2\pi} \sup_{f \in \mathcal{F}} \int \left| \tfrac{\psi_{g_N}(t)}{\psi_{g_L}(t)} - 1 \right|^2 |\psi_f(t)|^2 \, dt \\ &= \tfrac{1}{2\pi} \sup_{f \in \mathcal{F}} \int \underbrace{\left| \frac{t^2+1}{\exp(t^2/2)} - 1 \right|^2}_{\substack{\to 1 \text{ für } |t| \to \infty \text{ und stetig} \\ \Rightarrow \text{ beschränkt bzgl. einem } S < \infty}} |\psi_f(t)|^2 \, dt \end{aligned}$$

$$\leq \frac{S}{2\pi} \sup_{f \in \mathcal{F}} \int |\psi_f(t)|^2 \, dt$$

$$= S \sup_{f \in \mathcal{F}} \|f\|^2_{L_2(\mathbb{R})} \, dt$$

$$< +\infty.$$

$\Rightarrow$ Das Supremum des MISE des Dekonvolutionskernschätzers ist zwar keine Nullfolge (Schätzersequenz nicht konsistent), aber besitzt wenigstens eine endliche obere Schranke.

Die Laplace-Dekonvolution mit g_L als Fehlerdichte ist hier also der Gauß-Dekonvolution mit g_N als Fehlerdichte unbedingt vorzuziehen, wenn beide Fehlerdichten a-priori gleichberechtigt sind. Eine Missspezifikation der Fehlerdichte besitzt bei der Laplace-Dekonvolution weitaus weniger gravierende Auswirkungen, wie eben gesehen. Der Robustheitsgrad der Laplace-Dekonvolution bezüglich Missspezifikation der Fehlerdichte ist also größer. Außerdem ist noch zu bedenken, dass im Falle der richtigen Spezifikation der Fehlerdichte zwar beide Arten von Dekonvolutionsschätzerfolgen Konsistenz aufweisen, die Laplace-Dekonvolution allerdings mit schnelleren (algebraischen) Konvergenzraten als die Gauß-Dekonvolution mit logarithmischen Raten (siehe z.B. *Fan* (1993), *Neumann* (1997)). Dieser Aspekt spricht also auch für die Entscheidung für die Laplace-Dekonvolution.

Ich führe noch ein weiteres praxisnahes Beispiel an. Es sei bekannt, dass die Fehlerdichte g eine Normalverteilungsdichte ist, jedoch können ihr Erwartungswert μ und ihre Varianz σ^2 nicht exakt spezifiziert werden. An deren Stelle werden falsche Werte $\tilde{\mu}$ und $\tilde{\sigma^2}$ verwendet. Was bedeutet das für $d_{\mathcal{F}}$? Die Fouriertransformierten der Normalverteilungsdichten ergeben sich als $\psi_g(t) = \exp(it\mu - (1/2) \cdot \sigma^2 t^2)$ und $\psi_{\tilde{g}}(t) = \exp(it\tilde{\mu} - (1/2) \cdot \tilde{\sigma^2} t^2)$. Unterscheiden wir zwei Fälle:

(a) Die Varianz wird richtig bestimmt ($\tilde{\sigma^2} = \sigma^2$), nur der Erwartungswert ist missspezifiziert. Der Abstand (2.7) beträgt

$$
\begin{aligned}
d_{\mathcal{F}}(g,\tilde{g}) &= \frac{1}{2\pi}\sup_{f\in\mathcal{F}}\int\left|\frac{\exp(i\mu t-(1/2)\cdot\sigma^2t^2)}{\exp(it\tilde{\mu}-(1/2)\cdot\sigma^2t^2)}-1\right|^2|\psi_f(t)|^2dt \\
&= \frac{1}{2\pi}\sup_{f\in\mathcal{F}}\int\underbrace{|\exp(i(\mu-\tilde{\mu})t)-1|^2}_{\leq 4}|\psi_f(t)|^2dt \\
&= 4\cdot\sup_{f\in\mathcal{F}}\|f\|^2_{L_2(\mathbb{R})}.
\end{aligned}
$$

Wenn $\mathcal{F}$ bzgl. der $L_2(\mathbb{R})$ -Norm gleichgradig beschränkt ist (und das sollte man voraussetzen, wenn man f in eben dieser Norm mit gleichmäßiger Konsistenz schätzen will), tritt in diesem Fall also eine noch harmlose Situation auf, der MISE ist nach oben beschränkt. Zudem ist anzumerken, dass $4|\psi_f|^2$ eine integrable Majorante für den Integranden in obigem Integral darstellt, wenn $\{|\psi_f| \mid f\in\mathcal{F}\}$ gleichgradig beschränkt ist, so dass nach dem Satz von der majorisierten Konvergenz $d_{\mathcal{F}}(g,\tilde{g})$ zumindest gegen Null konvergiert, wenn $\tilde{\mu}$ gegen μ strebt. Ist die Abweichung $\tilde{\mu}$ von μ also klein, konvergiert auch der MISE gegen eine kleine positive Zahl.

(b) Die Varianz wird missspezifiziert, (2.7) berechnet man

$$
d_{\mathcal{F}}(g,\tilde{g}) = \frac{1}{2\pi}\sup_{f\in\mathcal{F}}\int\left|\frac{\exp(i\mu t-(1/2)\cdot\sigma^2t^2)}{\exp(it\tilde{\mu}-(1/2)\cdot\tilde{\sigma^2}t^2)}-1\right|^2|\psi_f(t)|^2dt
$$

$$
= \frac{1}{2\pi}\sup_{f\in\mathcal{F}}\int\left|\exp(i(\mu-\tilde{\mu})t)\cdot\exp((1/2)\tilde{\sigma^2}-(1/2)\sigma^2)-1\right|^2|\psi_f(t)|^2dt.
$$

Ist $\tilde{\sigma^2} > \sigma^2$, so divergiert der erste Faktor des Integranden mit exponentieller Geschwindigkeit gegen $+\infty$. Es genügt schon, dass eine Dichte, die wie eine Potenzfunktion abklingt, als zu schätzende Dichte in Frage kommt, und schon divergiert der MISE gegen $d_{\mathcal{F}}(g,\tilde{g}) = +\infty$ und auch hier tritt eine Katastrophe ein. Wenn dagegen $\tilde{\sigma^2} < \sigma^2$ gilt, so ist der erste Teil des Integranden nach oben beschränkt, und eine Explosion des MISE kann somit verhindert werden.

Hierbei zeigt sich auch, dass der Abstand $d_{\mathcal{F}}$ nicht $\|\cdot\|_{L_1(\mathbb{R})}$-stetig ist, d.h. $\|g_n - g\|_{L_1(\mathbb{R})} \stackrel{n\to\infty}{\longrightarrow} 0$ für $g, g_n \in \mathcal{G}$ für alle $n \in \mathbb{N}$ impliziert im Allgemeinen nicht $d_{\mathcal{F}}(g, g_n) \stackrel{n\to\infty}{\longrightarrow} 0$. Denn eine beliebige reelle Folge $\sigma_n^2 \downarrow \sigma^2$ bewirkt zwar, dass $\mathcal{N}(\mu, \sigma_n^2)$ im $L_1(\mathbb{R})$-Sinne gegen $\mathcal{N}(\mu, \sigma^2)$ konvergiert, $d_{\mathcal{F}}(\mathcal{N}(\mu, \sigma^2), \mathcal{N}(\mu, \sigma_n^2))$ ist nach Punkt (b) der vorangegangenen Überlegung unendlich unter nicht unrealistischen Voraussetzungen an $\mathcal{F}$. ($\mathcal{N}(\mu, \sigma^2)$ bezeichne die Normalverteilungsdichte mit Erwartungswert μ und Varianz σ^2.)
Eine beliebig kleine Abweichung der verwendeten Fehlerdichte von der tatsächlichen im $L_1(\mathbb{R})$-Sinne kann also unter Umständen eine gravierende Verzerrung des asymptotischen Verhaltens der Schätzersequenz bewirken bzw. die Konsistenz zerstören. Hier wird die hohe Sensitivität des Dekonvolutionsschätzungsverfahrens bezüglich der Wahl der richtigen Fehlerdichte deutlich.

Als Regeln für Praktiker und Anwender der Dekonvolutionsdichteschätzung kann man folglich festhalten:

- **Je „glatter“ eine Fehlerdichte ist, d.h. je schneller ihre Fouriertransformierte abklingt, um so größere Gefahren birgt sie, wenn sie möglicherweise falsch gewählt wird.**

- **Bei normalverteilten Fehlern sollte die Varianz lieber zu klein als zu groß gewählt werden.**

Unter der Voraussetzung, dass alle Elemente aus $\mathcal{F}$ bezüglich ihrer $L_2(\mathbb{R})$-Norm beschränkt sind, also

$$\sup_{f\in\mathcal{F}} \|f\|^2_{L_2(\mathbb{R})} \leq C \tag{2.21}$$

gilt, existiert eine Möglichkeit, eine Unbeschränktheit des Supremum des MISE wie im ersten Fall des obigen Beispieles zu vermeiden und gleichzeitig die Konsistenz des Schätzverfahrens im Falle richtiger Spezifikation der Fehlerdichte zu wahren. Die Grundidee liefert folgendes Lemma

Lemma 2.4 *Es sei* $\tilde{f} \in L_2(\mathbb{R})$ *mit* $\|\tilde{f}\|^2_{L_2(\mathbb{R})} \geq C > 0$ *und* $f \in L_2(\mathbb{R})$ *mit* $\|f\|^2_{L_2(\mathbb{R})} \leq C$. *Sei* $\hat{f} := \frac{\sqrt{C}}{\|\tilde{f}\|_{L_2(\mathbb{R})}} \cdot \tilde{f}$. *Dann gilt*

$$\|\hat{f} - f\|^2_{L_2(\mathbb{R})} \;\leq\; \|\tilde{f} - f\|^2_{L_2(\mathbb{R})} \tag{2.22}$$

und

$$\|\hat{f} - f\|^2_{L_2(\mathbb{R})} \;\leq\; 4C \tag{2.23}$$

Beweis: ad (2.22):

$$\sqrt{C} \geq \|f\|$$

mit der Cauchy-Schwarz'schen Ungleichung gilt:

$$\Rightarrow \quad 2\sqrt{C} \geq 2\, Re\langle \tfrac{\tilde{f}}{\|\tilde{f}\|}, f\rangle$$

aus $\sqrt{C} \leq \|\tilde{f}\|$ *folgt weiter*

$$\Rightarrow \quad \sqrt{C} + \|\tilde{f}\| \geq 2\, Re\langle \tfrac{\tilde{f}}{\|\tilde{f}\|}, f\rangle$$

Erweitern mit $\sqrt{C} - \|\tilde{f}\| \leq 0$ *liefert*

$$\Rightarrow \quad C - \|\tilde{f}\|^2 \leq \tfrac{2}{\|\tilde{f}\|}(\sqrt{C} - \|\tilde{f}\|)\, Re\langle \tilde{f}, f\rangle$$

$$\Rightarrow \quad C\tfrac{\|\tilde{f}\|^2}{\|\tilde{f}\|^2} - 2\tfrac{\sqrt{C}}{\|\tilde{f}\|} Re\langle \tilde{f}, f\rangle + \|f\|^2 \leq \|\tilde{f}\|^2 - 2Re\langle \tilde{f}, f\rangle + \|f\|^2$$

$$\Rightarrow \quad \|\tfrac{\sqrt{C}}{\|\tilde{f}\|}\tilde{f} - f\|^2 \leq \|\tilde{f} - f\|^2$$

$$\Rightarrow \quad \|\hat{f} - f\|^2 \leq \|\tilde{f} - f\|^2$$

ad (2.23):

$$\begin{aligned}
&\|\hat{f} - f\|^2_{L_2(\mathbb{R})} \\
&= \|\sqrt{C}\tfrac{\tilde{f}}{\|\tilde{f}\|} - f\|^2 \\
&\textit{(Dreiecksungleichung)} \\
&\leq (\sqrt{C} + \|f\|)^2 \\
&\leq 4C
\end{aligned}$$

■

Sei $(\hat{f}_n)_n$ eine Schätzersequenz für f, deren L_2-Norm nicht gleichgradig beschränkt zu sein braucht, etwa die Dekonvolutionskernschätzer (2.16) oder (2.1) mit möglicherweise missspezifizierten Fehlerdichten. Dann definieren wir den davon abgeleiteten Schätzer

$$\tilde{f}_n := \begin{cases} \hat{f}_n & , \text{ wenn } \|\hat{f}_n\|^2 \leq C \\ \sqrt{C}\frac{\hat{f}_n}{\|\hat{f}_n\|_{L_2(\mathbb{R})}} & , \text{ sonst }. \end{cases} \tag{2.24}$$

Für den Schätzer (2.24) folgt folgender Satz

Satz 2.3 *Für die Schätzersequenz $(\tilde{f}_n)_n$ aus (2.24), abgeleitet von einem beliebigen Schätzer $(\hat{f}_n)_n$ mit jeweils endlicher, aber nicht zwingend gleichgradig beschränkter L_2-Norm, gilt für alle $f \in \mathcal{F}, g \in \mathcal{G}$, wenn $\mathcal{F}$ in der quadrierten L_2-Norm bzgl. eines $C > 0$ gleichgradig beschränkt ist,*

$$E_{f,g}\|\tilde{f}_n - f\|^2_{L_2(\mathbb{R})} \leq 4C \tag{2.25}$$

sowie

$$E_{f,g}\|\tilde{f}_n - f\|^2_{L_2(\mathbb{R})} \leq E_{f,g}\|\hat{f}_n - f\|^2_{L_2(\mathbb{R})} \tag{2.26}$$

Beweis: Der Beweis geschieht durch Konditionierung des Schätzers (2.24).
ad (2.25):

$$\begin{aligned}
&E_{f,g}\|\tilde{f}_n - f\|^2_{L_2(\mathbb{R})} \\
&= E_{f,g}\left(\|\tilde{f}_n - f\|^2_{L_2(\mathbb{R})} \quad \middle| \quad \|\hat{f}_n\|^2_{L_2(\mathbb{R})} \leq C\right) \cdot P_{f,g}(\|\hat{f}_n\|^2_{L_2(\mathbb{R})} \leq C) \\
&+ E_{f,g}\left(\|\tilde{f}_n - f\|^2_{L_2(\mathbb{R})} \quad \middle| \quad \|\hat{f}_n\|^2_{L_2(\mathbb{R})} > C\right) \cdot P_{f,g}(\|\hat{f}_n\|^2_{L_2(\mathbb{R})} > C)
\end{aligned}$$

$$
\begin{aligned}
&= E_{f,g}\left(\underbrace{\|\hat{f}_n - f\|^2_{L_2(\mathbb{R})}}_{\leq 4C} \mid \|\hat{f}_n\|^2_{L_2(\mathbb{R})} \leq C\right) \cdot P_{f,g}(\|\hat{f}_n\|^2_{L_2(\mathbb{R})} \leq C) \\
&+ E_{f,g}\left(\underbrace{\|\sqrt{C}\frac{\hat{f}_n}{\|\hat{f}_n\|_{L_2(\mathbb{R})}} - f\|^2_{L_2(\mathbb{R})}}_{\leq 4C \text{ nach } (2.23)} \mid \|\hat{f}_n\|^2_{L_2(\mathbb{R})} > C\right) \cdot P_{f,g}(\|\hat{f}_n\|^2_{L_2(\mathbb{R})} > C) \\
&\leq 4C\left(P_{f,g}(\|\hat{f}_n\|^2_{L_2(\mathbb{R})} \leq C) + P_{f,g}(\|\hat{f}_n\|^2_{L_2(\mathbb{R})} > C)\right) \\
&= 4C.
\end{aligned}
$$

ad (2.26):

$$
\begin{aligned}
&E_{f,g}\|\tilde{f}_n - f\|^2_{L_2(\mathbb{R})} \\
&= E_{f,g}\left(\|\hat{f}_n - f\|^2_{L_2(\mathbb{R})} \quad \mid \quad \|\hat{f}_n\|^2_{L_2(\mathbb{R})} \leq C\right) \cdot P_{f,g}(\|\hat{f}_n\|^2_{L_2(\mathbb{R})} \leq C) \\
&+ E_{f,g}\left(\underbrace{\|\sqrt{C}\frac{\hat{f}_n}{\|\hat{f}_n\|_{L_2(\mathbb{R})}} - f\|^2_{L_2(\mathbb{R})}}_{\leq \|\hat{f}_n - f\|^2_{L_2(\mathbb{R})} \text{ nach } (2.22)} \mid \|\hat{f}_n\|^2_{L_2(\mathbb{R})} > C\right) \cdot P_{f,g}(\|\hat{f}_n\|^2_{L_2(\mathbb{R})} > C) \\
&\leq E_{f,g}\left(\|\hat{f}_n - f\|^2_{L_2(\mathbb{R})} \mid \|\hat{f}_n\|^2_{L_2(\mathbb{R})} \leq C\right) \cdot P_{f,g}(\|\hat{f}_n\|^2_{L_2(\mathbb{R})} \leq C) \\
&+ E_{f,g}\left(\|\hat{f}_n - f\|^2_{L_2(\mathbb{R})} \mid \|\hat{f}_n\|^2_{L_2(\mathbb{R})} > C\right) \cdot P_{f,g}(\|\hat{f}_n\|^2_{L_2(\mathbb{R})} > C) \\
&\qquad\qquad = E_{f,g}\|\hat{f}_n - f\|^2_{L_2(\mathbb{R})}.
\end{aligned}
$$

■

Satz 2.3 sagt also aus, dass der Schätzer (2.24) alle Vorzüge des Schätzers $\hat{f}_n$, von dem er abgeleitet ist, übernimmt, d.h. das Risiko ist für alle $f \in \mathcal{F}$ und $g \in \mathcal{G}$ kleiner oder gleich als das des Schätzers $\hat{f}_n$, insbesondere schätzt $\tilde{f}_n$ immer dann konsistent in $L_2(\mathbb{R})$, wenn $\hat{f}_n$ konsistent schätzt. Zudem ist der

Katastrophenfall eines über alle Grenzen wachsenden L_2-Risikos bei dem Schätzer $\tilde{f}_n$ ausgeschlossen, indem ihm die Sicherheitsschranke C für die L_2-Norm des Schätzers auferlegt wurde. Immer wenn $\mathcal{F}$ in der L_2-Norm gleichgradig beschränkt ist und eine obere Schranke C bestimmt werden kann, sollte daher bei Ungewissheit bezüglich der Fehlerdichte der Schätzer (2.24) verwendet werden. Dies ist für alle „wesentlichen“ Dichteklassen $\mathcal{F}$ der Fall, wie später in Abschnitt 3.3 im Anschluss an Lemma 3.4 gezeigt wird. Der zusätzliche Aufwand bei der Berechnung von $\tilde{f}_n$ im Vergleich zu dem Ausgangsschätzer $\hat{f}_n$ ist überschaubar, denn er besteht darin, die L_2-Norm des Schätzers $\hat{f}_n$ zu berechnen und diese mit $\sqrt{C}$ zu vergleichen; C kann zu Beginn der Schätzung in einem einmaligen Vorgang allein aus Kenntnis von $\mathcal{F}$ bestimmt werden. Dennoch kann auch dieser Schätzer nicht zufriedenstellend gut schätzen, da die nichtparametrische Dichteschätzung zwecks sehr genauer Schätzungen einer Dichte gebraucht wird.

Im ersten Kapitel wird bereits der Begriff der gleichmäßig robusten $d-k$- Konsistenz für Schätzerfolgen eingeführt. d sei jetzt speziell die $L_2(\mathbb{R})$-Metrik und $k = 2$. Die Frage ist nun, unter welchen Bedingungen an $\mathcal{F}$ und $\mathcal{G}$ es eine Wahl von $\tilde{g} \in \mathcal{G}$ gibt, so dass der mittels dieser Fehlerdichte konstruierte Schätzer (2.16) gleichmäßig robust konsistent ist.
Nach Satz 2.2 ist dies äquivalent dazu, dass

$$\exists \tilde{g} \in \mathcal{G} \, : \, d_{\mathcal{F}}(g, \tilde{g}) = 0 \, , \, \forall g \in \mathcal{G} \tag{2.27}$$

gilt. Dies ist eine sehr starke Forderung an $\mathcal{F}$ und $\mathcal{G}$. Wenn $\mathcal{G}$ mehrelementig und der Abstand $d_{\mathcal{F}}$ positiv definit ist, wofür im zweiten Abschnitt dieses Kapitels hinreichende Bedingungen gefunden wurden, ist eine solche Wahl von $\tilde{g}$, die zu einer gleichmäßig robust konsistenten Schätzung führt, bereits nicht mehr möglich. Damit ist freilich noch nicht beantwortet, ob es in solchen Fällen überhaupt eine gleichmäßig robust konsistente Schätzung geben kann - möglicherweise ganz anderer Bauart. Zu diesem Zwecke sollen im nächsten Kapitel theoretische Fragestellungen nach der Existenz robust konsistenter und im übernächsten Kapitel gleichmäßig robust konsistenter Schätzer aufgegriffen werden.

Kapitel 3

Robust konsistente Schätzbarkeit

In diesem Kapitel wird die Frage aufgegriffen, unter welchen notwendigen und hinreichenden Bedingungen an die Mengen $\mathcal{F}$ und $\mathcal{G}$ robust d^k-konsistente Schätzbarkeit besteht (siehe (1.5)). d ist - wenn nichts anderes ausgesagt ist - eine beliebige Metrik und (X,d) der zugehörige metrische Raum. Um den Begriff der robusten Konsistenz überprüfen zu können, muss $\mathcal{F} \subseteq X$ gefordert werden, wie im ersten Kapitel bereits erwähnt. Dies wird zum Beispiel relevant, wenn d die Supremumnormmetrik ist; von allgemeinen Dichten, die ja Äquivalenzklassen von LB-fast überall übereinstimmenden Funktionen sind, kann keine Supremumnorm betrachtet werden. Eine Dichte hätte sonst zu sich selbst positiven Abstand. Hier muss beispielsweise noch die Forderung der Stetigkeit aller Dichten aus $\mathcal{F}$ gestellt werden, um die Inklusion von $\mathcal{F}$ in X zu garantieren.

3.1 Überlappen der Klassen $\mathcal{F}$ und $\mathcal{G}$

Zunächst definieren wir den Begriff der sich überlappenden Dichtemengen $\mathcal{F}$ und $\mathcal{G}$:
Die Mengen $\mathcal{F}$ und $\mathcal{G}$ überlappen sich genau dann, wenn gilt

$$\exists f, \tilde{f} \in \mathcal{F} \, \exists g, \tilde{g} \in \mathcal{G} \, : \, f \neq \tilde{f} \, \wedge \, f * g = \tilde{f} * \tilde{g} \tag{3.1}$$

Dafür kann man leicht äquivalente Formulierungen erkennen, so zum Beispiel:

$$(3.1) \Leftrightarrow \exists f, \tilde{f} \in \mathcal{F} \text{ mit } f \neq \tilde{f} \, : \, f * G \cap \tilde{f} * G \neq \emptyset \tag{3.2}$$

Intuitiv bedeutet das Überlappen von $\mathcal{F}$ und $\mathcal{G}$, dass es (mindestens) eine verrauschte Dichte gibt, die von zwei verschiedenen unverfälschten zu schätzenden Dichten herrühren kann. Für sich überlappende Klassen lässt sich ein zentraler Satz beweisen

Satz 3.1 *Sei (X,d) ein beliebiger metrischer Raum. Sei $\mathcal{D}$ die Menge aller Dichten (1.1). Es gelte $\mathcal{F} \subseteq \mathcal{D} \cap X$, $\mathcal{G} \subseteq \mathcal{D}$. $\mathcal{F}$ und $\mathcal{G}$ sollen sich überlappen.*
Dann existiert **keine** *robust d^k-konsistente auf n uiv verrauschten Beobachtungen $Y_1, \ldots, Y_n$ basierende Schätzersequenz $(\hat{f}_n)_{n\in\mathbb{N}} \subseteq X$ gemäß (1.5) für irgendein $k > 0$.*

Um den Satz beweisen zu können, benötigt man noch das folgende kleine technische Lemma

Lemma 3.1 *Seien $a, b \geq 0$ und $k > 0$. Dann gilt die Ungleichung*

$$a^k + b^k \geq \min\{2^{1-k}, 1\}(a+b)^k.$$

Beweis: Für $a = b = 0$ ist die Ungleichung trivialerweise erfüllt. Daher setzen wir $a^2 + b^2 \neq 0$ voraus und betrachten

$$\frac{a^k + b^k}{(a+b)^k} = \left(\frac{a}{a+b}\right)^k + \left(\frac{b}{a+b}\right)^k.$$

Nun substituieren wir $\lambda = \frac{a}{a+b}$. Dann lässt sich obiger Ausdruck darstellen als reelle Funktion

$$f(\lambda) = \lambda^k + (1-\lambda)^k \quad , \quad \lambda \in [0,1].$$

Ziel ist es jetzt, das Minimum der Funktion zu finden. Für die Randwerte gilt $f(0) = f(1) = 1$. Für $k = 1$ ist f konstant 1. Für $k \neq 1$ ist die Funktion im Inneren des Definitionsbereichs stetig differenzierbar und durch Nullsetzen der ersten Ableitung erhält man $\lambda = 1/2$ als einzigen in Frage kommenden Extremwert. Da f zudem auf ganz $[0,1]$ stetig ist, kommen nur 1 und $f(1/2) = 2^{1-k}$ als Minima in Frage, daher gilt $f(\lambda) \geq \min\{2^{1-k}, 1\}$ für jedes $\lambda \in [0,1]$. Durch Resubstitution folgt schließlich die Ungleichung

$$a^k + b^k \geq \min\{2^{1-k}, 1\}(a+b)^k.$$

■

Damit können wir jetzt zum Beweis von Satz 3.1 schreiten
Beweis: Wir nehmen an, es existiere eine robust d^k-konsistente Schätzerfolge $(\hat{f}_n)_n$ für eine beliebige Metrik und $k > 0$, d.h.

$$E_{f,g} d(\hat{f}_n(Y_1, \ldots, Y_n), f)^k \overset{n\to\infty}{\longrightarrow} 0, \forall f \in \mathcal{F}, g \in \mathcal{G}$$

Da $\mathcal{F}$ und $\mathcal{G}$ sich überlappen, existieren $f, \tilde{f} \in \mathcal{F}$ und $g, \tilde{g} \in \mathcal{G}$ mit $f \neq \tilde{f}$ und $f * g = \tilde{f} * \tilde{g}$, und da die Summe zweier konvergenter Folgen konvergiert, folgt

$$E_{f,g} d(\hat{f}_n(Y_1, \ldots, Y_n), f)^k + E_{\tilde{f},\tilde{g}} d(\hat{f}_n(Y_1, \ldots, Y_n), \tilde{f})^k \overset{n\to\infty}{\longrightarrow} 0$$

$$= E_{h=f*g=\tilde{f}*\tilde{g}} \Big(d(\hat{f}_n(Y_1, \ldots, Y_n), f)^k + d(\hat{f}_n(Y_1, \ldots, Y_n), \tilde{f})^k \Big)$$

(verwende Lemma 3.1)

$$\geq E_h \left(\min\{2^{1-k}, 1\} \cdot \Big(d(\hat{f}_n(Y_1, \ldots, Y_n), f) + d(\hat{f}_n(Y_1, \ldots, Y_n), \tilde{f}) \Big)^k \right)$$

(Dreiecksungleichung für die Metrik d)

$$\geq E_h \left(\min\{2^{1-k}, 1\} \cdot d(f, \tilde{f})^k \right)$$

$$= \min\{2^{1-k}, 1\} \cdot d(f, \tilde{f})^k$$

(da d als Metrik positiv definit und $f \neq \tilde{f}$ ist, gilt $d(f, \tilde{f}) > 0$ und der Term ist unabhängig von n.)

$$> 0.$$

Damit ist ein Widerspruch zur Annahme erreicht. ■

3.2 Beispiele sich überlappender Dichteklassen

Mittels Satz 3.1 ist also klar, dass es für sich überlappende Mengentupel $(\mathcal{F}, \mathcal{G})$ keine robust d^k-konsistente Schätzung für irgendeine Metrik d und somit natürlich auch keine gleichmäßig robust d^k-konsistente Schätzung gemäß (1.6) geben kann. Wir betrachten im Folgenden einige Beispiele für sich überlappende Mengentupel, um in diesen Fällen nicht unnötig nach nicht existierenden robust konsistenten Schätzern zu suchen:

(1) Der Schnitt der Mengen $\mathcal{F}$ und $\mathcal{G}$ enthalte mindestens zwei Elemente, d.h. $card(\mathcal{F} \cap \mathcal{G}) \geq 2$. $\mathcal{F}$ und $\mathcal{G}$ überlappen sich, denn für $f, g \in \mathcal{F} \cap \mathcal{G}$ und $f \neq g$ gilt

$$\underbrace{f}_{\in\mathcal{F}} * \underbrace{g}_{\in\mathcal{G}} = \underbrace{g}_{\in\mathcal{F}} * \underbrace{f}_{\in\mathcal{G}}.$$

Dass es hier keinen robust konsistenten Schätzer geben kann, ist auch intuitiv klar: Wenn man nur die additiv aus Fehler und unverfälschter Größe zusammengesetzten Beobachtungen zur Verfügung hat und mindestens zwei Fehlerdichten auch als zu schätzende Dichte auftreten können, kann man nicht unterscheiden, welcher Effekt der Beobachtungen aus dem Fehler und welcher Effekt aus der unverrauschten Größe stammt.
Durch dieses Beispiel wird auch klar, dass ein robust konsistenter Schätzer - wenn er denn existiert - die Deterministik der Definitionen von $\mathcal{F}$ und $\mathcal{G}$ ausnützen muss (etwa Sobolev- oder Hölderparameter). Ansonsten ist es aussichtslos, nach robust konsistenten Schätzverfahren zu suchen.

(2) Jetzt betrachten wir den Fall:

$$\exists f \in \mathcal{F}\, \exists g \in \mathcal{G}\, \exists a \in \mathbb{R}\backslash\{0\} \;:\; f(\cdot - a) \in \mathcal{F} \wedge g(\cdot + a) \in \mathcal{G} \tag{3.3}$$

Dann gilt für alle $x \in \mathbb{R}$

$$(f(\cdot - a) * g(\cdot + a))(x) \;=\; \int f(x-y-a)\, g(y+a)\, dy$$

$$= \int f(x-(y+a))\, g(y+a)\, dy \;=\; \int f(x-y)\, g(y)\, dy$$

$$= (f * g)(x)$$

durch eine einfache lineare Substitution im vorletzten Schritt.
Außerdem gilt aber auch $f \neq f(\cdot - a)$, denn:
Annahme: Es gelte $f = f(\cdot - a)$. Dann gilt für die Fouriertransformierten $\psi_f(t) = \psi_{f(\cdot - a)}(t) = e^{ita} \cdot \psi_f(t)$ für alle $t \in \mathbb{R}$. Für jedes $t \in \mathbb{R}$ gilt also $\psi_f(t) = 0$ oder $e^{ita} = 1$. Die zweite Bedingung gilt genau dann, wenn ta ganzzahliges Vielfaches von 2π bzw. da $a \neq 0$ vorausgesetzt ist, wenn t ganzzahliges Vielfaches von $\frac{2\pi}{a}$ ist. Das bedeutet, dass ψ_f auf dem halboffenen Intervall $(0, \frac{\pi}{a}]$ verschwindet. Da f eine Dichte ist, gilt $\psi_f(0) = 1$.

ψ_f besäße im Nullpunkt einen Sprung und wäre also nicht stetig. Die Fouriertransformierte einer Dichte ist jedoch immer stetig, so dass wir hier bei einem Widerspruch angekommen sind.

Damit ist gezeigt, dass sich die Mengen $\mathcal{F}$ und $\mathcal{G}$ überlappen. Zu dieser Rubrik gehören Mengen $\mathcal{F}$ und $\mathcal{G}$, die vollständige Translationsfamilien von einer Dichte enthalten, also wenn es ein $f \in \mathcal{F}$ gibt, so dass $\{f(\cdot+a) \mid a \in \mathbb{R}\}$ Teilmenge von $\mathcal{F}$ ist, <u>und</u> es ein $g \in \mathcal{G}$ gibt, so dass $\{g(\cdot + a) \mid a \in \mathbb{R}\}$ Teilmenge von $\mathcal{G}$ ist. Dazu zählen Mengen, die aufgrund Bedingungen an den Betrag der Fouriertransformierten oder Hölderbedingungen definiert sind, d.h. wenn also sowohl $\mathcal{F}$ als auch $\mathcal{G}$ Mengen der Bauart

$$\{f \in \mathcal{D}_2 \mid |\psi_f(t)| = \varphi(t)\}$$

oder

$$\{f \in \mathcal{D}_2 \mid |f^{(l)}(x) - f^{(l)}(y)| = \chi(x-y)\}$$

als Teilmenge besitzen und diese Mengen nicht leer sind für irgendwelche Funktionen $\varphi, \chi : \mathbb{R} \to \mathbb{R}$. Dies ist leicht nachprüfbar:

$$|\psi_{f(\cdot-a)}(t)| = |e^{ita}\psi_f(t)| = |\psi_f(t)|\,,\ \forall t, a \in \mathbb{R}$$

sowie

$$|f^l(x-a) - f^l(y-a)| = \chi(x-a-y+a) = \chi(x-y).$$

Bedingungen an den Betrag der Fouriertransformierten wie etwa Sobolev-Bedingungen oder obere und/oder untere Schrankenfunktionen für $|\psi_\bullet|$ oder Hölderbedingungen sowohl an die zu schätzende Dichte als auch an die Fehlerdichte tauchen z.B. bei *Fan* (1993), *Neumann* (1997) und *Hesse und Meister* (2001) auf.

Sobald die Fehlerdichte nicht eindeutig spezifizierbar ist, muss $\mathcal{G}$ als mindestens zweielementig angenommen werden. An $\mathcal{G}$ soll nun keine weitere Forderung gestellt werden. Folgendes Lemma wird in dieser sehr allgemein gehaltenen Situation alle möglichen Bestrebungen, robust konsistente Schätzerfolgen bzgl. irgendeiner Metrik d für die gängigen nichtparametrischen Definitionen für $\mathcal{F}$ zu finden, zum Scheitern verurteilen

Lemma 3.2 *$\mathcal{G}$ besitze mindestens zwei Elemente. Für $\mathcal{F}$ gelte $\mathcal{F} * \mathcal{G} =: \mathcal{H} \subseteq \mathcal{F}$ und $\exists f \in \mathcal{F} : \psi_f(t) \neq 0\,, \forall t \in \mathbb{R}$.*
Dann überlappen sich $\mathcal{F}$ und $\mathcal{G}$.

Beweis: Sei $f \in \mathcal{F}$, so dass $\psi_f(t) \neq 0$ für alle $t \in \mathbb{R}$ gilt, und seien $g, \tilde{g} \in \mathcal{G}$ mit $g \neq \tilde{g}$.

$$g * f \quad \in \mathcal{H} \subseteq \mathcal{F}$$

$$\tilde{g} * f \quad \in \mathcal{H} \subseteq \mathcal{F}$$

Da die Faltung von Dichten assoziativ und kommutativ ist (siehe z.B. *N. Schmitz* [14]), folgt

$$\Rightarrow \underbrace{(g * f)}_{\in \mathcal{F}} * \underbrace{\tilde{g}}_{\in \mathcal{G}} = \underbrace{g}_{\in \mathcal{G}} * \underbrace{(f * \tilde{g})}_{\in \mathcal{F}}.$$

Wenn jetzt noch gezeigt wird, dass $g * f \neq f * \tilde{g}$ gilt, ist das Lemma bewiesen. Zu diesem Zwecke nehmen wir an, es gelte $g * f = \tilde{g} * f$.
Dann folgt, dass für alle $t \in \mathbb{R}$ die Fouriertransformierten $\psi_g(t) \cdot \psi_f(t) = \psi_{\tilde{g}}(t) \cdot \psi_f(t)$ übereinstimmen und weiter, da ψ_f als nirgends verschwindend vorausgesetzt ist, dass $\psi_g(t) = \psi_{\tilde{g}}(t)$ für alle $t \in \mathbb{R}$ gilt. Die Fouriertransformierten von g und $\tilde{g}$ sind also identisch. Nach dem Eindeutigkeitssatz der Fouriertransformierten von W-Maßen (z.B. *Schmitz* [14]) folgt daraus, dass $g = \tilde{g}$ gilt (Widerspruch zur obigen Annahme für g und $\tilde{g}$).
■
Die Konsequenzen von Lemma 3.2 kann man in folgendem Satz ausdrücken

Satz 3.2 *Für $\mathcal{F}$ kommen folgende Modellräume in Frage:*

$$\mathcal{F}^{(1)}_{(C,\beta)} := \{f \in \mathcal{D}_2 \mid |\psi_f(t)| \leq C|t|^{-\beta}\,, \forall t \in \mathbb{R}\} \quad (\beta > 1, C > 0)$$

$$\mathcal{F}^{(2)}_{(C,\beta)} := \{f \in \mathcal{D}_2 \mid \int\limits_{\omega}^{\infty} |\psi_f(t)|^2 dt \leq C\omega^{1-2\beta}\,, \forall \omega > \omega_0\} \quad (\beta > 1, C > 0)$$

$$\mathcal{F}^{(3)}_{(C,\beta)} := \{f \in \mathcal{D}_2 \mid \int |\psi_f(t)|^2 (1 + |t|)^{2\beta} dt \leq C\} \quad (\beta \in \mathbb{R}, C > 0)$$

$$\mathcal{F}^{(4)}_{(l,\alpha,B)} := \{f \in \mathcal{D} \mid |f^{(l)}(x) - f^{(l)}(y)| \leq B|x-y|^{\alpha}\} (l \in \mathbb{N}_0, B > 0, \alpha \in [0,1))$$

Wenn $\mathcal{F}$ eine dieser vier Mengen ist und $card(\mathcal{G}) \geq 2$ gilt, überlappen sich $\mathcal{F}$ und $\mathcal{G}$.

Anmerkung: Die ersten drei Dichteklassen sind durch Fourier bzw. Sobolevbedingungen definiert. $\mathcal{F}^{(1)}_{(C,\beta)}$ und $\mathcal{F}^{(2)}_{(C,\beta)}$ kommen in *Meister* (2001) sowie in *Hesse und Meister* (2001) vor. $\mathcal{F}^{(3)}_{(C,\beta)}$ wird in *Neumann* (1997) verwendet. Die Dichteklasse $\mathcal{F}^{(4)}_{(C,\beta)}$, welche durch Hölderbedingungen definiert ist, taucht in *Fan* (1993) auf.

Um den Satz für die Dichteklasse $\mathcal{F}^{(4)}_{(l,\alpha,B)}$ beweisen zu können, benötigt man folgendes Lemma

Lemma 3.3 *Für alle* $f \in \mathcal{F}^{(4)}_{(l,\alpha,B)}$ *und alle* $x \in \mathbb{R}$ *gilt für jedes* $j \in \{0,1,\ldots,l\}$*:*

$$|f^{(j)}(x)| \leq S_j$$

mit der umgekehrt-rekursiv definierten Schranke:

$$\begin{aligned} S_l &= B \cdot 3^{l\alpha} + 2^{l+1} 3^{(1-l)l/2} \\ S_{j-1} &= 2^j \cdot 3^{(2-j)(j-1)/2} + S_j \cdot 3^{j-1} \quad , \quad \forall j = 1,\ldots,l \end{aligned}$$

Beweisen wir zunächst Lemma 3.3

Beweis: Behauptung: Für alle $j \in \{0,1,\ldots,l\}$ existiert in jedem kompakten reellen Intervall I der Länge $3^j \cdot \epsilon$ ein $x \in I$, so dass $|f^{(j)}(x)| \leq \left(\frac{2}{\epsilon}\right)^{j+1} \cdot 3^{(1-j)j/2}$ gilt. ($\epsilon > 0$ beliebig)

Der Beweis geschieht durch vollständige Induktion nach j:

Induktionsanfang ($j = 0$): Da $f^{(0)} = f$ eine Dichte ist, folgt

$$\begin{aligned} 1 &= \int f(t)\,dt \\ &\geq \int_I f(t)\,dt \\ &\geq \epsilon \cdot \inf_{t\in I} f(t). \end{aligned}$$

Also ist $\inf_{t\in I} f(t) \leq 1/\epsilon$ und es existiert ein $x \in I$ mit $|f(x)| = f(x) \leq 2/\epsilon$.

Induktionsschritt ($j < l, j \to j+1$): Sei I ein beliebiges reelles kompaktes Intervall der Länge $3^{j+1}\epsilon$. Unterteilen wir I in drei gleich große disjunkte Teilintervalle J, K, L mit einer jeweiligen Länge $3^j\epsilon$. K bezeichne das mittlere Intervall. Nach Induktionsvoraussetzung gibt es $y \in \overline{J}$ und $z \in \overline{L}$ mit $|f^{(j)}(y)| \leq \left(\frac{2}{\epsilon}\right)^{j+1} \cdot 3^{(1-j)j/2}$ und $|f^{(j)}(z)| \leq \left(\frac{2}{\epsilon}\right)^{j+1} \cdot 3^{(1-j)j/2}$. Da $y \in \overline{J}$ und $z \in \overline{L}$ gilt, folgt $|y - z| \geq 3^j \epsilon$. Mit dem Mittelwertsatz der Differentialrechnung folgt

$$\begin{aligned} 2 \cdot \left(\tfrac{2}{\epsilon}\right)^{j+1} \cdot 3^{(1-j)j/2} &\geq |f^{(j)}(y)| + |f^{(j)}(z)| \\ &\geq |f^{(j)}(y) - f^{(j)}(z)| \\ &\geq |f^{(j+1)}(\xi)| \cdot |y - z| \\ &\geq |f^{(j+1)}(\xi)| \cdot 3^j \epsilon. \end{aligned}$$

ξ liegt zwischen y und z und somit in I. Es gilt weiter

$$\begin{aligned} |f^{(j+1)}(\xi)| &\leq 2 \cdot \left(\tfrac{2}{\epsilon}\right)^{j+1} \cdot 3^{(1-j)j/2} \cdot 3^{-j} \epsilon^{-1} \\ &= \left(\tfrac{2}{\epsilon}\right)^{j+2} \cdot 3^{(-j)(j+1)/2}. \end{aligned}$$

Damit ist die Behauptung bewiesen.

Sei $x \in \mathbb{R}$ beliebig. Nach der obigen Behauptung existiert ein y mit $|y-x| \leq 3^l$ und $|f^{(l)}(y)| \leq 2^{l+1} \cdot 3^{(1-l)l/2}$, wenn man $\epsilon = 1$ setzt. Da $f \in \mathcal{F}^{(4)}_{(l,\alpha,B)}$ ist, gilt

$$\begin{aligned} |f^{(l)}(x) - f^{(l)}(y)| &\leq B\,|x - y|^\alpha \\ &\leq B\,3^{l\alpha} \end{aligned}$$

Daraus folgt

$$\begin{aligned} |f^{(l)}(x)| &\leq |f^{(l)}(y)| + |f^{(l)}(x) - f^{(l)}(y)| \\ &\leq 2^{l+1} \cdot 3^{(1-l)l/2} \;+\; B\,3^{l\alpha} \\ &= S_l. \end{aligned}$$

Noch zu zeigen ist die Rekursionsformel. Auch hier existiert für ein beliebiges $x \in \mathbb{R}$ ein y mit $|y - x| \leq 3^{j-1}$ und $|f^{(j-1)}(y)| \leq 2^j \cdot 3^{(2-j)(j-1)/2}$. Mit dem Mittelwertsatz der Differentialrechnung schließt man

$$
\begin{aligned}
|f^{(j-1)}(x) - f^{(j-1)}(y)| &= |f^{(j)}(\xi)| \cdot |x-y| \\
&\leq S_j \cdot |x-y| \\
&\leq S_j \cdot 3^{j-1}
\end{aligned}
$$

Daher gilt

$$
\begin{aligned}
|f^{(j-1)}(x)| &\leq |f^{(j-1)}(y)| + |f^{(j-1)}(x) - f^{(j-1)}(y)| \\
&\leq 2^j \cdot 3^{(2-j)(j-1)/2} + S_j \cdot 3^{j-1} \\
&= S_{j-1}.
\end{aligned}
$$

■

Sicher könnte man die hier gefundenen gleichgradigen Schranken S_j noch verschärfen, insbesondere durch Minimierung bzgl. ϵ anstelle des 1-Setzens von ϵ. Für die Thematik dieser Dissertation genügt jedoch die Existenz und die rekursive Berechenbarkeit einer gleichgradigen Schranke.

Schreiten wir nun zum Beweis von Satz 3.2

Beweis: Das Werkzeug für den Beweis des Satzes liefert Lemma 3.2. Es müssen folglich die Voraussetzungen des Lemmas für jede der vier Dichteklassen überprüft werden. Zunächst konstruieren wir eine Familie von Dichten mit nirgends verschwindender Fouriertransformierten. Sei f_{σ^2} die Dichte der Normalverteilung mit Erwartungswert 0 und Varianz σ^2 $(\mathcal{N}(0,\sigma^2))$. Jede Normalverteilungsdichte ist beschränkt und liegt somit in $L_2(\mathbb{R})$. Für die Fouriertransformierte von f_{σ^2} gilt

$$\Rightarrow \psi_{f_{\sigma^2}}(t) = \exp(-\frac{1}{2}\sigma^2 t^2)$$

(s. *Schmitz*, [14]) und daher verschwindet $\psi_{f_{\sigma^2}}(t)$ für jedes $\sigma > 0$ nirgends. Es ist jetzt das Ziel, zu zeigen, dass es für jede der vier möglichen Dichteklassen für $\mathcal{F}$ eine Wahl von $\sigma^2 > 0$ gibt, dass f_{σ^2} in $\mathcal{F}$ liegt:

$\underline{\mathcal{F}^{(1)}_{(C,\beta)}:}$

$$|\psi_{f_{\sigma^2}}(t)| \;=\; \exp(-\tfrac{1}{2}\sigma^2 t^2) \;\overset{!}{\leq}\; C|t|^{-\beta}\,,\ \forall t$$

$$\Leftrightarrow\ \underbrace{\exp(-\sigma^2 t^2/2)\,|t|^{\beta}}_{=:\Gamma(t)} \;\overset{!}{\leq}\; C\,,\ \forall t$$

(Γ ist achsensymmetrisch)

$$\Leftrightarrow\ \Gamma(t) \leq C\,,\ \forall t \geq 0$$

Nun gilt es also das Supremum von Γ in $[0,\infty)$ zu bestimmen. Die Randwerte 0 und $+\infty$ scheiden als Maxima aus, da $\Gamma(0) = \lim_{t\to\infty}\Gamma(t) = 0$ und $\Gamma(t) \geq 0$ für alle nichtnegativen t gilt. Damit ist aber auch klar, dass für die stetige Funktion Γ ein globales Maximum existieren muss. Da Γ stetig differenzierbar ist, lässt sich dieses Maximum elementar durch Nullsetzen der ersten Ableitung bestimmen:

$$\begin{aligned}
\Gamma'(t) = 0 \quad \Leftrightarrow \quad & -\sigma^2 t\exp(-\sigma^2t^2/2)t^{\beta} \;+\; \exp(-\sigma^2t^2/2)\beta t^{\beta-1} \;=\; 0 \\
& = \; t^{\beta-1}\exp(-\sigma^2t^2/2)\,(-\sigma^2t^2+\beta) \\
\Leftrightarrow \quad & \underbrace{t=0}_{\text{da } \beta>1} \;\vee\, \sigma^2t^2 = \beta
\end{aligned}$$

$t=0$ scheidet als mögliches Maximum aus (s.o.), folglich muss an der Stelle $t = \sqrt{\beta}/\sigma$ die Funktion ihr globales Maximum in $[0,\infty)$ annehmen.

$$\Gamma(\sqrt{\beta}/\sigma) \;=\; \sqrt{\beta}^{\beta}\,\sigma^{-\beta}\exp(-\beta/2)$$

Deshalb gilt

$$\begin{aligned}
\Gamma(t) \leq C\,,\ \forall t \quad \Leftrightarrow \quad & \beta^{\beta/2}\sigma^{-\beta}e^{-\beta/2} \;\leq\; C \\
\Leftrightarrow \quad & \sigma^{\beta} \;\geq\; C^{-1}e^{-\beta/2}\beta^{\beta/2} \\
\Leftrightarrow \quad & \sigma \;\geq\; C^{-1/\beta}e^{-1/2}\beta^{1/2}(<\infty)
\end{aligned}$$

Wenn also σ größer oder gleich $C^{-1/\beta}e^{-1/2}\beta^{1/2}$ gewählt wird, liegt f_{σ^2} in $\mathcal{F}^{(1)}_{(C,\beta)}$.

$\underline{\mathcal{F}^{(2)}_{(C,\beta)}:}$

Durch einfaches Quadrieren und anschließendes Integrieren erkennt man, dass $\mathcal{F}^{(1)}_{(C,\beta)} \subseteq \mathcal{F}^{(2)}_{(C^2,\beta)}$ bzw. $\mathcal{F}^{(1)}_{(\sqrt{C},\beta)} \subseteq \mathcal{F}^{(2)}_{(C,\beta)}$ gilt. Daher liegt mit obigen Ergebnissen für $\sigma \geq C^{-1/2\beta} e^{-1/2} \beta^{1/2}$ f_{σ^2} auch in $\mathcal{F}^{(2)}_{(C,\beta)}$.

$\underline{\mathcal{F}^{(3)}_{(C,\beta)}:}$

$$\int \underbrace{\exp(-\sigma^2 t^2)(1+|t|)^{2\beta}}_{0<\cdot\leq \exp(-\sigma_0^2 t^2)(1+|t|)^{2\beta}, \forall t} dt \overset{!}{\leq} C$$

wobei $\sigma > \sigma_0 > 0$ gewählt ist. Also ist $\exp(-\sigma_0^2 \cdot^2)(1+|\cdot|)^{2\beta}$ eine integrable Majorante. Außerdem gilt für alle $t \neq 0$ und somit für LB- fast alle $t \in \mathbb{R}$ $\lim_{\sigma\to\infty} \left(\exp(-\sigma^2 t^2)(1+|t|)^{2\beta}\right) = 0$, was durch elementare Analysis einzusehen ist. Mit dem Satz von der majorisierten Konvergenz folgt

$$\int \exp(-\sigma^2 t^2)(1+|t|)^{2\beta}\, dt \overset{\sigma\to\infty}{\longrightarrow} 0.$$

Daher kann man σ in Abhängigkeit von β groß genug wählen, so dass $\int \exp(-\sigma^2 t^2)(1+|t|)^{2\beta} dt \leq C$ ist und folglich $f_{\sigma^2} \in \mathcal{F}^{(3)}_{(C,\beta)}$ gilt.

$\underline{\mathcal{F}^{(4)}_{(l,\alpha,B)}:}$

f_{σ^2} ist als Normalverteilungsdichte für jedes $\sigma > 0$ reell analytisch und unendlich oft differenzierbar. Die j-te Ableitung von f_1 besitzt die Form

$$\frac{d^j}{dx^j} f_1(x) \;=\; P_j(x) \exp(-x^2/2) \tag{3.4}$$

wobei P_j ein nicht genauer spezifiziertes Polynom vom Grad höchstens j ist. Dies kann man durch vollständige Induktion zeigen:

Induktionsanfang: $j = 0$: $\quad f_1(x) = \exp(-x^2/2) \cdot \frac{1}{\sqrt{2\pi}}$

Induktionsschritt: $j \to j+1$: $\quad \frac{d}{dx}\left(\frac{d^j}{dx^j} f(x)\right)$

$$= \frac{d}{dx}\left(P_j(x)\exp(-x^2/2)\right)$$

$$= \left(P_j'(x) - xP_j(x)\right)\exp(-x^2/2)$$

wobei $P_j'(\bullet) - \bullet P_j(\bullet)$ ein Polynom vom Grade höchstens $j+1$ ist und wir $P_{j+1} = P_j'(\bullet) - \bullet P_j(\bullet)$ definieren können. Damit ist (3.4) bewiesen.

Daher ist durch elementare Analysis klar, dass $|f_1^{(j)}(t)| = |P_j(t)| \cdot \exp(-t^2/2)$ für $|t| \to \infty$ gegen 0 konvergiert. Da zudem $f_1^{(j)}$ als j-te Ableitung einer beliebig oft differenzierbaren Funktion stetig ist, lässt sich für $f_1^{(j)}$ eine obere Schranke finden, die von j abhängen kann:

$$|f_1^{(j)}(t)| \leq C_j \qquad \forall t \in \mathbb{R}$$

Mit dem Mittelwertsatz der Differentialrechnung gilt weiter, wenn $|x-y| \leq 1$ ist,

$$\begin{aligned}
|f_1^{(l)}(x) - f_1^{(l)}(y)| &\leq |f_1^{(l+1)}(\xi)| \cdot |x-y| \\
&\leq |f_1^{(l+1)}(\xi)| \cdot |x-y|^\alpha \\
&\leq C_{l+1} \cdot |x-y|^\alpha
\end{aligned}$$

Für $|x-y| > 1$, was auch $|x-y|^\alpha > 1$ impliziert, gilt mit der Dreiecksungleichung

$$\begin{aligned}
|f_1^{(l)}(x) - f_1^{(l)}(y)| &\leq |f_1^{(l)}(x)| + |f_1^{(l)}(y)| \\
&\leq C_l + C_l \\
&\leq 2C_l \cdot |x-y|^\alpha.
\end{aligned}$$

Insgesamt gilt für alle $x, y \in \mathbb{R}$

$$|f_1^{(l)}(x) - f_1^{(l)}(y)| \leq (2C_l + C_{l+1}) \cdot |x-y|^\alpha. \tag{3.5}$$

Die l-te Ableitung von f_{σ^2} berechnet man durch l-faches Anwenden der Kettenregel

$$f_{\sigma^2}^{(l)}(x) = \frac{d^l}{dx^l}\left(\sigma^{-1} f_1(\sigma^{-1}x)\right) = \sigma^{-l-1} f_1^{(l)}(\sigma^{-1}x)$$

und schließlich folgt

$$\begin{aligned}
|f_{\sigma^2}^{(l)}(x) - f_{\sigma^2}^{(l)}(y)| &= \sigma^{-l-1}|f_1^{(l)}(\sigma^{-1}x) - f_1^{(l)}(\sigma^{-1}y)| \\
&(\textit{Verwende } (3.5)) \\
&\leq \sigma^{-l-1}(2C_l + C_{l+1})|\sigma^{-1}x - \sigma^{-1}y|^{\alpha} \\
&= \underbrace{\sigma^{-(l+\alpha+1)}(2C_l + C_{l+1})}_{\overset{!}{\leq} B} |x-y|^{\alpha}
\end{aligned}$$

Die definierende Bedingung ist also genau dann erfüllt, wenn $\sigma^{l+\alpha+1} \geq \frac{2C_l+C_{l+1}}{B}$ bzw. $\sigma \geq \left(\frac{2C_l+C_{l+1}}{B}\right)^{1/(l+\alpha+1)}$ gilt. Für eine solche Wahl von σ liegt f_{σ^2} also in $\mathcal{F}^{(4)}_{(l,\alpha,B)}$.

Damit ist für alle vier Dichteklassen gezeigt, dass sie eine Dichte mit nirgends verschwindender Fouriertransformierten enthalten. Noch zu zeigen ist, dass in allen vier Fällen $\mathcal{H}$ in $\mathcal{F}$ als Teilmenge einbettet.

$\underline{\mathcal{F}^{(1)}_{(C,\beta)}}, \underline{\mathcal{F}^{(2)}_{(C,\beta)}}, \underline{\mathcal{F}^{(3)}_{(C,\beta)}}$:
Für jedes $h \in \mathcal{H}$ existieren ein $f \in \mathcal{F}$ und ein $g \in \mathcal{G}$, so dass $h = f * g$ gilt. Daraus folgt

$$|\psi_h(t)| \leq |\psi_f(t)| \cdot \underbrace{|\psi_g(t)|}_{\leq 1} \leq |\psi_f(t)| \quad \forall t \in \mathbb{R},$$

da g eine Dichte ist. Nach der Definition der ersten drei Dichteklassen ist dadurch klar, dass $\mathcal{H} \subseteq \mathcal{F}$ gilt, denn eine Dichte gefalten mit einer $L_2(\mathbb{R})$-Dichte ergibt erneut eine $L_2(\mathbb{R})$-Dichte (siehe Lemma 1.1) und es existiert für jedes $h \in \mathcal{H}$ ein $f \in \mathcal{F}$, so dass

$$|\psi_h(t)| \leq |\psi_f(t)| \leq C|t|^{-\beta}, \forall t \in \mathbb{R}$$

$$\int_{\omega}^{\infty} |\psi_h(t)|^2 dt \leq \int_{\omega}^{\infty} |\psi_f(t)|^2 dt \leq C\omega^{1-2\beta}$$

$$\int |\psi_h(t)|^2 (1+|t|)^{2\beta}\, dt \leq \int |\psi_f(t)|^2 (1+|t|)^{2\beta}\, dt \leq C.$$

$\underline{\mathcal{F}^{(4)}_{(l,\alpha,B)}:}$

Sei $h \in \mathcal{H}$ beliebig. Also kann man h darstellen als $h = g * f = \int g(y) f(\cdot - y)\,dy$ für ein bestimmtes $f \in \mathcal{F}$ und ein $g \in \mathcal{G}$. Nach Lemma 3.3 wissen wir, dass $f^{(j)}$ gleichmäßig beschränkt ist bezüglich S_j. $\int g(y) f^{(j)}(\cdot - y)\,dy$ existiert, da mit $g(\cdot)\cdot S_j$ eine integrable Majorante existiert und der Integrand messbar ist. Wir wollen jetzt zeigen, dass

$$h^{(j)} = \int g(y) f^{(j)}(\cdot - y)\,dy$$

für jedes $j = 0, 1, \ldots, l$ gilt. Der Beweis dafür geschieht durch Induktion nach j:

<u>Induktionsanfang</u>: Für $j = 0$ ist die Aussage wahr.

<u>Induktionsschritt</u> $(j \to j+1)$: Der Differenzenquotient von $h^{(j)}$ für $j < l$ an einer beliebigen Stelle x lautet

$$\frac{h^{(j)}(x+\Delta x) - h^{(j)}(x)}{\Delta x}$$

$$= \int g(y) \underbrace{\frac{f^{(j)}(x + \Delta x - y) - f^{(j)}(x - y)}{\Delta x}}_{|\cdot| \le |f^{(j+1)}(x+\xi-y)| \le S_{j+1}} dy$$

Die Abschätzung erfolgt mittels Verwendung des Mittelwertsatzes der Differentialrechnung. Damit ist $g(\cdot) \cdot S_{(j+1)}$ integrable Majorante. Da $f^{(j)}$ für $j < l$ differenzierbar ist, konvergiert der Integrand punktweise für alle $y \in \mathbb{R}$ gegen $f^{(j+1)}(x-y)g(y)$ für $\Delta x \to 0$. Nach dem Satz von der majorisierten Konvergenz konvergiert damit der Differenzenquotient von $h^{(j)}$ für $\Delta x \to 0$ gegen $\int g(y) f^{(j+1)}(x-y)\,dy$, womit der Induktionsbeweis abgeschlossen ist.

Mit dieser Differentiationsregel für h folgern wir

$$\begin{aligned}
|h^{(l)}(x) - h^{(l)}(y)| &= |\textstyle\int g(z)(f^{(l)}(x-z) - f^{(l)}(y-z))dz| \\
&\le \textstyle\int g(z) \underbrace{|f^{(l)}(x-z) - f^{(l)}(y-z)|}_{\le B|x-z-y+z|^\alpha} dz \\
&\le B\,|x-y|^\alpha \cdot \underbrace{\int g(z)dz}_{=1} \\
&= B|x-y|^\alpha.
\end{aligned}$$

Damit schließen wir $h \in \mathcal{F}$ und somit $\mathcal{H} \subseteq \mathcal{F}$ und der Satz ist bewiesen. ■

Für vier wesentliche Dichteklassen $\mathcal{F}$ lässt sich also mittels Überlappung zeigen, dass es keine robust konsistente Schätzersequenz geben kann, sobald die Fehlerdichte nicht eindeutig spezifizierbar ist.

3.3 Hinreichende und notwendige Kriterien

In diesem Abschnitt wird nach hinreichenden und notwendigen Kriterien für die robust d^k-konsistente Schätzbarkeit geforscht. In Satz 3.1 wird gezeigt, dass ein Überlappen der Mengen $\mathcal{F}$ und $\mathcal{G}$ die Existenz eines robust konsistenten Schätzers unmöglich macht. Deshalb kommen nur sich nicht überlappende Mengentupel $(\mathcal{F}, \mathcal{G})$ in Betracht.

Die Bedingung des Nicht-Überlappens lässt sich mathematisch verschiedentlich ausdrücken. Zunächst kann man die Definition (3.1) einfach negieren und erhält

$$\forall f, \tilde{f} \in \mathcal{F} \, \forall g, \tilde{g} \in \mathcal{G} \; : \; f * g = \tilde{f} * \tilde{g} \; \Rightarrow \; f = \tilde{f}. \tag{3.6}$$

Nach Definition von $\mathcal{H}$ gilt

$$\forall h \in \mathcal{H} \, \exists g \in \mathcal{G}, f \in \mathcal{F} \; : \; h = f * g.$$

(3.6) garantiert die Eindeutigkeit bzgl. $f \in \mathcal{F}$:

$$\forall h \in \mathcal{H} \, \exists g \in \mathcal{G} \, \exists! f \in \mathcal{F} \; : \; h = f * g.$$

Mit dieser Aussage lässt sich eine Abbildung Φ von $\mathcal{H}$ nach $\mathcal{F}$ durch

$$\Phi(h) = f :\Leftrightarrow \exists g \in \mathcal{G} : h = g * f \tag{3.7}$$

definieren. Φ existiert also immer, wenn sich $\mathcal{F}$ und $\mathcal{G}$ nicht überlappen, und ist in diesem Fall durch (3.7) eindeutig definiert. Im Falle der Einelementigkeit von $\mathcal{G}$ entspricht Φ der Inversen des Faltungsoperators mit der einzigen möglichen Fehlerdichte. Sobald aber $\mathcal{G}$ mehrelementig ist, handelt es sich bei Φ im Allgemeinen um keinen linearen Operator!

Weiterhin kann man (3.6) auch äquivalent durch Faltungsklassen formulieren (siehe (3.2)):

$$\forall f, \tilde{f} \in \mathcal{F} \; : \; f * \mathcal{G} \cap \tilde{f} * \mathcal{G} \neq \emptyset \; \Rightarrow \; f = \tilde{f} \tag{3.8}$$

Aufgrund der Definition von $\mathcal{H}$ gilt zudem $\mathcal{H} = \bigcup_{f \in \mathcal{F}} f * \mathcal{G}$. $\mathcal{G}$ und somit auch alle Mengen $f * \mathcal{G}$ sind nicht leer. Aus diesen Aussagen folgt, dass die Menge

$$\{f * \mathcal{G} \mid f \in \mathcal{F}\} \tag{3.9}$$

eine Partition von $\mathcal{H}$ ist. Diese Bedingungen können also bei der weiteren Betrachtung gefordert werden.
Mit der Definition der Abbildung Φ können auch die robuste d^k-Konsistenz und die gleichmäßig robuste d^k-Konsistenz äquivalent formuliert werden:

$$(1.5) \Leftrightarrow E_h\, d(\hat{f}_n(Y_1, \ldots, Y_n), \Phi(h))^k \stackrel{n\to\infty}{\longrightarrow} 0\,, \forall h \in \mathcal{H} \tag{3.10}$$

sowie

$$(1.6) \Leftrightarrow \sup_{h \in \mathcal{H}} E_h\, d(\hat{f}_n(Y_1, \ldots, Y_n), \Phi(h))^k \stackrel{n\to\infty}{\longrightarrow} 0. \tag{3.11}$$

Wird an die Abbildung Φ noch eine Stetigkeitsforderung gestellt, so findet man folgende hinreichende Bedingung, um die Exsistenz einer robust konsistenten Schätzung zu zeigen

Satz 3.3 *Seien (X, d), (Y, b) metrische Räume, $\mathcal{D}$ die Menge aller Dichten. Es gelte $\mathcal{F} \subseteq X \cap \mathcal{D}$, $\mathcal{G} \subseteq \mathcal{D}$ und $\mathcal{H} = \mathcal{F} * \mathcal{G} \subseteq Y$*

(1) Es gelte $\sup_{f, \tilde{f} \in \mathcal{F}} d(f, \tilde{f}) \leq C < +\infty$.

(2) Es existiere eine b^l-konsistente Schätzerfolge $(\tilde{h}_n)_{n \in \mathbb{N}}$, basierend auf unabhängigen, identisch verteilten Beobachtungen aus $h \in \mathcal{H}$ $(l > 0)$

$$E_h\, b(\tilde{h}_n, h)^l \stackrel{n\to\infty}{\longrightarrow} 0 \quad \forall h \in \mathcal{H}$$

(3) Die Mengen $\mathcal{F}$ und $\mathcal{G}$ sollen sich nicht überlappen, und die folglich existierende durch (3.7) definierte Abbildung Φ sei (b, d)-stetig.

(a) Dann existiert eine $\mathcal{H}$-immanente b^l-konsistente Schätzersequenz $(\hat{h}_n)_{n \in \mathbb{N}}$, basierend auf unabhängigen, identisch verteilten Beobachtungen aus $h \in \mathcal{H}$.
(b) Dann existiert eine robust d^k-konsistente $\mathcal{F}$-immanente Schätzerfolge von $f \in \mathcal{F}$, nämlich $(\Phi(\hat{h}_n))_{n \in \mathbb{N}}$.

Anmerkung: (a) $\mathcal{H}$-*immanent* bedeutet, dass für die Schätzerfolge $(\hat{h}_n)_n$ für alle möglicherweise auftretenden Realisierungen $y_1, \ldots, y_n$ und alle $n \in \mathbb{N}$

$$\hat{h}_n(y_1, \ldots, y_n) \in \mathcal{H}$$

gilt. $\mathcal{F}$-Immanenz sei analog definiert.
(b) Wenn die Abbildung Φ nur in manchen Stellen h stetig ist, liefert der Beweis des Satzes auch die robuste Konsistenz des Schätzers $(\Phi(\hat{h}_n))_n$ in diesen Stellen. Dies wird bei Satz 3.4 benötigt.

Beweis: ad (a): Hier gilt es nun, aufgrund des in Bedingung (2) gegebenen Schätzers $(\tilde{h}_n)_n$, einen $\mathcal{H}$-immanenten zu finden und gleichzeitig die b^l-Konsistenz des Schätzers zu bewahren. Dazu wählen wir eine reelle Folge $(o_n)_n$ aus, die gegen 0 konvergiert und deren Folgeglieder alle strikt positiv sind, also z.B. $o_n = 1/n$. Definieren wir die Menge

$$\mathcal{H}(n, y_1, \ldots, y_n) := \{h \in \mathcal{H} \mid b(\tilde{h}_n(y_1, \ldots, y_n), h)^l \le o_n + \inf_{\tilde{h} \in \mathcal{H}} b(\tilde{h}_n(y_1, \ldots, y_n), \tilde{h})^l\}$$

Das dabei beschriebene Infimum existiert in $\mathbb{R}_0^+$. $\mathcal{H}(n, y_1, \ldots, y_n)$ ist für alle $n \in \mathbb{N}$ und alle $(y_1, \ldots, y_n) \in \mathbb{R}^{(n)}$ nicht leer, sonst wäre $\frac{1}{2}o_n + \inf_{\tilde{h} \in \mathcal{H}} b(\tilde{h}_n(y_1, \ldots, y_n), \tilde{h})^l$ untere Schranke von $\{b(\tilde{h}_n(y_1, \ldots, y_n), \tilde{h})^l \mid \tilde{h} \in \mathcal{H}\}$ und somit nicht größer als $\inf_{\tilde{h} \in \mathcal{H}} b(\tilde{h}_n(y_1, \ldots, y_n), \tilde{h})^l$. Nach dem Auswahlaxiom existiert daher eine Auswahlabbildung $\hat{h}$ von $\mathbb{N} \times \mathbb{R}^{(n)}$ nach $\bigcup_{(n, y_1, \ldots, y_n) \in \mathbb{N} \times \mathbb{R}^{(n)}} \mathcal{H}(n, y_1, \ldots, y_n)$ mit $\hat{h}(n, y_1, \ldots, y_n) \in \mathcal{H}(n, y_1, \ldots, y_n) \subseteq \mathcal{H}$. Diese Auswahlabbildung $\hat{h}$ fungiert jetzt als $\mathcal{H}$-immanenter Dichteschätzer. Zu zeigen ist noch seine b^l-Konsistenz. Wegen $\hat{h}(n, y_1, \ldots, y_n) \in \mathcal{H}(n, y_1, \ldots, y_n)$ gilt

$$b(\hat{h}_n(y_1, \ldots, y_n), \tilde{h}_n(y_1, \ldots, y_n))^l \le o_n + \inf_{\tilde{h} \in \mathcal{H}} b(\tilde{h}_n(y_1, \ldots, y_n), \tilde{h})^l.$$

Daraus folgt unter Verwendung der Dreiecksungleichung und von Lemma 3.1

$$E_h\, b(\hat{h}_n(Y_1, \ldots, Y_n), h)^l$$

$$\leq \Big(E_h\, b(\tilde{h}_n(Y_1,\ldots,Y_n),\hat{h}_n(Y_1,\ldots,Y_n))^l \,+\, E_h\, b(\tilde{h}_n(Y_1,\ldots,Y_n),h)^l\Big) \cdot (\min\{2^{1-l},1\})^{-1}$$

$$\leq \Big(o_n \,+\, E_h \inf_{\tilde{h}\in\mathcal{H}} b(\tilde{h}_n(Y_1,\ldots,Y_n),\tilde{h})^l \,+\, E_h\, b(\tilde{h}_n(Y_1,\ldots,Y_n),h)^l\Big) \cdot (\min\{2^{1-l},1\})^{-1}$$

$$\leq \Big(o_n \,+\, 2\, E_h\, b(\tilde{h}_n(Y_1,\ldots,Y_n),h)^l\Big) \cdot (\min\{2^{1-l},1\})^{-1}$$

$$\stackrel{n\to\infty}{\longrightarrow} 0.$$

für alle $h \in \mathcal{H}$. Der Beweis ist nicht konstruktiv, und die Existenz der Schätzerfolge ist zwar gezeigt, aber im Allgemeinen kann sie nicht explizit aus $(\tilde{h}_n)_n$ konstruiert werden.

ad (b): Es ist zu zeigen, dass (3.10) für $\hat{f}_n = \Phi(\hat{h}_n)$ gilt. Die $\mathcal{F}$-Immanenz von $(\hat{f}_n)_n$ ist somit klar, da Φ nach $\mathcal{F}$ abbildet. Mittels Konditionierung gilt für ein beliebiges, aber festes $\epsilon > 0$ und ein beliebiges $h \in \mathcal{H}$

$$E_h d(\Phi(\hat{h}_n),\Phi(h))^k$$

$$\begin{aligned}
&= E_h\Big(d(\Phi(\hat{h}_n),\Phi(h))^k \mid d(\Phi(\hat{h}_n),\Phi(h)) \geq \epsilon\Big)\cdot P_h(d(\Phi(\hat{h}_n),\Phi(h)) \geq \epsilon)\\
&\quad + E_h\Big(d(\Phi(\hat{h}_n),\Phi(h))^k \mid d(\Phi(\hat{h}_n),\Phi(h)) < \epsilon\Big)\cdot P_h(d(\Phi(\hat{h}_n),\Phi(h)) < \epsilon)
\end{aligned}$$

$$\begin{aligned}
&\leq E_h\Big(C^k \mid d(\Phi(\hat{h}_n),\Phi(h)) \geq \epsilon\Big)\cdot P_h(d(\Phi(\hat{h}_n),\Phi(h)) \geq \epsilon)\\
&\quad + E_h\Big(\epsilon^k \mid d(\Phi(\hat{h}_n),\Phi(h)) < \epsilon\Big)\cdot \underbrace{P_h(d(\Phi(\hat{h}_n),\Phi(h)) < \epsilon)}_{\leq 1}
\end{aligned}$$

$$\leq \epsilon^k + C^k \cdot P_h(d(\Phi(\hat{h}_n),\Phi(h)) \geq \epsilon).$$

◀ Φ ist (b,d)-stetig, d.h. für das obige $\epsilon > 0$ existiert ein $\delta(\epsilon,h) > 0$, so dass für alle $\tilde{h} \in \mathcal{H}$ mit $b(\tilde{h},h) < \delta(\epsilon,h)$ gilt $d(\Phi(\tilde{h}),\Phi(h)) < \epsilon$. Also ist

$$\begin{aligned}
&\{\omega\in\Omega | b(\hat{h}_n(Y_1(\omega),\ldots,Y_n(\omega)),h) < \delta(\epsilon,h)\}\\
&\qquad \subseteq \{\omega\in\Omega | d(\Phi(\hat{h}_n(Y_1(\omega),\ldots,Y_n(\omega))),\Phi(h)) < \epsilon\},
\end{aligned}$$

woraus wegen der Monotonie eines jeden (Wahrscheinlichkeits-)maßes folgt

$$P_h\left(b(\hat{h}_n(Y_1,\ldots,Y_n),h) < \delta(\epsilon,h)\right) \;\leq\; P_h\left(d(\Phi(\hat{h}_n(Y_1,\ldots,Y_n)),\Phi(h)) < \epsilon\right)$$

und folglich

$$P_h\left(b(\hat{h}_n(Y_1,\ldots,Y_n),h) \geq \delta(\epsilon,h)\right) \;\geq\; P_h\left(d(\Phi(\hat{h}_n(Y_1,\ldots,Y_n)),\Phi(h)) \geq \epsilon\right).$$

▶

Mit dieser Überlegung fahren wir mit der obigen Abschätzung fort

$$\epsilon^k + C^k \cdot P_h(d(\Phi(\hat{h}_n),\Phi(h)) \geq \epsilon)$$

$$\leq \epsilon^k + C^k \cdot P_h(b(\hat{h}_n,h) \geq \delta(\epsilon,h))$$

$$\leq \epsilon^k + C^k \cdot \delta(\epsilon,h)^{-l}\left(\delta(\epsilon,h)^l P_h(b(\hat{h}_n,h) \geq \delta(\epsilon,h))\right)$$

(Markov'sche Ungleichung)

$$\leq \epsilon^k + C^k \cdot \delta(\epsilon,h)^{-l}\, \underbrace{E_h\, b(\hat{h}_n,h)^l}_{\longrightarrow 0 \text{ für } n\to\infty}$$

$$\leq 2\cdot\epsilon^k.$$

Der letzte Schritt der Ungleichungskette gilt, wenn n in Abhängigkeit von ϵ und h groß genug gewählt wird, also für alle $n > N(\epsilon,h)$. Man muss folglich zunächst $\epsilon > 0$ klein genug wählen und anschließend $N(\epsilon,h)$ groß genug, um $E_h d(\Phi(\hat{h}_n),\Phi(h))^k$ bzgl. einer beliebig kleinen, aber festen positiven Schranke nach oben für alle n , die größer als ein gewisses N sind, abzuschätzen. Damit ist die robuste d^k -Konsistenz der Schätzerfolge bewiesen. ■

Wenn es sich bei (Y,b) um $Y = L_1(\mathbb{R})$ mit der kanonischen $L_1(\mathbb{R})$ -Metrik handelt, kann auf $\mathcal{H} \subseteq Y$ und die Voraussetzung (2) verzichtet werden, denn (2) ist erfüllt nach dem Äquivalenztheorem von *Devroye* (1983) , wie es z.B. in *Devroye und Györfi* [10] nachzulesen ist. Das Theorem sagt aus, dass bei geeigneter Wahl der Bandbreitenfolge ein Kernschätzer mit einer Dichte als Kern im L_1 -Sinne jede Dichte h konsistent schätzt. Dies fassen wir in folgendem Korollar zu Satz 3.3 zusammen

Korollar 3.1 *Sei* (X,d) *ein metrischer Raum,* $\mathcal{D}$ *die Menge aller Dichten. Es gelte* $\mathcal{F} \subseteq X \cap \mathcal{D}$, $\mathcal{G} \subseteq \mathcal{D}$

(1) Es gelte $\sup\limits_{f,\tilde{f}\in\mathcal{F}} d(f,\tilde{f}) \leq C < +\infty$.

(2) Die Mengen $\mathcal{F}$ *und* $\mathcal{G}$ *sollen sich nicht überlappen, und die folglich existierende durch (3.7) definierte Abbildung* Φ *sei* $(L_1(\mathbb{R}),d)$ *-stetig.*

(a) Dann existiert eine $\mathcal{H}$ *-immanente* $L_1(\mathbb{R})$ *-konsistente Schätzersequenz* $(\hat{h}_n)_{n\in\mathbb{N}}$ *, basierend auf unabhängigen, identisch verteilten Beobachtungen aus* $h \in \mathcal{H}$ *.*
(b) Dann existiert eine robust d^k *-konsistente* $\mathcal{F}$ *-immanente Schätzerfolge von* $f \in \mathcal{F}$ *, nämlich* $(\Phi(\hat{h}_n))_{n\in\mathbb{N}}$ *.*

Die Bedingung (1) darf dabei nicht vergessen werden. Im Falle $(X,d) = (L_1(\mathbb{R}), \|\cdot\|_{L_1(\mathbb{R})})$ dagegen ist (1) erfüllt, da die $L_1(\mathbb{R})$ -Distanz zweier Dichten stets kleiner oder gleich 2 ist.
Für die vier Dichteklassen aus Satz 3.2 lässt sich Bedingung (1) auch für weitere $L_p(\mathbb{R})$ -Räume verifizieren. Dazu benötigt man folgendes Lemma

Lemma 3.4 *(a) Sei* $f \in L_p(\mathbb{R}) \cap L_q(\mathbb{R})$ *mit* $p < q < \infty$ *. Dann gilt* $f \in L_r(\mathbb{R})$ *und* $\|f\|_{L_r(\mathbb{R})} \leq \left(\|f\|^p_{L_p(\mathbb{R})} + \|f\|^q_{L_q(\mathbb{R})}\right)^{1/r}$ *für alle* $p < r < q$ *.*

(b) Sei $f \in L_p(\mathbb{R})$ *und* $|f(t)| \leq C , \forall t \in \mathbb{R}$ *für ein* $C < \infty$ *. Dann ist* $f \in L_r(\mathbb{R})$ *und es gilt* $\|f\|_{L_r(\mathbb{R})} \leq \|f\|^{p/r}_{L_p(\mathbb{R})} (1+C^r)^{1/r}$ *für alle* $\infty > r > p$ *.*

Beweis: Zunächst definieren wir $G := \{t \in \mathbb{R} \mid |f(t)| \leq 1\}$. Diese Menge ist LB-messbar, da in (a) und (b) die Mitgliedschaft von f in einem $L_p(\mathbb{R})$ -Raum vorausgesetzt ist. Aus der LB-Messbarkeit der Funktion $|f|^p$ folgt auch die LB-Messbarkeit von $|f|^r$. Man braucht für $|f|^r$ also nur noch die Existenz einer integrablen Majorante zu zeigen, was simultan mit dem Beweis der Ungleichungen geschehen wird.
ad (a): Es gilt

$$\begin{aligned} \int_G |f(t)|^r dt &\leq \int_G |f(t)|^p dt &&\leq \|f\|^p_{L_p(\mathbb{R})} \\ \int_{G^\complement} |f(t)|^r dt &\leq \int_{G^\complement} |f(t)|^q dt &&\leq \|f\|^q_{L_q(\mathbb{R})} \end{aligned}$$

Also ist $\chi_G|f|^p + \chi_{G^\complement}|f|^q$ integrable Majorante von $|f|^r$ und es gilt

$$\|f\|^r_{L_r(\mathbb{R})} \le \|f\|^p_{L_p(\mathbb{R})} + \|f\|^q_{L_q(\mathbb{R})}$$

Durch anschließendes Potenzieren mit $1/r$ ist der Beweis erbracht.
ad (b): Für das LB-Maß von $G^\complement$, $\mu(G^\complement)$, gilt

$$\infty > \|f\|^p_{L_p(\mathbb{R})} \ge \int_{G^\complement} \underbrace{|f(t)|^p}_{\ge 1} dt \ge \mu(G^\complement).$$

$G^\complement$ besitzt also endliches LB-Maß und $\chi_G|f|^p + \chi_{G^\complement}C^r$ ist integrable Majorante von $|f|^r$. Weiter folgt

$$\begin{aligned}
\|f\|^r_{L_r(\mathbb{R})} &= \int |f(t)|^r \, dt \\
&= \int_G \underbrace{|f(t)|^r}_{\le |f(t)|^p} dt + \int_{G^\complement} \underbrace{|f(t)|^r}_{\le C^r} dt \\
&\le \|f\|^p_{L_p(\mathbb{R})} + C^r \cdot \mu(G^\complement) \\
&\le \|f\|^p_{L_p(\mathbb{R})} (1 + C^r) .
\end{aligned}$$

Erneut liefert einfaches Potenzieren mit $1/r$ die gewünschte Ungleichung.
■

Für die ersten drei Dichteklassen $\mathcal{F}^{(1)}_{(C,\beta)}, \mathcal{F}^{(2)}_{(C,\beta)}, \mathcal{F}^{(3)}_{(C,\beta)}$ lassen sich gleichgradige Schranken für die $L_2(\mathbb{R})$ -Norm der Fouriertransformierten der Dichten erkennen: im Falle $\mathcal{F}^{(1)}_{(C,\beta)}$ durch Erkennen der integrablen Majorante $\min\{1, C^2|\bullet|^{-2\beta}\}$, wenn man $\beta > 1$ bedenkt, für $\mathcal{F}^{(2)}_{(C,\beta)}$ ist $2\omega_0 + 2C\omega_0^{1-2\beta}$ eine Schranke für die $L_2(\mathbb{R})$ -Norm und schließlich für $\mathcal{F}^{(3)}_{(C,\beta)}$ ist $|\psi_f|^2(1+|\bullet|)^{2\beta}$ integrable Majorante für $|\psi_f|^2$ und die $L_2(\mathbb{R})$ -Norm der Fouriertransformierten somit durch C beschränkt.
Mit der Parseval-Identität kann man sofort daraus gleichgradige Schranken für die $L_2(\mathbb{R})$ -Norm der Dichten f durch Multiplikation der entsprechenden Schranke von $\|\psi_f\|_{L_2(\mathbb{R})}$ mit $\frac{1}{\sqrt{2\pi}}$ herleiten. Wendet man nun Lemma 3.4(a) an, erhält man, dass alle Dichtemengen $\mathcal{F}^{(1)}_{(C,\beta)}, \mathcal{F}^{(2)}_{(C,\beta)}, \mathcal{F}^{(3)}_{(C,\beta)}$ jeweils bezüglich der $L_p(\mathbb{R})$ -Norm für alle $1 \le p \le 2$ gleichgradig beschränkt sind und für diese Metriken Bedingung (1) von Satz 3.3 und Korollar 3.1 also erfüllt

ist.
Für die Dichteklasse $\mathcal{F}^{(4)}_{(l,\alpha,B)}$ wird in Lemma 3.3 ermittelt, dass $|f(t)| \leq S_0$ für alle $t \in \mathbb{R}$ und alle $f \in \mathcal{F}^{(4)}_{(l,\alpha,B)}$ gilt. Unter Anwendung von Lemma 3.4(b) erkennt man schließlich eine gleichgradige Schranke der $L_p(\mathbb{R})$ -Norm sogar für alle $p \geq 1$ und für diese Metriken ist Bedingung (1) damit erfüllt. Für diese Dichteklassen nach robust konsistenten Schätzern zu suchen ist nicht sinnvoll, da durch Satz 3.2 die Möglichkeit einer solchen Schätzung ausgeschlossen ist, sobald $\mathcal{G}$ mehrelementig ist. Die Bedingung (1) von Satz 3.3 und Korollar 3.1 vererbt sich jedoch auf jede Teilmenge dieser Dichteklassen. In Kapitel 5 wird eine (sogar gleichmäßig) robust konsistente Schätzung für eine solche Teilmenge von $\mathcal{F}^{(1)}_{(C,\beta)}$ als Dichteklasse $\mathcal{F}$ durchgeführt.

Anzumerken ist auch, dass aus der Bedingung des Nicht-Überlappens bzw. der Existenz der Abbildung Φ nach (3.7) im Allgemeinen nicht ihre $(L_1(\mathbb{R}), d)$ -Stetigkeit für irgendeine Metrik d folgt. Betrachten wir zum Beweis der Aussage folgendes Gegenbeispiel: Es sei

$$\begin{aligned} &\mathcal{F} := \{\tilde{f}, f\} \quad \text{und} \\ &\mathcal{G} := \{\lambda f + (1-\lambda)\tilde{f} \mid \lambda \in (0;1]\}. \end{aligned} \tag{3.12}$$

Dabei sei $f \neq \tilde{f}$ und ψ_f und $\psi_{\tilde{f}}$ sollen nirgends verschwinden. Dann gilt für die Faltungsklassen

$$\begin{aligned} f * \mathcal{G} &= \{\lambda(f * f) + (1-\lambda)(\tilde{f} * f) \mid \lambda \in (0;1]\} \\ \tilde{f} * \mathcal{G} &= \{\lambda(\tilde{f} * f) + (1-\lambda)(\tilde{f} * \tilde{f}) \mid \lambda \in (0;1]\} \end{aligned}$$

sowie $\mathcal{H} = f * \mathcal{G} \cup \tilde{f} * \mathcal{G}$.

$\mathcal{F}$ und $\mathcal{G}$ überlappen sich nicht. Zum Beweis nehmen wir an, es gebe ein $h \in f * \mathcal{G} \cap \tilde{f} * \mathcal{G}$. Also gilt

$$\begin{aligned} &h = \lambda(f * f) + (1-\lambda)(f * \tilde{f}) \quad \text{für ein } \lambda \in (0;1] \quad \text{und} \\ &h = \mu(f * \tilde{f}) + (1-\mu)(\tilde{f} * \tilde{f}) \quad \text{für ein } \mu \in (0;1] \end{aligned}$$

Durch Subtraktion der Gleichungen erhält man

$$\lambda(f * f) + (1 - \lambda - \mu)(f * \tilde{f}) - (1 - \mu)(\tilde{f} * \tilde{f}) \;=\; 0.$$

Durch Fouriertransformieren dieser Gleichung folgt

$$\lambda\psi_f(t)^2 + (1 - \lambda - \mu)\psi_f(t) \cdot \psi_{\tilde{f}}(t) - (1 - \mu)\psi_{\tilde{f}}(t)^2 \;=\; 0 \quad , \quad \forall t \in \mathbb{R}.$$

Da $\psi_{\tilde{f}}$ nirgends verschwindet, darf durch $\psi_{\tilde{f}}(t)^2$ dividiert werden. Setzen wir dann $\Psi(t) := \frac{\psi_f(t)}{\psi_{\tilde{f}}(t)}$, so lautet die Gleichung

$$\lambda\Psi(t)^2 + (1 - \lambda - \mu)\Psi(t) - (1 - \mu) \;=\; 0 \quad , \quad \forall t \in \mathbb{R}.$$

Da $\lambda \neq 0$ nach Definition ist, besitzt die quadratische Gleichung höchstens zwei komplexe Lösungen. Da für jedes $t \in \mathbb{R}$ $\Psi(t)$ Lösung dieser Gleichung (mit t-unabhängigen Koeffizienten!) ist, kann der Wertebereich der Funktion Ψ höchstens zwei Elemente enthalten. Ψ ist als Quotient zweier stetiger Funktionen mit nirgends verschwindendem Nenner selbst eine stetige Funktion $\mathbb{R} \to \mathbb{C}$, und deshalb kann die Mächtigkeit des Wertebereiches von Ψ entweder 1 oder unendlich sein. Folglich ist der Wertebereich von Ψ einelementig, und Ψ ist somit konstant. Es ist $\Psi(0) = \frac{\psi_f(0)}{\psi_{\tilde{f}}(0)} = 1$ und somit

$$\Psi \equiv 1.$$

Aus der Definition von Ψ folgt somit

$$\psi_f(t) = \psi_{\tilde{f}}(t) \quad , \quad \forall t \in \mathbb{R}.$$

Nach dem Eindeutigkeitssatz der Fouriertransformierten (siehe *Schmitz, N.* [14]) impliziert dies $f = \tilde{f}$. Das ist aber ein Widerspruch zu der Definition von $\mathcal{F}$.
$\Longrightarrow$ $f * \mathcal{G} \cap \tilde{f} * \mathcal{G} = \emptyset.$

Die Abbildung $\Phi : \mathcal{H} \longrightarrow \mathcal{F}$ ist bzgl. keiner Metrik d auf $\mathcal{F}$ $(L_1(\mathbb{R}), d)$-stetig. Zum Beweis betrachten wir die Folge

$$(h_n)_{n\in\mathbb{N}} := \left(\frac{1}{n}(f * f) + \left(1 - \frac{1}{n}\right)(f * \tilde{f})\right)_{n\in\mathbb{N}}$$

$(h_n)_n$ konvergiert in der $L_1(\mathbb{R})$ -Norm gegen $f * \tilde{f}$, denn:

$$\begin{aligned}
\|h_n - f * \tilde{f}\|_{L_1(\mathbb{R})} &= \|\tfrac{1}{n}(f * f) + (1 - \tfrac{1}{n} - 1)(f * \tilde{f})\|_{L_1(\mathbb{R})} \\
&\leq \tfrac{1}{n}\,\|f * f - f * \tilde{f}\|_{L_1(\mathbb{R})} \\
&\leq \tfrac{2}{n} \overset{n\to\infty}{\longrightarrow} 0.
\end{aligned}$$

Man erkennt, dass für alle $n \in \mathbb{N}$ h_n in $f * \mathcal{G}$ liegt, also gilt $\Phi(h_n) = f$ für alle $n \in \mathbb{N}$; andererseits liegt $\tilde{f} * f$ in $\tilde{f} * \mathcal{G}$, und es gilt $\Phi(\tilde{f} * f) = \tilde{f}$. Das impliziert

$$d(\Phi(h_n), \Phi(f * \tilde{f})) \;=\; \underbrace{d(f, \tilde{f})}_{\text{unabh. von } n} \;>\; 0,$$

da $f \neq \tilde{f}$ ist. Somit konvergiert $\Phi(h_n)$ nicht in der Metrik d gegen $\Phi(f*\tilde{f})$. Folglich kann ϕ nicht $(L_1(\mathbb{R}), d)$ -stetig sein.

Dennoch kann man zeigen, dass in dieser Situation bezüglich jeder Metrik d ein robust konsistenter Schätzer existiert. Die Abbildung Φ ist nur an einer Stelle unstetig und zwar in $h = \tilde{f} * f$. Dies kann man leicht erkennen, indem man den $\|\cdot\|_{L_1(\mathbb{R})}$ -Abschluss der Dichteklassen betrachtet

$$\begin{aligned}
\overline{f * \mathcal{G}}^{L_1(\mathbb{R})} &= \{\lambda(f * f) + (1-\lambda)(\tilde{f} * f) \mid \lambda \in [0;1]\} \\
\overline{\tilde{f} * \mathcal{G}}^{L_1(\mathbb{R})} &= \{\lambda(\tilde{f} * f) + (1-\lambda)(\tilde{f} * \tilde{f}) \mid \lambda \in [0;1]\}.
\end{aligned}$$

Die Schnittmenge $\overline{\tilde{f} * \mathcal{G}}^{L_1(\mathbb{R})} \cap \overline{f * \mathcal{G}}^{L_1(\mathbb{R})}$ ist somit $\{f * \tilde{f}\}$. Diese beiden Dichteklassen sind als stetiges Bild eines kompakten reellen Intervalls in $L_1(\mathbb{R})$ kompakt. Damit ist der L_1 -Abstand eines $h \in f * \mathcal{G}$ zu $\overline{\tilde{f} * \mathcal{G}}^{L_1(\mathbb{R})}$ positiv, ebenso der L_1 -Abstand eines $h \in (\tilde{f} * \mathcal{G})\backslash\{\tilde{f} * f\}$ zu $\overline{f * \mathcal{G}}^{L_1(\mathbb{R})}$. Für jede Folge $(h_n)_n$ in $\mathcal{H}$ mit $\|h_n - h\|_{L_1(\mathbb{R})} \overset{n\to\infty}{\longrightarrow} 0$ und $h \neq \tilde{f} * f$ gilt also für alle bis auf endlich viele $n \in \mathbb{N}$ $\Phi(h_n) = \Phi(h)$ und folglich besteht in allen Dichten aus $\mathcal{H}$ bis auf $\tilde{f} * f$ Stetigkeit.
Die robust konsistente Schätzbarkeit in der vorliegenden Situation kann mit folgendem Satz gefolgert werden:

Satz 3.4 *Seien* (X, d) *ein metrischer Raum,* $\mathcal{F} \subseteq \mathcal{D} \cap X$, $\mathcal{G} \subseteq \mathcal{D}$. *Es gelte* $\sup_{f,\tilde{f} \in \mathcal{F}} d(f, \tilde{f}) \leq C < \infty$. $\mathcal{F}$ *und* $\mathcal{G}$ *überlappen sich nicht und die Menge*

$$\mathcal{H}_0 := \{h \in \mathcal{H} \mid \Phi \textit{ ist nicht } (L_1(\mathbb{R}), d)\textit{-stetig in } h\} \tag{3.13}$$

sei höchstens abzählbar.
Dann existiert eine $\mathcal{F}$-immanente d^k-robust konsistente Schätzerfolge $(\hat{f}_n)_{n\in\mathbb{N}}$ von f.

Der Schätzer $\hat{f}_n$ wird im Beweis unter (3.15) angegeben. Zunächst wird noch ein technisches Lemma gebraucht.

Lemma 3.5 *Es sei $\{b_{m,n} \mid n \in \mathbb{N}, m \in \mathbb{N}\}$ eine Menge nichtnegativer reeller Zahlen mit $b_{m,n} \overset{n\to\infty}{\longrightarrow} 0$ für jedes $m \in \mathbb{N}$. Dann existiert eine Folge $(M_n)_{n\in\mathbb{N}}$ natürlicher Zahlen mit $M_n \overset{n\to\infty}{\longrightarrow} \infty$ und $b_{M_n,n} \overset{n\to\infty}{\longrightarrow} 0$.*

Beweisen wir das Lemma
Beweis: Die Folge $(M_n)_{n\in\mathbb{N}}$ werde rekursiv definiert durch

$$M_1 := 1,$$

$$M_{n+1} := \begin{cases} M_n + 1 & , \text{ falls } b_{M_n+1,j} \leq 1/M_n \,,\ \forall j \geq n \\ M_n & , \text{ sonst.} \end{cases}$$

Nach dieser Definition ist klar, dass $(M_n)_{n\in\mathbb{N}}$ monoton wächst (nicht zwingend streng) und dass $M_n \leq n$ für alle $n \in \mathbb{N}$ gilt. Zweiteres ist leicht mit Induktion einzusehen. Nehmen wir an, $(M_n)_{n\in\mathbb{N}}$ sei nach oben beschränkt (bzgl. $S > 0$). Nach der Rekursionsformel folgt daraus, dass es ein N gibt, ab dem die Folge stagniert, also $M_n = M_N$ für alle $n \geq N$ gilt. Deshalb muss es für alle $n \geq N$ ein $j \geq n$ geben mit $b_{M_n+1,j} > 1/M_n$. Dabei kann $M_n = M_N$ ersetzt werden und die Aussage lautet, dass für alle $n \geq N$ ein $j \geq n$ existieren muss mit $b_{M_N+1,j} > 1/M_N$. Da $1/M_N \geq 1/S > 0$ gilt, konvergiert eine Teilfolge von $(b_{M_N+1,j})_{j\in\mathbb{N}}$ nicht gegen 0, was einen Widerspruch zur Voraussetzung des Lemmas darstellt. Folglich ist $(M_n)_{n\in\mathbb{N}}$ nach oben unbeschränkt und zusammen mit der Monotonie dieser Folge erhält man

$$M_n \overset{n\to\infty}{\longrightarrow} \infty.$$

Betrachten wir jetzt die Folge $(b_{M_n,n})_{n\in\mathbb{N}}$. Dazu wählen wir $\epsilon > 0$ beliebig. Nach Obigem existiert ein N, so dass $1/M_n < \epsilon$ für alle $n \geq N$ gilt. Ebenso schließt die bestimmte Divergenz von $(M_n)_n$ die Stagnation der Folge aus, so dass sicherlich ein $\tilde{N} \geq N$ existiert mit

$$0 \leq b_{M_{\tilde{N}}+1,j} \leq 1/M_{\tilde{N}} < \epsilon$$

für alle $j \geq \tilde{N}$ und damit $M_{\tilde{N}} + 1 = M_{\tilde{N}+1}$ gilt. Daraus folgt wiederum

$$0 \leq b_{M_{\tilde{N}+1},j} \leq 1/M_{\tilde{N}} < \epsilon \tag{3.14}$$

für alle $j \geq \tilde{N}+1$. Weiter kann jetzt behauptet werden, dass für alle $n \geq \tilde{N}$

$$0 \leq b_{M_{n+1},j} \leq \epsilon \, , \forall j \geq n+1$$

gilt. Dies lässt sich durch vollständige Induktion nach n zeigen. Der Induktionsanfang $(n = \tilde{N})$ ist durch (3.14) bereits bewiesen. Im Induktionsschritt schließt man jetzt von n auf $n+1$. Betrachten wir den

1. Fall: $M_{n+2} = M_{n+1} + 1$. Daraus folgt

$$\begin{aligned}
&\Rightarrow b_{M_{n+1}+1,j} \leq 1/M_{n+1} \, , \forall j \geq n+1 \\
&\Rightarrow b_{M_{n+2},j} \leq 1/M_{n+1} \leq 1/M_{\tilde{N}} < \epsilon, \forall j \geq n+1 \\
&\Rightarrow 0 \leq b_{M_{n+2},j} < \epsilon \, , \forall j \geq n+2.
\end{aligned}$$

2. Fall: $M_{n+2} = M_{n+1}$. Dann gilt

$$b_{M_{n+2},j} = b_{M_{n+1},j} \leq \epsilon \, , \forall j \geq n+1, \text{ also auch für } j \geq n+2$$

Damit ist der Induktionsbeweis in beiden Fällen erbracht; für alle $n \geq \tilde{N}$ gilt also $0 \leq b_{M_{n+1},n+1} \leq \epsilon$ bzw. für alle $n \geq \tilde{N}+1$ $0 \leq b_{M_n,n} \leq \epsilon$. Da $\epsilon > 0$ beliebig gewählt ist, folgt direkt nach Definition

$$b_{M_n,n} \stackrel{n\to\infty}{\longrightarrow} 0.$$

■

Mit diesem Hintergrund nehmen wir den Beweis von Satz (3.4) in Angriff
Beweis: $\mathcal{H}_0$ ist nach Voraussetzung höchstens abzählbar, also ist

$$\mathcal{H}_0 = \{\dot{h}_l \mid l \in \mathbb{N}\}$$

natürlich indizierbar. Gehen wir zunächst davon aus, $\mathcal{H}_0$ sei abzählbar, also nicht endlich, dann kann oBdA angenommen werden, die Zuordnung $l \mapsto \dot{h}_l$ sei injektiv. Davon lassen sich die endlichen Mengen

$$\mathcal{H}_0^M = \{\dot{h}_l \mid l \in \{1, \ldots, M\}\}$$

ableiten. Definieren wir weiter

$$\epsilon_M := \min_{\substack{j,k \in \{1,\ldots,M\} \\ \wedge \dot{h}_j \neq \dot{h}_k}} \|\dot{h}_j - \dot{h}_k\|_{L_1(\mathbb{R})}$$

ϵ_M ist dabei stets existent und positiv, da das Minimum über eine endliche Menge gebildet wird. Seien $(M_n)_{n\in\mathbb{N}}$ und $(\delta_n)_{n\in\mathbb{N}}$ noch zu bestimmende Folgen, $M_n \in \mathbb{N}$ und $\delta_n > 0$. Dann ist der Dichteschätzer durch

$$\hat{f}_n := \begin{cases} \Phi(\dot{h}_1) & \text{, wenn } \|\hat{h}_n - \dot{h}_1\| < \delta_n \\ \vdots & \vdots \\ \Phi(\dot{h}_{M_n}) & \text{, wenn } \|\hat{h}_n - \dot{h}_{M_n}\| < \delta_n \\ \Phi(\hat{h}_n) & \text{, sonst.} \end{cases} \tag{3.15}$$

erklärt. $(\hat{h})_{n\in\mathbb{N}}$ sei ein $\mathcal{H}$-immanenter $L_1(\mathbb{R})$-konsistenter Schätzer von h. Ein solcher existiert nach dem Äquivalenztheorem von *Devroye* (1983) und dem Satz 3.3(a).
Es stellt sich zuerst die Frage, ob der Schätzer (3.15) wohldefiniert ist, denn schließlich können sich die L_1-Kugeln um die Elemente aus $\mathcal{H}_0^{M_n}$ mit Radius $\delta_n > 0$ auch schneiden. Durch die Forderung

$$0 < \delta_n < \epsilon_{M_n}/3 \tag{3.16}$$

lässt sich dies ausschließen, denn angenommen, es existiere ein $\tilde{h} \in \mathcal{H}$ mit $\|\tilde{h} - \dot{h}_l\|_{L_1(\mathbb{R})} < \delta_n$ und $\|\tilde{h} - \dot{h}_j\|_{L_1(\mathbb{R})} < \delta_n$. Es folgt

$$\begin{aligned}
\|\dot{h}_l - \dot{h}_j\|_{L_1(\mathbb{R})} &\leq \|\tilde{h} - \dot{h}_l\|_{L_1(\mathbb{R})} + \|\tilde{h} - \dot{h}_j\|_{L_1(\mathbb{R})} \\
&< 2\delta_n \\
&< \frac{2}{3}\epsilon_{M_n} \\
&= \frac{2}{3} \min_{\substack{j,k \in \{1,\dots,M_n\} \\ \wedge \dot{h}_j \neq \dot{h}_k}} \|\dot{h}_j - \dot{h}_k\|_{L_1(\mathbb{R})} \\
&\leq \frac{2}{3}\|\dot{h}_l - \dot{h}_j\|_{L_1(\mathbb{R})}
\end{aligned}$$

Die positive Definitheit einer Norm reicht aus, um hier $\|\dot{h}_l - \dot{h}_j\|_{L_1(\mathbb{R})} = 0$ und somit $\dot{h}_l = \dot{h}_j$ folgern zu können. Daraus folgt selbstverständlich auch $\Phi(\dot{h}_l) = \Phi(\dot{h}_j)$. (3.16) genügt also, um die Wohldefiniertheit des Schätzers zu sichern.

Nun sei $h \in \mathcal{H}$ fest, aber beliebig und die Risikoabschätzung beginnt

$$E_h\, d(\hat{f}_n, \Phi(h))^k$$

(Konditionierung)

$$\begin{aligned}
&= \sum_{j=1}^{M_n} E_h\big(d(\hat{f}_n, \Phi(h))^k \mid \|\hat{h}_n - \dot{h}_j\|_{L_1(\mathbb{R})} < \delta_n\big) \cdot P_h(\|\hat{h}_n - \dot{h}_j\|_{L_1(\mathbb{R})} < \delta_n) \\
&\quad + E_h\big(d(\hat{f}_n, \Phi(h))^k \mid \|\hat{h}_n - \dot{h}_j\|_{L_1(\mathbb{R})} \geq \delta_n,\ \forall j = 1,\dots,M_n\big) \\
&\qquad \cdot P_h(\|\hat{h}_n - \dot{h}_j\|_{L_1(\mathbb{R})} \geq \delta_n,\ \forall j = 1,\dots,M_n) \\
&= \sum_{j=1}^{M_n} d(\Phi(\dot{h}_j), \Phi(h))^k \cdot P_h(\|\hat{h}_n - \dot{h}_j\|_{L_1(\mathbb{R})} < \delta_n) \\
&\quad + E_h\big(d(\Phi(\hat{h}_n), \Phi(h))^k \mid \|\hat{h}_n - \dot{h}_j\|_{L_1(\mathbb{R})} \geq \delta_n,\ \forall j = 1,\dots,M_n\big) \\
&\qquad \cdot P_h(\|\hat{h}_n - \dot{h}_j\|_{L_1(\mathbb{R})} \geq \delta_n,\ \forall j = 1,\dots,M_n)
\end{aligned}$$

An dieser Stelle erfolgt eine Fallunterscheidung. Zunächst wird der Fall betrachtet, dass h in $\mathcal{H}_0$ liegt. Dann gibt es ein $l \in \mathbb{N}$, so dass $h = \dot{h}_l$ gilt. M_n sollte groß genug sein, damit der Schätzer $\hat{f}_n$ die Unstetigkeitsstelle $\dot{h}_l$

erfassen kann, also soll $M_n > l$ gelten. Daraus begründet sich die zweite Forderung

$$M_n \stackrel{n\to\infty}{\longrightarrow} \infty. \tag{3.17}$$

Da das asymptotische Verhalten des Risikos untersucht wird, kann $n > N$ angenommen werden, wobei N wegen (3.17) groß genug gewählt werden kann, so dass $M_n \geq l$ für alle $n > N$ gilt. Kehren wir zur Risikoabschätzung im diesem Fall zurück

$$\begin{aligned}
&\sum_{j=1}^{M_n} d(\Phi(\dot{h}_j), \Phi(h))^k \cdot P_h(\|\hat{h}_n - \dot{h}_j\|_{L_1(\mathbb{R})} < \delta_n) \\
&+ E_h\big(d(\Phi(\hat{h}_n), \Phi(h))^k \mid \|\hat{h}_n - \dot{h}_j\|_{L_1(\mathbb{R})} \geq \delta_n,\ \forall j = 1, \ldots, M_n\big) \\
&\qquad \cdot P_h(\|\hat{h}_n - \dot{h}_j\|_{L_1(\mathbb{R})} \geq \delta_n,\ \forall j = 1, \ldots, M_n) \\
&= \sum_{\substack{j=1 \\ \wedge \dot{h}_j \neq \dot{h}_l}}^{M_n} d(\Phi(\dot{h}_j), \Phi(\dot{h}_l))^k \cdot P_{\dot{h}_l}(\|\hat{h}_n - \dot{h}_j\|_{L_1(\mathbb{R})} < \delta_n) \\
&+ E_h\big(d(\Phi(\hat{h}_n), \Phi(\dot{h}_l))^k \mid \|\hat{h}_n - \dot{h}_j\|_{L_1(\mathbb{R})} \geq \delta_n,\ \forall j = 1, \ldots, M_n\big) \\
&\qquad \cdot P_{\dot{h}_l}(\|\hat{h}_n - \dot{h}_j\|_{L_1(\mathbb{R})} \geq \delta_n,\ \forall j = 1, \ldots, M_n).
\end{aligned}$$

Da Φ nach $\mathcal{F}$ abbildet, kann man $d(\Phi(\hat{h}_n), \Phi(\dot{h}_l))$ und $d(\Phi(\dot{h}_j), \Phi(\dot{h}_l))$ nach oben durch C abschätzen. Damit lässt sich die obige Ungleichungskette fortsetzen

$$\begin{aligned}
&\leq C^k \cdot \sum_{\substack{j=1 \\ \wedge \dot{h}_j \neq \dot{h}_l}}^{M_n} P_{\dot{h}_l}(\|\hat{h}_n - \dot{h}_j\|_{L_1(\mathbb{R})} < \delta_n) \\
&\quad + C^k \cdot P_{\dot{h}_l}(\|\hat{h}_n - \dot{h}_j\|_{L_1(\mathbb{R})} \geq \delta_n\,,\ \forall\, j \in \{1, \ldots, M_n\}) \\
&\leq C^k \sum_{\substack{j=1 \\ \wedge \dot{h}_j \neq \dot{h}_l}}^{M_n} P_{\dot{h}_l}(\|\hat{h}_n - \dot{h}_j\|_{L_1(\mathbb{R})} < \delta_n) + C^k \cdot P_{\dot{h}_l}(\|\hat{h}_n - \dot{h}_l\|_{L_1(\mathbb{R})} \geq \delta_n),
\end{aligned}$$

wobei in diesem Schritt (3.17) in der Form $M_n \geq l$ gebraucht wird. Durch die Forderung (3.16) ist gewährleistet, dass die Mengen $\{\omega \in \Omega \mid \|\hat{h}_n(\omega) -$

$\dot{h}_j\|_{L_1(\mathbb{R})} < \delta_n\}$ für $j \in \{1,\ldots,M_n\}$ disjunkt sind und aufgrund der Injektivität der Indizierung der Menge $\mathcal{H}_0$ folgt mit der Additivität von Maßen

$$C^k \cdot \sum_{\substack{j=1\\ \wedge \dot{h}_j \neq \dot{h}_l}}^{M_n} P_{\dot{h}_l}(\|\hat{h}_n - \dot{h}_j\|_{L_1(\mathbb{R})} < \delta_n) + C^k \cdot P_{\dot{h}_l}(\|\hat{h}_n - \dot{h}_l\|_{L_1(\mathbb{R})} \geq \delta_n)$$

$$\begin{aligned} &= C^k \cdot P_{\dot{h}_l}(\|\hat{h}_n - \dot{h}_l\|_{L_1(\mathbb{R})} \geq \delta_n) \\ &\quad + C^k \cdot P_{\dot{h}_l}\Big(\bigcup_{\substack{j=1\\ \wedge j \neq l}}^{M_n} \{\omega \in \Omega \mid \|\hat{h}_n(\omega) - \dot{h}_j\|_{L_1(\mathbb{R})} < \delta_n\}\Big). \end{aligned}$$

Für ein $j \in \{1,\ldots,M_n\}\backslash\{l\}$ gelte $\|\hat{h}_n - \dot{h}_j\|_{L_1} < \delta_n$, daraus folgt mit der Dreiecksungleichung und Forderung (3.16) für alle $\omega \in \Omega$.

$$\begin{aligned} \|\hat{h}_n(\omega) - \dot{h}_l\|_{L_1(\mathbb{R})} &\geq \|\dot{h}_l - \dot{h}_j\|_{L_1(\mathbb{R})} - \|\hat{h}_n(\omega) - \dot{h}_j\|_{L_1(\mathbb{R})} \\ &\geq \epsilon_{M_n} - \|\hat{h}_n(\omega) - \dot{h}_j\|_{L_1(\mathbb{R})} \\ &\geq \epsilon_{M_n} - \delta_n \\ &\geq 3\delta_n - \delta_n = 2\delta_n > \delta_n. \end{aligned}$$

Daher ist also $\{\omega \in \Omega \mid \|\hat{h}_n(\omega) - \dot{h}_j\|_{L_1(\mathbb{R})} < \delta_n\}$ Teilmenge von $\{\omega \in \Omega \mid \|\hat{h}_n(\omega) - \dot{h}_l\|_{L_1(\mathbb{R})} \geq \delta_n\}$ für jedes $j \in \{1,\ldots,M_n\}\backslash\{l\}$. Somit inkludiert auch die Vereinigung dieser Mengen und mit der Monotonie eines Maßes geht nun die Risikoabschätzung weiter unter Verwendung der Markov'schen Ungleichung

$$
\begin{aligned}
& C^k \cdot P_{\dot{h}_l}(\|\hat{h}_n - \dot{h}_l\|_{L_1(\mathbb{R})} \ \geq \ \delta_n) \\
& \qquad + C^k \cdot P_{\dot{h}_l}\Big(\bigcup_{\substack{j=1 \\ \wedge j \neq l}}^{M_n} \{\omega \in \Omega \mid \|\hat{h}_n(\omega) - \dot{h}_j\|_{L_1(\mathbb{R})} < \delta_n\}\Big) \\
& \leq \ 2\, C^k \cdot P_{\dot{h}_l}(\|\hat{h}_n - \dot{h}_l\|_{L_1(\mathbb{R})} \ \geq \ \delta_n) \\
& \leq \ 2\, C^k\, \delta_n^{-1}\, E_{\dot{h}_l} \|\hat{h}_n - \dot{h}_l\|_{L_1(\mathbb{R})} \\
& \leq \ 2\, C^k\, \delta_n^{-1} \max_{j=1,\ldots,M_n} E_{\dot{h}_j} \|\hat{h}_n - \dot{h}_j\|_{L_1(\mathbb{R})}
\end{aligned}
$$

Dieser Ausdruck soll gegen 0 konvergieren, dann ist die Konsistenzaussage im ersten Fall bewiesen, dies ist die dritte Forderung

$$
\delta_n^{-1} \max_{j=1,\ldots,M_n} E_{\dot{h}_j} \|\hat{h}_n - \dot{h}_j\|_{L_1(\mathbb{R})} \stackrel{n\to\infty}{\longrightarrow} 0. \tag{3.18}
$$

Kommen wir jetzt zum zweiten Fall $(h \in \mathcal{H}\backslash\mathcal{H}_0)$. Φ ist also $(L_1(\mathbb{R}), d)$-stetig in h. Kehren wir in der Risikoabschätzung zu der Stelle zurück, an der die beiden Fälle unterschieden wurden:

$$
\begin{aligned}
& \sum_{j=1}^{M_n} d(\Phi(\dot{h}_j), \Phi(h))^k \cdot P_h(\|\hat{h}_n - \dot{h}_j\|_{L_1(\mathbb{R})} < \delta_n) \\
& + E_h\big(d(\Phi(\hat{h}_n), \Phi(h))^k \mid \|\hat{h}_n - \dot{h}_j\|_{L_1(\mathbb{R})} \geq \delta_n\,,\ \forall j = 1, \ldots, M_n\big) \\
& \qquad P_h(\|\hat{h}_n - \dot{h}_j\|_{L_1(\mathbb{R})} \geq \delta_n\,,\ \forall j = 1, \ldots, M_n) \\
& \leq \ \sum_{j=1}^{M_n} d(\Phi(\dot{h}_j), \Phi(h))^k \cdot P_h(\|\hat{h}_n - \dot{h}_j\|_{L_1(\mathbb{R})} < \delta_n) \\
& + E_h d(\Phi(\hat{h}_n), \Phi(h))^k
\end{aligned}
$$

$E_h d(\Phi(\hat{h}_n), \Phi(h))^k$ konvergiert gegen 0 wegen der Stetigkeit von Φ in h, dies lässt sich analog der Beweisführung von Satz 3.3 bzw. Korollar 3.1 zeigen (siehe Satz 3.3, Anmerkung (b)). Damit gilt es nun noch zu zeigen, dass der Summenterm gegen 0 strebt. Dazu wähle man ein $\varepsilon > 0$ beliebig. Wegen der Stetigkeit von Φ existiert ein $\eta(h,\varepsilon) > 0$, so dass $\|\tilde{h} - h\| < \eta(h,\varepsilon)$ $d(\Phi(\tilde{h}), \Phi(h)) < \varepsilon$ impliziert. Damit folgt

$$\sum_{j=1}^{M_n} d(\Phi(\dot h_j),\Phi(h))^k \cdot P_h(\|\hat h_n - \dot h_j\|_{L_1(\mathbb{R})} < \delta_n)$$

$$= \sum_{\substack{j=1\\ \wedge \|\dot h_j - h\|_{L_1(\mathbb{R})} \le \eta(h,\varepsilon)}}^{M_n} \underbrace{d(\Phi(\dot h_j),\Phi(h))^k}_{\le \varepsilon^k} \cdot P_h(\|\hat h_n - \dot h_j\|_{L_1(\mathbb{R})} < \delta_n)$$

$$+ \sum_{\substack{j=1\\ \wedge \|\dot h_j - h\|_{L_1(\mathbb{R})} > \eta(h,\varepsilon)}}^{M_n} d(\Phi(\dot h_j),\Phi(h))^k \cdot P_h(\|\hat h_n - \dot h_j\|_{L_1(\mathbb{R})} < \delta_n)$$

$$\le \varepsilon^k \cdot \sum_{j=1}^{M_n} P_h(\|\hat h_n - \dot h_j\|_{L_1(\mathbb{R})} < \delta_n)$$

$$+ C^k \cdot \sum_{\substack{j=1\\ \wedge \|\dot h_j - h\|_{L_1(\mathbb{R})} > \eta(h,\varepsilon)}}^{M_n} P_h(\|\hat h_n - \dot h_j\|_{L_1(\mathbb{R})} < \delta_n).$$

Wegen der bereits im ersten Fall erwähnten Disjunktheit der Mengen $\{\omega \in \Omega \mid \|\hat h_n(\omega) - \dot h_j\|_{L_1} < \delta_n\}$ folgt weiter

$$\le \varepsilon^k P_h\Big(\bigcup_{j=1}^{M_n} \{\omega \in \Omega \mid \|\hat h_n - \dot h_j\|_{L_1(\mathbb{R})} < \delta_n\}\Big)$$

$$+ C^k \cdot P_h\Big(\bigcup_{\substack{j=1\\ \wedge \|\dot h_j - h\|_{L_1(\mathbb{R})} > \eta(h,\varepsilon)}}^{M_n} \{\omega \in \Omega \mid \|\hat h_n(\omega) - \dot h_j\|_{L_1(\mathbb{R})} < \delta_n\}\Big).$$

Die erste hier auftretende Wahrscheinlichkeit wird grob gegen 1 abgeschätzt, für die zweite kann man sich überlegen: Für ein ω aus dieser Vereinigungsmenge existiert ein $j \in \{1,\dots,M_n\}$ mit $\|\dot h_j - h\|_{L_1(\mathbb{R})} > \eta(h,\varepsilon)$, so dass

$$\delta_n + \|\hat h_n(\omega) - h\|_{L_1} \ \ge\ \|\hat h_n(\omega) - h\|_{L_1} + \|\hat h_n(\omega) - \dot h_j\|_{L_1} \ \ge\ \|\dot h_j - h\|_{L_1} \ >\ \eta(h,\varepsilon)$$

Daraus folgt $\|\hat h_n(\omega) - h\|_{L_1} \ \ge\ \eta(h,\varepsilon) - \delta_n \ \ge\ \eta(h,\varepsilon)/2 \ >\ 0$, wenn $\delta_n < \eta(h,\varepsilon)/2$ gilt. Es kann garantiert werden, dass dies für alle n ab einer gewissen natürlichen Zahl gilt, wenn die vierte Forderung

$$\delta_n \overset{n\to\infty}{\longrightarrow} 0 \tag{3.19}$$

postuliert wird. Daraus folgt also für ausreichend großes n

$$\bigcup_{\substack{j=1 \\ \wedge \|\dot{h}_j - h\|_{L_1(\mathbb{R})} > \eta(h,\varepsilon)}}^{M_n} \{\omega \in \Omega \mid \|\hat{h}_n(\omega) - \dot{h}_j\|_{L_1(\mathbb{R})} < \delta_n\} \subseteq \{\omega \in \Omega \mid \|\hat{h}_n - h\|_{L_1} \geq \eta(h,\varepsilon)/2\}$$

und somit die Abschätzung

$$P_h\Big(\bigcup_{\substack{j=1 \\ \wedge \|\dot{h}_j - h\|_{L_1(\mathbb{R})} > \eta(h,\varepsilon)}}^{M_n} \{\omega \in \Omega \mid \|\hat{h}_n(\omega) - \dot{h}_j\|_{L_1(\mathbb{R})} < \delta_n\}\Big) \leq P_h(\|\hat{h}_n - h\|_{L_1} \geq \eta(h,\varepsilon)/2).$$

Kehren wir wieder zur Risikoabschätzung zurück

$$\begin{aligned}
&\varepsilon^k P_h\Big(\bigcup_{j=1}^{M_n}\{\omega \in \Omega \mid \|\hat{h}_n - \dot{h}_j\|_{L_1(\mathbb{R})} < \delta_n\}\Big) \\
&+ C^k \cdot P_h\Big(\bigcup_{\substack{j=1 \\ \wedge \|\dot{h}_j - h\|_{L_1(\mathbb{R})} > \eta(h,\varepsilon)}}^{M_n} \{\omega \in \Omega \mid \|\hat{h}_n(\omega) - \dot{h}_j\|_{L_1(\mathbb{R})} < \delta_n\}\Big) \\
&\leq \varepsilon^k + C^k \cdot P_h(\|\hat{h}_n - h\|_{L_1(\mathbb{R})} \geq \eta(h,\varepsilon)/2) \\
&\textit{(Markov'sche Ungleichung)} \\
&\leq \varepsilon^k + C^k \tfrac{2}{\eta(h,\varepsilon)} \underbrace{E_h\|\hat{h}_n - h\|_{L_1(\mathbb{R})}}_{\underset{n\to\infty}{\longrightarrow} 0} \\
&\leq 2\varepsilon^k
\end{aligned}$$

für alle n, die größer als eine gewisse von ϵ und h abhängende natürliche Zahl sind. Damit ist auch der zweite Fall bewiesen und die robuste d^k-Konsistenz des Dichteschätzers $(\hat{f}_n)_n$ unter dem Vorbehalt der Erfüllbarkeit der vier Forderungen gezeigt. Es gilt jetzt also noch zu klären, ob es eine geeignete Wahl der Parameterfolgen $(M_n)_n$ und $(\delta_n)_n$ gibt, die (3.16), (3.17), (3.18) und (3.19) genügt. Dazu betrachten wir

$$b_{m,n} := \max_{j\in\{1,\ldots,m\}} E_{\dot{h}_j}\|\hat{h}_n - \dot{h}_j\|_{L_1(\mathbb{R})} \Big/ \min_{\substack{j,k\in\{1,\ldots,m\}\\ \wedge \dot{h}_j\neq \dot{h}_k}} \|\dot{h}_j - \dot{h}_k\|_{L_1(\mathbb{R})} \;\geq\; 0.$$

Aufgrund der individuellen L_1-Konsistenz des Dichteschätzers $(\hat{h}_n)_n$ ist gewährleistet, dass

$$b_{m,n} \overset{n\to\infty}{\longrightarrow} 0 \quad , \quad \forall m \in \mathbb{N}.$$

Nun kann gemäß Lemma 3.5 die Folge $(M_n)_{n\in\mathbb{N}}$ so gewählt werden, dass sie bestimmt divergiert und $b_{M_n,n} \overset{n\to\infty}{\longrightarrow} 0$ gilt. Damit ist die Forderung (3.17) bereits erfüllt. Wenn $\frac{1}{6} \min\limits_{\substack{j,k\in\{1,\ldots,M_n\}\\ \wedge \dot{h}_j\neq \dot{h}_k}} \|\dot{h}_j - \dot{h}_k\|_{L_1(\mathbb{R})}$ gegen 0 konvergiert, wählt man $(\delta_n)_n$ als eben diese Folge. Dann sind (3.16) und (3.19) erfüllt. Zudem ist

$$\delta_n^{-1} \max_{j\in\{1,\ldots,M_n\}} E_{\dot{h}_j}\|\hat{h}_n - \dot{h}_j\|_{L_1(\mathbb{R})} \;=\; 6 b_{M_n,n} \overset{n\to\infty}{\longrightarrow} 0$$

und damit (3.18) erfüllt. Gilt dagegen $\frac{1}{6} \min\limits_{\substack{j,k\in\{1,\ldots,m\}\\ \wedge \dot{h}_j\neq \dot{h}_k}} \|\dot{h}_j - \dot{h}_k\|_{L_1(\mathbb{R})} \geq K > 0$ für alle $m \in \mathbb{N}$, so wähle man $\delta_n := \min\{(b_{M_n,n})^{1/2}, K/2\}$. Dadurch ist (3.19) erfüllt und auch (3.16) gilt wegen

$$0 \;<\; \delta_n \;<\; 2K \;\leq\; \frac{1}{3} \min_{\substack{j,k\in\{1,\ldots,M_n\}\\ \wedge \dot{h}_j\neq \dot{h}_k}} \|\dot{h}_j - \dot{h}_k\|_{L_1(\mathbb{R})}.$$

Weiterhin konvergiert $\delta_n^{-1} b_{M_n,n} \;=\; \sqrt{b_{M_n,n}}$, wenn n groß genug ist, gegen 0. Da $\min\limits_{\substack{j,k\in\{1,\ldots,M\}\\ \wedge \dot{h}_j\neq \dot{h}_k}} \|\dot{h}_j - \dot{h}_k\|_{L_1(\mathbb{R})}$ monoton fallend bezüglich M

und somit nach oben beschränkt ist, konvergiert auch $\delta_n^{-1} \max_{j\in\{1,\ldots,M_n\}} E_{\dot{h}_j} \|\hat{h}_n - \dot{h}_j\|_{L_1(\mathbb{R})}$ gegen 0 und (3.18) ist erfüllt. Damit ist gezeigt, wie die beiden Parameterfolgen zu wählen sind, um die robuste d^k-Konsistenz von $(\hat{f}_n)_{n\in\mathbb{N}}$ zu erhalten.

Zu guter Letzt betrachten wir noch den Fall der Endlichkeit der Menge $\mathcal{H}_0 = \{\dot{h}_j \mid j \in \{1,\ldots,M\}\}$. Auch hier sei die Indizierung injektiv. Dann wird auf die Forderung (3.17) verzichtet und die Folge M_n konstant M, der Mächtigkeit von $\mathcal{H}_0$, gesetzt und

$$\delta_n := \min\Big\{\big(\max_{j\in\{1,\ldots,M\}} E_{\dot{h}_j}\|\hat{h}_n - \dot{h}_j\|_{L_1(\mathbb{R})}\big)^{1/2}, \frac{1}{6} \min_{\substack{j,k\in\{1,\ldots,M\}\\ \wedge \dot{h}_j \neq \dot{h}_k}} \|\dot{h}_j - \dot{h}_k\|_{L_1(\mathbb{R})}\Big\}$$

Dann sind auch die übrigen drei Forderungen erfüllt und die gesamte Risikoabschätzung ist gültig.

■

Nachdem die $(L_1(\mathbb{R}), d)$-Stetigkeit von Φ durch Satz 3.4 als nicht-notwendiges Kriterium für die robust d^k-konsistente Schätzbarkeit ausgewiesen wurde, stellt sich folglich die Frage, ob die in Satz 3.1 als notwendiges Kriterium erkannte Bedingung des Nicht-Überlappens auch schon hinreichend für die robust konsistente Schätzbarkeit ist. Wir betrachten ab jetzt keinen allgemeinen metrischen Raum (X, d) mehr, sondern einen reflexiven separablen Banachraum X mit Norm $\|\cdot\|_X$ oder $(L_1(\mathbb{R}), \|\cdot\|_{L_1})$. Unter erstere Kategorie fallen die $L_p(\mathbb{R})$-Räume für $p > 1$. (siehe z.B. *Werner, D.* [15])

Satz 3.5 *Sei $(X, \|\cdot\|_X)$ ein reflexiver separabler Banachraum oder $(L_1(\mathbb{R}), \|\cdot\|_{L_1})$. Sei $\mathcal{F} \subseteq X \cap \mathcal{D}$. Sei $\mathcal{G} \subseteq \mathcal{D}$. Sei $\sup_{f\in\mathcal{F}} \|f\|_X \leq C$ (im Falle $X = L_1(\mathbb{R})$ setze man $C = 1$). Es existiere eine robust X^k-konsistente Schätzersequenz $(\hat{f}_n)_n$ mit $k \geq 1$.*
(a) Dann kann der Schätzer $(\hat{f}_n)_n$ oBdA als $\mathcal{F}$-immanent angenommen werden.
(b) Dann gibt es eine Folge $(\Phi_m)_{m\in\mathbb{N}}$ von Abbildungen $\mathcal{H}$ nach $\{x \in X \mid \|x\|_X \leq C\}$, die gleichmäßig $(L_1, \|\cdot\|_X)$-stetig sind und die in (3.7) definierte Abbildung Φ punktweise approximieren, d.h.

$$\|\Phi(h) - \Phi_m(h)\|_X \stackrel{m\to\infty}{\longrightarrow} 0 \quad , \quad \forall h \in \mathcal{H}.$$

Die gleichmäßige Stetigkeit lässt sich durch die Ungleichung

$$\|\Phi_m(h) - \Phi_m(\tilde{h})\|_X \leq C \cdot m \cdot \|h - \tilde{h}\|_{L_1(\mathbb{R})} \,,\ \forall h, \tilde{h} \in \mathcal{H} \tag{3.20}$$

verschärfen.

Diese Eigenschaft, welche die Schlussfolgerung in Satz 3.5 ist, möchte ich dergestalt bezeichnen, dass Φ *gleichmäßig* X *-stetig approximierbar* sei.

Beweis: Die Abbildung Φ existiert aufgrund der vorausgesetzten robust konsistenten Schätzbarkeit (Satz 3.1). ad (a): Für $\hat{f}_n$ kann oBdA $\|\hat{f}_n\|_X \leq C$ vorausgesetzt werden, denn wie in Satz 3.3 können für einen nach Voraussetzung existierenden robust X^k-konsistenten Schätzer $\tilde{f}_n$ die nicht leeren Mengen

$$F(n, y_1, \ldots, y_n) := \{\tilde{f} \in \mathcal{F} \mid \|\tilde{f} - \tilde{f}_n\|_X^k \leq \inf_{f \in \mathcal{F}} \|f - \tilde{f}_n\|_X^k + o_n\}$$

mit einer positiven Folge $(o_n)_n \downarrow 0$ definiert werden, aus denen eine beliebige Auswahlfunktion $\hat{f}_n(y_1, \ldots, y_n)$ als $\mathcal{F}$-immanenten Schätzer herausgegriffen werden kann. Diese ist auch robust X^k-konsistent, denn

$$\begin{aligned}
E_{f,g}\|\hat{f}_n - f\|_X^k &\leq (\min\{1, 2^{1-k}\})^{-1} \cdot \Big(E_{f,g}\|\tilde{f}_n - \hat{f}_n\|_X^k + \|f - \tilde{f}_n\|_X^k\Big) \\
&\leq (\min\{1, 2^{1-k}\})^{-1} \cdot \Big(E_{f,g} \inf_{\tilde{f} \in \mathcal{F}} \|\tilde{f}_n - \tilde{f}\|_X^k + E_{f,g}\|f - \tilde{f}_n\|_X^k + o_n\Big) \\
&\leq (\min\{1, 2^{1-k}\})^{-1} \cdot \Big(2\, E_{f,g}\|\tilde{f}_n - \tilde{f}\|_X^k + o_n\Big) \\
&\stackrel{n \to \infty}{\longrightarrow} 0
\end{aligned}$$

ad (b): Sei X ein reflexiver separabler Banachraum. Eine Bochner-integrable Zufallsvariable $\hat{x}$ mit Werten in X (d.h. $E\|\hat{x}\|_X < \infty$) ist auch schwach integrierbar und es gilt $\|E\hat{x}\|_X \leq E\|\hat{x}\|_X$. Zu der dahintersteckenden Theorie kann das Buch von *Ledoux und Talagrand* [13] herangezogen werden. Das schwache Integral $E\hat{x}$ wird dabei als Element des Bidualraumes von X gedeutet, die Reflexivität von X garantiert die Mitgliedschaft von $E\hat{x}$ in X. Mit diesem Hintergrund wird gefolgert

$$
\begin{aligned}
& E_h \|\hat{f}_n - \Phi(h)\|_X^k \overset{n\to\infty}{\longrightarrow} 0 \\
& \textit{(Jensen'sche Ungleichung)} \\
& \geq (E_h \|\hat{f}_n - \Phi(h)\|_X)^k \\
& \geq \|E_h \hat{f}_n - \Phi(h)\|_X^k \\
& \textit{(da } \Phi(h) \textit{ deterministisch ist)} \\
& = \|\Phi_n(h) - \Phi(h)\|_X
\end{aligned}
$$

mit $\Phi_n(h) := E_h \hat{f}_n$ für jedes $h \in \mathcal{H}$. Φ_n bildet somit von $\mathcal{H}$ nach $\{x \in X \mid \|x\| \leq C\}$ ab, da $\|E\hat{f}_n\|_X \leq E\|\hat{f}_n\|_X \leq C$ ist. $\Phi_n(h) = E_h(\hat{f}(n))$ ist nach Definition des schwachen Integrals die Abbildung $x' \mapsto \int \cdots \int x'(\hat{f}_n(y_1, \ldots, y_n)) h(y_1) \cdots h(y_n)\, dy_1 \cdots dy_n$, wobei x' ein Element des Dualraumes von X sei. Entsprechend ist

$$
\begin{aligned}
& |\Big(\Phi_n(h) - \Phi_n(\tilde{h})\Big)(x')| \\
& = |\int \cdots \int x'(\hat{f}_n(y_1, \ldots, y_n))\, (h(y_1) \cdots h(y_n) - \tilde{h}(y_1) \cdots \tilde{h}(y_n))\, dy_1 \cdots dy_n| \\
& \leq \int \cdots \int |x'(\hat{f}_n(y_1, \ldots, y_n))|\, |h(y_1) \cdots h(y_n) - \tilde{h}(y_1) \cdots \tilde{h}(y_n)|\, dy_1 \cdots dy_n \\
& \leq \int \cdots \int \|x'\|_{X'} \|\hat{f}_n(y_1, \ldots, y_n)\|_X\, |h(y_1) \cdots h(y_n) - \tilde{h}(y_1) \cdots \tilde{h}(y_n)| \\
& \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad dy_1 \cdots dy_n \\
& \leq C \cdot \|x'\|_{X'} \cdot \int \cdots \int |h(y_1) \cdots h(y_n) - \tilde{h}(y_1) \cdots \tilde{h}(y_n)|\, dy_1 \cdots dy_n \\
& \leq C \cdot \|x'\|_{X'} \cdot n \cdot \|h - \tilde{h}\|_{L_1(\mathbb{R})}
\end{aligned}
$$

Für die Bidualnorm von X, die wegen der Reflexivität von X, d.h. der isometrischen Isomorphie von X zu X'' zu $\|\cdot\|_X$ äquivalent ist, gilt

$$
\begin{aligned}
& \|\Phi_n(h) - \Phi_n(\tilde{h})\|_X = \|\Phi_n(h) - \Phi_n(\tilde{h})\|_{X''} \\
& = \sup_{\|x'\|_{X'}=1} |\Big(\Phi_n(h) - \Phi_n(\tilde{h})\Big)(x')| \leq C \cdot n \cdot \|h - \tilde{h}\|_{L_1(\mathbb{R})}
\end{aligned}
$$

nach obiger Ungleichung. Damit ist die gleichmäßige $(L_1(\mathbb{R}), X)$-Stetigkeit gezeigt (allerdings nicht gleichgradig in n).
Schließen wir jetzt noch den Fall, dass es sich bei X um den separablen nicht reflexiven Banachraum $L_1(\mathbb{R})$ handelt, an. Hier gilt

$$E_h\|\hat{f}_n - \Phi(h)\|_{L_1}^k \overset{n\to\infty}{\longrightarrow} 0$$

(Jensen'sche Ungleichung)

$$\geq (E_h\|\hat{f}_n - \Phi(h)\|_{L_1})^k$$
$$= \Big(\int\cdots\int\int|\hat{f}_n(t,y_1,\ldots,y_n) - \big(\Phi(h)\big)(t)|dth(y_1)\cdots h(y_n)dy_1\cdots dy_n\Big)^k$$

(Satz von Fubini)

$$= \Big(\int\int\cdots\int|\hat{f}_n(t,y_1,\ldots,y_n) - \big(\Phi(h)\big)(t)|h(y_1)\cdots h(y_n)dy_1\cdots dy_n dt\Big)^k$$

$$\geq \Big(\int|\int\cdots\int(\hat{f}_n(t,y_1,\ldots,y_n) - \big(\Phi(h)\big)(t))h(y_1)\cdots h(y_n)dy_1\cdots dy_n|dt\Big)^k$$

$$= \Big(\int|\big(\int\cdots\int\hat{f}_n(t,y_1,\ldots,y_n)h(y_1)\cdots h(y_n)dy_1\cdots dy_n - \big(\Phi(h)\big)(t)\big)|dt\Big)^k$$

$$= (\|\underbrace{\int\cdots\int\hat{f}_n(\bullet,y_1,\ldots,y_n)h(y_1)\cdots h(y_n)dy_1\cdots dy_n}_{=:\Phi_n(h)} - \Phi(h)\|_{L_1(\mathbb{R})})^k.$$

Φ_n bildet somit von $\mathcal{H}$ in die Einheitssphäre von $L_1(\mathbb{R})$ ab (Bei der L_1-Betrachtung kann stets $C = 1$ vorausgesetzt werden). Die gleichmäßige Stetigkeit zeigt man schließlich auf weniger abstrakte Weise als im vorherigen Fall der reflexiven Räume:

$$\|\Phi_n(h) - \Phi_n(\tilde{h})\|_{L_1(\mathbb{R})}$$

$$= \int|\int\cdots\int\hat{f}_n(t,y_1,\ldots,y_n)(h(y_1)\cdots h(y_n) - \tilde{h}(y_1)\cdots\tilde{h}_{y_n})dy_1\cdots dy_n|dt$$

$$\leq \int\int\cdots\int|\hat{f}_n(t,y_1,\ldots,y_n)||h(y_1)\cdots h(y_n) - \tilde{h}(y_1)\cdots\tilde{h}(y_n)|dy_1\cdots dy_n dt$$

(Satz von Fubini)

$$= \int\cdots\int\|\hat{f}_n(\bullet,y_1,\ldots,y_n)\|_{L_1(\mathbb{R})}|h(y_1)\cdots h(y_n) - \tilde{h}(y_1)\cdots\tilde{h}(y_n)|dy_1\cdots dy_n$$

$$\leq n\cdot\|h - \tilde{h}\|_{L_1(\mathbb{R})}.$$

■

Nachdem die gleichmäßige X-stetige Approximierbarkeit von Φ als notwendige Bedingung der robusten X-konsistenten Schätzbarkeit nun gesichert ist, gebe ich mit diesem theoretischen Hintergrund folgendes Beispiel:

Zunächst wird auf abstrakte Weise eine reelle Teilmenge $\mathcal{R}$ konstruiert. Es ist leicht zu erkennen, dass $(\mathbb{Q},+)$ eine Untergruppe der abelschen Gruppe $(\mathbb{R},+)$ ist. Daher existieren eine Äquivalenzrelation

$$x \sim y \;:\Leftrightarrow\; x-y \in \mathbb{Q}$$

auf $\mathbb{R}$ und die induzierte Partition von $\mathbb{R}$ in Äquivalenzklassen von $\sim$.

$$\{M_i \mid i \in I\}, \quad I \text{ nicht leere Indexmenge}$$

$$M_i \neq \emptyset\,, \quad \forall i \in I$$

sei diese Partition. Man definiere weiter

$$A_i := M_i \cap [0,1]\,, \quad \forall i \in I.$$

Dabei ist zu beachten, dass $A_i \neq \emptyset$ ist, denn für jedes $x \in M_i$ existiert ein $q \in \mathbb{Q}$, so dass $x+q \in M_i \cap [0,1]$ gilt, da $\mathbb{Q}$ dicht in $\mathbb{R}$ liegt.
Nach dem Auswahlaxiom existiert eine Abbildung

$$f \,:\, I \longrightarrow \bigcup_{i \in I} A_i \subseteq [0,1]$$

mit $f(i) \in A_i, \forall i \in I$.
f ist injektiv, denn

$$i \neq j \wedge f(i) = f(j)$$

$$\Rightarrow\; f(i) \in A_i \wedge f(i) = f(j) \in A_j$$

$$\Rightarrow\; f(i) \in A_i \cap A_j \;\subseteq\; M_i \cap M_j = \emptyset \text{ , da } i \neq j$$

Widerspruch!

Doch es gilt noch mehr:

$$\begin{aligned} f(i) \sim f(j) &\Rightarrow f(i) \in A_i \subseteq M_i \wedge f(i) \sim f(j) \in A_j \subseteq M_j \\ &\Rightarrow f(i) \in M_i \wedge f(i) \in M_j \\ &\Rightarrow M_i \cap M_j \neq \emptyset \\ &\Rightarrow i = j. \end{aligned}$$

Definieren wir schließlich

$$\mathcal{R} := \{f(i) \mid i \in I\} \subseteq [0,1].$$

Mit dieser Definition kann man nun Dichteklassen $\mathcal{F}$ und $\mathcal{G}$ definieren und somit ein spezielles Schätzproblem betrachten:

$$\begin{aligned} \mathcal{F} &:= \{\mathcal{N}(\mu,1) \mid \mu \in \mathcal{R}\} \\ \mathcal{G} &:= \{\mathcal{N}(q,1) \mid q \in \mathbb{Q}\}. \end{aligned} \tag{3.21}$$

Wir stellen fest, dass sich $\mathcal{F}$ und $\mathcal{G}$ nicht überlappen, denn aus der Annahme, es existiere ein $h \in \mathcal{H}$ mit $h = \mathcal{N}(\mu,1) * \mathcal{N}(q,1) = \mathcal{N}(\tilde{\mu},1) * \mathcal{N}(\tilde{q},1)$ und $\mu, \tilde{\mu} \in \mathcal{R}$ sowie $q, \tilde{q} \in \mathbb{Q}$, folgt wegen der Faltungsregel für Normalverteilungsdichten (z.B. *Schmitz, N.* [14],S.194)

$$\begin{aligned} & \mathcal{N}(\mu+q,2) = \mathcal{N}(\tilde{\mu}+\tilde{q},2) \\ \Rightarrow\ & \mu + q = \tilde{\mu} + \tilde{q} \\ \Rightarrow\ & \mu - \tilde{\mu} = \tilde{q} - q \in \mathbb{Q} \\ \Rightarrow\ & \mu \sim \tilde{\mu} \\ & \textit{(wegen } \mu, \tilde{\mu} \in \mathcal{R} \textit{ folgt)} \\ \Rightarrow\ & \mu = f(i) \sim f(j) = \tilde{\mu} \\ \overset{\text{s.o.}}{\Rightarrow}\ & i = j \\ \Rightarrow\ & \mu = \tilde{\mu} \\ \Rightarrow\ & \mathcal{N}(\mu,1) = \mathcal{N}(\tilde{\mu},1). \end{aligned}$$

Nun soll gezeigt werden, dass die unverrauschte Dichte $\mathcal{N}(\mu,1)$ trotz der Existenz der Abbildung Φ nicht robust X^k-konsistent schätzbar ist, wenn wir nun speziell $X = L_2(\mathbb{R})$ wählen. Zunächst erkennt man, dass

$$\mathcal{H} = \mathcal{F} * \mathcal{G} = \{\mathcal{N}(x,2) \mid x \in \mathbb{R}\}$$

gilt. Die Inklusion $\mathcal{H} \subseteq \cdots$ ist nach der oben zitierten Faltungsregel klar. Um die Umkehrung zu zeigen, greifen wir ein beliebiges $\mathcal{N}(x,2)$ mit $x \in \mathbb{R}$ heraus. Da $\{M_i \mid i \in I\}$ Partition von $\mathbb{R}$ ist, existiert ein $i \in I$, so dass $x \in M_i$ gilt. Somit ist $x \sim f(i)$,d.h. $\exists q \in \mathbb{Q} : x = f(i) + q$. Das impliziert

$$\mathcal{N}(x,2) = \underbrace{\mathcal{N}(f(i),1)}_{\in\mathcal{F}} * \underbrace{\mathcal{N}(q,1)}_{\in\mathcal{G}}.$$

und die Inklusion ist gezeigt. Die Abbildung Φ bildet dann $\mathcal{N}(x,2)$ nach $\mathcal{N}(f(i),1)$ ab. Dies induziert eine eindeutig bestimmte Abbildung $\tilde{\Phi}$ von $\mathbb{R}$ nach $\mathcal{R}$, definiert durch

$$\Phi(\mathcal{N}(x,2)) =: \mathcal{N}(\tilde{\Phi}(x),1) \quad \forall x \in \mathbb{R}.$$

Nehmen wir an, $\tilde{\Phi}$ sei in einem Punkt $x \in \mathbb{R}$ stetig. Es ist $y := \tilde{\Phi}(x) \in \mathcal{R}$. Also existiert ein $q \in \mathbb{Q}$, so dass $\mathcal{N}(y,1) * \mathcal{N}(q,1) = \mathcal{N}(x,2)$ und folglich $x = q + y$ gilt. Man wähle nun ein $z \in \mathcal{R}$ mit $z \not\sim y$ aus (Die Existenz mehrerer Äquivalenzklassen bzgl. $\sim$, die aus der Existenz irrationaler Zahlen resultiert, gewährleistet die Wohldefiniertheit von z). Aufgrund der Dichtheit von $\mathbb{Q}$ in $\mathbb{R}$ existiert eine Folge $(q_n)_n$ rationaler Zahlen, die gegen $x - z$ konvergiert. Somit folgt $x_n := z + q_n \overset{n\to\infty}{\longrightarrow} x$. Aufgrund der angenommenen Stetigkeit von $\tilde{\Phi}$ in x folgt $\tilde{\Phi}(x_n) \overset{n\to\infty}{\longrightarrow} \tilde{\Phi}(x) = y$. Da $z \in \mathcal{R}$ liegt und kein weiteres zu z äquivalentes Element in $\mathcal{R}$ liegt, folgt aber $\tilde{\Phi}(x_n) = z$ für alle $n \in \mathbb{N}$ und somit $z = y$, was im Widerspruch zu $z \not\sim y$ steht. Folglich ist gezeigt, dass $\tilde{\Phi}$ nirgends stetig ist.

Nehmen wir weiter an, ein robust L_2^2-konsistenter Schätzer existiere. Dieser kann oBdA als $\mathcal{F}$-immanent (siehe Satz 3.5) angenommen werden und daher in der Form $\hat{f}_n = \mathcal{N}(\hat{\mu}_n, 1)$ mit einem reellwertigen Schätzer $\hat{\mu}_n$, der für alle $x \in \mathbb{R}$ $\hat{\mu}_n \in [0,1]$ fast sicher erfüllt, dargestellt werden. Die robuste L_2^2-Konsistenz des Schätzers bedeutet

$$E_{\mathcal{N}(x,2)} \left\| \mathcal{N}(\hat{\mu}_n, 1) - \mathcal{N}(\tilde{\Phi}(x), 1) \right\|_{L_2(\mathbb{R})} \overset{n\to\infty}{\longrightarrow} 0.$$

Für alle $\mu, \tilde{\mu} \in [0,1]$ gilt die Ungleichung

$$\|\mathcal{N}(\tilde{\mu},1)-\mathcal{N}(\mu,1)\|^2_{L_2(\mathbb{R})}$$

(Parseval-Identität)

$$= \tfrac{1}{2\pi}\int \left|\exp\left(i\mu t-\tfrac{1}{2}t^2\right)-\exp\left(i\tilde{\mu}t-\tfrac{1}{2}t^2\right)\right|^2 dt$$

$$= \tfrac{1}{2\pi}\int \left|\exp(i(\mu-\tilde{\mu})t)-1\right|^2 \exp(-t^2)dt$$

$$\geq \tfrac{1}{\pi}\int(1-\cos((\mu-\tilde{\mu})t))^2\exp(-t^2)dt$$

$$= (\mu-\tilde{\mu})^4\tfrac{1}{\pi}\int \left(\tfrac{1-\cos((\mu-\tilde{\mu})t)}{(\mu-\tilde{\mu})^2t^2}\right)^2 t^4\exp(-t^2)\,dt$$

$$\geq (\mu-\tilde{\mu})^4\tfrac{1}{\pi}\int_{-1}^{1} \left(\tfrac{1-\cos((\mu-\tilde{\mu})t)}{(\mu-\tilde{\mu})^2t^2}\right)^2 t^4\exp(-t^2)\,dt$$

$$\geq (\mu-\tilde{\mu})^4\tfrac{1}{\pi}\inf_{s\in[-1,1]\setminus\{0\}}\left(\tfrac{1-\cos s}{s^2}\right)^2\int_{-1}^{1} t^4\exp(-t^2)\,dt$$

Das Integral $\int_{-1}^{1} t^4\exp(-t^2)\,dt$ ist positiv. Die Funktion $(1-\cos\bullet)/(\bullet^2)$ ist auf $[-1,1]\setminus\{0\}$ stetig und positiv und konvergiert gegen $1/2$, wenn das Argument der Funktion gegen 0 strebt. Damit ist auch $\inf_{s\in[-1,1]\setminus\{0\}}\left(\frac{1-\cos s}{s^2}\right)^2$ gegen eine positive Konstante nach unten beschränkt. Insgesamt folgt

$$\|\mathcal{N}(\tilde{\mu},1)-\mathcal{N}(\mu,1)\|^2_{L_2(\mathbb{R})} \geq c\cdot(\mu-\tilde{\mu})^4$$

mit einem $c>0$ unabhängig von μ, $\tilde{\mu}$. Damit gilt für den MISE für alle $x\in\mathbb{R}$

$$E_{\mathcal{N}(x,2)}\,\|\mathcal{N}(\hat{\mu}_n,1)-\mathcal{N}(\tilde{\Phi}(x),1)\|_{L_2(\mathbb{R})} \stackrel{n\to\infty}{\longrightarrow} 0$$

$$\geq c\cdot E_{\mathcal{N}(x,2)}\,(\hat{\mu}_n-\tilde{\Phi}(x))^4$$

(Jensen'sche Ungleichung)

$$\geq c\cdot((E_{\mathcal{N}(x,2)}\hat{\mu}_n)-\tilde{\Phi}(x))^4.$$

Man definiere $\tilde{\Phi}_n(x) := E_{\mathcal{N}(x,2)}\hat{\mu}_n$. Nach der oben stehenden Aussage konvergiert die Funktionenfolge $(\tilde{\Phi}_n)_n$ punktweise überall gegen $\tilde{\Phi}$. Die Funktionen $\tilde{\Phi}_n$ sind individuell stetig (nicht gleichgradig bezüglich n), da

$$
\begin{aligned}
&|\tilde{\Phi}_n(x) - \tilde{\Phi}_n(y)| \\
&\le |E_{\mathcal{N}(x,2)}\hat{\mu}_n - E_{\mathcal{N}(y,2)}\hat{\mu}_n| \\
&\le \int \cdots \int \underbrace{|\hat{\mu}_n(y_1, \ldots, y_n)|}_{\le 1} \\
&\cdot |(\mathcal{N}(x,2))(y_1) \cdots (\mathcal{N}(x,2))(y_n) - (\mathcal{N}(y,2))(y_1) \cdots (\mathcal{N}(y,2))(y_n)| dy_1 \cdots dy_n \\
&\le n \cdot \int |(\mathcal{N}(x,2))(t) - (\mathcal{N}(y,2))(t)| dt \\
&= n \cdot \|\mathcal{N}(x,2) - \mathcal{N}(y,2)\|_{L_1(\mathbb{R})}.
\end{aligned}
\tag{3.22}
$$

Die Abbildung

$$
\begin{array}{ccc}
\mathbb{R} & \longrightarrow & \mathcal{H} \\
x & \longmapsto & \mathcal{N}(x,2)
\end{array}
\tag{3.23}
$$

ist bijektiv, da jede Normalverteilung mit Varianz 2 durch ihren Erwartungswert eineindeutig bestimmt ist. Weiterhin ist die Abbildung $L_1(\mathbb{R})$-stetig, denn für beliebige $x \in \mathbb{R}$, $\delta > 0$ gilt für alle y mit $|y - x| \le \delta$

$$
\begin{aligned}
&\left| \tfrac{1}{2\sqrt{\pi}} \exp\left(-\tfrac{(t-x)^2}{4}\right) - \tfrac{1}{2\sqrt{\pi}} \exp\left(-\tfrac{(t-y)^2}{4}\right) \right| \\
&\qquad \le \tfrac{1}{2\sqrt{\pi}} \exp\left(-\tfrac{(t-(x-\delta))^2}{4}\right), \text{ falls } t < x - \delta \\
&\qquad \le \tfrac{1}{2\sqrt{\pi}}, \text{ falls } t \in [x-\delta, x+\delta] \\
&\qquad \le \tfrac{1}{2\sqrt{\pi}} \exp\left(-\tfrac{(t-(x+\delta))^2}{4}\right), \text{ falls } t > x + \delta.
\end{aligned}
$$

Damit ist eine abschnittsweise definierte integrable Majorante für den Integranden gefunden. Punktweise konvergiert für $y \to x$ der Integrand für alle $t \in \mathbb{R}$ gegen 0, da dieser für festes x, t bezüglich y stetig ist. Mit dem Satz von der majorisierten Konvergenz folgt

$$\int \Big| \frac{1}{2\sqrt{\pi}} \exp\big(-\frac{(t-x)^2}{4}\big) - \frac{1}{2\sqrt{\pi}} \exp\big(-\frac{(t-y)^2}{4}\big)\Big| dt \overset{y\to x}{\longrightarrow} 0$$

und somit ist die L_1-Stetigkeit der Abbildungen (3.23) bewiesen. Zusammen mit (3.22) folgt damit die Stetigkeit der Funktionen $\tilde{\Phi}_n$. Es liegt also die Begebenheit vor, dass eine Folge stetiger reeller Funktionen $(\tilde{\Phi}_n)_n$ punktweise gegen eine nirgends stetige Funktion $\tilde{\Phi}$ konvergiert. Nach dem Satz über den Stetigkeitstransport bei punktweiser Konvergenz (siehe *Heuser* [12], Seite 260), der aus dem Satz von *Baire* aus der topologischen Funktionalanalysis stammt, ist dies unmöglich, da die Menge der Stetigkeitspunkte von $\tilde{\Phi}$ sogar dicht in jedem kompakten Intervall liegen müßte. Somit ist ein Widerspruch erreicht, trotz der Existenz der Abbildung Φ existiert in diesem Beispiel kein robust L_2^2-konsistenter Schätzer. Auch wenn dieses Schätzproblem sehr konstruiert erscheinen mag, widerlegt es doch, dass die Nicht-Überlappungseigenschaft (3.1) alleine hinreichendes Kriterium für die robust d^k-konsistente Schätzbarkeit ist.

3.4 Spezieller Äquivalenzsatz

Unter einer bestimmten Forderung an die Schätzbarkeit der verrauschten Dichte mittels direkter Beobachtungen kann mit Hilfe der gleichmäßig X-stetig approximierenden Folge $(\Phi_n)_n$ aus Satz 3.5 ein robust X^k-konsistenter Schätzer konstruiert werden. Diese Forderung möchte ich als *quasi-gleichmäßig konsistente* $L_1(\mathbb{R})$-Schätzbarkeit von h bezeichnen:

$$E_h \|\hat{h}_n - h\|_{L_1} \le D(h) \cdot o_n \quad , \quad \forall\, h \in \mathcal{H} \tag{3.24}$$

mit einer reellen Funktion $D : \mathcal{H} \to \mathbb{R}^+$ und einer von h unabhängigen positiven Nullfolge $(o_n)_n$. Für Dichteklassen $\mathcal{H}$, die dieser Forderung genügen, erhält man einen Äquivalenzsatz für die robust X^k-konsistente Schätzbarkeit:

Satz 3.6 (Äquivalenzsatz über quasi-gleichmäßig konsistent schätzbare Dichteklassen) *Sei* $(X, \|\cdot\|_X)$ *ein reflexiver separabler Banachraum oder* $(L_1(\mathbb{R}), \|\cdot\|_{L_1(\mathbb{R})})$. *Sei* $\mathcal{F} \subseteq \mathcal{D} \cap X$, $\mathcal{G} \subseteq \mathcal{D}$, $\mathcal{H} := \mathcal{F} * \mathcal{G} \subseteq \mathcal{D}$. *Es gelte* $\sup_{f\in\mathcal{F}} \|f\|_X \le C < \infty$. *Sei* $k \ge 1$. *Es existiere eine quasi-gleichmäßig* $L_1(\mathbb{R})$ *-konsistente Schätzerfolge* $(\hat{h}_n)_n$ *von* h *(siehe (3.24)). Dieser kann*

oBdA als $\mathcal{H}$ -immanent angenommen werden.
Dann existiert **genau dann** *eine robust X^k -konsistente Schätzerfolge $(\hat{f}_n)_n$, wenn die durch (3.7) definierte Abbildung Φ gleichmäßig X -stetig approximierbar durch eine Folge $(\Phi_m)_{m\in\mathbb{N}}$ ist, welche (3.20) erfüllen (zum Begriff siehe Satz 3.5).*
Sofern die Forderungen

(a) $m_n \stackrel{n\to\infty}{\longrightarrow} \infty$

(b) $m_n \cdot o_n \stackrel{n\to\infty}{\longrightarrow} 0$

erfüllt sind, ist $(\hat{f}_n)_n = (\Phi_{m_n}(\hat{h}_n))_n$ eine robust X^k -konsistente Schätzerfolge, wann immer eine solche Folge existiert. $(o_n)_{n\in\mathbb{N}}$ ist durch (3.24) definiert.

Beweis: Zuerst muss gezeigt werden, dass die Voraussetzung der $\mathcal{H}$ -Immanenz von $\hat{h}_n$ die Allgemeinheit nicht zerstört. Dabei kann die Argumentation des Beweises von Satz 3.3(a) übernommen werden. Die Metrik b kann dabei durch den speziellen metrischen Raum $L_1(\mathbb{R})$ ersetzt werden, l kann 1 gesetzt werden. Der $\mathcal{H}$ -immanente Schätzer kann analog konstruiert werden, der Unterschied besteht darin, dass die dort auftretende Nullfolge (o_n) nicht mehr beliebig ist, sondern mit eben der im Satz 3.6 durch die quasi-gleichmäßig konsistenten Schätzbarkeit auftretenden gleichzusetzen ist. Wir haben also den $\mathcal{H}$ -immanenten Schätzer $\hat{h}_n$ mit

$$\|\hat{h}_n(y_1,\ldots,y_n) - \tilde{h}_n(y_1,\ldots,y_n))\|_{L_1(\mathbb{R})} \le o_n + \inf_{\tilde{h}\in\mathcal{H}} \|\tilde{h}_n(y_1,\ldots,y_n) - \tilde{h}\|_{L_1(\mathbb{R})}.$$

Die darauffolgende Ungleichungskette kann ebenso übernommen werden, wenn $\tilde{h}_n$ einen nach Voraussetzung dieses Satzes existierenden, nicht zwingend $\mathcal{H}$ -immanenten quasi-gleichmäßig $L_1(\mathbb{R})$ -konsistenten Schätzer bezeichne.

$$\begin{aligned}
&E_h \|\hat{h}_n(Y_1,\ldots,Y_n) - h\|_{L_1(\mathbb{R})} \\
&\le o_n + 2\,E_h\|\tilde{h}_n(Y_1,\ldots,Y_n) - h\|_{L_1(\mathbb{R})} \\
&\le o_n + 2D(h)\cdot o_n \\
&= (2D(h)+1)\cdot o_n.
\end{aligned}$$

Somit schätzt der $\mathcal{H}$-immanente Schätzer ebenso quasi-gleichmäßig $L_1(\mathbb{R})$-konsistent mit reellwertiger Funktion $D^*(h) := 2D(h) + 1$.
Dass aus der Existenz eines robust X^k-konsistenten Schätzers die gleichmäßige X-stetige Approximierbarkeit von Φ folgt, steht durch Satz 3.5 fest. Zu zeigen ist also nur noch die andere Richtung:
Dazu beginne man mit dem üblichen Ansatz zur Risikoabschätzung

$$E_h \|\hat{f}_n - \Phi(h)\|_X^k$$

$$= E_h \|\Phi_{m_n}(\hat{h}_n) - \Phi(h)\|_X^k$$

(Lemma 3.1)

$$\leq 2^{k-1} \cdot \left(\|\Phi_{m_n}(h) - \Phi(h)\|_X^k + E_h \|\Phi_{m_n}(\hat{h}_n) - \Phi_{m_n}(h)\|_X^k\right)$$

Wegen Voraussetzung (a) konvergiert die deterministische Verzerrung aufgrund des Approximationsverhaltens der Folge $(\Phi_m)_m$ gegen 0, d.h.

$$\|\Phi_{m_n}(h) - \Phi(h)\|_X^k \overset{n\to\infty}{\longrightarrow} 0.$$

Es gilt nun noch zu zeigen, dass der Term $E_h \|\Phi_{m_n}(\hat{h}_n) - \Phi_{m_n}(h)\|_X^k$ gegen 0 strebt. Dieser kann abgeschätzt werden für ein beliebiges $\epsilon > 0$ durch

$$\begin{aligned}
&E_h \|\Phi_{m_n}(\hat{h}_n) - \Phi_{m_n}(h)\|_X^k \\
&\leq E_h\big(\|\Phi_{m_n}(\hat{h}_n) - \Phi_{m_n}(h)\|_X^k \mid \|\Phi_{m_n}(\hat{h}_n) - \Phi_{m_n}(h)\|_X \geq \epsilon\big) \\
&\quad \cdot P_h(\|\Phi_{m_n}(\hat{h}_n) - \Phi_{m_n}(h)\|_X \geq \epsilon) \\
&\quad + E_h\big(\|\Phi_{m_n}(\hat{h}_n) - \Phi_{m_n}(h)\|_X^k \mid \|\Phi_{m_n}(\hat{h}_n) - \Phi_{m_n}(h)\|_X < \epsilon\big) \\
&\quad \cdot \underbrace{P_h(\|\Phi_{m_n}(\hat{h}_n) - \Phi_{m_n}(h)\|_X < \epsilon)}_{\leq 1} \\
&\leq (2C)^k \cdot P_h(\|\Phi_{m_n}(\hat{h}_n) - \Phi_{m_n}(h)\|_X \geq \epsilon) + (2\epsilon)^k
\end{aligned}$$

Da Φ_{m_n} (3.20) erfüllt, gilt für jedes $h \in \mathcal{H}$

$$\begin{aligned}
&\{\omega \in \Omega \mid \|\hat{h}_n(\omega) - h\|_{L_1(\mathbb{R})} < \epsilon/(m_n C)\} \\
&\qquad \subseteq \{\omega \in \Omega \mid \|\Phi_{m_n}(\hat{h}_n(\omega)) - \Phi_{m_n}(h)\|_X < \epsilon\}
\end{aligned}$$

und folglich

$$
\begin{aligned}
&P_h(\{\omega \in \Omega \mid \|\hat{h}_n(\omega) - h\|_{L_1(\mathbb{R})} < \epsilon/(m_n C)\}) \\
&\leq\ P_h(\{\omega \in \Omega \mid \|\Phi_{m_n}(\hat{h}_n(\omega)) - \Phi_{m_n}(h)\|_X < \epsilon\})
\end{aligned}
$$

und durch Komplementbildung

$$
\begin{aligned}
&P_h(\{\omega \in \Omega \mid \|\hat{h}_n(\omega) - h\|_{L_1(\mathbb{R})} \geq \epsilon/(m_n \cdot C)\}) \\
&\geq\ P_h(\{\omega \in \Omega \mid \|\Phi_{m_n}(\hat{h}_n(\omega)) - \Phi_{m_n}(h)\|_X \geq \epsilon\}).
\end{aligned}
$$

Daher folgt für die Risikoabschätzung

$$(2C)^k \cdot P_h(\|\Phi_{m_n}(\hat{h}_n) - \Phi_{m_n}(h)\|_X \geq \epsilon)\ +\ (2\epsilon)^k$$

$$\leq\ (2C)^k \cdot P_h(\{\omega \in \Omega \mid \|\hat{h}_n(\omega) - h\|_{L_1(\mathbb{R})} \geq \epsilon/(m_n \cdot C)\})\ +\ (2\epsilon)^k$$

(Markov'sche Ungleichung)

$$\leq\ (2C)^k \cdot (\epsilon/(m_n \cdot C))^{-1}\, E_h\|\hat{h}_n - h\|_{L_1(\mathbb{R})}\ +\ (2\epsilon)^k$$

$$\leq\ (2C)^k \cdot \epsilon^{-1} \cdot m_n \cdot C \cdot D^*(h) \cdot o_n\ +\ (2\epsilon)^k$$

Dieser Ausdruck lässt sich gegen $2(2\epsilon)^k$, wenn n größer als ein gewisses N ist, nach oben abschätzen, wenn $o_n \cdot m_n$ gegen 0 konvergiert, dies wird durch Voraussetzung (b) gewährleistet.

■

Als Beispiel für Dichteklassen, für welche man einen quasi-gleichmäßig konsistenten Schätzer finden kann, möchte ich

$$
\begin{aligned}
\mathcal{H}\ :=\ \{h \in \mathcal{D} \mid &\, supp(\psi_h) \text{ beschränkt },\ \psi_h^{(0)}, \ldots, \psi_h^{(s-1)} \text{ existieren} \\
&\text{und sind gleichmäßig stetig},\ \textstyle\int |\psi_h^{(s)}(t)|^2 dt < +\infty\}
\end{aligned}
$$

anführen. In *Devroye und Györfi* [10] wird gezeigt, dass sich die Bandbreitenfolge eines Kernschätzers $\hat{h}_n$ so wählen lässt, dass für die dort als

$$
\begin{aligned}
A_{T,s,C}\ :=\ \{h \in \mathcal{D} \mid &\, supp(\psi_h) \subseteq [-T, T],\ \psi_h^{(0)}, \ldots, \psi_h^{(s-1)} \text{ existieren} \\
&\text{und sind gleichmäßig stetig},\ \textstyle\int |\psi_h^{(s)}(t)|^2 dt \leq C\}
\end{aligned}
$$

bezeichneten Dichteklassen

$$\sup_{h \in A_{T,s,C}} E_h\|\hat{h}_n - h\|_{L_1(\mathbb{R})}\ \leq\ Const. \cdot n^{-1/2}$$

gilt. Dabei bezeichnet $Const.$ – wie auch im weiteren Text der Dissertation – eine positive nicht genauer spezifizierte Konstante, die ihren exakten Wert während Abschätzungen ändern kann. Wegen $\bigcup\limits_{T,C\in\mathbb{R}^+} A_{T,s,C} = \mathcal{H}$ ist die quasi-gleichmäßige Konsistenz gezeigt. Klar ist auch, dass sich die quasi-gleichmäßige Konsistenz dann auf jede Teilmenge von $\mathcal{H}$ überträgt.

Sicherlich gibt es hier noch viele andere Beispiele von Dichteklassen $\mathcal{F}$ und $\mathcal{G}$, die noch auf robust konsistente Schätzbarkeit untersucht werden können. Da manche Beispiele eher unrealistisch erscheinen, werde ich diese Untersuchung nicht weiter verfolgen. Im nächsten Kapitel möchte ich mich noch der gleichmäßig robusten Konsistenz zuwenden.

Kapitel 4

Gleichmäßig robust konsistente Schätzbarkeit

In diesem Kapitel soll wie im vorhergehenden die robust konsistente Schätzbarkeit nun die gleichmäßig robust konsistente Schätzbarkeit beleuchtet werden (siehe (1.6)). Dabei sei (X,d) wieder ein beliebiger metrischer Raum, und $\mathcal{F} \subseteq X$ werde vorausgesetzt.

4.1 Ein notwendiges Kriterium

Wenn sich $\mathcal{F}$ und $\mathcal{G}$ überlappen, kann es keinen robust d^k-konsistenten Schätzer und folglich auch keinen gleichmäßig robust d^k-konsistenten Schätzer geben. Die Bedingung des Nicht-Überlappens, die nach (3.7) durch die Existenz der Abbildung Φ ausgedrückt werden kann, ist also notwendige Bedingung der gleichmäßig robust d^k-konsistenten Schätzbarkeit. Doch anders als bei der robust d^k-konsistenten Schätzbarkeit ist hier auch die gleichmäßige $(L_1(\mathbb{R}), d)$-Stetigkeit der Abbildung Φ notwendige Voraussetzung, was folgender Satz belegt:

Satz 4.1 *Sei (X,d) beliebiger metrischer Raum, $\mathcal{D}$ die Menge aller Dichten. Sei $\mathcal{F} \subseteq X \cap \mathcal{D}$, $\mathcal{G} \subseteq \mathcal{D}$.*
Es gelte $\sup\limits_{f,\tilde{f} \in \mathcal{F}} d(f,\tilde{f}) \leq C < +\infty$ (Beschränktheit von $\mathcal{F}$ bzgl. d).
Es existiere eine gleichmäßig robust d^k-konsistente Schätzersequenz.

(a) Dann existiert auch eine $\mathcal{F}$-immanente gleichmäßig d^k-robust konsistente Schätzersequenz.

(b) *Dann existiert die Abbildung* Φ *gemäß (3.7)* **und** Φ *ist gleichmäßig* $(L_1(\mathbb{R}), d)$ *-stetig.*

Beweis: ad (a): Diese Aussage ähnelt Satz 3.3(a). Es existiert also eine Schätzerfolge mit

$$\sup_{f\in\mathcal{F}}\sup_{g\in\mathcal{G}} E_{f,g} d(\tilde{f}_n, f)^k \overset{n\to\infty}{\longrightarrow} 0.$$

Man kann auch hier mit dem Auswahlaxiom eine Funktion $\hat{f}_n(y_1,\ldots,y_n)$, die nach $\mathcal{F}$ abbildet, finden, so dass gilt

$$d(\hat{f}_n(y_1,\ldots,y_n), \tilde{f}_n(y_1,\ldots,y_n))^k \leq o(n) + \inf_{f\in\mathcal{F}} d(f, \tilde{f}_n(y_1,\ldots,y_n))^k$$

für alle $y_i \in \mathbb{R}$ gilt (zur Def. von $o(n)$ siehe Satz 3.3(a)). Damit folgt

$$\begin{aligned}
&\sup_{f\in\mathcal{F}}\sup_{g\in\mathcal{G}} E_{f,g} d(\hat{f}_n(Y_1,\ldots,Y_n), f)^k \\
&\leq (\min\{2^{1-k},1\})^{-1}\Big(\sup_{f\in\mathcal{F}}\sup_{g\in\mathcal{G}} E_{f,g} d(\hat{f}_n(Y_1,\ldots,Y_n), \tilde{f}_n(Y_1,\ldots,Y_n))^k \\
&\quad + \sup_{f\in\mathcal{F}}\sup_{g\in\mathcal{G}} E_{f,g} d(\tilde{f}_n, f)^k\Big) \\
&\leq (\min\{2^{1-k},1\})^{-1}\left(o(n) \,+\, 2\cdot \sup_{f\in\mathcal{F}}\sup_{g\in\mathcal{G}} E_{f,g} d(\tilde{f}_n(Y_1,\ldots,Y_n), f)^k\right) \\
&\overset{n\to\infty}{\longrightarrow} 0.
\end{aligned}$$

Somit existiert mit $(\hat{f}_n)_n$ eine $\mathcal{F}$-immanente gleichmäßig robust d^k-konsistente Schätzerfolge.

ad (b): Zu zeigen ist die gleichmäßige $(L_1(\mathbb{R}), d)$-Stetigkeit von Φ, welche bedeutet

$$\forall \epsilon > 0\, \exists \delta > 0 \,:\, \|h - \tilde{h}\|_{L_1(\mathbb{R})} < \delta \,\Rightarrow\, d(\Phi(h), \Phi(\tilde{h})) < \epsilon.$$

Nehmen wir nun an, dies gelte nicht, und führen wir das zu einem Widerspruch, also

$$\exists \epsilon > 0 \left(\forall \delta > 0\, \exists h, \tilde{h} \in \mathcal{H} : \|h - \tilde{h}\|_{L_1(\mathbb{R})} < \delta \,\wedge\, d(\Phi(h), \Phi(\tilde{h})) \geq \epsilon\right).$$

Somit existieren für die reelle Zahlenfolge $(\delta_n)_{n\in\mathbb{N}}$ mit $\delta_n = \frac{1}{n^2}$ Dichtefolgen $(h_n)_n$ und $(\tilde{h}_n)_n$ in $\mathcal{H}$ mit

$$\|h_n - \tilde{h}_n\|_{L_1(\mathbb{R})} < \tfrac{1}{n^2}$$

$$\text{und } d(\Phi(h_n), \Phi(\tilde{h}_n)) \geq \epsilon\,,\ \forall n \in \mathbb{N}.$$

$(\hat{f}_n)_n$ bezeichne die nach (a) existente $\mathcal{F}$-immanente gleichmäßig robust konsistente Schätzerfolge. Damit folgt

$$\sup_{f\in\mathcal{F}} \sup_{g\in\mathcal{G}} E_{f,g}\, d(\hat{f}_n(Y_1,\dots,Y_n), f)^k \stackrel{n\to\infty}{\longrightarrow} 0$$

(nach (3.11))

$$= \sup_{h\in\mathcal{H}} E_h\, d(\hat{f}_n(Y_1,\dots,Y_n), \Phi(h))^k$$

$$\geq \tfrac{1}{2}\left(E_{h_n} d(\hat{f}_n(Y_1,\dots,Y_n), \Phi(h_n))^k \,+\, E_{\tilde{h}_n} d(\hat{f}_n(Y_1,\dots,Y_n), \Phi(\tilde{h}_n))^k\right)$$

$$= \tfrac{1}{2}\left(E_{h_n} d(\hat{f}_n(Y_1,\dots,Y_n), \Phi(h_n))^k \,+\, E_{h_n} d(\hat{f}_n(Y_1,\dots,Y_n), \Phi(\tilde{h}_n))^k\right)$$
$$+\, \tfrac{1}{2}\left(-E_{h_n} d(\hat{f}_n(Y_1,\dots,Y_n), \Phi(\tilde{h}_n))^k \,+\, E_{\tilde{h}_n} d(\hat{f}_n(Y_1,\dots,Y_n), \Phi(\tilde{h}_n))^k\right)$$

$$= \tfrac{1}{2} E_{h_n}\left(d(\hat{f}_n(Y_1,\dots,Y_n), \Phi(h_n))^k \,+\, d(\hat{f}_n(Y_1,\dots,Y_n), \Phi(\tilde{h}_n))^k\right)$$
$$+\, \tfrac{1}{2}\left(-E_{h_n} \underbrace{d(\hat{f}_n(Y_1,\dots,Y_n), \Phi(\tilde{h}_n))^k}_{\leq C^k} + E_{\tilde{h}_n} \underbrace{d(\hat{f}_n(Y_1,\dots,Y_n), \Phi(\tilde{h}_n))^k}_{\leq C^k}\right)$$

(mit Lemma (3.1))

$$\geq \tfrac{1}{2}\min\{2^{1-k},1\}\cdot d(\Phi(h_n),\Phi(\tilde{h}_n))^k$$
$$-\tfrac{1}{2}C^k \int |h_n(y_1)\cdots h_n(y_n) - \tilde{h}_n(y_1)\cdots\tilde{h}_n(y_n)|\,dy_1\cdots dy_n$$

(mit n-facher Anwendung der Dreiecksungleichung)

$$\geq \tfrac{1}{2}\min\{2^{1-k},1\}\cdot\epsilon^k - \tfrac{1}{2}C^k\int|h_n(y_1)-\tilde{h}_n(y_1)|h_n(y_2)\cdots h_n(y_n)\,dy_1\cdots dy_n$$
$$-\tfrac{1}{2}C^k\int\tilde{h}_n(y_1)|h_n(y_2)-\tilde{h}_n(y_2)|h_n(y_3)\cdots h_n(y_n)\,dy_1\cdots dy_n$$
$$\vdots$$
$$-\tfrac{1}{2}C^k\int\tilde{h}_n(y_1)\cdots\tilde{h}_n(y_{n-1})|h_n(y_n)-\tilde{h}_n(y_n)|dy_1\cdots dy_n$$

(da alle h_i und $\tilde{h}_i$ Dichten sind)

$$\geq \tfrac{1}{2}\min\{2^{1-k},1\}\cdot\epsilon^k \;-\; \tfrac{1}{2}C^k\cdot n\cdot\|h_n-\tilde{h}_n\|_{L_1(\mathbb{R})}$$

$$\geq \tfrac{1}{2}\min\{2^{1-k},1\}\cdot\epsilon^k \;-\; \tfrac{1}{2}C^k\cdot n\cdot\tfrac{1}{n^2}$$

$$\geq \tfrac{1}{2}\min\{2^{1-k},1\}\cdot\epsilon^k \;-\; \tfrac{1}{2}C^k\cdot\tfrac{1}{n}$$

$$\stackrel{n\to\infty}{\longrightarrow} \tfrac{1}{2}\min\{2^{1-k},1\}\cdot\epsilon^k > 0.$$

Somit ist ein Widerspruch erreicht.

■

Die Bedingung der gleichmäßigen $(L_1(\mathbb{R}),d)$-Stetigkeit von Φ lässt sich auch durch die gleichmäßige Folgenstetigkeit formulieren: Für Dichtenfolgen $(f_n)_n, (\tilde{f}_n)_n$ in $\mathcal{F}$ und $(g_n)_n, (\tilde{g}_n)_n$ in $\mathcal{G}$ gilt

$$\|f_n * g_n - \tilde{f}_n * \tilde{g}_n\|_{L_1(\mathbb{R})} \stackrel{n\to\infty}{\longrightarrow} 0 \Longrightarrow d(f_n,\tilde{f}_n) \stackrel{n\to\infty}{\longrightarrow} 0. \tag{4.1}$$

Eine weitere Folgerung lässt sich durch Eigenschaften der Partition (3.9) ausdrücken. Dazu definieren wir

$$dist_{\mathcal{G}}^{L_1(\mathbb{R})}(f,\tilde{f}) := \inf_{g,\tilde{g}\in\mathcal{G}}\|f*g-\tilde{f}*\tilde{g}\|_{L_1(\mathbb{R})} \tag{4.2}$$

Aufgrund der Definition ist klar, dass $dist_{\mathcal{G}}^{L_1(\mathbb{R})}$ als Abstand positiv semidefinit ist. Weiter gilt

Lemma 4.1 *Φ sei gleichmäßig $(L_1(\mathbb{R}), d)$ -stetig. Dann gilt für alle $f, \tilde{f} \in \mathcal{F}$*

$$dist_{\mathcal{G}}^{L_1(\mathbb{R})}(f, \tilde{f}) = 0 \Leftrightarrow f = \tilde{f}$$

(d.h. $dist_{\mathcal{G}}^{L_1(\mathbb{R})}$ ist positiv definit) .

Beweis: Es sei $dist_{\mathcal{G}}^{L_1(\mathbb{R})}(f, \tilde{f}) = 0$. Dann existieren $(g_n)_n, (\tilde{g}_n)_n$ in $\mathcal{G}$ mit $\|f * g_n - \tilde{f} * \tilde{g}_n\|_{L_1(\mathbb{R})} \overset{n\to\infty}{\longrightarrow} 0$. Gemäß (4.1) folgt daraus $d(f, \tilde{f}) \overset{n\to\infty}{\longrightarrow} 0$. Da hier keine Abhängigkeit von n mehr besteht, folgt $d(f, \tilde{f}) = 0$ und mit der positiven Definitheit der Metrik d schließlich $f = \tilde{f}$.

■

Das bedeutet also, dass jeweils zwei verschiedene Dichteklassen $f * \mathcal{G}$ und $\tilde{f} * \mathcal{G}$ nicht nur disjunkt sind, sondern sogar positiven L_1 -Abstand voneinander haben.

Eine Folgerung von Satz 4.1 kann man als Korollar 4.1 festhalten:

Korollar 4.1 *Sei (X, d) beliebiger metrischer Raum, $\mathcal{D}$ die Menge aller Dichten. Sei $\mathcal{F} \subseteq X \cap \mathcal{D}$, $\mathcal{G} \subseteq \mathcal{D}$.*
Es gelte $\sup\limits_{f, \tilde{f} \in \mathcal{F}} d(f, \tilde{f}) \le C < +\infty$.
Es existiere eine gleichmäßig robust d^k -konsistente Schätzersequenz.

Dann dominiert in $\mathcal{F}$ die $L_1(\mathbb{R})$ -Metrik die d -Metrik, d.h. wenn der Abstand von Folgen $(f_n)_n, (\tilde{f}_n)_n$ in $\mathcal{F}$ bzgl. der $L_1(\mathbb{R})$ -Metrik gegen 0 konvergiert, dann auch bzgl. der d -Metrik.

Beweis: Es gelte

$$\|f_n - \tilde{f}_n\|_{L_1(\mathbb{R})} \overset{n\to\infty}{\longrightarrow} 0$$

mit $f_n, \tilde{f}_n \in \mathcal{F}$. Daraus folgt für ein beliebiges $g \in \mathcal{G}$ mit der Young'schen Ungleichung (siehe *Werner* [15])

$$0 \le \|g * f_n - g * \tilde{f}_n\|_{L_1(\mathbb{R})} \le \underbrace{\|g\|_{L_1(\mathbb{R})}}_{=1} \cdot \|f_n - \tilde{f}_n\|_{L_1(\mathbb{R})} \overset{n\to\infty}{\longrightarrow} 0$$

Wegen (4.1) folgt daraus

$$\begin{aligned} & d(\Phi(g * f_n), \Phi(g * \tilde{f}_n)) \quad \overset{n\to\infty}{\longrightarrow} 0 \\ & = d(f_n, \tilde{f}_n) \end{aligned}$$

■

4.2 Äquivalenzsätze

Nun suchen wir nach hinreichenden Bedingungen, um die gleichmäßig robuste d^k-Konsistenz zu gewährleisten. Man kann sich die Frage nach einem Analogon für Satz 3.3 stellen. Dabei wird vorausgesetzt, dass die verrauschte Dichte $h \in \mathcal{H}$ aufgrund direkter Beobachtungen in einer Metrik b konsistent schätzbar ist. Diese Voraussetzung erscheint realistisch, zumal in anschließenden Korollar 3.1 darauf verwiesen wird, dass dies stets erfüllt ist, wenn b die $L_1(\mathbb{R})$-Metrik ist. Bei der Betrachtung der gleichmäßigen Konsistenz jedoch existiert kein Analogon zum Äquivalenztheorem von *Devroye*. Insofern bietet es sich an, die Bedingung der gleichmäßig konsistenten Schätzbarkeit der verrauschten Dichte abzuschwächen zu einer gleichmäßig konsistenten Schätzbarkeit der Dichte h modulo der Dichteklasse $f * \mathcal{G}$, in der h liegt. Dies geschieht durch die Voraussetzung (a) in dem folgenden Satz. Wie gesagt, wird (a) durch die Forderung

$$\sup_{h\in\mathcal{H}} E_h\, b(\hat{h}_n, h)^l \overset{n\to\infty}{\longrightarrow} 0$$

mit $l > 0$ impliziert.

Satz 4.2 *Seien (X,d), (Y,b) beliebige metrische Räume und $\mathcal{F} \subseteq X \cap \mathcal{D}$, $\mathcal{G} \subseteq \mathcal{D}$, $\mathcal{H} \subseteq Y \cap \mathcal{D}$.*
Es gelte $\sup\limits_{f,\tilde{f}\in\mathcal{F}} d(f,\tilde{f}) \leq C < +\infty$.
Unter den Voraussetzungen

(a) Es existiert eine Schätzersequenz $(\hat{h}_n)_n$, basierend auf n unabhängigen identisch verteilten direkten Beobachtungen aus $h \in \mathcal{H}$, so dass

$$\sup_{h\in\mathcal{H}} E_h \inf_{\tilde{h}\in\Phi(h)*\mathcal{G}} b(\hat{h}_n, \tilde{h})^l \overset{n\to\infty}{\longrightarrow} 0.$$

gilt (oBdA ist $\hat{h}_n$ $\mathcal{H}$-immanent). ($l > 0$)

(b) Die Abbildung Φ, definiert durch (3.7), existiert und ist gleichmäßig (b,d)-stetig.

gibt es eine $\mathcal{F}$-immanente gleichmäßig robust d^k-konsistente Schätzersequenz, nämlich $(\Phi(\hat{h}_n))_{n\in\mathbb{N}}$. ($k > 0$)

Beweis: Zunächst gehe ich kurz auf die oBdA postulierte $\mathcal{H}$-Immanenz von $(\hat{h}_n)_n$ ein. Wie in den Beweisen der Sätze 3.3 und 4.1 ist die Menge

$$\{h \in \mathcal{H} \mid b(\hat{h}_n(\omega), h)^l \leq \inf_{\tilde{h} \in \mathcal{H}} b(\hat{h}_n(\omega), \tilde{h})^l + o(n)\},$$

wobei $o(n)$ eine positive Nullfolge sei, nicht leer für jedes $n \in \mathbb{N}$, $\omega \in \Omega$. Nach dem Auswahlaxiom existiert daher eine Schätzersequenz $(\tilde{h}_n)_n \subseteq \mathcal{H}$ mit

$$b(\tilde{h}_n(\omega), \hat{h}_n(\omega))^l \;\leq\; \inf_{\tilde{h} \in \mathcal{H}} b(\hat{h}_n(\omega), \tilde{h})^l \;+\; o(n) \quad \forall n, \omega$$

Damit gilt

$$\sup_{h \in \mathcal{H}} E_h \inf_{\tilde{h} \in \Phi(h) * \mathcal{G}} b(\tilde{h}_n, \tilde{h})^l$$

(mit Lemma 3.1 und Dreiecksungleichung)

$$\leq\; (\min\{2^{1-l}, 1\})^{-1} \sup_{h \in \mathcal{H}} E_h \inf_{\tilde{h} \in \Phi(h) * \mathcal{G}} b(\hat{h}_n, \tilde{h})^l + b(\hat{h}_n, \tilde{h}_n)^l$$

$$= (\min\{2^{1-l}, 1\})^{-1} \sup_{h \in \mathcal{H}} \Big(E_h \inf_{\tilde{h} \in \Phi(h) * \mathcal{G}} b(\hat{h}_n, \tilde{h})^l \;+\; E_h \underbrace{b(\hat{h}_n, \tilde{h}_n)^l}_{\substack{\leq \inf_{\tilde{h} \in \mathcal{H}} b(\hat{h}_n, \tilde{h})^l + o(n) \\ \leq \inf_{\tilde{h} \in \Phi(h) * \mathcal{G}} b(\hat{h}_n, \tilde{h})^l + o(n)}} \Big)$$

$$\leq (\min\{2^{1-l}, 1\})^{-1} \big(2 \cdot \sup_{h \in \mathcal{H}} E_h \inf_{\tilde{h} \in \Phi(h) * \mathcal{G}} b(\hat{h}_n, \tilde{h})^l \;+\; o(n) \big)$$

$$\xrightarrow{n \to \infty} 0.$$

Die Schätzersequenz $(\tilde{h}_n)_n$ ist also $\mathcal{H}$-immanent und erfüllt (a). Dies rechtfertigt, die Schätzer $(\hat{h}_n)_n$ oBdA als $\mathcal{H}$-immanent anzusehen. Damit ist die Schätzerfolge $(\Phi(\hat{h}_n))_n$ wohldefiniert und $\mathcal{F}$-immanent. Zu zeigen ist noch ihre gleichmäßig robuste d^k-Konsistenz.

$$\sup_{h\in\mathcal{H}} E_h\, d(\Phi(\hat{h}_n), \Phi(h))^k$$

(Sei $\epsilon > 0$ beliebig, aber fest.)

$$\leq \sup_{h\in\mathcal{H}}(E_h \left(\underbrace{d(\Phi(\hat{h}_n), \Phi(h))^k}_{\leq \epsilon^k} \mid d(\Phi(\hat{h}_n), \Phi(h)) < \epsilon \right) \cdot P_h(d(\Phi(\hat{h}_n), \Phi(h)) < \epsilon)$$

$$+ E_h \left(\underbrace{d(\Phi(\hat{h}_n), \Phi(h))^k}_{\leq C^k} \mid d(\Phi(\hat{h}_n), \Phi(h)) \geq \epsilon \right) \cdot P_h(d(\Phi(\hat{h}_n), \Phi(h)) \geq \epsilon))$$

$$\leq \epsilon^k + C^k \cdot \sup_{h\in\mathcal{H}} P_h(d(\Phi(\hat{h}_n), \Phi(h)) \geq \epsilon)$$

◀ Dazu eine Nebenrechnung:
Da Φ glm. (b,d)-stetig ist, existiert ein $\delta(\epsilon) > 0$, so dass

$$b(\tilde{h}, \hat{h}_n(\omega)) \leq \delta(\epsilon) \;\Rightarrow\; d(\Phi(\tilde{h}), \Phi(\hat{h}_n(\omega))) < \epsilon$$

für alle $\tilde{h} \in \mathcal{H}$, $n \in \mathbb{N}$, $\omega \in \Omega$ gilt. Für ein $n \in \mathbb{N}$, $\omega \in \Omega$ gelte für ein beliebiges $h \in \mathcal{H}$

$$\inf_{\tilde{h}\in\Phi(h)*\mathcal{G}} b(\hat{h}_n(\omega), \tilde{h}) < \frac{\delta(\epsilon)}{2} =: \delta'(\epsilon)(> 0).$$

Also existiert ein $\overline{h}_n(\omega) \in \Phi(h)*\mathcal{G}$ mit $b(\overline{h}_n(\omega), \hat{h}_n(\omega)) < \delta(\epsilon)$. Daraus folgt mit der gleichmäßigen Stetigkeitsbedingung

$$d(\Phi(\hat{h}_n(\omega)), \underbrace{\Phi(\overline{h}_n(\omega)))}_{=\Phi(h)} \;<\; \epsilon.$$

Dies impliziert die Inklusion

$$\{\omega \in \Omega \mid \inf_{\tilde{h}\in\Phi(h)*\mathcal{G}} b(\hat{h}_n(\omega), \tilde{h}) < \delta'(\epsilon)\} \subseteq \{\omega \in \Omega \mid d(\Phi(\hat{h}_n(\omega)), \Phi(h)) < \epsilon\}.$$

Mit der Monotonie des (Wahrscheinlichkeits-)maßes (P_h) folgt

$$P_h\left(\inf_{\tilde{h}\in\Phi(h)*\mathcal{G}} b(\hat{h}_n(\omega),\tilde{h}) < \delta'(\epsilon)\right) \le P_h\left(d(\Phi(\hat{h}_n(\omega)),\Phi(h)) < \epsilon\right)$$

und weiter für jenes beliebige $h \in \mathcal{H}$

$$P_h\left(\inf_{\tilde{h}\in\Phi(h)*\mathcal{G}} b(\hat{h}_n(\omega),\tilde{h}) \ge \delta'(\epsilon)\right) \ge P_h\left(d(\Phi(\hat{h}_n(\omega)),\Phi(h)) \ge \epsilon\right)$$

Mit dieser Ungleichung kann man mit obiger Abschätzung fortfahren. ▶

$$\epsilon^k + C^k \cdot \sup_{h\in\mathcal{H}} P_h(d(\Phi(\hat{h}_n),\Phi(h)) \ge \epsilon)$$

$$\le \epsilon^k + C^k \cdot \sup_{h\in\mathcal{H}} P_h\left(\inf_{\tilde{h}\in\Phi(h)*\mathcal{G}} b(\hat{h}_n(\omega),\tilde{h}) \ge \delta'(\epsilon)\right)$$

$$= \epsilon^k + C^k\delta'(\epsilon)^{-l}\delta'(\epsilon)^l \cdot \sup_{h\in\mathcal{H}} P_h\left(\inf_{\tilde{h}\in\Phi(h)*\mathcal{G}} b(\hat{h}_n(\omega),\tilde{h}) \ge \delta'(\epsilon)\right)$$

(Markov'sche Ungleichung)

$$\le \epsilon^k + C^k\delta'(\epsilon)^{-l} \cdot \underbrace{\sup_{h\in\mathcal{H}} E_h \inf_{\tilde{h}\in\Phi(h)*\mathcal{G}} b(\hat{h}_n(\omega),\tilde{h})^l}_{\overset{n\to\infty}{\longrightarrow} 0 \text{ , unabh. von } \epsilon}$$

Folglich kann man $\epsilon > 0$ beliebig klein wählen und findet in jedem Fall ein $N \in \mathbb{N}$, so dass für alle $n > N$ der obige Term kleiner als 2ϵ ist.
■

Es stellt sich nun die Frage, ob umgekehrt auch aus der gleichmäßigen robusten d^k-Konsistenz die Voraussetzung (a) gefolgert werden kann. Diese Frage beantwortet der folgende Satz für die L_1- und die L_2-Metrik als Metrik b positiv

Satz 4.3 *(a) Sei $\mathcal{D}$ die Menge aller Dichten, $\mathcal{F} \subseteq \mathcal{D}$, $\mathcal{G} \subseteq \mathcal{D}$.*
Es existiere eine glm. robust $L_1(\mathbb{R})$-konsistente Schätzerfolge $(\hat{f}_n)_n$, die oBdA $\mathcal{F}$-immanent ist (siehe Satz 4.1).

Dann existiert eine ($\mathcal{H}$-immanente) Schätzerfolge $(\hat{h}_n)_n$, basierend auf n unabhängigen identisch verteilten direkten Beobachtungen aus h, so dass

$$\sup_{h\in\mathcal{H}} E_h \inf_{\tilde{h}\in\Phi(h)*\mathcal{G}} \|\hat{h}_n - \tilde{h}\|_{L_1(\mathbb{R})} \overset{n\to\infty}{\longrightarrow} 0$$

gilt.

(b) *Sei* $\mathcal{D}_2$ *die Menge aller Dichten,* $\mathcal{F} \subseteq \mathcal{D}_2$ *,* $\mathcal{G} \subseteq \mathcal{D}$ *.*
Es existiere eine glm. robust $L_2(\mathbb{R})^2$ *-konsistente Schätzerfolge* $(\hat{f}_n)_n$ *, die oBdA* $\mathcal{F}$ *-immanent ist.*
Dann existiert eine ($\mathcal{H}$*-immanente) Schätzerfolge* $(\hat{h}_n)_n$ *, basierend auf* n *unabhängigen identisch verteilten direkten Beobachtungen aus* h *, so dass*

$$\sup_{h\in\mathcal{H}} E_h \inf_{\tilde{h}\in\Phi(h)*\mathcal{G}} \|\hat{h}_n - \tilde{h}\|^2_{L_2(\mathbb{R})} \stackrel{n\to\infty}{\longrightarrow} 0$$

gilt.

Beweis: ad (a):

$$\sup_{h\in\mathcal{H}} E_h \,\|\hat{f}_n - \Phi(h)\|_{L_1(\mathbb{R})} \stackrel{n\to\infty}{\longrightarrow} 0.$$

Nun wähle man ein $g \in \mathcal{G}$ beliebig aus und definiere den folglich $\mathcal{H}$-immanenten Schätzer

$$\hat{h}_n := \hat{f}_n * g$$

Damit folgt

$$\begin{aligned}
&\sup_{h\in\mathcal{H}} E_h \underbrace{\inf_{\tilde{h}\in\Phi(h)*\mathcal{G}} \|\hat{h}_n - \tilde{h}\|_{L_1(\mathbb{R})}}_{\leq \|\hat{h}_n - \Phi(h)*g\|_{L_1(\mathbb{R})}} \\
&\leq \sup_{h\in\mathcal{H}} E_h \|\hat{f}_n * g - \Phi(h) * g\|_{L_1(\mathbb{R})} \\
&= \sup_{h\in\mathcal{H}} E_h \|g * (\hat{f}_n - \Phi(h))\|_{L_1(\mathbb{R})} \\
&\leq \sup_{h\in\mathcal{H}} E_h \|\hat{f}_n - \Phi(h)\|_{L_1(\mathbb{R})} \cdot \underbrace{\|g\|_{L_1(\mathbb{R})}}_{=1} \\
&\stackrel{n\to\infty}{\longrightarrow} 0.
\end{aligned}$$

ad (b): Die Argumentation aus (a) lässt sich vollständig übernehmen, wenn man bedenkt, dass mit der Parseval-Identität für jede $L_2(\mathbb{R})$-Funktion f gilt

$$\|g * f\|^2_{L_2(\mathbb{R})} = \frac{1}{2\pi}\|\psi_f \cdot \underbrace{\psi_g}_{|\cdot|\leq 1}\|^2_{L_2(\mathbb{R})} \leq \frac{1}{2\pi}\|\psi_f\|^2_{L_2(\mathbb{R})} = \|f\|^2_{L_2(\mathbb{R})}.$$

■

Damit sind alle Ergebnisse zusammengestellt, um den folgende Äquivalenzsatz beweisen zu können:

Satz 4.4 (Hauptsatz über gleichmäßige robuste $L_1(\mathbb{R})$ -Konsistenz) *Sei $\mathcal{D}$ die Menge aller Dichten, $\mathcal{F}, \mathcal{G} \subseteq \mathcal{D}$.*

Eine gleichmäßig robust $L_1(\mathbb{R})$ -konsistente Schätzerfolge $(\hat{f}_n)_{n\in\mathbb{N}}$ von f existiert **genau dann**, *wenn*

(a) eine Schätzersequenz $(\hat{h}_n)_{n\in\mathbb{N}}$, basierend auf n unabhängigen identisch verteilten direkten Beobachtungen aus $h \in \mathcal{H}$ existiert, so dass

$$\sup_{h\in\mathcal{H}} E_h \inf_{\tilde{h}\in\Phi(h)*\mathcal{G}} \|\hat{h}_n - \tilde{h}\|_{L_1(\mathbb{R})} \overset{n\to\infty}{\longrightarrow} 0$$

gilt (OBdA ist $(\hat{h}_n)_n$ $\mathcal{H}$ -immanent.)
und

(b) die Abbildung Φ, definiert durch (3.7), existiert und gleichmäßig $(L_1(\mathbb{R}), L_1(\mathbb{R}))$ -stetig ist.

$(\Phi(\hat{h}_n))_{n\in\mathbb{N}}$ ist eine glm. robust $L_1(\mathbb{R})$ -konsistente Schätzerfolge (und zudem $\mathcal{F}$ -immanent), wann immer eine solche existiert.

Anmerkung: Leider ist der Beweis nicht konstruktiv, was den $\mathcal{H}$ -immanenten Schätzer $(\hat{h}_n)_n$ anbelangt, so dass die Existenzaussage zunächst von theoretischem Interesse ist.

Beweis: Der metrische Raum (X, d), in dem die unverrauschte Dichte geschätzt werden soll, ist hier der Banachraum $(L_1(\mathbb{R}), \|\cdot\|_{L_1(\mathbb{R})})$. Die Bedingungen $\mathcal{F} \subseteq X$ und $\sup_{f,\tilde{f}\in\mathcal{F}} d(f,\tilde{f}) \le C < \infty$ sind hier erfüllt, denn aus $\mathcal{F} \subseteq \mathcal{D} \subset L_1(\mathbb{R})$ folgt die Beschränktheit für $C = 2$.

$\underline{\Rightarrow}$: Ein glm. robust $L_1(\mathbb{R})$ -konsistenter Schätzer existiere. Nach Satz 4.1 folgt sofort Voraussetzung (b) und gemäß Satz 4.3(a) Voraussetzung (a).

$\underline{\Leftarrow}$: Die Voraussetzungen (a) und (b) sollen gelten. Dann liefert Satz 4.2 sofort die Existenz der glm. robust $L_1(\mathbb{R})$ -konsistenten Schätzersequenz $(\Phi(\hat{h}_n))_n$.

■

Der Hauptsatz 4.4 sagt auch aus, dass - wenn es überhaupt einen glm. robust $L_1(\mathbb{R})$ -konsistenten Schätzer gibt - $(\Phi(\hat{h}_n))_n$ mit der Schätzfunktion

$(\hat{h}_n)_n$ mit der im Satz geforderten Eigenschaft, Voraussetzung (a), glm. robust $L_1(\mathbb{R})$ -konsistent schätzt.
So weit so gut, doch häufig wird bei der Dekonvolutionsdichteschätzung nicht das $L_1(\mathbb{R})$ -, sondern das $L_2(\mathbb{R})$ -Risiko der Schätzung der unverrauschten Dichte betrachtet. Satz 4.4 lässt sich nicht analog auf $L_2(\mathbb{R})$ -Betrachtungen übertragen, da sich in Satz 4.1 $L_1(\mathbb{R})$ nicht durch $L_2(\mathbb{R})$ ersetzen lässt. Hier sind also weitere Überlegungen nötig. Betrachten wir die Situation unter der Voraussetzung, dass der Träger der Dichten in $\mathcal{F}$ gleichgradig beschränkt sei. Solche Dichteklassen werden beispielsweise in *Efromovich* (1997) verwendet. Dann kann man den nächsten Äquivalenzsatz zeigen:

Satz 4.5 (Hauptsatz über gleichmäßig robuste $L_2(\mathbb{R})^2$ -Konsistenz für Dichteklassen mit gleichgradig kompaktem Träger) *Sei $\mathcal{D}$ die Menge aller Dichten, $\mathcal{F}, \mathcal{G} \subseteq \mathcal{D}$ sowie $\mathcal{F} \subseteq L_2(\mathbb{R})$ und $\sup_{f\in\mathcal{F}} \|f\|_{L_2(\mathbb{R})} \leq C < \infty$.*
Es existiere ein $R > 0$, so dass supp$(f) \subseteq [-R, R]$ für alle $f \in \mathcal{F}$ gelte. (supp(f) bezeichne wie üblich den Träger von f).
Eine gleichmäßig robust $L_2(\mathbb{R})^2$ -konsistente Schätzerfolge $(\hat{f})_{n\in\mathbb{N}}$ von f existiert **genau dann**, *wenn*

(a) eine Schätzerfolge $(\hat{h}_n)_{n\in\mathbb{N}}$, basierend auf n unabhängigen identisch verteilten direkten Beobachtungen aus $h \in \mathcal{H}$, existiert, so dass

$$\sup_{h\in\mathcal{H}} E_h \inf_{\tilde{h}\in\Phi(h)*\mathcal{G}} \|\hat{h}_n - \tilde{h}\|_{L_1(\mathbb{R})} \overset{n\to\infty}{\longrightarrow} 0$$

gilt (OBdA ist $(\hat{h}_n)_{n\in\mathbb{N}}$ $\mathcal{H}$ -immanent)
und

(b) die Abbildung Φ, definiert durch (3.7), existiert und gleichmäßig $(L_1(\mathbb{R}), L_2(\mathbb{R}))$ -stetig ist.

Im Falle der glm. robusten $L_2(\mathbb{R})^2$ -konsistenten Schätzbarkeit ist $(\Phi(\hat{h}_n))_{n\in\mathbb{R}}$ ein Schätzer, der dies leistet.

Beweis: Aus den Voraussetzungen (a) und (b) folgt mit Satz 4.2 unmittelbar die glm. robuste $L_2(\mathbb{R})^2$ -Konsistenz des besagten Schätzers. Nun zur Rückrichtung:
Es gilt für alle Dichtefolgen $(f_n)_n, (\tilde{f}_n)_n$ in $\mathcal{F}$

$$\|f_n - \tilde{f}_n\|_{L_1(\mathbb{R})} \overset{n\to\infty}{\longrightarrow} 0 \Leftrightarrow \|f_n - \tilde{f}_n\|_{L_2(\mathbb{R})} \overset{n\to\infty}{\longrightarrow} 0. \tag{4.3}$$

Die Richtung $\Leftarrow$ folgt aus Korollar 4.1. Die Richtung $\Rightarrow$ folgt aus der Voraussetzung der gleichgradigen Beschränktheit des Trägers der Dichten in $\mathcal{F}$ unter Verwendung der Cauchy-Schwarz'schen Ungleichung:

$$\|f_n - \tilde{f}_n\|_{L_1(\mathbb{R})} = \int_{-R}^{R} |f_n(t) - \tilde{f}_n(t)| \cdot 1 \, dt \leq \|f_n - \tilde{f}_n\|_{L_2(\mathbb{R})} \cdot \sqrt{2R} \stackrel{n\to\infty}{\longrightarrow} 0. \quad (4.4)$$

Aus (4.3) folgt, dass die glm. $(L_1(\mathbb{R}), L_1(\mathbb{R}))$ -Stetigkeit und die glm. $(L_1(\mathbb{R}), L_2(\mathbb{R}))$ -Stetigkeit von Φ äquivalent sind. Ebenso impliziert mit (4.4) die gleichmäßige robuste $L_2(\mathbb{R})^2$ -Konsistenz einer $\mathcal{F}$ -immanenten Schätzerfolge $(\hat{f}_n)_{n\in\mathbb{N}}$ die gleichmäßige robuste $L_1(\mathbb{R})$ -Konsistenz, denn

$$0 \leq \left(\sup_{h\in\mathcal{H}} E_h \|\hat{f}_n - f\|_{L_1(\mathbb{R})}\right)^2 \leq \sup_{h\in\mathcal{H}} E_h \|\hat{f}_n - f\|^2_{L_1(\mathbb{R})} \leq 2R \cdot \sup_{h\in\mathcal{H}} E_h \|\hat{f}_n - f\|^2_{L_2(\mathbb{R})}.$$

Dann kann man Satz 4.4 anwenden und man erhält die Voraussetzungen (a) und (b).

■

Kapitel 5

Anwendungen

In den beiden vorigen Kapiteln wurden die Kriterien für die robuste bzw. gleichmäßig robuste Konsistenz ermittelt. Dabei erhielt der Schätzer $(\Phi(\hat{h}_n))_{n\in\mathbb{N}}$ eine große Bedeutung für beide Arten von Konsistenz. Korollar 3.1 sagt aus, dass dieser Schätzer unter der (allerdings nicht notwendigen) Annahme der $(L_1(\mathbb{R}), d)$ -Stetigkeit von Φ robust d^k -konsistent schätzt; die Äquivalenzsätze 4.4 und 4.5 liefern die gleichmäßig robuste Konsistenz unter schwachen Voraussetzungen.
Das Problem ist jedoch, dass der Schätzer $(\Phi(\hat{h}_n))_{n\in\mathbb{N}}$ im Allgemeinen nicht explizit bestimmt werden kann. Die $\mathcal{H}$ -Immanenz des Schätzers $(\Phi(\hat{h}_n))_{n\in\mathbb{N}}$ wird nur durch Verwendung des Auswahlaxioms gewährleistet. Außerdem ist die Abbildung Φ , deren Existenz zwar durch die Nicht-Überlappungsforderung (3.7) garantiert wird, allein aus expliziter Kenntnis der Menge $\mathcal{F}$ und $\mathcal{G}$ im Allgemeinen nicht kalkulierbar. Um die theoretischen Ergebnisse nun auf konkrete Dichteklassen $\mathcal{F}$, $\mathcal{G}$ anwenden zu können, bedarf es also weiterer Überlegungen.

5.1 Doppelte Ausnutzung der Empirik

Betrachten wir nun folgende Situation:

$$\begin{aligned} \mathcal{F} &:= \{f\, L_2(\mathbb{R})\text{-Dichte} \mid \tfrac{1}{t^2} \geq |\psi_f(t)| \geq \tfrac{1}{2t^2}\,,\ \forall t \text{ mit } |t| \geq t_0 \geq 1\} \\ \mathcal{G} &:= \{g_L, g_N\}, \end{aligned} \tag{5.1}$$

wobei g_L die Laplace-Dichte und g_N die Standardnormalverteilungsdichte bezeichne. t_0 sei fest und bekannt. $\mathcal{G}$ ist also zweielementig, $\mathcal{F}$ enthält die

Laplace-Dichte (leicht einzusehen, da $\psi_{g_L}(t) = \frac{1}{1+t^2}$ gilt). Darüberhinaus sind auch alle argument-transferierten Laplace-Dichten $g_L(\bullet + a)$ für jedes konstante $a \in \mathbb{R}$ Elemente von $\mathcal{F}$. Somit enthält $\mathcal{F}$ unendlich viele Dichten (siehe auch Abschnitt 3.2).
Eine gleichmäßig robust $L_2(\mathbb{R})^2$ -konsistente Schätzung durch geeignete Wahl einer Fehlerdichte $\tilde{g} \in \mathcal{G}$ ist unmöglich, denn der Abstand $d_{\mathcal{F}}$ ist positiv definit ($\mathcal{F}$ enthält die Laplace-Dichte mit nirgends verschwindender Fouriertransformierten) und $\mathcal{G}$ ist zweielementig (siehe (2.27)).
Jedoch existiert eine Möglichkeit, die Dichte f glm. robust $L_2(\mathbb{R})^2$ -konsistent zu schätzen, womit auch die offene Frage am Ende des Kapitels 2 beantwortet wäre. Die dazu verwendete Methode möchte ich *doppelte Ausnutzung der Empirik* nennen, da die gegebenen Beobachtungen aus der verrauschten Dichte zweifach verwendet werden: einmal zum Schätzen der nicht direkt beobachtbaren Fehlerdichte g und zum anderen zur Schätzung der verrauschten Dichte h, deren Verteilung die Zufallsvariablen direkt entstammen. Die existierende Abbildung Φ ist die Zusammensetzung des Inversen des Faltungsoperators B_{g_N} auf der Dichteklassen $g_N * \mathcal{F}$ und der Inversen des Faltungsoperators B_{g_L} auf der Dichteklasse $g_L * \mathcal{F}$.

Bei dieser Situation der Existenz einer überall positiven Funktion als gleichgradige untere Schranke des Betrages der Fouriertransformierten von $\mathcal{F}$ müssen die Dichteklassen $g{*}\mathcal{F}$ und $\tilde{g}{*}\mathcal{F}$ für verschiedene g und $\tilde{g}$ stets disjunkt sein. Ansonsten gibt es keinen glm. robust d^k -konsistenten Schätzer, denn:

$$
\begin{aligned}
& g * \mathcal{F} \cap \tilde{g} * \mathcal{F} \neq \emptyset \\
& \Rightarrow\ \exists f, \tilde{f} \in \mathcal{F} :\ g * f = \tilde{g} * \tilde{f} \\
& \textit{(nach Satz 3.1)} \\
& \Rightarrow\ f = \tilde{f} \\
& \Rightarrow\ g * f = \tilde{g} * f \\
& \Rightarrow\ |\psi_g(t)|\, \underbrace{|\psi_f(t)|}_{>0} = |\psi_{\tilde{g}}(t)|\, \underbrace{|\psi_f(t)|}_{>0},\ \forall t \in \mathbb{R} \\
& \Rightarrow\ \psi_g = \psi_{\tilde{g}} \\
& \Rightarrow\ g = \tilde{g}
\end{aligned}
$$

Nun muss anhand der Empirik überprüft werden, ob g_N oder g_L die Fehlerdichte ist. Dies geschieht hier aufgrund der Definition der Menge $\mathcal{F}$ mittels der Fouriertransformierten folgendermaßen:

Für alle $h \in g_L * \mathcal{F}$ gilt

$$|\psi_h(t)| = |\psi_{g_L}(t)| \cdot |\psi_f(t)| \geq \frac{1}{2t^2} \cdot \frac{1}{1+t^2} = \frac{1}{2(t^2+t^4)} =: O(t)$$

für alle $|t| \geq t_0$.
Für alle $h \in g_N * \mathcal{F}$ gilt

$$|\psi_h(t)| = |\psi_{g_N}(t)| \cdot |\psi_f(t)| \leq \frac{1}{t^2} \cdot \exp(-t^2/2) =: U(t)$$

für alle $|t| \geq t_0$.
Nach der Definition der Funktionen O, U ist klar, dass $O(T) > U(T)$ gilt, wenn $T \in \mathbb{R}$ groß genug gewählt wird, notwendigerweise auch größer als t_0 (z.B. für $t_0 = 1 : \ T = 3$)
Mit Hilfe der empirischen Fouriertransformierten lässt sich ein Schätzer für die Fehlerdichte konstruieren

$$\hat{g}_{T,n}(Y_1, \ldots, Y_n) := \begin{cases} g_L & , \text{ falls } \left| \frac{1}{n} \sum_{j=1}^{n} e^{iTY_j} \right| \geq \frac{1}{2}(O(T)+U(T)) \\ \\ g_N & , \text{ falls } \left| \frac{1}{n} \sum_{j=1}^{n} e^{iTY_j} \right| < \frac{1}{2}(O(T)+U(T)) \end{cases} \tag{5.2}$$

Da $\hat{g}_{T,n}$ nur zwei Werte annehmen kann, darf man die Begebenheit auch als Problem der Testtheorie auffassen. Fehler erster und zweiter Art konvergieren für $n \to \infty$ in der Ordnung $\sim n^{-1}$ gegen 0 gleichmäßig über $\mathcal{F}$, was das folgende Lemma zeigen wird

Lemma 5.1 *Es gilt:*

$$\sup_{f \in \mathcal{F}} P_{g_N * f}\left(\hat{g}_{T,n} = g_L\right) \leq \tfrac{4(O(T)-U(T))^{-2}}{n},$$

$$\sup_{f \in \mathcal{F}} P_{g_L * f}\left(\hat{g}_{T,n} = g_N\right) \leq \tfrac{4(O(T)-U(T))^{-2}}{n},$$

Beweis: zur ersten Ungleichung:

$$\sup_{f\in\mathcal{F}} P_{g_N*f}\left(\hat{g}_{T,n}=g_L\right)$$

$$= \sup_{f\in\mathcal{F}} P_{g_N*f}\left(\left|\frac{1}{n}\sum_{j=1}^{n}e^{iTY_j}\right| \geq \frac{1}{2}(O(T)+U(T))\right)$$

$$\leq \sup_{f\in\mathcal{F}} P_{g_N*f}\left(\left|\frac{1}{n}\sum_{j=1}^{n}e^{iTY_j} - \psi_{g_N*f}(T)\right| + \underbrace{|\psi_{g_N*f}(T)|}_{\leq U(T)} \geq \frac{1}{2}(O(T)+U(T))\right)$$

$$\leq \sup_{f\in\mathcal{F}} P_{g_N*f}\left(\left|\frac{1}{n}\sum_{j=1}^{n}e^{iTY_j} - \psi_{g_N*f}(T)\right| \geq \frac{1}{2}(O(T)-U(T)) > 0\right)$$

(Tschebyscheff'sche Ungleichung)

$$\leq \sup_{f\in\mathcal{F}} 4(O(T)-U(T))^{-2}\, var_{g_N*f}\left(\frac{1}{n}\sum_{j=1}^{n} e^{iTY_j}\right)$$

$$= 4(O(T)-U(T))^{-2}\,\frac{1}{n}\sup_{f\in\mathcal{F}}\left(1-|\psi_{f*g_N}(T)|^2\right)$$

$$\leq \frac{4(O(T)-U(T))^{-2}}{n}.$$

zur zweiten Ungleichung:

$$\sup_{f\in\mathcal{F}} P_{h=g_L*f}\left(\hat{g}_{T,n}=g_N\right)$$

$$= \sup_{f\in\mathcal{F}} P_{h=g_L*f}\left(\left|\frac{1}{n}\sum_{j=1}^{n}e^{iTY_j}\right| < \frac{O(T)+U(T)}{2}\right)$$

$$= \sup_{f\in\mathcal{F}} P_{h=g_L*f}\left(\left|\frac{1}{n}\sum_{j=1}^{n}e^{iTY_j}\right| - |\psi_h(T)| < \frac{O(T)+U(T)}{2} - \underbrace{|\psi_h(T)|}_{\geq O(T)}\right)$$

$$\leq \sup_{f\in\mathcal{F}} P_{h=g_L*f}\left(\left|\tfrac{1}{n}\sum_{j=1}^{n} e^{iTY_j}\right| - |\psi_h(T)| < \tfrac{U(T)-O(T)}{2} < 0\right)$$

$$\leq \sup_{f\in\mathcal{F}} P_{h=g_L*f}\left(\left|\left|\tfrac{1}{n}\sum_{j=1}^{n} e^{iTY_j}\right| - |\psi_h(T)|\right| \geq \tfrac{O(T)-U(T)}{2}\right)$$

(Dreiecksungleichung nach unten)

$$\leq \sup_{f\in\mathcal{F}} P_{h=g_L*f}\left(\left|\tfrac{1}{n}\sum_{j=1}^{n} e^{iTY_j} - \psi_h(T)\right| \geq \tfrac{O(T)-U(T)}{2}\right)$$

(Tschebyscheff'sche Ungleichung)

$$\leq \sup_{f\in\mathcal{F}} 4(O(T)-U(T))^{-2}\, var_{h=g_L*f}\left(\tfrac{1}{n}\sum_{j=1}^{n} e^{iTY_j}\right)$$

$$\leq \tfrac{4(O(T)-U(T))^{-2}}{n}.$$

■

Daraus lässt sich nun ein Schätzer bauen. Die Wahl der Bandbreitenfolge $(\omega_n)_n$ bleibt vorerst offen.

$$\hat{f}_n := \left(B^{\dagger}_{\hat{g}_{T,n}} B_{\hat{g}_{T,n}}\right)^{-1} B^{\dagger}_{\hat{g}_{T,n}} \hat{h}_{\omega_n}. \tag{5.3}$$

$\hat{h}_{\omega_n}$ ist dabei der Dichtekernschätzer, der den sync-Kern verwendet (siehe (2.1)). $\hat{f}_n$ ist in jedem Fall für $0 < \omega_n < \infty$ wohldefiniert, da die Fouriertransformierte von $\hat{h}_{\omega_n}$ außerhalb von $[-\omega_n, \omega_n]$ verschwindet. (5.3) kann gemäß (2.16) dargestellt werden durch

$$\hat{f}_n(x) = \frac{1}{2\pi}\int_{-\omega_n}^{\omega_n} \exp(-itx)\frac{1}{n}\sum_{j=1}^{n}\exp(itY_j)/\psi_{\hat{g}_{T,n}}(t)\,dt. \tag{5.4}$$

Nun zur Konsistenzaussage

Satz 5.1 *Der durch (5.3) definierte Schätzer ist gleichmäßig robust* $L_2(\mathbb{R})^2$ *-konsistent in der Situation (5.1). Die Konvergenzrate wird bei Wahl der Bandbreitenfolge* $(\omega_n)_n = (\frac{1}{2}\sqrt{\ln n})_n$ *nach oben beschränkt durch*

$$\sup_{g\in\mathcal{G}}\sup_{f\in\mathcal{F}} E_{f,g}\|\hat{f}_n - f\|^2_{L_2(\mathbb{R})} \leq Const. \cdot (\ln n)^{-3/2}$$

Beweis:

$$\sup_{g\in\mathcal{G}}\sup_{f\in\mathcal{F}} E_{f,g}\|\hat{f}_n - f\|^2_{L_2(\mathbb{R})}$$

$$\leq \sup_{f\in\mathcal{F}} E_{f,g_N}\|\hat{f}_n - f\|^2_{L_2(\mathbb{R})} + \sup_{f\in\mathcal{F}} E_{f,g_L}\|\hat{f}_n - f\|^2_{L_2(\mathbb{R})}$$

(Parseval-Identität)

$$= \tfrac{1}{2\pi}\sup_{f\in\mathcal{F}} E_{f,g_N}\int \left|\frac{\chi_{[-\omega_n,\omega_n]}}{\psi_{\hat{g}_{T,n}}(t)}\frac{1}{n}\sum_{j=1}^{n} e^{itY_j} - \psi_f(t)\right|^2 dt$$
$$+ \tfrac{1}{2\pi}\sup_{f\in\mathcal{F}} E_{f,g_L}\int \left|\frac{\chi_{[-\omega_n,\omega_n]}}{\psi_{\hat{g}_{T,n}}(t)}\frac{1}{n}\sum_{j=1}^{n} e^{itY_j} - \psi_f(t)\right|^2 dt$$

$$\leq \tfrac{1}{2\pi}\sup_{f\in\mathcal{F}} E_{f,g_N}\int 2\left|\frac{\chi_{[-\omega_n,\omega_n]}}{\psi_{\hat{g}_{T,n}}(t)}\frac{1}{n}\sum_{j=1}^{n} e^{itY_j} - \frac{\chi_{[-\omega_n,\omega_n]}}{\psi_{g_N}(t)}\frac{1}{n}\sum_{j=1}^{n} e^{itY_j}\right|^2 dt$$
$$+ \tfrac{1}{2\pi}\sup_{f\in\mathcal{F}} E_{f,g_N}\int 2\left|\frac{\chi_{[-\omega_n,\omega_n]}}{\psi_{g_N}(t)}\frac{1}{n}\sum_{j=1}^{n} e^{itY_j} - \psi_f(t)\right|^2 dt$$
$$+ \tfrac{1}{2\pi}\sup_{f\in\mathcal{F}} E_{f,g_L}\int 2\left|\frac{\chi_{[-\omega_n,\omega_n]}}{\psi_{\hat{g}_{T,n}}(t)}\frac{1}{n}\sum_{j=1}^{n} e^{itY_j} - \frac{\chi_{[-\omega_n,\omega_n]}}{\psi_{g_L}(t)}\frac{1}{n}\sum_{j=1}^{n} e^{itY_j}\right|^2 dt$$
$$+ \tfrac{1}{2\pi}\sup_{f\in\mathcal{F}} E_{f,g_L}\int 2\left|\frac{\chi_{[-\omega_n,\omega_n]}}{\psi_{g_L}(t)}\frac{1}{n}\sum_{j=1}^{n} e^{itY_j} - \psi_f(t)\right|^2 dt$$

$$\leq \frac{1}{\pi} \sup_{f\in\mathcal{F}} \int_{-\omega_n}^{\omega_n} E_{f,g_N} \left(\underbrace{\left| \frac{1}{n} \sum_{j=1}^{n} e^{itY_j} \right|^2}_{\leq 1} \left| \frac{1}{\psi_{\hat{g}_{T,N}(t)}} - \frac{1}{\psi_{g_N}(t)} \right|^2 \right) dt$$

$$+ \frac{1}{\pi} \sup_{f\in\mathcal{F}} E_{f,g_N} \int \left| \frac{\chi_{[-\omega_n,\omega_n]}}{\psi_{g_N}(t)} \frac{1}{n} \sum_{j=1}^{n} e^{itY_j} - \psi_f(t) \right|^2 dt$$

$$+ \frac{1}{\pi} \sup_{f\in\mathcal{F}} \int_{-\omega_n}^{\omega_n} E_{f,g_L} \left(\underbrace{\left| \frac{1}{n} \sum_{j=1}^{n} e^{itY_j} \right|^2}_{\leq 1} \left| \frac{1}{\psi_{\hat{g}_{T,N}(t)}} - \frac{1}{\psi_{g_L}(t)} \right|^2 \right) dt$$

$$+ \frac{1}{\pi} \sup_{f\in\mathcal{F}} E_{f,g_L} \int \left| \frac{\chi_{[-\omega_n,\omega_n]}}{\psi_{g_L}(t)} \frac{1}{n} \sum_{j=1}^{n} e^{itY_j} - \psi_f(t) \right|^2 dt$$

$$\leq \frac{1}{\pi} \sup_{f\in\mathcal{F}} \int_{-\omega_n}^{\omega_n} \left| \frac{1}{\psi_{g_L}(t)} - \frac{1}{\psi_{g_N}(t)} \right|^2 dt \cdot P_{h=g_N*f} \left(\hat{g}_{T,n} = g_L \right)$$

$$+ \frac{1}{\pi} \sup_{f\in\mathcal{F}} E_{f,g_N} \int \left| \frac{\chi_{[-\omega_n,\omega_n]}}{\psi_{g_N}(t)} \frac{1}{n} \sum_{j=1}^{n} e^{itY_j} - \psi_f(t) \right|^2 dt$$

$$+ \frac{1}{\pi} \sup_{f\in\mathcal{F}} \int_{-\omega_n}^{\omega_n} \left| \frac{1}{\psi_{g_L(t)}} - \frac{1}{\psi_{g_N}(t)} \right|^2 dt \cdot P_{h=g_L*f} \left(\hat{g}_{T,n} = g_N \right)$$

$$+ \frac{1}{\pi} \sup_{f\in\mathcal{F}} E_{f,g_L} \int \left| \frac{\chi_{[-\omega_n,\omega_n]}}{\psi_{g_L}(t)} \frac{1}{n} \sum_{j=1}^{n} e^{itY_j} - \psi_f(t) \right|^2 dt$$

(Lemma 5.1)

$$\leq \frac{8(O(T)-U(T))^{-2}}{\pi n} \cdot \int_{-\omega_n}^{\omega_n} \left| \frac{1}{\psi_{g_L(t)}} - \frac{1}{\psi_{g_N}(t)} \right|^2 dt$$

$$+ \frac{1}{\pi} \sup_{f \in \mathcal{F}} E_{f,g_N} \int \left| \frac{\chi_{[-\omega_n,\omega_n]}}{\psi_{g_N}(t)} \frac{1}{n} \sum_{j=1}^{n} e^{itY_j} - \psi_f(t) \right|^2 dt$$
$$+ \frac{1}{\pi} \sup_{f \in \mathcal{F}} E_{f,g_L} \int \left| \frac{\chi_{[-\omega_n,\omega_n]}}{\psi_{g_L}(t)} \frac{1}{n} \sum_{j=1}^{n} e^{itY_j} - \psi_f(t) \right|^2 dt$$

$$\leq \frac{8(O(T)-U(T))^{-2}}{\pi n} \cdot \int_{-\omega_n}^{\omega_n} \left| \frac{1}{\psi_{g_L(t)}} - \frac{1}{\psi_{g_N}(t)} \right|^2 dt + \frac{4}{\pi} \sup_{f \in \mathcal{F}} \int_{\omega_n}^{\infty} |\psi_f(t)|^2 dt$$
$$+ \frac{1}{\pi} \sup_{f \in \mathcal{F}} \int_{-\omega_n}^{\omega_n} var_{f,g_N} \left(\frac{1}{n} \sum_{j=1}^{n} e^{itY_j} / \psi_{g_N}(t) \right) dt$$
$$+ \frac{1}{\pi} \sup_{f \in \mathcal{F}} \int_{-\omega_n}^{\omega_n} var_{f,g_L} \left(\frac{1}{n} \sum_{j=1}^{n} e^{itY_j} / \psi_{g_L}(t) \right) dt$$

$$\leq \frac{8(O(T)-U(T))^{-2}}{\pi n} \cdot \int_{-\omega_n}^{\omega_n} \left| \frac{1}{\psi_{g_L(t)}} - \frac{1}{\psi_{g_N}(t)} \right|^2 dt + \frac{4}{\pi} \sup_{f \in \mathcal{F}} \int_{\omega_n}^{\infty} |\psi_f(t)|^2 dt$$
$$+ \frac{2}{n\pi} \int_0^{\omega_n} |\psi_{g_N}(t)|^{-2} dt + \frac{2}{n\pi} \int_0^{\omega_n} |\psi_{g_L}(t)|^{-2} dt$$

$$\leq \frac{32(O(T)-U(T))^{-2}}{\pi n} \cdot \int_0^{\omega_n} \left(\left| \frac{1}{\psi_{g_L(t)}} \right|^2 + \left| \frac{1}{\psi_{g_N}(t)} \right|^2 dt \right) + \frac{4}{\pi} \sup_{f \in \mathcal{F}} \int_{\omega_n}^{\infty} |\psi_f(t)|^2 dt$$
$$+ \frac{2}{n\pi} \int_0^{\omega_n} |\psi_{g_N}(t)|^{-2} dt + \frac{2}{n\pi} \int_0^{\omega_n} |\psi_{g_L}(t)|^{-2} dt$$

$$\leq \frac{4}{\pi} \sup_{f \in \mathcal{F}} \int_{\omega_n}^{\infty} |\psi_f(t)|^2 dt + \left(\frac{2}{n\pi} + \frac{32(O(T)-U(T))^{-2}}{\pi n} \right) \int_0^{\omega_n} |\psi_{g_N}(t)|^{-2} dt$$
$$+ \left(\frac{2}{n\pi} + \frac{32(O(T)-U(T))^{-2}}{\pi n} \right) \int_0^{\omega_n} |\psi_{g_L}(t)|^{-2} dt$$

Dieser Term konvergiert genau dann gegen 0, wenn alle drei nichtnegativen Summanden dies tun. Die Konvergenzrate des Gesamttermes gleicht dann der des Summanden mit der langsamsten Konvergenz.
Der erste Summand ist nach Definition der Menge $\mathcal{F}$ für $\omega_n > 1$ nach oben modulo konstanten Vorfaktoren abschätzbar (da $\omega_n \to \infty$) durch

$$\int_{\omega_n}^{\infty} t^{-4} dt = \frac{1}{3} \omega_n^{-3}.$$

Der zweite Term ist nach oben modulo Konstanten abschätzbar durch

$$n^{-1}\omega_n \exp(\omega_n^2)$$

und der dritte durch

$$n^{-1}\omega_n(1+\omega_n^2)^2.$$

Durch die Konvergenzforderung an den ersten Summanden ist klar, dass $\omega_n \overset{n\to\infty}{\longrightarrow} \infty$ gelten muss. Daher konvergiert der dritte Term sicherlich schneller gegen 0 als der zweite, so dass der dritte nicht betrachtet werden muss. Setzt man nun die Wahl der Bandbreitenfolge $(\omega_n)_n$ ein, so erhält man für die beiden relevanten Summanden Konvergenzraten von $(\ln n)^{-3/2}$ und $(\ln n)^{1/2}\cdot n^{-3/4}$. Eine Logarithmusfunktion wächst langsamer an als jede Potenzfunktion mit positivem Exponenten. Die langsamere Konvergenzrate ist also $(\ln n)^{-3/2}$, dies ist somit auch die Konvergenzordnung des Gesamttermes.

■

Somit ist also eine obere Schranke für die Konsistenzrate des Schätzers (5.3) gefunden. Diese entspricht der Konsistenzrate des Dekonvolutionsdichteschätzers im Falle $\mathcal{F} := \{f\ L_2\text{-Dichte} \mid |\psi_f(t)| \le C|t|^{-2}, \forall t \in \mathbb{R}\}$ und einer bekannten superglatte Fehlerdichte mit geeigneten Parametern für die Standardnormalverteilungsdichte, wie sie in *Fan* (1993) bestimmt wird. Dekonvolutionsschätzer mit superglatten Fehlerdichten liefern leider stets nur logarithmische Konvergenzraten (siehe *Fan* (1993)), und eine Verbesserung der Konvergenzrate durch Unkenntnis der Fehlerdichte kann man natürlich nicht erwarten.

Dass die beiden konkurrierenden Fehlerdichten g_N und g_L die speziellen Dichten der Standardnormalverteilung und der Laplace-Verteilung sind, ist nur an zwei Stellen entscheidend: bei der Wahl von T bzw. bei der Existenz eines T, das die Bedingung $O(T) > U(T)$ gewährleistet, und bei der Ermittlung der Konvergenzrate. Modifizieren wir die Situation dergestalt, dass zwei Normalverteilungsdichten ($g = \mathcal{N}(\mu, \sigma^2)$ und $\tilde{g} = \mathcal{N}(\tilde{\mu}, \tilde{\sigma^2})$) mit $\sigma^2 < \tilde{\sigma^2}$ als mögliche Fehlerdichten zur Auswahl stehen. Ihre Fouriertransformierten ergeben sich entsprechend als $\psi_g(t) = \exp(i\mu t - \frac{1}{2}\sigma^2 t^2)$ und $\psi_{\tilde{g}}(t) = \exp(i\tilde{\mu}t - \frac{1}{2}\tilde{\sigma^2}t^2)$. Daraus folgt:

$$|\psi_h(t)| = |\psi_f(t)|\cdot|\psi_g(t)| \ge \frac{1}{2t^2}\cdot\exp(-\frac{1}{2}\sigma^2 t^2) =: O(t) \text{ für } h \in g * \mathcal{F}$$
$$|\psi_h(t)| = |\psi_f(t)|\cdot|\psi_{\tilde{g}}(t)| \le \frac{1}{t^2}\cdot\exp(-\frac{1}{2}\tilde{\sigma^2} t^2) =: U(t) \text{ für } h \in \tilde{g} * \mathcal{F}$$

Auch nach dieser Definition ist aufgrund $\sigma^2 \leq \tilde{\sigma^2}$ klar, dass man ein $T > t_0$ finden kann, so dass $O(T) > U(T)$ gilt. Die Konvergenzraten der drei relevanten Summanden bei der Abschätzung im Beweis von Satz 5.1 sind modulo konstante Vorfaktoren nach oben beschränkt durch

$$\omega_n^{-3}$$

$$n^{-1}\omega_n \exp(\sigma^2\omega_n^2)$$

$$n^{-1}\omega_n \exp(\tilde{\sigma^2}\omega_n^2).$$

Die Bandbreitenfolge $(\omega_n)_n$ ist als $\omega_n = \frac{1}{2\sqrt{\tilde{\sigma^2}}}\sqrt{\ln n}$ zu wählen. Dann konvergiert der MISE des Schätzers (5.3) gleichmäßig in der Ordnung $O((\ln n)^{-3/2})$. Das Verfahren lässt sich also auf mehrere Situationen anwenden. Die untere Schranke des Betrages der Fouriertransformierten der Dichteklasse $\mathcal{F}$ darf aber nicht gänzlich entfallen.

5.2 Kontinuierliche Fehlerdichteklassen

Die Methode der doppelten Ausnutzung der Empirik soll nun verallgemeinert werden für den Fall, dass nicht nur zwei Fehlerdichten zur Auswahl stehen, sondern ein Kontinuum von Dichten als Fehlerdichte g a-priori möglich ist. Dazu wird folgendes Schätzproblem betrachtet:

$$\begin{aligned}\mathcal{F} &:= \{f\ L_2(\mathbb{R})\text{-Dichte} \mid C_2|t|^{-\beta} \geq |\psi_f(t)| \geq C_1|t|^{-\beta},\ \forall t \text{ mit } |t| \geq T > 0\} \\ \mathcal{G} &:= \{\mathcal{N}(\mu, \sigma^2) \mid 0 < \sigma^2 \leq \sigma_0^2\}.\end{aligned} \tag{5.5}$$

Dabei seien $C_2 > C_1 > 0$, $\beta > 1$, $T > 0$, μ, $\sigma_0^2 > 0$ bekannt, aber beliebig bis auf die Tatsache, dass C_1, C_2, T so gewählt seien, dass $\mathcal{F}$ nicht leer ist. Dies ist für alle $\beta > 0$ praktikabel, man betrachte zum Beispiel die Gammaverteilungsdichten, deren Fouriertransformierte im Betrag $|\psi_{f_\beta}(t)| = (1+t^2)^{-\beta/2}$ ist. Der Parameter μ kann im Übrigen nicht als unbekannt angenommen werden, sonst kann es in Anbetracht der Dichteklasse $\mathcal{F}$ keinen robust konsistenten Schätzer geben (siehe Abschnitt 3.2, Punkt *(2)*).

Die Abbildung Φ gemäß (3.7) ist hier eine Komposition aus den Inversen überabzählbar unendlich vieler Faltungsoperatoren mit Normalverteilungsdichten, die jeweils auf eine Teilmenge von $\mathcal{H} = \mathcal{F} * \mathcal{G}$ wirken. Diese Teilmengen sind die Dichteklassen $g * \mathcal{F}$. Die Zugehörigkeit einer geschätzten verrauschten Dichte zu jenen Klassen kann empirisch mit ein und demselben Datensatz $Y_1, \ldots, Y_n$ untersucht werden. Zunächst definieren wir die absolute empirische Fouriertransformierte

$$\tilde{\varphi}_n(t) := \left| \frac{1}{n} \sum_{j=1}^{n} \exp(itY_j) \right|. \tag{5.6}$$

Diese kann als Schätzer für den Betrag der Fouriertransformierten $|\psi_h(t)|$ dienen. Allerdings kennen wir aufgrund der gegebenen Parameter für $|\psi_h(t)|$ eine untere Schranke, ohne h zu kennen, und zwar:

$$|\psi_h(t)| = |\psi_f(t)| \cdot |\psi_g(t)| \geq C_1 |t|^{-\beta} \exp\Big(- \frac{1}{2}\sigma_0^2 t^2 \Big), \quad \forall |t| \geq T. \tag{5.7}$$

Mit dieser Erkenntnis kann man den Schätzer aus (5.6) dahingehend verbessern, dass (5.7) berücksichtigt wird. Dies führt zu

$$\hat{\varphi}_n(t) := \begin{cases} C_1 |t|^{-\beta} \exp\big(- \frac{1}{2}\sigma_0^2 t^2 \big) & \text{für } \tilde{\varphi}_n(t) < C_1 |t|^{-\beta} \exp\big(- \frac{1}{2}\sigma_0^2 t^2 \big) \\ & \text{und } |t| \geq T. \\ \tilde{\varphi}_n(t) & \text{sonst.} \end{cases} \tag{5.8}$$

Also gilt für alle t mit $|t| \geq T$ für jedes $h \in \mathcal{H}$

$$\hat{\varphi}_n(t) \geq C_1 |t|^{-\beta} \exp(-\frac{1}{2}\sigma_0^2 t^2)$$

fast sicher. Weiter definieren wir jetzt den Schätzer für σ^2:

$$\hat{\sigma_n^2} := \max\Big(- \frac{2}{t_n^2} \cdot \ln\left(\frac{\hat{\varphi}_n(t_n)}{C_1 t_n^{-\beta}} \right), 0 \Big). \tag{5.9}$$

Dabei sei $(t_n)_{n \in \mathbb{N}}$ eine positive reelle Zahlenfolge mit $t_n > T$, die später geeignet gewählt wird. Der Positivität der Varianz ist in diesem Schätzer Rechnung getragen. Für diesen Schätzer erhält man wegen der Modifikation (5.8) des Schätzers eine obere Schranke:

$$\begin{aligned}
\hat{\sigma_n^2} &\leq \max\left(-\tfrac{2}{t_n^2}\cdot\ln\left(\tfrac{C_1|t_n|^{-\beta}\exp\left(-\frac{1}{2}\sigma_0^2t_n^2\right)}{C_1t_n^{-\beta}}\right),0\right)\\
&= \max(-\tfrac{2}{t_n^2}\cdot\ln(\exp\big(-\tfrac{1}{2}\sigma_0^2t_n^2\big)),0)\\
&= \max(-\tfrac{2}{t_n^2}\cdot\big(-\tfrac{1}{2}\sigma_0^2t_n^2\big),0)\\
&= \sigma_0^2,\ \text{da } \sigma_0^2>0.
\end{aligned}$$

Also gilt für alle $h \in \mathcal{H}$ fast sicher

$$0 \leq \hat{\sigma_n^2} \leq \sigma_0^2. \tag{5.10}$$

Da auch $\sigma^2 \in (0, \sigma_0^2]$ gilt, folgt $|\hat{\sigma_n^2} - \sigma^2| \leq \sigma_0^2$ fast sicher für alle $h \in \mathcal{H}$. Damit kommen wir zu folgendem Satz:

Satz 5.2 *Sei das Schätzproblem (5.5) gegeben. Es seien $Y_1, \ldots, Y_n$ unabhängige identisch verteilte Beobachtungen aus $h \in \mathcal{H}$ gegeben. Man wähle $(\omega_n)_{n\in\mathbb{N}} = (u_1 \cdot (\ln n)^{1/2})_{n\in\mathbb{N}}$ und $(t_n)_{n\in\mathbb{N}} = (u_2 \cdot (\ln n)^{1/2})_{n\in\mathbb{N}}$ mit positiven Konstanten u_1, u_2, für welche gilt $u_1^2 + u_2^2 < \frac{1}{\sigma_0^2}$. Man definiere den Schätzer*

$$\hat{f}_n(x) := \frac{1}{2\pi}\int_{-\omega_n}^{\omega_n} \exp(-it(x+\mu) + \frac{1}{2}\hat{\sigma_n^2}t^2)\,\frac{1}{n}\sum_{j=1}^{n}\exp(itY_j)\,dt, \tag{5.11}$$

wobei $\hat{\sigma_n^2}$ in (5.9) definiert wird. Dann gilt für das L_2-Risiko dieses Schätzers ab einem gewissen $N \in \mathbb{N}$

$$\sup_{g\in\mathcal{G}}\sup_{f\in\mathcal{F}} E_{f,g}\|\hat{f}_n - f\|^2_{L_2(\mathbb{R})} \leq Const.\cdot\begin{cases}(\ln n)^{(1/2)-\beta} & ,\ \textit{falls } \beta < 5/2\\ (\ln n)^{-2}\cdot\ln(\ln(n)) & ,\ \textit{falls } \beta = 5/2\\ (\ln n)^{-2} & ,\ \textit{falls } \beta > 5/2\end{cases}$$

für alle $n \geq N$.

Anmerkung: Wir haben also die Situation einer nicht exakt spezifizierbaren Fehlerdichte, die daher geschätzt werden muss. Anders als bei *Neumann* (1997) und *Efromovich* (1997) ist der Schätzer der Varianz der Fehlerdichte $\hat{\sigma_n^2}$ hier von $Y_1, \ldots, Y_n$ stochastisch abhängig, da nur ein Datensatz $Y_1, \ldots, Y_n$ vorliegt, und die Schätzung der Fehlerdichte erfolgt nicht aufgrund direkter Beobachtungen, sondern aufgrund Ausnutzung der unteren Schranke für den Betrag der Fouriertransformierten der Dichten aus $\mathcal{F}$.

Beweis: Generell kann man feststellen, dass wegen der bestimmten Divergenz von $(\omega_n)_n$ und $(t_n)_n$ ab einem bestimmten N $\omega_n, t_n > T$ für alle $n \geq N$ vorausgesetzt werden kann und also die definierende Bedingung an $\mathcal{F}$ (obere und untere Schranke an den Betrag der Fouriertransformierten) verwendet werden darf.
Beginnen wir mit der Risikobetrachtung

$$\sup_{g\in\mathcal{G}} \sup_{f\in\mathcal{F}} E_{f,g} \|\hat{f}_n - f\|^2_{L_2(\mathbb{R})}$$

(Parseval-Identität)

$$= \tfrac{1}{2\pi} \sup_{g\in\mathcal{G}} \sup_{f\in\mathcal{F}} E_{f,g} \int \left| \chi_{[-\omega_n,\omega_n]}(t) \tfrac{1}{n} \sum_{j=1}^{n} \exp(it(Y_j - \mu)) \exp(\tfrac{1}{2}\hat{\sigma_n^2} t^2) \; - \; \psi_f(t) \right|^2 dt$$

$$= \tfrac{1}{2\pi} \sup_{g\in\mathcal{G}} \sup_{f\in\mathcal{F}} E_{f,g} \left(\int_{-\omega_n}^{\omega_n} \left| \tfrac{1}{n} \sum_{j=1}^{n} \exp(it(Y_j - \mu)) \exp(\tfrac{1}{2}\hat{\sigma_n^2} t^2) \; - \; \psi_f(t) \right|^2 dt \right.$$
$$\left. + \int_{|t|\geq\omega_n} |\psi_f(t)|^2 dt \right)$$

(Satz von Fubini)

$$= \tfrac{1}{2\pi} \sup_{g\in\mathcal{G}} \sup_{f\in\mathcal{F}} \left(\int_{-\omega_n}^{\omega_n} E_{f,g} \left| \tfrac{1}{n} \sum_{j=1}^{n} \exp(it(Y_j - \mu)) \exp(\tfrac{1}{2}\hat{\sigma_n^2} t^2) \; - \; \psi_f(t) \right|^2 dt \right.$$
$$\left. + \int_{|t|\geq\omega_n} |\psi_f(t)|^2 dt \right)$$

$$\leq \tfrac{1}{2\pi} \left\{ \underbrace{\sup_{g\in\mathcal{G}} \sup_{f\in\mathcal{F}} 2 \int_{\omega_n}^{\infty} |\psi_f(t)|^2 dt}_{=:B} \right.$$
$$+ \underbrace{\sup_{g\in\mathcal{G}} \sup_{f\in\mathcal{F}} \int_{-\omega_n}^{\omega_n} 2\, E_{f,g} \left| \exp(-i\mu t + \frac{1}{2}\hat{\sigma_n^2} t^2) \big(\frac{1}{n} \sum_{j=1}^{n} \exp(itY_j) \; - \; \psi_h(t) \big) \right|^2 dt}_{=:V}$$

$$+ \underbrace{\sup_{g\in\mathcal{G}}\sup_{f\in\mathcal{F}} \int_{-\omega_n}^{\omega_n} 2\, E_{f,g}\left| \frac{\psi_h(t)}{\exp(it\mu - \frac{1}{2}\hat{\sigma_n^2}t^2)} \;-\; \psi_f(t)\right|^2 dt}_{=:E} \Bigg\}$$

Der Verzerrungsterm B ist derselbe wie bei der Dekonvolutionsschätzung mit bekannter Fehlerdichte und kann mit der oberen Schranke $|\psi_f(t)|^2 \leq C_2^2|t|^{-2\beta}$ leicht nach oben abgeschätzt werden, da $(\omega_n)_n$ gegen unendlich strebt und dabei größer als T wird. Damit folgt mit der Wahl von $(\omega_n)_n$ aus dem Satz

$$B \;\leq\; \frac{2C_2^2}{2\beta-1}\omega_n^{1-2\beta} \;\leq\; Const. \cdot (\ln n)^{(1/2)-\beta}$$

Den Ausdruck V kann man mit Hilfe von (5.10) nach oben abschätzen durch

$$\sup_{g\in\mathcal{G}}\sup_{f\in\mathcal{F}} \int_{-\omega_n}^{\omega_n} 2\,\exp(\sigma_0^2t^2)E_{f,g}\left|\frac{1}{n}\sum_{j=1}^n \exp(itY_j) \;-\; \psi_h(t)\right|^2 dt$$

Dadurch stellt die stochastische Abhängigkeit des Schätzers $\hat{\sigma_0^2}$ von $Y_1,\ldots,Y_n$ kein Problem mehr da. Nun tritt auch hier dieselbe Situation wie bei der gewöhnlichen Dekonvolutionsschätzung auf. Da $\frac{1}{n}\sum\limits_{j=1}^n \exp(itY_j)$ ein unverfälschter Schätzer für die Fouriertransformierte $\psi_h(t)$ ist, gilt

$$\begin{aligned} E_{f,g}\left|\tfrac{1}{n}\sum_{j=1}^n \exp(itY_j) \;-\; \psi_h(t)\right|^2 \;=\; var\left(\tfrac{1}{n}\sum_{j=1}^n \exp(itY_j)\right) \;&=\; \tfrac{1}{n}\, var\, \exp(itY_1) \\ &\leq\; \tfrac{1}{n}. \end{aligned} \tag{5.12}$$

Damit ist V kleiner oder gleich

$$\begin{aligned} \tfrac{2}{n}\int_{-\omega_n}^{\omega_n} \exp(\sigma_0^2t^2)dt \leq \tfrac{4}{n}\omega_n \cdot \exp(\sigma_0^2\omega_n^2) &\leq Const. \cdot (\ln n)^{1/2}\exp(\sigma_0^2u_1^2(\ln n))/n \\ &\leq Const. \cdot (\ln n)^{1/2} n^{\sigma_0^2u_1^2-1}. \end{aligned}$$

Wegen der Voraussetzung $u_1^2 + u_2^2 \;<\; \sigma_0^{-2}$ ist $\sigma_0^2u_1^2 < 1$, und damit konvergiert der Term V mit algebraischer Rate, also schneller als Term B.

Der eigentlich interessante Term E ist grundsätzlich neu und tritt in der Risikoabschätzung der gewöhnlichen Dekonvolutionsschätzung nicht auf, da er die Schätzung der Fehlerdichte miteinbezieht. Wenn man $\psi_h(t) = \psi_f(t) \cdot \psi_g(t) = \psi_f(t) \cdot \exp(i\mu t - (1/2)\sigma^2 t^2)$ verwendet, ist E gleich

$$\sup_{g\in\mathcal{G}} \sup_{f\in\mathcal{F}} \int_{-\omega_n}^{\omega_n} 2\, E_{f,g} \Big| \frac{\psi_f(t)}{\exp(i\mu t - i\mu t - \frac{1}{2}(\hat{\sigma_n^2} - \sigma^2)t^2)} - \psi_f(t)\Big|^2 dt$$

$$= \sup_{g\in\mathcal{G}} \sup_{f\in\mathcal{F}} \int_{-\omega_n}^{\omega_n} 2\, E_{f,g} \Big| \exp(\frac{1}{2}(\hat{\sigma_n^2} - \sigma^2)t^2) - 1\Big|^2 |\psi_f(t)|^2 dt$$

(mit linearer Substitution: $s = \omega_n^{-1} t$)

$$= \sup_{g\in\mathcal{G}} \sup_{f\in\mathcal{F}} \omega_n \int_{-1}^{1} E_{f,g} |\exp((1/2)(\hat{\sigma_n^2} - \sigma^2)\omega_n^2 s^2) - 1|^2 |\psi_f(\omega_n s)|^2 ds$$

$$\leq \sup_{g\in\mathcal{G}} \sup_{f\in\mathcal{F}} \omega_n \int_{-1}^{1} E_{f,g} |\exp((1/2)|\hat{\sigma_n^2} - \sigma^2|\omega_n^2 s^2) - 1|^2 |\psi_f(\omega_n s)|^2 ds, \tag{5.13}$$

wobei im letzten Schritt die elementare Beziehung $|\exp(x)-1| \geq |\exp(-x)-1|$ für jedes $x > 0$ verwendet wurde. Für ein beliebiges $s \in [-1, 1]$ und beliebige $f \in \mathcal{F}$, $g \in \mathcal{G}$ kann man den Term $E_{f,g}|\exp((1/2)|\hat{\sigma_n^2} - \sigma^2|\omega_n^2 s^2) - 1|^2$ weiter abschätzen. Dazu führen wir die positive Zahlenfolge $(d_n)_{n\in\mathbb{N}}$ durch $d_n = u_2^{-2}(\ln n)^{-1} \cdot 2\ln(2C_2/C_1)$ ein und erhalten

$$E_{f,g} |\exp((1/2)|\hat{\sigma_n^2} - \sigma^2|\omega_n^2 s^2) - 1|^2$$

$$
\begin{aligned}
&= E_{f,g}\big(|\exp((1/2)|\hat{\sigma_n^2}-\sigma^2|\omega_n^2 s^2)-1|^2 \mid |\hat{\sigma_n^2}-\sigma^2|\le d_n\big)\\
&\qquad \cdot P_{f,g}(|\hat{\sigma_n^2}-\sigma^2|\le d_n)\\
&\quad + E_{f,g}\big(|\exp((1/2)|\hat{\sigma_n^2}-\sigma^2|\omega_n^2 s^2)-1|^2 \mid |\hat{\sigma_n^2}-\sigma^2|> d_n\big)\\
&\qquad \cdot P_{f,g}(|\hat{\sigma_n^2}-\sigma^2|> d_n)\\[1em]
&\le E_{f,g}\big(|\exp((1/2)d_n\omega_n^2 s^2)-1|^2 \mid |\hat{\sigma_n^2}-\sigma^2|\le d_n\big)\\
&\quad + E_{f,g}\big(|\exp((1/2)\underbrace{|\hat{\sigma_n^2}-\sigma^2|}_{\le\sigma_0^2}\omega_n^2 s^2)-1|^2 \mid |\hat{\sigma_n^2}-\sigma^2|> d_n\big)\\
&\qquad \cdot P_{f,g}(|\hat{\sigma_n^2}-\sigma^2|> d_n)\\[1em]
&\le |\exp((1/2)d_n\omega_n^2 s^2)-1|^2\\
&\quad + \exp(\sigma_0^2\omega_n^2)\cdot P_{f,g}(|\hat{\sigma_n^2}-\sigma^2|> d_n)
\end{aligned}
$$

Die Folge $(d_n\omega_n^2)_{n\in\mathbb{N}}$ ist nach oben gegen $2u_1^2u_2^{-2}\ln(2C_2/C_1)$ beschränkt, was nach Definition der Folgen leicht zu sehen ist. Wegen $s^2\le 1$ besitzt auch die Folge $(d_n\omega_n^2 s^2)_{n\in\mathbb{N}}$ eben diese Schranke unabhängig von s. Die Funktion $(\exp(\bullet)-1)/\bullet$ ist auf jedem Intervall $[0,c]$ für $c>0$ als stetige (in 0 stetig ergänzbare) Funktion nach oben beschränkt. Somit ist auch die Folge $(|(\exp((1/2)d_n\omega_n^2 s^2)-1)/((1/2)d_n\omega_n^2 s^2)|^2)_n$ unabhängig von s nach oben beschränkt und damit lässt sich die obige Ungleichungskette fortsetzen

$$
\begin{aligned}
&|\exp((1/2)d_n\omega_n^2 s^2)-1|^2\\
&\quad + \exp(\sigma_0^2\omega_n^2)\cdot P_{f,g}(|\hat{\sigma_n^2}-\sigma^2|> d_n)\\[1em]
&\le Const.\cdot d_n^2\omega_n^4 s^4 \;+\; \exp(\sigma_0^2\omega_n^2)\cdot P_{f,g}(|\hat{\sigma_n^2}-\sigma^2|> d_n).
\end{aligned}
\tag{5.14}
$$

Nun muss noch die Wahrscheinlichkeit $\sup_{g\in\mathcal{G}}\sup_{f\in\mathcal{F}} P_{f,g}(|\hat{\sigma_n^2}-\sigma^2|\;\ge\; d_n)$ abgeschätzt werden. Es ist

$$
\begin{aligned}
\sup_{g\in\mathcal{G}}\sup_{f\in\mathcal{F}} P_{f,g}(|\hat{\sigma_n^2}-\sigma^2|\;\ge\; d_n) \;\le\; &\sup_{g\in\mathcal{G}}\sup_{f\in\mathcal{F}} P_{f,g}(\hat{\sigma_n^2}-\sigma^2\;\ge\; d_n)\\
&+\sup_{g\in\mathcal{G}}\sup_{f\in\mathcal{F}} P_{f,g}(\hat{\sigma_n^2}-\sigma^2\;\le\; -d_n).
\end{aligned}
$$

Die beiden hier in Erscheinung tretenden Summanden sollen jetzt abgeschätzt werden. Aus der Definition der Folge $(d_n)_n$ ist ersichtlich, dass

$$t_n^2 d_n \;=\; 2 \cdot (\ln \frac{2C_2}{C_1}) > 0 \tag{5.15}$$

gilt. Beginnen wir mit dem ersten Summanden

$$\begin{aligned}
&\sup_{g\in\mathcal{G}}\sup_{f\in\mathcal{F}} P_{f,g}(\hat{\sigma_n^2} - \sigma^2 \;\geq\; d_n) \\
&= \sup_{g\in\mathcal{G}}\sup_{f\in\mathcal{F}} P_{f,g}(\max\left(-\tfrac{2}{t_n^2}\cdot\ln\left(\tfrac{\hat{\varphi}_n(t_n)}{C_1 t_n^{-\beta}}\right), 0\right) - \sigma^2 \;\geq\; d_n) \\
&= \sup_{g\in\mathcal{G}}\sup_{f\in\mathcal{F}} \Bigg\{ P_{f,g}(-\tfrac{2}{t_n^2}\cdot\ln\left(\tfrac{\hat{\varphi}_n(t_n)}{C_1 t_n^{-\beta}}\right) - \sigma^2 \;\geq\; d_n \mid -\tfrac{2}{t_n^2}\cdot\ln\left(\tfrac{\hat{\varphi}_n(t_n)}{C_1 t_n^{-\beta}}\right) \geq 0) \\
&\quad \cdot P_{f,g}(-\tfrac{2}{t_n^2}\cdot\ln\left(\tfrac{\hat{\varphi}_n(t_n)}{C_1 t_n^{-\beta}}\right) \geq 0) \\
&\quad + \underbrace{P_{f,g}(-\sigma^2 \;\geq\; d_n \mid -\frac{2}{t_n^2}\cdot\ln\left(\frac{\hat{\varphi}_n(t_n)}{C_1 t_n^{-\beta}}\right) < 0)}_{=0,\text{ da } d_n > 0} \cdot P_{f,g}(-\tfrac{2}{t_n^2}\cdot\ln\left(\tfrac{\hat{\varphi}_n(t_n)}{C_1 t_n^{-\beta}}\right) < 0) \Bigg\} \\
&= \sup_{g\in\mathcal{G}}\sup_{f\in\mathcal{F}} P_{f,g}(-\tfrac{2}{t_n^2}\cdot\ln\left(\tfrac{\hat{\varphi}_n(t_n)}{C_1 t_n^{-\beta}}\right) - \sigma^2 \;\geq\; d_n \mid -\tfrac{2}{t_n^2}\cdot\ln\left(\tfrac{\hat{\varphi}_n(t_n)}{C_1 t_n^{-\beta}}\right) \geq 0) \\
&\quad \cdot P_{f,g}(-\tfrac{2}{t_n^2}\cdot\ln\left(\tfrac{\hat{\varphi}_n(t_n)}{C_1 t_n^{-\beta}}\right) \geq 0) \\
&\leq \sup_{g\in\mathcal{G}}\sup_{f\in\mathcal{F}} P_{f,g}(-\tfrac{2}{t_n^2}\cdot\ln\left(\tfrac{\hat{\varphi}_n(t_n)}{C_1 t_n^{-\beta}}\right) - \sigma^2 \;\geq\; d_n) \\
&= \sup_{g\in\mathcal{G}}\sup_{f\in\mathcal{F}} P_{f,g}(\ln\left(\tfrac{\hat{\varphi}_n(t_n)}{C_1 t_n^{-\beta}}\right) + \tfrac{t_n^2}{2}\sigma^2 \;\leq\; -\tfrac{t_n^2}{2}d_n) \\
&\leq \sup_{g\in\mathcal{G}}\sup_{f\in\mathcal{F}} P_{f,g}(\tfrac{\hat{\varphi}_n(t_n)}{C_1 t_n^{-\beta}}\cdot\exp(\tfrac{t_n^2}{2}\sigma^2) \;\leq\; \exp(-\tfrac{t_n^2}{2}d_n)) \\
&= \sup_{g\in\mathcal{G}}\sup_{f\in\mathcal{F}} P_{f,g}(\hat{\varphi}_n(t_n) \;\leq\; \underbrace{C_1 t_n^{-\beta}\cdot\exp(-\frac{t_n^2}{2}\sigma^2)}_{\leq|\psi_h(t_n)|}\cdot\exp(-\tfrac{t_n^2}{2}d_n)) \\
&\leq \sup_{g\in\mathcal{G}}\sup_{f\in\mathcal{F}} P_{f,g}(\hat{\varphi}_n(t_n) \;\leq\; |\psi_h(t_n)|\cdot\exp(-\tfrac{t_n^2}{2}d_n))
\end{aligned}$$

$$= \sup_{g\in\mathcal{G}}\sup_{f\in\mathcal{F}} P_{f,g}(\hat{\varphi}_n(t_n) - |\psi_h(t_n)| \;\leq\; |\psi_h(t_n)| \cdot (\exp(-\tfrac{t_n^2}{2}d_n) - 1) < 0)$$

$$\leq \sup_{g\in\mathcal{G}}\sup_{f\in\mathcal{F}} P_{f,g}(|\hat{\varphi}_n(t_n) - |\psi_h(t_n)|| \;\geq\; |\psi_h(t_n)| \cdot (-\exp(-\tfrac{t_n^2}{2}d_n) + 1))$$

(Markov'sche Ungleichung)

$$\leq\; (1-\exp(-\tfrac{t_n^2}{2}d_n))^{-2} \sup_{g\in\mathcal{G}}\sup_{f\in\mathcal{F}} |\psi_h(t_n)|^{-2} E_{f,g}|\hat{\varphi}_n(t_n) - |\psi_h(t_n)||^2$$

(wegen (5.7) und (5.8))

$$\leq\; (1-\exp(-\tfrac{t_n^2}{2}d_n))^{-2} \sup_{g\in\mathcal{G}}\sup_{f\in\mathcal{F}} |\psi_h(t_n)|^{-2} E_{f,g}|\tilde{\varphi}_n(t_n) - |\psi_h(t_n)||^2$$

(Anwendung der Dreiecksungleichung nach unten)

$$\leq\; (1-\exp(-\tfrac{t_n^2}{2}d_n))^{-2} \sup_{g\in\mathcal{G}}\sup_{f\in\mathcal{F}} |\psi_h(t_n)|^{-2} E_{f,g}\left|\tfrac{1}{n}\sum_{j=1}^{n} \exp(it_nY_j) - \psi_h(t_n)\right|^2$$

(5.12)

$$\leq\; (1-\exp(-\tfrac{t_n^2}{2}d_n))^{-2} \sup_{g\in\mathcal{G}}\sup_{f\in\mathcal{F}} |\psi_h(t_n)|^{-2}\tfrac{1}{n}$$

(5.7)

$$\leq\; (1-\exp(-\tfrac{t_n^2}{2}d_n))^{-2} \cdot \tfrac{C_1^{-2}t_n^{2\beta}\exp(\sigma_0^2t_n^2)}{n}$$

(5.15)

$$= \left(1-\tfrac{C_1}{2C_2}\right)^{-2} \cdot \tfrac{C_1^{-2}t_n^{2\beta}\exp(\sigma_0^2t_n^2)}{n}$$

Wegen $C_2 \geq C_1$ ist der erste Summand dann nach oben abschätzbar durch

$$4 \cdot \frac{t_n^{2\beta}\exp(\sigma_0^2t_n^2)}{n}. \tag{5.16}$$

Kommen wir jetzt zur Abschätzung des zweiten Summanden.

$$
\begin{aligned}
&\sup_{g\in\mathcal{G}}\sup_{f\in\mathcal{F}} P_{f,g}(\hat{\sigma_n^2} - \sigma^2 \leq -d_n) \\
&= \sup_{g\in\mathcal{G}}\sup_{f\in\mathcal{F}} P_{f,g}\Big(\max\Big(-\tfrac{2}{t_n^2}\cdot\ln\Big(\tfrac{\hat{\varphi}_n(t_n)}{C_1 t_n^{-\beta}}\Big),0\Big) - \sigma^2 \leq -d_n\Big) \\
&\leq \sup_{g\in\mathcal{G}}\sup_{f\in\mathcal{F}} P_{f,g}\Big(-\tfrac{2}{t_n^2}\cdot\ln\Big(\tfrac{\hat{\varphi}_n(t_n)}{C_1 t_n^{-\beta}}\Big) - \sigma^2 \leq -d_n\Big) \\
&= \sup_{g\in\mathcal{G}}\sup_{f\in\mathcal{F}} P_{f,g}\Big(\ln\Big(\tfrac{\hat{\varphi}_n(t_n)}{C_1 t_n^{-\beta}}\Big) + \tfrac{t_n^2}{2}\sigma^2 \geq \tfrac{1}{2}t_n^2 d_n\Big) \\
&= \sup_{g\in\mathcal{G}}\sup_{f\in\mathcal{F}} P_{f,g}\Big(\tfrac{\hat{\varphi}_n(t_n)}{C_1 t_n^{-\beta}}\cdot\exp(\tfrac{t_n^2}{2}\sigma^2) \geq \exp(\tfrac{1}{2}t_n^2 d_n)\Big) \\
&= \sup_{g\in\mathcal{G}}\sup_{f\in\mathcal{F}} P_{f,g}\Big(\hat{\varphi}_n(t_n) \geq C_1 t_n^{-\beta}\exp(\tfrac{1}{2}t_n^2 d_n)\exp(-\tfrac{t_n^2}{2}\sigma^2)\Big) \\
&= \sup_{g\in\mathcal{G}}\sup_{f\in\mathcal{F}} P_{f,g}\Big(\hat{\varphi}_n(t_n) \geq \tfrac{C_1}{C_2}\underbrace{C_2 t_n^{-\beta}\exp(-\frac{t_n^2}{2}\sigma^2)}_{\geq|\psi_h(t_n)|}\exp(\tfrac{1}{2}t_n^2 d_n)\Big) \\
&\leq \sup_{g\in\mathcal{G}}\sup_{f\in\mathcal{F}} P_{f,g}\Big(\hat{\varphi}_n(t_n) \geq \tfrac{C_1}{C_2}|\psi_h(t_n)|\exp(\tfrac{1}{2}t_n^2 d_n)\Big) \\
&= \sup_{g\in\mathcal{G}}\sup_{f\in\mathcal{F}} P_{f,g}\Big(\hat{\varphi}_n(t_n) - |\psi_h(t_n)| \geq \underbrace{(\frac{C_1}{C_2}\exp(\frac{1}{2}t_n^2 d_n) - 1)}_{=1 \text{ wegen } (5.15)}\cdot|\psi_h(t_n)| > 0\Big) \\
&\leq \sup_{g\in\mathcal{G}}\sup_{f\in\mathcal{F}} P_{f,g}\big(|\hat{\varphi}_n(t_n) - |\psi_h(t_n)|| \geq |\psi_h(t_n)|\big)
\end{aligned}
$$

(Markov'sche Ungleichung)

$$
\leq \sup_{g\in\mathcal{G}}\sup_{f\in\mathcal{F}} |\psi_h(t_n)|^{-2} E_{f,g}|\hat{\varphi}_n(t_n) - |\psi_h(t_n)||^2
$$

(analog der Abschätzung des ersten Summanden)

$$\leq \tfrac{C_1^{-2} t_n^{2\beta} \exp(\sigma_0^2 t_n^2)}{n}.$$

Somit kann der zweite Summand genau wie der erste durch die obere Schranke (5.16) abgeschätzt werden. Und insgesamt folgt für die betrachtete Wahrscheinlichkeit

$$\sup_{g\in\mathcal{G}} \sup_{f\in\mathcal{F}} P_{f,g}(|\hat{\sigma_n^2} - \sigma^2| \geq d_n) \leq Const. \cdot \frac{t_n^{2\beta} \exp(\sigma_0^2 t_n^2)}{n}. \tag{5.17}$$

Dieses Resultat kann man verwenden, um (5.14) weiter abzuschätzen

$$\begin{aligned} &Const. \cdot d_n^2 \omega_n^4 s^4 + \exp(\sigma_0^2 \omega_n^2) \cdot P_{f,g}(|\hat{\sigma_n^2} - \sigma^2| > d_n) \\ &\leq Const. \cdot \Big(d_n^2 \omega_n^4 s^4 + \exp(\sigma_0^2 \omega_n^2) \cdot \tfrac{t_n^{2\beta} \exp(\sigma_0^2 t_n^2)}{n}\Big) \\ &\leq Const. \cdot \Big(d_n^2 \omega_n^4 s^4 + \tfrac{\exp(\sigma_0^2(\omega_n^2 + t_n^2))}{n} \cdot t_n^{2\beta}\Big). \end{aligned}$$

Dieses Ergebnis wiederum kann man verwenden, um den Term E weiter abzuschätzen. Die Ungleichung (5.13) wird also fortgesetzt

$$\begin{aligned} &\sup_{g\in\mathcal{G}} \sup_{f\in\mathcal{F}} \omega_n \int_{-1}^{1} E_{f,g} |\exp((1/2)|\hat{\sigma_n^2} - \sigma^2|\omega_n^2 s^2) - 1|^2 |\psi_f(\omega_n s)|^2 ds \\ &\leq Const. \cdot \omega_n \sup_{f\in\mathcal{F}} \int_{-1}^{1} \Big(d_n^2 \omega^4 s^4 + \exp(\sigma_0^2(\omega_n^2 + t_n^2)) \cdot \tfrac{1}{n} t_n^{2\beta}\Big) \cdot |\psi_f(\omega_n s)|^2 \, ds \\ &= Const. \cdot \sup_{f\in\mathcal{F}} \Big(\omega_n^5 d_n^2 \int_{-1}^{1} s^4 |\psi_f(\omega_n s)|^2 ds \\ &\qquad + \exp(\sigma_0^2(\omega_n^2 + t_n^2)) \tfrac{1}{n} t_n^{2\beta} \omega_n \int_{-1}^{1} \underbrace{|\psi_f(\omega_n s)|^2}_{\leq 1} ds\Big) \\ &\leq Const. \cdot \sup_{f\in\mathcal{F}} \Big(\omega_n^5 d_n^2 \int_{-1}^{1} s^4 |\psi_f(\omega_n s)|^2 ds + 2\exp(\sigma_0^2(\omega_n^2 + t_n^2)) \tfrac{1}{n} t_n^{2\beta} \omega_n\Big). \end{aligned}$$

An dieser Stelle muss nun eine Fallunterscheidung erfolgen:

<u>1. Fall:</u> $1 < \beta < 5/2$

Nach Definition von $\mathcal{F}$ wissen wir, dass für jedes $f \in \mathcal{F}$ $|\psi_f(t)| \leq C_2|t|^{-\beta}$ für alle $|t| \geq T$ gilt. Wegen $|\psi_f(t)| \leq 1$ für alle $|t| \leq T$ gilt für diese t $|\psi_f(t)| \leq T^\beta |t|^{-\beta}$ und somit $|\psi_f(t)| \leq C\,|t|^{-\beta}$ mit $C := \max(T^\beta, C_2)$ für alle $t \in \mathbb{R}$. Dies führt zu

$$\sup_{f\in\mathcal{F}} \Big(\omega_n^5 d_n^2 \int_{-1}^{1} s^4 \underbrace{|\psi_f(\omega_n s)|^2}_{\leq C^2 s^{-2\beta}\omega_n^{-2\beta}} ds \;+\; 2\exp(\sigma_0^2(\omega_n^2+t_n^2))\tfrac{1}{n}t_n^{2\beta}\omega_n\Big)$$

$$\leq\; Const. \cdot \Big(\omega_n^{5-2\beta} d_n^2 \int_{-1}^{1} s^{4-2\beta} ds \;+\; 2\exp(\sigma_0^2(\omega_n^2+t_n^2))\tfrac{1}{n}t_n^{2\beta}\omega_n\Big)$$

$\int_{-1}^1 s^{4-2\beta} ds$ existiert als endliches uneigentliches Integral, da die Fallbedingung $4 - 2\beta > -1$ impliziert. Damit kann der von n unabhängige Wert dieses Integrals von $Const.$ absorbiert werden und die Ungleichung zu Ende geführt werden

$$\leq\; Const. \cdot \Big(\omega_n^{5-2\beta} d_n^2 \;+\; 2\exp(\sigma_0^2(\omega_n^2+t_n^2))\frac{1}{n}t_n^{2\beta}\omega_n\Big).$$

Setzt man $\omega_n = u_1 \cdot (\ln n)^{1/2}$, $t_n = u_2 \cdot (\ln n)^{1/2}$ und $d_n = u_2^{-2}(\ln n)^{-1} \cdot 2\ln(2C_2/C_1)$ ein, so erhält man

$$\leq\; Const. \cdot \Big((\ln n)^{(1/2)-\beta} \;+\; 2\exp(\sigma_0^2(u_1^2+u_2^2)(\ln n))\tfrac{1}{n}(\ln n)^{\beta+1/2}\Big)$$

$$\leq\; Const. \cdot \Big((\ln n)^{(1/2)-\beta} \;+\; 2n^{\sigma_0^2(u_1^2+u_2^2)-1}(\ln n)^{\beta+1/2}\Big)$$

und wegen der Forderung $u_1^2 + u_2^2 < \sigma_0^{-2}$ lässt sich der Term E bezüglich $Const. \cdot (\ln n)^{(1/2)-\beta}$ nach oben beschränken. Damit liegt diesselbe Konvergenzrate wie beim Term B vor und insgesamt ist in diesem 1. Fall die Konvergenzordnung des MISE $(\ln n)^{(1/2)-\beta}$.

<u>2. Fall:</u> $\beta > 5/2$

In diesem Fall substituiert man das Integral mit $t = s\omega_n$:

$$\sup_{f\in\mathcal{F}} \Big(\omega_n^5 d_n^2 \int_{-1}^{1} s^4 |\psi_f(\omega_n s)|^2 ds \;+\; 2\exp(\sigma_0^2(\omega_n^2+t_n^2))\tfrac{1}{n}t_n^{2\beta}\omega_n\Big)$$

$$\leq\; Const. \cdot \sup_{f\in\mathcal{F}} \Big(d_n^2 \int_{-\omega_n}^{\omega_n} t^4|\psi_f(t)|^2 dt \;+\; \tfrac{2}{n}t_n^{2\beta}\omega_n \exp(\sigma_0^2(\omega_n^2+t_n^2))\Big)$$

$$=\; Const. \cdot \sup_{f\in\mathcal{F}} \Big(d_n^2 \int_{-\omega_n}^{\omega_n} t^4|\psi_f(t)|^2 dt \;+\; n^{\sigma_0^2(u_1^2+u_2^2)-1}(\ln n)^{\beta+1/2}$$

Betrachten wir jetzt das Integral

$$\int_{-\omega_n}^{\omega_n} t^4|\psi_f(t)|^2 dt$$

$$\leq\; \int_{-T}^{T} t^4 \underbrace{|\psi_f(t)|^2}_{\leq 1} dt \;+\; \int_{T<|t|<\omega_n} t^4 \underbrace{|\psi_f(t)|^2}_{\leq C_2^2|t|^{-2\beta}} dt$$

$$\leq\; Const. \;+\; 2C_2^2 \int_{T}^{\omega_n} t^{4-2\beta} dt$$

Da für $\beta > 5/2$ $\;4-2\beta < -1$ gilt, ist die Funktion $\bullet^{4-2\beta}$ über ganz $\mathbb{R}$ integrabel und somit ist das Integral unabhängig bezüglich n nach oben beschränkt. Also kann der Term E in diesem Fall gegen $Const. \cdot d_n^{-2} = Const. \cdot (\ln n)^{-2}$ nach oben abgeschätzt werden, zumal der zweite Summand im Term E algebraisch und somit schneller als d_n^{-2} gegen 0 konvergiert. Aufgrund der Fallbedingung ist $-2 > (1/2) - \beta$ und somit konvergiert der Term E effizient langsamer als der Term B. Folglich ergibt sich als obere Schranke für die Konvergenzrate des MISE $Const. \cdot (\ln n)^{-2}$. Es bleibt noch der

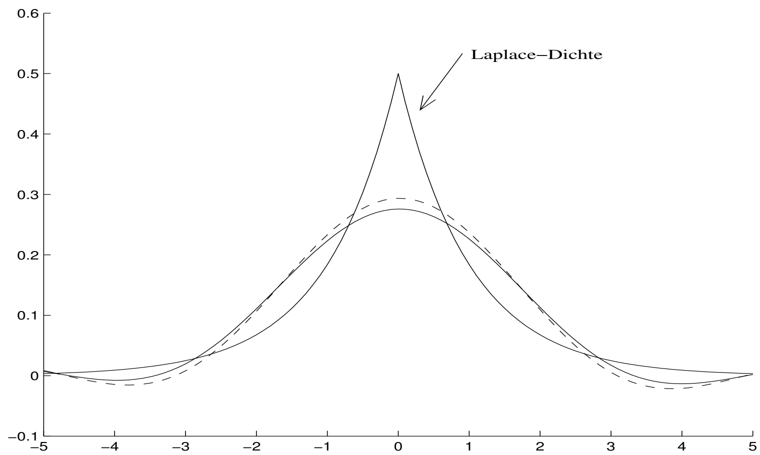
<u>3. Fall:</u> $\beta = 5/2$
Die Abschätzung kann aus dem dritten Fall bis zu der Stelle der Begrenzung des Integrals

$$\begin{aligned}\int_T^{\omega_n} t^{4-2\beta}dt = \int_T^{\omega_n} t^{-1}dt &\\ &\leq (\ln(\omega_n) \,-\, \ln T) \\ &\leq Const. \cdot \ln(\sqrt{\ln(n)}) \\ &\leq Const. \cdot \ln(\ln(n))\end{aligned}$$

übernommen werden, wenn n groß genug ist. Damit ergibt sich als obere Schranke der Konvergenzrate von E $Const. \cdot (\ln n)^{-2} \cdot \ln(\ln(n))$. Diese Konvergenzgeschwindigkeit ist auch hier geringer als die von B ($(\ln n)^{-2}$) und damit ist die Konvergenzgeschwindigkeit des MISE nicht kleiner als $(\ln n)^{-2} \cdot \ln(\ln(n))$. ■

Die hier gefundene obere Schranke des Schätzers (5.11) entspricht damit im Falle $1 < \beta < 5/2$ der Konvergenzrate des Dekonvolutionsschätzers mit bekannter Fehlerdichte. Für $\beta > 5/2$ dagegen verursacht die Schätzung von σ^2 eine Verschlechterung der Konvergenzrate gegenüber der des gewöhnlichen Dekonvolutionskernschätzer, dessen MISE für alle $\beta > 1$ die Rate $(\ln n)^{1/2-\beta}$ erzielt (siehe z.B. *Fan* (1993)). In diesem Bereich stagniert die Rate bei -2 als Exponent von $\ln n$ und größere β bewirken keine besseren Konvergenzraten. Die Optimalität der Raten wird in Abschnitt 5.3 gezeigt.

Der Schätzer (5.11) wird nun an einem konkreten Beispiel simuliert. Für die beim Schätzproblem (5.5) auftretenden Parameter wird vorausgesetzt, dass $C_2 = 1$, $C_1 = 1/2$, $T = 1$, $\mu = 0$, $\sigma_0^2 = 1$ und $\beta = 2$ ist. Dann liegt die Laplace-Dichte in $\mathcal{F}$ (siehe auch Abschnitt 5.1). Die Fehlerdichte sei $\mathcal{N}(0, 1/2)$. Anzumerken ist noch, dass nur der Realteil des Schätzers (5.11) zu betrachten ist. Wird $Re(\hat{f}_n)$ statt $\hat{f}_n$ als Schätzer verwendet, vergrößert sich der MISE nicht, da die zu schätzende Funktion f eine Dichte und somit reellwertig ist. Dies wird auch in der folgenden Simulation getan. Es stehen 1000 unabhängige Beobachtungen aus der verunreinigten Dichte, die hier folglich die Konvolution aus $\mathcal{N}(0, 1/2)$ und der Laplace-Dichte ist, zur Verfügung. Für die Konstanten u_1 und u_2 zur Wahl der Parameterfolgen des Schätzers (5.11) wird $u_1 = 0.5$ und $u_2 = 0.5$, so dass die Bedingung des Satzes 5.2 $0.5^2 + 0.5^2 = 0.25 < 1$ erfüllt ist. Die in MATLAB durchgeführte Simulation ergab das folgende Diagramm:



Dabei sind die zu schätzende Laplace-Dichte, die geschätzte Dichte nach (5.11) sowie als gestrichelte Kurve der gewöhnliche Dekonvolutionskernschätzer, der die Fehlerdichte $\mathcal{N}(0, 1/2)$ bereits exakt kennt und in seiner Konstruktion verwendet, dargestellt. Für diesen ist $\hat{\sigma_n^2}$ gleich konstant $1/2$ gesetzt und seine Bandbreitenfolge ist – wie bei Dekonvolutionsschätzern mit bekannter Normalverteilungsdichte als Fehlerdichte üblich – $\omega_n = 0.5\sqrt{\ln(n)}$ gewählt. Zu erkennen ist, dass der Schätzer (5.11) nur geringfügig von dem gewöhnlichen Dekonvolutionsschätzer abweicht.

5.3 Optimale Konvergenzraten

Nun soll die Frage aufgegriffen werden, ob die Konvergenzrate des Schätzers (5.11) in Satz 5.2 noch verbesserungsfähig ist. Es soll also eine untere Schranke für den MISE des Schätzproblems (5.5) für irgendeinen Schätzer $\hat{f}_n$ von f gefunden werden. Dies geschieht im nachfolgenden Satz. Bis auf den Fall $\beta = 5/2$, bei dem sich die obere Schranke aus Satz 5.2 lediglich durch Multiplikation eines iterierten Logarithmus von der unteren unterscheidet, erhält man tatsächlich Optimalität der Konvergenzraten unter einer technischen Voraussetzung an T, C_2.

Satz 5.3 *Sei das Schätzproblem (5.5) unter der zusätzlichen Voraussetzung $(T/e)^\beta \geq C_2$ gegeben. Sei $\hat{f}_n$ eine beliebige Schätzfunktion für f, d.h. eine Abbildung von $\mathbb{R}^n$ nach $L_2(\mathbb{R})$. Seien $Y_1, \ldots, Y_n$ unabhängige und identisch verteilte Beobachtungen aus der verrauschten Dichte h. Dann gilt für alle n, die größer als ein gewisses N sind,*

$$\sup_{g\in\mathcal{G}} \sup_{f\in\mathcal{F}} E_{f,g}\|\hat{f}_n(Y_1,\ldots,Y_n)-f\|^2_{L_2(\mathbb{R})} \geq Const.\cdot \begin{cases} (\ln n)^{(1/2)-\beta}, & \textit{falls } \beta < 5/2 \\ (\ln n)^{-2} & , \textit{falls } \beta \geq 5/2. \end{cases}$$

mit einer vom Schätzer $\hat{f}_n$ unabhängigen positiven Konstanten $Const.$.

Beweis: Aufgrund der gleichgradigen oberen Schranke an die Beträge der Fouriertransformierten der Elemente aus $\mathcal{F}$ und der Parseval-Identität ist die L_2-Norm der Funktion

$$S(t) = \begin{cases} 1 & , \text{ für } |t| \leq T \\ C_2|t|^{-\beta} & , \text{ sonst.} \end{cases}$$

größer oder gleich der L_2-Norm aller Dichten in $\mathcal{F}$. Somit sind die Elemente von $\mathcal{F}$ bezüglich ihrer L_2-Norm gleichgradig nach oben beschränkt, nennen wir diese Schranke C. Daher kann oBdA angenommen werden, dass die L_2-Norm eines Schätzers $\hat{f}_n$, für den eine untere Schranke des MISE ermittelt werden soll, für alle möglichen Realisierungen durch C nach oben beschränkt ist (siehe dazu Satz 2.3).
Wir wählen jetzt ein $\sigma^2 \in (0, \sigma_0^2)$ beliebig, aber fest aus sowie eine Folge $(\sigma_n^2)_n$ in $(0, \sigma_0^2]$, die streng monoton fallend gegen σ_0^2 konvergiert und deren exakte Definition an dieser Stelle noch offen gehalten wird. Entsprechend betrachten wir für $g_{\sigma^2} = \mathcal{N}(\mu, \sigma^2)$ bzw. $g_{\sigma_n^2} = \mathcal{N}(\mu, \sigma_n^2)$ und für vorerst nicht genauer spezifizierte $f, f_n \in \mathcal{F}$ das Risiko mit der Notation $h_n = f * g_{\sigma_n^2}$ und $\tilde{h}_n = f_n * g_{\sigma^2}$

$$\sup_{g\in\mathcal{G}}\sup_{f\in\mathcal{F}} E_{f,g}\|\hat{f}_n(Y_1,\ldots,Y_n)-f\|^2_{L_2(\mathbb{R})}$$

$$\geq \tfrac{1}{2}\Big(E_{f_n,g_{\sigma^2}}\|\hat{f}_n(Y_1,\ldots,Y_n)-f_n\|^2_{L_2(\mathbb{R})} + E_{f,g_{\sigma^2_n}}\|\hat{f}_n(Y_1,\ldots,Y_n)-f\|^2_{L_2(\mathbb{R})}\Big)$$

$$= \tfrac{1}{2}\Big(E_{f_n,g_{\sigma^2}}\|\hat{f}_n(Y_1,\ldots,Y_n)-f_n\|^2_{L_2(\mathbb{R})} + E_{f_n,g_{\sigma^2}}\|\hat{f}_n(Y_1,\ldots,Y_n)-f\|^2_{L_2(\mathbb{R})}$$
$$- E_{f_n,g_{\sigma^2}}\|\hat{f}_n(Y_1,\ldots,Y_n)-f\|^2_{L_2(\mathbb{R})} + E_{f,g_{\sigma^2_n}}\|\hat{f}_n(Y_1,\ldots,Y_n)-f\|^2_{L_2(\mathbb{R})}\Big)$$

$$\geq \tfrac{1}{2}\Big(E_{f_n,g_{\sigma^2}}(\|\hat{f}_n(Y_1,\ldots,Y_n)-f_n\|^2_{L_2(\mathbb{R})} + \|\hat{f}_n(Y_1,\ldots,Y_n)-f\|^2_{L_2(\mathbb{R})})$$
$$-\big| - E_{f_n,g_{\sigma^2}}\|\hat{f}_n(Y_1,\ldots,Y_n)-f\|^2_{L_2(\mathbb{R})} + E_{f,g_{\sigma^2_n}}\|\hat{f}_n(Y_1,\ldots,Y_n)-f\|^2_{L_2(\mathbb{R})}\big|\Big)$$

$$= \tfrac{1}{2}\Big(E_{f_n,g_{\sigma^2}}(\|\hat{f}_n(Y_1,\ldots,Y_n)-f_n\|^2_{L_2(\mathbb{R})} + \|\hat{f}_n(Y_1,\ldots,Y_n)-f\|^2_{L_2(\mathbb{R})})$$
$$-\Big| \int\cdots\int \|\hat{f}_n(y_1,\ldots,y_n)-f\|^2_{L_2(\mathbb{R})}(h_n(y_1)\cdots h_n(y_n)-\tilde{h}_n(y_1)\cdots\tilde{h}_n(y_n))$$
$$dy_1\cdots dy_n\Big|\Big)$$

(Dreiecksungleichung)

$$\geq \tfrac{1}{2}\Big(\tfrac{1}{2}E_{f_n,g_{\sigma^2}}\|\hat{f}_n(Y_1,\ldots,Y_n)-f_n-\hat{f}_n(Y_1,\ldots,Y_n)+f\|^2_{L_2(\mathbb{R})}$$
$$-\int\cdots\int \|\hat{f}_n(y_1,\ldots,y_n)-f\|^2_{L_2(\mathbb{R})}|h_n(y_1)\cdots h_n(y_n)-\tilde{h}_n(y_1)\cdots\tilde{h}_n(y_n)|$$
$$dy_1\cdots dy_n\Big)$$

(da $\|\hat{f}_n\|^2_{L_2(\mathbb{R})} \leq C^2$ und $\|f\|^2_{L_2(\mathbb{R})} \leq C^2$ gilt, folgt)

$$\geq \tfrac{1}{2}\Big(\tfrac{1}{2}E_{f_n,g_{\sigma^2}}\|f_n-f\|^2_{L_2(\mathbb{R})}$$
$$-C^2\int\cdots\int |h_n(y_1)\cdots h_n(y_n)-\tilde{h}_n(y_1)\cdots\tilde{h}_n(y_n)|dy_1\cdots dy_n\Big)$$

(da h_n, $\tilde{h}_n$ Dichten sind, gilt)

$$\geq \tfrac{1}{2}\Big(\tfrac{1}{2}\|f_n-f\|^2_{L_2(\mathbb{R})} - C^2\cdot n\cdot\int |h_n(y)-\tilde{h}_n(y)|dy\Big)$$

Jetzt definieren wir die Dichte

$$\xi(t) := \begin{cases} c & \text{, falls } |t| \leq 1 \\ c \cdot |t|^{-3/2} & \text{, falls } |t| > 1 \end{cases}$$

mit geeignet gewählter Normierungskonstanten $c > 0$. Da ξ eine strikt positive Funktion ist, lässt sich die obigen Ungleichungskette fortsetzen:

$$\frac{1}{2}\Big(\frac{1}{2}\|f_n - f\|^2_{L_2(\mathbb{R})} - C^2 \cdot n \cdot \int |h_n(y) - \tilde{h}_n(y)| dy\Big)$$

$$= \frac{1}{2}\Big(\frac{1}{2}\|f_n - f\|^2_{L_2(\mathbb{R})} - C^2 \cdot n \cdot \int \sqrt{\xi(y)}|h_n(y) - \tilde{h}_n(y)|/\sqrt{\xi(y)} dy\Big)$$

(Cauchy-Schwarz'sche Ungleichung)

$$\geq \frac{1}{2}\Big(\frac{1}{2}\|f_n - f\|^2_{L_2(\mathbb{R})} - C^2 \cdot n \cdot \underbrace{(\int \xi(y)dy)^{1/2}}_{=1} \cdot (\int |h_n(y) - \tilde{h}_n(y)|^2/\xi(y)dy)^{1/2}\Big)$$

Ziel ist es jetzt, die Funktionen f und f_n sowie die reelle Folge $(\sigma_n^2)_n$ so zu wählen, dass der Term $n \cdot (\int |h_n(y) - \tilde{h}_n(y)|^2/\xi(y)dy)^{1/2}$ schneller gegen 0 konvergiert als $\|f_n - f\|^2_{L_2(\mathbb{R})}$, d.h. dass

$$\frac{n \cdot (\int |h_n(y) - \tilde{h}_n(y)|^2/\xi(y)dy)^{1/2}}{\|f_n - f\|^2_{L_2(\mathbb{R})}} \stackrel{n\to\infty}{\longrightarrow} 0 \tag{5.18}$$

gilt. Dann ist der Gesamtterm für alle n ab einem gewissen N größer oder gleich $\frac{1}{8}\|f_n - f\|^2_{L_2(\mathbb{R})}$.

Wir definieren die reelle Funktion

$$\varphi(t) := \begin{cases} \exp(a|t|) & , \quad |t| \leq T \\ C_2|t|^{-\beta} & , \quad |t| > T \end{cases}$$

mit $a := \frac{1}{T}\ln(C_2 T^{-\beta}) < 0$, da nach Voraussetzung $(T/e)^\beta \geq C_2$ gilt. Die Definition von a impliziert auch $\exp(aT) = C_2 T^{-\beta}$ und somit die Stetigkeit von φ. Weiterhin ist φ die Fouriertransformierte einer Wahrscheinlichkeitsdichte f. Um dies zu zeigen, kann das Kriterium von Polya (siehe *Durrett* [11]) angewandt werden. Dieses liefert als hinreichendes Kriterium, dass φ gerade, stetig, nichtnegativ, auf $(0,\infty)$ monoton fallend und konvex ist sowie $\varphi(0) = 1$ und $\lim_{t\to\infty} \varphi(t) = 0$ gilt. Die Parität, die Stetigkeit, die Nichtnegativität von φ sowie $\varphi(0) = 1$ und $\lim_{t\to\infty} \varphi(t) = 0$ sind leicht einzusehen. Die fallende Monotonie für $(0,\infty)\backslash\{T\}$ ist ebenso erkennbar, und die Stetigkeit von φ garantiert dann die fallende Monotonie auf ganz $(0,\infty)$. Ebenso kann die Konvexität durch die Positivität der zweiten Ableitung für $(0,\infty)\backslash\{T\}$ und durch Berücksichtigung, dass die linksseitige Ableitung $a\exp(aT)$ an der Stelle T kleiner oder gleich der rechtsseitigen $-\beta C_2 T^{-\beta-1}$ ist, was wiederum aus der Voraussetzung $C_2 \leq (T/e)^\beta$ resultiert, gezeigt werden. Also gilt $\varphi = \psi_f$. Aufgrund des Abklingverhaltens $\sim t^{-\beta}$ und der Beschränktheit nach oben bezüglich 1 liegt φ in $L_2(\mathbb{R})$. Da die Fouriertransformation (im Plancherel-Sinne) einen Isomorphismus des $L_2(\mathbb{R})$ darstellt, muss auch $f \in L_2(\mathbb{R})$ und somit $f \in \mathcal{D}_2$ gelten. Da für $|t| \geq T$ $C_2|t|^{-\beta} = |\varphi(t)| \geq C_1|t|^{-\beta}$ nach Definition von φ gilt, ist sogar $f \in \mathcal{F}$.

Weiter definiere man die Funktion

$$\varphi_n(t) = \begin{cases} \varphi(t)\exp((1/2)(\sigma^2 - \sigma_n^2)t^2) & , \quad |t| \leq t_n \\ \varphi(t)\exp((1/2)(\sigma^2 - \sigma_n^2)t_n^2) & , \quad |t| > t_n \end{cases}$$

mit $t_n^2 = \frac{2}{\sigma_n^2 - \sigma^2}\ln(C_2/C_1) > 0$, da $\sigma_n^2 > \sigma^2$. Somit muss $t_n \overset{n\to\infty}{\longrightarrow} +\infty$ gelten. Auch bei dieser Funktion kann man zeigen, dass es sich um die Fouriertransformierte einer Dichte f_n handelt, indem man wieder Polyas Kriterium verwendet: Dass φ_n gerade, nichtnegativ ist, $\varphi_n(0) = 1$ und $\lim_{t\to\infty} \varphi_n(t) = 0$ gilt, ist leicht erkennbar. Als Produkt stetiger Funktionen ist auch φ_n stetig, ebenso als Produkt zweier monoton fallender, positiver Funktionen auch wieder monoton fallend. Die Konvexität dagegen ist weniger intuitiv. Dazu betrachte man zuerst den Fall $0 \leq t \leq T$, in dem man $\varphi_n(t) = \exp(at + \frac{1}{2}(\sigma^2 - \sigma_n^2)t^2)$ ableitet:

$$\varphi_n'(t) = \exp(at + \frac{1}{2}(\sigma^2 - \sigma_n^2)t^2) \cdot (a + (\sigma^2 - \sigma_n^2)t)$$
$$\varphi_n''(t) = \exp(at + \frac{1}{2}(\sigma^2 - \sigma_n^2)t^2) \cdot ((\sigma^2 - \sigma_n^2) + (a + (\sigma^2 - \sigma_n^2)t)^2).$$

Wegen $\sigma_n^2 > \sigma^2$ folgt $\varphi_n''(t) \geq \exp(at + \frac{1}{2}(\sigma^2 - \sigma_n^2)t^2) \cdot ((\sigma^2 - \sigma_n^2) + a^2)$ und wegen $\sigma_n^2 \downarrow \sigma^2$ somit, dass $\varphi_n''(t) > 0$ ist für alle n, die größer als ein gewisses N sind. Damit ist die Konvexität im Bereich $(0, T)$ gezeigt. Der nächste zu betrachtende Bereich ist (T, t_n). In diesem gilt $\varphi_n(t) = \exp((1/2)(\sigma^2 - \sigma_n^2)t^2)C_2 t^{-\beta}$ und für die Ableitungen

$$\begin{aligned}
\varphi_n'(t) &= \exp((1/2)(\sigma^2 - \sigma_n^2)t^2) \cdot (-\beta C_2 t^{-\beta-1} + C_2(\sigma^2 - \sigma_n^2)t^{-\beta+1}) \\
\varphi_n''(t) &= \exp((1/2)(\sigma^2 - \sigma_n^2)t^2) \cdot \big(- \beta(-\beta - 1)C_2 t^{-\beta-2} \\
&\qquad + C_2(1-\beta)t^{-\beta}(\sigma^2 - \sigma_n^2) - \beta C_2 t^{-\beta}(\sigma^2 - \sigma_n^2) + C_2 t^{-\beta+2}(\sigma^2 - \sigma_n^2)^2\big) \\
&= C_2 t^{-\beta-2} \exp((1/2)(\sigma^2 - \sigma_n^2)t^2) \\
&\qquad \cdot (\beta(\beta+1) + (\sigma^2 - \sigma_n^2)(1 - 2\beta)t^2 + t^4(\sigma^2 - \sigma_n^2)^2).
\end{aligned}$$

Da $(\sigma^2 - \sigma_n^2)(1 - 2\beta) > 0$ gilt, ist φ_n'' positiv. Also haben wir auch in dem Intervall (T, t_n) Konvexität. Es bleibt noch der Bereich (t_n, ∞) übrig. Dort ist $\varphi_n(t) = C_2 t^{-\beta} \exp((1/2)(\sigma^2 - \sigma_n^2)t_n^2) = C_1 t^{-\beta}$ und

$$\begin{aligned}
\varphi_n'(t) &= -C_1 \beta t^{-\beta-1} \\
\varphi_n''(t) &= \beta(\beta+1)C_1 t^{-\beta-2}.
\end{aligned}$$

Auch hier ist die Positivität der zweiten Ableitung sofort ersichtlich. Nun ist aber noch zu prüfen, ob in den Anschlussstellen T und t_n die linksseitige Ableitung stets kleiner als die rechtsseitige ist. Beschäftigen wir uns zuerst mit T. Hier ist die linksseitige Ableitung $\varphi_n'(T-) = \exp(aT)\exp((1/2)(\sigma^2 - \sigma_n^2)T^2) \cdot (a + (\sigma^2 - \sigma_n^2)T)$, wobei $\exp(aT)$ durch $C_2 T^{-\beta}$ ersetzt werden kann (s.o.). Setzt man nun T in die rechtsseitige Ableitung ein, erhält man $\varphi_n'(T+) = C_2 T^{-\beta} \exp((1/2)(\sigma^2 - \sigma_n^2)T^2) \cdot (-\beta T^{-1} + T(\sigma^2 - \sigma_n^2))$. Wegen $aT \leq -\beta$, was direkt aus der Voraussetzung $C_2 \leq (T/e)^\beta$ folgt, gilt $\varphi_n'(T-) \leq \varphi_n'(T+)$. In t_n gilt $\varphi_n'(t_n-) = -\beta C_1 t_n^{-\beta-1} + C_1 t_n^{-\beta+1}(\sigma^2 - \sigma_n^2)$ und $\varphi_n'(t_n+) = -\beta C_1 t_n^{-\beta-1}$. Wegen $\sigma^2 < \sigma_n^2$ folgt $\varphi_n'(t_n-) \leq \varphi_n'(t_n+)$. Damit ist insgesamt die Konvexität von φ_n auf $(0, \infty)$ bewiesen.
Mit $0 \leq \varphi_n(t) \leq \varphi(t)$ für alle $t \in \mathbb{R}$ folgt, dass φ_n in $L_2(\mathbb{R})$ liegen muss. Analog der Vorgehensweise bei φ kann daraus geschlossen werden, dass auch f_n eine $L_2(\mathbb{R})$-Dichte ist. Nach der Definition von φ_n gilt für alle $|t| \geq T$

$$|\varphi_n(t)| \geq |\varphi(t)| \exp((1/2)(\sigma^2 - \sigma_n^2)t_n^2) = C_1 |t|^{-\beta},$$

da $\exp((1/2)(\sigma^2 - \sigma_n^2)t^2) \geq \exp((1/2)(\sigma^2 - \sigma_n^2)t_n^2)$ für $|t| \leq t_n$. Außerdem gilt für alle $|t| \leq T$

$$|\varphi_n(t)| \leq |\varphi(t)| = C_2|t|^{-\beta}.$$

Somit folgt schließlich $f_n \in \mathcal{F}$.

Betrachten wir jetzt

$$\psi_{h_n}(t) = \psi_f(t) \cdot \psi_{g_{\sigma_n^2}}(t) = \varphi(t)\exp(i\mu t)\exp(-(1/2)\sigma_n^2 t^2)$$

und

$$\psi_{\tilde{h}_n}(t) = \psi_{f_n}(t)\cdot\psi_{g_{\sigma^2}}(t) = \begin{cases} \varphi(t)\exp(i\mu t)\exp(-(1/2)\sigma_n^2 t^2) & , |t| \leq t_n \\ \varphi(t)\exp(i\mu t)\exp(-(1/2)\sigma^2 t^2)\cdot(C_1/C_2) & , |t| > t_n. \end{cases}$$

Man erkennt, dass ψ_{h_n} und $\psi_{\tilde{h}_n}$ in ihrer Einschränkung auf $[-t_n, t_n]$ übereinstimmen. So kann man mit Hilfe der Parseval-Identität den L_2-Abstand von h_n und $\tilde{h}_n$ nach oben abschätzen

$$\begin{aligned}
\|h_n - \tilde{h}_n\|^2_{L_2(\mathbb{R})} &= \frac{1}{2\pi}\|\psi_{h_n} - \psi_{\tilde{h}_n}\|^2_{L_2(\mathbb{R})} \\
&= \frac{1}{2\pi}\int\limits_{|t|>t_n} |\psi_{h_n}(t) - \psi_{\tilde{h}_n}(t)|^2 dt \\
&\leq \frac{1}{\pi}\int\limits_{|t|>t_n} (|\psi_{h_n}(t)|^2 + |\psi_{\tilde{h}_n}(t)|^2) dt \\
&\quad \textit{(da } \sigma_n^2 > \sigma^2 \textit{ und } C_1/C_2 < 1\textit{)} \\
&\leq \frac{4}{\pi}\int\limits_{t_n}^{\infty} \varphi(t)^2\exp(-\sigma^2 t^2) dt \\
&\leq \frac{4}{\pi}\exp(-\sigma^2 t_n^2)\int\limits_{t_n}^{\infty}\varphi(t)^2 dt
\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{4}{\pi}\|\varphi\|^2_{L_2(\mathbb{R})} \cdot \exp(-\sigma^2 t_n^2) \\
&= \frac{4}{\pi}\|\varphi\|^2_{L_2(\mathbb{R})} \cdot \exp(-2\sigma^2(\ln(C_2/C_1))(\sigma_n^2 - \sigma^2)^{-1}). \qquad (5.19)
\end{aligned}
$$

Diese obere Schranke ist hilfreich bei der weiteren Abschätzung von $n \cdot (\int |h_n(y) - \tilde{h}_n(y)|^2/\xi(y)dy)^{1/2}$, womit der Zusammenhang zu der anfänglichen Ungleichungskette hergestellt ist. Für eine zunächst nicht genauer definierte Folge $(R_n)_n$ mit $R_n \overset{n\to\infty}{\longrightarrow} \infty$ gilt

$$
\begin{aligned}
&n \cdot (\int |h_n(y) - \tilde{h}_n(y)|^2/\xi(y)dy)^{1/2} \\
&= n \cdot (\int_{|y|\leq R_n} |h_n(y) - \tilde{h}_n(y)|^2/\xi(y)dy + \int_{|y|>R_n} |h_n(y) - \tilde{h}_n(y)|^2/\xi(y)dy)^{1/2} \\
&\leq n \cdot (\int_{|y|\leq R_n} |h_n(y) - \tilde{h}_n(y)|^2/\underbrace{\xi(y)}_{\geq c R_n^{-3/2}} dy)^{1/2} \\
&\quad + n \cdot (\int_{|y|>R_n} |h_n(y) - \tilde{h}_n(y)|^2/\xi(y)dy)^{1/2} \\
&\leq n \cdot c^{-1/2} R_n^{3/4} \cdot \|h_n - \tilde{h}_n\|_{L_2(\mathbb{R})} + n \cdot (\int_{|y|>R_n} |h_n(y) - \tilde{h}_n(y)|^2/\xi(y)dy)^{1/2} \\
&\leq Const. \cdot n \cdot R_n^{3/4} \cdot \exp(-\sigma^2(\ln(C_2/C_1))(\sigma_n^2 - \sigma^2)^{-1}) \\
&\quad + n \cdot (\int_{|y|>R_n} |h_n(y) - \tilde{h}_n(y)|^2/\xi(y)dy)^{1/2} \\
&\leq Const. \cdot n \cdot R_n^{3/4} \cdot \exp(-\sigma^2(\ln(C_2/C_1))(\sigma_n^2 - \sigma^2)^{-1}) \\
&\quad + 2n \cdot (\int_{|y|>R_n} |h_n(y)|^2/\xi(y)dy + \int_{|y|>R_n} |\tilde{h}_n(y)|^2/\xi(y)dy)^{1/2}.
\end{aligned}
$$

Die Funktion $\psi_{h_n}(t) = \varphi(t)\exp(i\mu t - (1/2)\sigma_n^2 t^2)$ ist stetig differenzierbar in $\mathbb{R}\backslash\{-T, 0, T\}$ und auf ganz $\mathbb{R}$ stetig. In $\mathbb{R}\backslash\{-T, 0, T\}$ kann man die Ableitung von ψ_{h_n} angeben durch $(\varphi'(t) + (i\mu - \sigma_n^2 t)\varphi(t)) \cdot \exp(i\mu t - (1/2)\sigma_n^2 t^2)$. Diese ist eine L_2-Funktion (leicht zu sehen aufgrund ihrer Beschränktheit und ihres exponentiellen Abklingverhaltens) und bleibt das auch, wenn man an den Stellen $-T$ und T endliche Funktionswerte für ψ'_{h_n} ergänzt. Die so auf ganz $\mathbb{R}$ definierte Abbildung ψ'_{h_n} ist dann schwache Ableitung von ψ_{h_n}, d.h. $\int_0^T \psi'_{h_n}(t)dt = \psi_{h_n}(T) - \psi_{h_n}(0)$ für alle $T \in \mathbb{R}$. Mit Hilfe partieller Integration folgt, dass die so definierte Ableitung ψ'_{h_n} auch Ableitung von ψ_{h_n} im Sobolev-Sinne ist, d.h. dass

$$\langle \psi'_{h_n}, \phi \rangle_{L_2(\mathbb{R})} = \langle \psi_{h_n}, \phi' \rangle_{L_2(\mathbb{R})}$$

für alle beliebig oft differenzierbaren Funktionen ϕ mit kompaktem Träger gilt. Wie oben erwähnt, gilt $\psi'_{h_n} \in L_2(\mathbb{R})$ und somit ist ψ_{h_n} ein Element des Sobolevraumes $W^1(\mathbb{R})$ (siehe *Werner* [15], Seite 164). Die L_2-Norm von ψ'_{h_n} kann man mit Verwendung von $\sigma_n^2 > \sigma^2$ nach oben abschätzen durch

$$\begin{aligned}
\|\psi'_{h_n}\|^2_{L_2(\mathbb{R})} &= \int |(\varphi'(t) + (i\mu - \sigma_n^2 t)\varphi(t)) \cdot \exp(i\mu t - (1/2)\sigma_n^2 t^2)|^2 dt \\
&\leq \int \exp(-\sigma^2 t^2)(|\varphi'(t)| + (|\mu| + \sigma_0^2 |t|)|\varphi(t)|)^2 dt \\
&< \infty
\end{aligned}$$

Das Integral existiert und ist endlich, da $|\varphi(t)| \sim |t|^{-\beta}$ und $|\varphi'(t)| \sim |t|^{-\beta-1}$ mit $\beta > 1$ für $|t| \to \infty$ gilt und $|\varphi|$ gegen 1 und $|\varphi'|$ gegen a nach oben beschränkt sind und somit der Integrand insgesamt exponentiell abklingt. Dieses Integral ist sogar unabhängig von n und damit sind die ψ'_{h_n} in ihrer L_2-Norm gleichgradig beschränkt.

Kommen wir nun zu $\psi_{\tilde{h}_n}$. Diese Funktion ist ebenfalls auf ganz $\mathbb{R}$ stetig und stetig differenzierbar in $\mathbb{R}\backslash\{-t_n, -T, 0, T, t_n\}$. Da wir dieselbe Situation einer stetigen und überall bis auf endlich viele Stellen stetig differenzierbaren Funktion haben, existiert auch hier eine schwache Ableitung, und zwar:

$$\psi'_{\tilde{h}_n}(t) = \begin{cases} (\varphi'(t) + (i\mu - \sigma_n^2 t)\varphi(t)) \cdot \exp(i\mu t - (1/2)\sigma_n^2 t^2) & , |t| < t_n \\ (\varphi'(t) + (i\mu - \sigma^2 t)\varphi(t)) \cdot \exp(i\mu t - (1/2)\sigma^2 t^2) \cdot (C_1/C_2), & |t| > t_n. \end{cases}$$

Auch diese schwache Ableitung ist eine L_2-Funktion und man kann ihre Norm abschätzen:

$$
\begin{aligned}
&\|\psi'_{\tilde{h}_n}\|^2_{L_2(\mathbb{R})} \\
&\le \int |(\varphi'(t) + (i\mu - \sigma_n^2 t)\varphi(t)) \cdot \exp(i\mu t - (1/2)\sigma_n^2 t^2)|^2 dt \\
&\qquad + \int |(\varphi'(t) + (i\mu - \sigma^2 t)\varphi(t)) \cdot \exp(i\mu t - (1/2)\sigma^2 t^2)|^2 dt \cdot (C_1/C_2) \\
&\le \int (|\varphi'(t)| + (|\mu| + \sigma_n^2|t|)\varphi(t))^2 \cdot \exp(-\sigma_n^2 t^2) dt \\
&\qquad + (C_1/C_2) \int (|\varphi'(t)| + (|\mu| + \sigma^2|t|)\varphi(t))^2 \cdot \exp(-\sigma^2 t^2) dt \\
&\le 2 \int (|\varphi'(t)| + (|\mu| + \sigma_0^2|t|)\varphi(t))^2 \cdot \exp(-\sigma^2 t^2) dt < \infty
\end{aligned}
$$

Hier erhalten wir also dasselbe von n unabhängige Integral wie bei der Abschätzung von $\|\psi'_{h_n}\|^2_{L_2(\mathbb{R})}$. Damit sind auch die Funktionen $\psi'_{\tilde{h}_n}$ in der L_2-Norm gleichgradig beschränkt. Auch hier gilt also $\psi_{\tilde{h}_n} \in W^1(\mathbb{R})$.

Mit dem Zusammenhang zwischen Glattheit der Funktion und Abklingverhalten ihrer Fouriertransformierten, der $\|\psi_{h'}\|_{L_2(\mathbb{R})} = \|\bullet\,\psi_h(\bullet)\|_{L_2(\mathbb{R})}$ für alle $h \in W^1(\mathbb{R})$ liefert, sowie der Tatsache, dass für alle $L_2(\mathbb{R})$-Funktionen h $\psi_{\psi_h} = 2\pi\, h(-\bullet)$ (siehe *Werner* [15], S. 170-173, insb. Lemma V.2.11) gilt, folgt weiter

$$
\begin{aligned}
+\infty > Const. &\ge \|\psi'_{h_n}\|^2_{L_2(\mathbb{R})} + \|\psi'_{\tilde{h}_n}\|^2_{L_2(\mathbb{R})} \\
&\textit{(Parseval-Identität)} \\
&= \frac{1}{2\pi}\|\psi_{\psi'_{h_n}}\|^2_{L_2(\mathbb{R})} + \frac{1}{2\pi}\|\psi_{\psi'_{\tilde{h}_n}}\|^2_{L_2(\mathbb{R})} \\
&= \frac{1}{2\pi}\int |t|^2|\psi_{\psi_{h_n}}(t)|^2 dt + \frac{1}{2\pi}\int |t|^2|\psi_{\psi_{\tilde{h}_n}}(t)|^2 dt \\
&= \int |t|^2 h_n(-t)^2 dt + \int |t|^2 \tilde{h}_n(-t)^2 dt \\
&= \int t^2 h_n(t)^2 dt + \int t^2 \tilde{h}_n(t)^2 dt. \qquad (5.20)
\end{aligned}
$$

Damit können wir zur Abschätzung der Terme
$\int\limits_{|y|>R_n} |h_n(y)|^2/\xi(y)dy + \int\limits_{|y|>R_n} |\tilde{h}_n(y)|^2/\xi(y)dy$ schreiten:

$$
\begin{aligned}
&\int\limits_{|y|>R_n} |h_n(y)|^2/\xi(y)dy \;+\; \int\limits_{|y|>R_n} |\tilde{h}_n(y)|^2/\xi(y)dy \\
&= \int\limits_{|y|>R_n} |h_n(y)|^2 c^{-1}|y|^{3/2}dy \;+\; \int\limits_{|y|>R_n} |\tilde{h}_n(y)|^2 c^{-1}|y|^{3/2}dy \\
&\le Const. \cdot \Big(\int\limits_{|y|>R_n} |h_n(y)|^2 y^2 \underbrace{|y|^{-1/2}}_{\le R_n^{-1/2}} dy \;+\; \int\limits_{|y|>R_n} |\tilde{h}_n(y)|^2 y^2 \underbrace{|y|^{-1/2}}_{\le R_n^{-1/2}} dy \Big) \\
&\le Const. \cdot \Big(R_n^{-1/2} \int\limits_{|y|>R_n} |h_n(y)|^2 y^2 dy \;+\; R_n^{-1/2} \int\limits_{|y|>R_n} |\tilde{h}_n(y)|^2 y^2 dy \Big) \\
&\le Const. \cdot R_n^{-1/2} \cdot \Big(\int h_n(y)^2 y^2 dy \;+\; \int \tilde{h}_n(y)^2 y^2 dy \Big) \\
&\textit{(5.20)} \\
&\le Const. \cdot R_n^{-1/2}
\end{aligned}
$$

Die Forderung (5.18) wird demnach erfüllt, wenn

$$
\frac{n \cdot \big(R_n^{3/4} \exp(-\sigma^2(\ln(C_2/C_1))(\sigma_n^2 - \sigma^2)^{-1}) \;+\; Const. \cdot R_n^{-1/4} \big)}{\|f - f_n\|^2_{L_2(\mathbb{R})}} \overset{n\to\infty}{\longrightarrow} 0
$$

gilt. Nun setze man für die Parameterfolgen $R_n = n^5$ und $\sigma_n^2 = \sigma^2 + d^{-1} \cdot (\ln n)^{-1}$ mit $d = \frac{19}{4\sigma^2 \ln(C_2/C_1)} + 1$. Daraus folgt $t_n = \sqrt{2\ln(C_2/C_1)d} \cdot (\ln n)^{1/2}$. Dann lautet die Forderung

$$
\begin{aligned}
&\frac{n \cdot (n^{15/4} \exp(-\sigma^2(\ln(C_2/C_1))d \ln n) + Const. \cdot n^{-5/4})}{\|f_n - f\|^2_{L_2(\mathbb{R})}} \\
&= \frac{n^{-\sigma^2 \ln(C_2/C_1)} + Const. \cdot n^{-1/4}}{\|f_n - f\|^2_{L_2(\mathbb{R})}} \overset{n\to\infty}{\longrightarrow} 0.
\end{aligned}
$$

Schätzen wir jetzt $\|f_n - f\|^2_{L_2(\mathbb{R})}$ nach unten ab. Für die noch zu bestimmende Folge $(m_n)_n$ werde vorerst nur vorausgesetzt, dass $T \le m_n \le t_n$ gilt, wenn n groß genug ist. Unter Ausnutzung der steigenden Monotonie der Funktion $(\exp((1/2)(\sigma^2 - \sigma_n^2)\bullet^2) - 1)^2$ im Bereich $(0, \infty)$ gilt

$$
\begin{aligned}
&\|f_n - f\|^2_{L_2(\mathbb{R})} \\
&= \tfrac{1}{2\pi}\|\varphi_n - \varphi\|^2_{L_2(\mathbb{R})} \\
&\geq \tfrac{1}{2\pi}\int_{m_n}^{t_n} |\varphi(t)|^2\,|\exp((1/2)(\sigma^2-\sigma_n^2)t^2)-1|^2\,dt \\
&\geq \tfrac{1}{2\pi}|\exp((1/2)(\sigma^2-\sigma_n^2)m_n^2)-1|^2 \int_{m_n}^{t_n} |\varphi(t)|^2\,dt \\
&\geq \tfrac{1}{2\pi}|\exp((1/2)(\sigma^2-\sigma_n^2)m_n^2)-1|^2\, C_1^2\,\tfrac{1}{2\beta-1}\left[m_n^{1-2\beta}-t_n^{1-2\beta}\right] \\
&\geq Const.\cdot(m_n^{1-2\beta}-t_n^{1-2\beta})\cdot(\exp((1/2)(\sigma^2-\sigma_n^2)m_n^2)-1)^2.
\end{aligned}
$$

Betrachten wir zunächst den Fall $1 < \beta < 5/2$. Dann definiere man $m_n = (1/2)\sqrt{2\ln(C_2/C_1)d}(\ln n)^{1/2}$, so dass $T < m_n \leq t_n$ erfüllt ist, wenn n groß genug ist. Weiter ist dann $(\sigma^2-\sigma_n^2)m_n^2$ konstant gleich $-\ln(C_2/C_1)/2$ und somit ist auch der Term $(\exp((1/2)(\sigma^2-\sigma_n^2)m_n^2)-1)^2$ konstant und positiv wegen $C_2 > C_1$. Damit ist

$$
\begin{aligned}
&\|f_n - f\|^2_{L_2(\mathbb{R})} \\
&\geq Const.\cdot(2\ln(C_2/C_1)d)^{(1/2)-\beta}(2^{2\beta-1}-1)(\ln n)^{(1/2)-\beta} \\
&\geq Const.\cdot(\ln n)^{(1/2)-\beta}.
\end{aligned}
$$

Hinreichend ist daher für die Erfüllung der Forderung (5.18), dass

$$
\frac{n^{-\sigma^2\ln(C_2/C_1)} + Const.\cdot n^{-1/4}}{(\ln n)^{(1/2)-\beta}} \stackrel{n\to\infty}{\longrightarrow} 0
$$

gilt. Da der Zähler algebraisch gegen 0 konvergiert, wogegen der Nenner eine logarithmische Rate aufweist, ist dies auch der Fall. Nachdem die Forderung (5.18) nun erfüllt ist, erhält man

$$
\sup_{g\in\mathcal{G}}\sup_{f\in\mathcal{F}} E_{f,g}\|\hat{f}_n(Y_1,\ldots,Y_n)-f\|^2_{L_2(\mathbb{R})} \geq Const.\cdot(\ln n)^{(1/2)-\beta}
$$

als untere Schranke und der Beweis ist in diesem Falle erbracht.
Kommen wir nun noch zum Fall $\beta \geq 5/2$. Hier definiere man $m_n = T$.

Da $(t_n^{1-2\beta})_n$ gegen 0 strebt, ist bei ausreichend großem n die Bedingung $T \leq m_n < t_n$ erfüllt und die Ungleichung lautet

$$
\begin{aligned}
&\|f_n - f\|^2_{L_2(\mathbb{R})} \\
&\geq \; Const. \cdot (\exp((1/2)(\sigma^2 - \sigma_n^2)T^2) - 1)^2 \\
&= \; Const. \cdot \underbrace{\left(\frac{\exp((1/2)(\sigma^2 - \sigma_n^2)T^2) - 1}{(1/2)(\sigma^2 - \sigma_n^2)T^2}\right)^2}_{\longrightarrow 1 \text{ wenn } n \to \infty} \cdot ((1/2)(\sigma^2 - \sigma_n^2)T^2)^2 \\
&\geq \; Const. \cdot (\sigma^2 - \sigma_n^2)^2 \\
&\geq \; Const. \cdot (\ln n)^{-2}
\end{aligned}
$$

Damit haben wir wieder dieselbe Situation wie im Fall $\beta < 5/2$, was die Forderung (5.18) betrifft (algebraische Konvergenzraten im Zähler und logarithmische im Nenner) und somit folgt für den MISE

$$\sup_{g \in \mathcal{G}} \sup_{f \in \mathcal{F}} E_{f,g} \|\hat{f}_n(Y_1, \ldots, Y_n) - f\|^2_{L_2(\mathbb{R})} \geq Const. \cdot (\ln n)^{-2}$$

und der Satz ist bewiesen. ∎

Anhang A

Index

Anhang B

Zusammenfassung

Die grundsätzliche Problemstellung in der Dekonvolutionsschätzung besteht darin, die Dichte f einer Zufallsvariablen aufgrund fehlerbehafteter Beobachtungen $Y_1, \ldots, Y_n$ zu schätzen. Dies wird mathematisch so modelliert, dass identisch verteilte Zufallsvariable $X_1, \ldots, X_n$ mit Verteilungsdichte f und ebenso identisch verteilte Zufallsvariable $\epsilon_1, \ldots, \epsilon_n$ mit Verteilungsdichte g gegeben seien, letztere stellen dabei den Fehler bzw. die Verunreinigung der Daten dar. Weiter seien $X_1, \epsilon_1, \ldots, X_n, \epsilon_n$ unabhängig. Beobachtet werden die Zufallsvariablen $Y_1, \ldots, Y_n$, definiert durch $Y_j = X_j + \epsilon_j$, für $j \in \{1, \ldots, n\}$. Die Dichte der Y_j ergibt sich somit als $h = f * g$, d.h. als Faltung der Dichten f und g. Ziel ist es jetzt, eine Schätzfunktion $\hat{f}_n$ für die Dichte f zu finden, die auf den Beobachtungen $Y_1, \ldots, Y_n$ basiert. Gewöhnlicherweise sind zudem noch deterministische Bedingungen nichtparametrischer Natur wie z.B. Forderungen an das Abklingverhalten der Fouriertransformierten für die Dichten f und g gestellt. Dies kann man durch Definition von Dichteklassen $\mathcal{F}$ und $\mathcal{G}$, die Teilmengen der Menge aller Dichte $\mathcal{D}$ sind, ausdrücken.
Im klassischen Ansatz der Dekonvolutionsschätzung wird die Fehlerdichte g als bekannt vorausgesetzt wie z.B. in *Stefanski und Carroll* (1990) oder *Devroye* (1989). Diese Annahme ist nicht immer realistisch. Die Arbeiten von *Efromovich* (1997) und *Neumann* (1997) befassen sich mit der Situation einer nicht exakt bekannten Fehlerdichte. Dabei wird die Fehlerdichte mittels zusätzlicher unabhängiger Beobachtungen, die direkt aus der Fehlerdichte stammen, geschätzt.
In dieser Dissertation wird erstmals der Fall beleuchtet, dass weder die Fehlerdichte exakt bekannt ist noch zusätzliche Beobachtungen vorhanden sind.

Die Informationen, mit denen die Dichte f geschätzt werden soll, sind lediglich der Datensatz $Y_1, \ldots, Y_n$ und die Mitgliedschaft von f und g in $\mathcal{F}$ bzw. $\mathcal{G}$. Ein anzustrebenden asymptotisches Qualitätsmerkmal eines Schätzers $\hat{f}_n$ in dieser Situation ist das der robust d^k-konsistenten Schätzbarkeit für einen metrischen Raum (X, d) ($k > 0$), für den $\mathcal{F} \subseteq X \cap \mathcal{D}$ postuliert wird. Dieses wird definiert als

$$E_{f,g}\, d(\hat{f}_n(Y_1, \ldots, Y_n), f)^k \quad \underset{n \to \infty}{\longrightarrow} 0 \qquad \forall f \in \mathcal{F}, \forall g \in \mathcal{G}.$$

Ein stärkeres Qualitätsmerkmal eines Schätzers $\hat{f}_n$ ist das der gleichmäßig robust d^k-konsistenten Schätzbarkeit, das definiert wird durch

$$\sup_{g \in \mathcal{G}} \sup_{f \in \mathcal{F}} E_{f,g}\, d(\hat{f}_n(Y_1, \ldots, Y_n), f)^k \quad \underset{n \to \infty}{\longrightarrow} 0.$$

Beiden Definitionen liegt die Voraussetzung zugrunde, dass es keine a-priori Gewichtung der Fehlerdichte gibt, sondern vor Bekanntwerden der empirischen Daten alle Fehlerdichten in $\mathcal{G}$ gleichberechtigt sind.

In **Kapitel 2** wird zunächst der gewöhnliche Dekonvolutionsschätzer $\hat{f}_n$ benannt, der von einer exakt bekannten Fehlerdichte g ausgeht und diese in seiner Konstruktion verwendet. Für den M(ean) I(ntegrated) S(quared) E(rror) $E\|\hat{f}_n - f\|^2_{L_2(\mathbb{R})}$ werden die asymptotischen Auswirkungen einer Missspezifikation der Fehlerdichte untersucht. Anstatt der tatsächlichen Fehlerdichte g wird eine falsche Fehlerdichte $\tilde{g}$ zur Konstruktion des Dekonvolutionsschätzers verwendet. Dabei werden sowohl an g als auch an $\tilde{g}$ entweder die Forderung

$$|\psi_g(t)| \neq 0 \quad \forall t \in \mathbb{R} \text{ und } |\psi_g(t)| \geq D|t|^{-\eta} \quad \forall t \text{ mit } |t| \geq T$$

für glatte Fehlerdichten oder für superglatte Fehlerdichten die Bedingung

$$|\psi_g(t)| \neq 0 \quad \forall t \in \mathbb{R} \text{ und } |\psi_g(t)| \geq B|t|^{\eta} \exp(-\delta|t|^{\xi}) \quad \forall t \text{ mit } |t| \geq T$$

gestellt. Diese Bedingungen sind hier definierende Eigenschaften der Dichteklasse $\mathcal{G}$ in zwei getrennten Fällen. Zudem wird noch verlangt, dass alle

Dichten aus $\mathcal{G}$, $\mathcal{F}$ quadrat-integrabel sind. $L_2(\mathbb{R})$ ist hier der metrische Raum X. Der Schnitt von $L_2(\mathbb{R})$ mit $\mathcal{D}$, der Menge aller Dichten, werde mit $\mathcal{D}_2$ bezeichnet. Satz 2.1 besagt dann, dass für den Schätzer

$$\hat{f}_{\omega_n} := (B_{\tilde{g}}^{\dagger} B_{\tilde{g}})^{-1} B_{\tilde{g}}^{t} R_{\omega_n} \hat{h}_{\omega_n}$$

gilt:

$$\sup_{f\in\mathcal{F}} E_{f,g} \, \|\hat{f}_{\omega_n} - f\|^2_{L_2(\mathbb{R})} \overset{n\to\infty}{\longrightarrow} d_{\mathcal{F}}(g,\tilde{g}),$$

wobei der Abstand $d_{\mathcal{F}}(\cdot,\cdot)$ definiert wird als

$$d_{\mathcal{F}}(g,\tilde{g}) := \frac{1}{2\pi} \sup_{f\in\mathcal{F}} \int_{\mathbb{R}} \left| \frac{\psi_g(t)}{\psi_{\tilde{g}}(t)} - 1 \right|^2 |\psi_f(t)|^2 dt.$$

Mit diesem Satz ist das asymptotische Verhalten des MISE geregelt. Die Wahl der Bandbreitenfolge $(\omega_n)_n$, die sich üblicherweise nach der Fehlerdichte richtet, um optimale Konvergenzraten zu erzielen, ist aus Sichtweise des Satzes freigestellt bis auf die Tatsache, dass im Falle der richtigen Spezifikation der Fehlerdichte ($\tilde{g} = g$) der Schätzer L_2-konsistent sein soll, also der MISE gegen 0 strebt. Auch wenn man eine Verschlechterung der Konvergenzraten im Falle der richtigen Spezifikation in Kauf nehmen wollte, kann man an der Konvergenz des MISE gegen den Abstand $d_{\mathcal{F}}(g,\tilde{g})$ nichts ändern. Der Abstandsbegriff $d_{\mathcal{F}}(g,\tilde{g})$ wird im Abschnitt 2.2 beleuchtet. Dort wird gezeigt, dass der Abstand unter sehr milden Bedingungen positiv definit ist, und diese Eigenschaft macht die Existenz eines gleichmäßig robust $\|\cdot\|^2_{L_2(\mathbb{R})}$-konsistenten Schätzers durch Auswahl einer geeigneten Fehlerdichte unmöglich. Im Abschnitt 2.3 wird dann die Frage aufgegriffen, ob eine Veränderung des Kernes (und nicht nur der Bandbreitenfolge) die asymptotische Qualität des Dekonvolutionsschätzers verbessern kann. Der Schätzer besitzt dann die allgemeine Form

$$\hat{f}_n(x) := \frac{1}{2\pi} \int \exp(-itx) L_n(t) \frac{1}{n} \sum_{j=1}^{n} e^{itY_j} / \xi(t) \, dt.$$

Dabei wird für $\psi_{\tilde{g}}$ nun auch nicht mehr zwingend die Fouriertransformierte einer Dichte aus $\mathcal{G}$ eingesetzt, sondern eine allgemeine Funktion $\xi : \mathbb{R} \to \mathbb{C}$, deren Infimum auf jedem kompakten Intervall $[-R, R]$, $R > 0$ positiv ist.

Eine solche Missspezifikation, die sogar aus der Menge der Fouriertransformierten der Dichten aus $\mathcal{G}$ hinausführt, kann zwar theoretisch vermieden werden, aber durch numerische Effekte bei Simulationen nicht vollständig ausgeschlossen werden. Satz 2.2 begräbt jedoch alle Hoffnungen, auf diese Weise eine Verbesserung der asymptotischen Qualität erreichen zu können: Unter technischen Voraussetzungen, die aus der Konsistenzforderung im Falle richtiger Spezifizierung stammen, erhält man, dass die Schätzersequenz $(\sup_{f\in\mathcal{F}} E_{f,g}\,\|\hat{f}_n - f\|^2_{L_2(\mathbb{R})})_{n\in\mathbb{N}}$ keinen Häufungspunkt besitzt, der kleiner als $\frac{1}{2\pi}\sup_{f\in\mathcal{F}} \int |\psi_f(t)|^2 \left|\frac{\psi_g(t)}{\xi(t)} - 1\right|^2 dt$ ist. Unter zusätzlichen Voraussetzungen an den Kern konvergiert bzw. divergiert $(\sup_{f\in\mathcal{F}} E_{f,g}\,\|\hat{f}_n - f\|^2_{L_2(\mathbb{R})})_{n\in\mathbb{N}}$ sogar gegen $\frac{1}{2\pi}\sup_{f\in\mathcal{F}} \int |\psi_f(t)|^2 \left|\frac{\psi_g(t)}{\xi(t)} - 1\right|^2 dt$. Der Abstand $d_{\mathcal{F}}(g,\tilde{g})$ kann auch unendlich als Wert annehmen. In Abschnitt 2.4 wird das an einem gängigen Beispiel illustriert: Die Dichteklasse $\mathcal{F}$ werde durch eine Forderung an das Abklingverhalten der Fouriertransformierten definiert, insbesondere befindet sich die Laplace-Dichte $f_L(t) = \exp(-|t|)/2$ in dieser Menge. Als Fehlerdichte stehen die Standardnormalverteilungsdichte und die Laplace-Dichte zur Auswahl. Ist die Standardnormalverteilungsdichte die tatsächliche Fehlerdichte und die Laplace-Dichte die fälscherweise angenommene, so konvergiert der MISE gegen einen endlichen Wert. Im umgekehrten Fall jedoch ist $d_{\mathcal{F}}(g,\tilde{g}) = \infty$ und der MISE divergiert somit gegen unendlich. Je mehr Beobachtungen verwendet werden, desto schlechter wird die Schätzung also. Im Zweifelsfall ist unter Robustheitsgesichtspunkten die Laplace-Dekonvolution der Normalverteilungsdekonvolution daher vorzuziehen. Ein ähnliches Problem tritt bei zwei zur Auswahl stehenden Normalverteilungsdichten mit verschiedenen Varianzen auf. Hier muss die Varianz lieber zu klein als zu groß gewählt werden, um eine Explosion des MISE zu verhindern. Wenn dagegen zwei Normalverteilungsdichten mit verschiedenen Erwartungswerten, aber derselben Varianz als Fehlerdichte konkurrieren, konvergiert der MISE in jedem Fall gegen einen endlichen Wert. Die Explosion des MISE kann bei gleichgradiger Beschränktheit der L_2-Norm der Dichten in $\mathcal{F}$ durch eine auferlegte Schranke an den Schätzer zwar vermieden werden (Satz 2.3), jedoch ist dies kein zufriedenstellendes Resultat, zumal es in der Dichteschätzung darauf ankommt, eine Dichte möglichst genau wiederzugeben.

Daran knüpft sich die Frage in **Kapitel 3** an, ob es überhaupt möglich ist, mit sinnvoll definierten Mengen $\mathcal{F}$ und $\mathcal{G}$ einen robust d^k-konsistenten Schätzer zu konstruieren – möglicherweise ganz anderer Bauart. In Abschnitt 3.1 wird die Bedingung des Nicht-Überlappens der Dichteklassen $\mathcal{F}$ und $\mathcal{G}$

$$\exists f, \tilde{f} \in \mathcal{F} \, \exists g, \tilde{g} \in \mathcal{G} \, : \, f \neq \tilde{f} \, \wedge \, f * g = \tilde{f} * \tilde{g}$$

als notwendige Bedingung der Existenz eines robust d^k-konsistenten Schätzers für irgendeine Metrik d und irgendein $k > 0$ ausgewiesen. In Abschnitt 3.2 werden Beispiele für sich überlappende Mengen $\mathcal{F}$ und $\mathcal{G}$ beschrieben, darunter befinden sich auch viele häufig auftretende Modellräume. Insbesondere wird für die Modellräume $\mathcal{F}$, die in *Fan* (1993), *Neumann* (1997), *Hesse und Meister* (2001) verwendet werden, gezeigt, dass bereits die Mehrelementigkeit der Menge $\mathcal{G}$, die äquivalent zu der nicht exakten Kenntnis der Fehlerdichte ist, bewirkt, dass es keinen robust konsistenten Schätzer bezüglich einer Metrik d geben kann. In Abschnitt 3.3 wird weiter nach notwendigen und hinreichenden Kriterien der robust konsistenten Schätzbarkeit gesucht. Zunächst wird gezeigt, dass das Postulat des Nicht-Überlappens der Dichteklassen $\mathcal{F}$ und $\mathcal{G}$ die Existenz einer dort definierten Abbildung $\Phi : \mathcal{H} \to \mathcal{F}$ mit $\mathcal{H} = \mathcal{F} * \mathcal{G}$, die den Inversen des Faltungsoperators im Falle einer bekannten Fehlerdichte ersetzt, allerdings im Allgemeinen kein linearer Operator ist, impliziert. Falls die verrauschte Dichte h bezüglich einer Metrik b konsistent schätzbar ist und die Abbildung Φ (b, d)-stetig ist, existiert ein robust d^k-konsistenter Schätzer (Satz 3.3). Gemäß dem Äquivalenztheorem von *Devroye* ist dies stets möglich, wenn es sich bei b um die $L_1(\mathbb{R})$-Metrik handelt. Daraus folgt dann Korollar 3.1, welches besagt, dass ein robust d^k-konsistenter Schätzer existiert, wenn Φ $(\|\cdot\|_{L_1(\mathbb{R})}, d)$-stetig ist. Allerdings ist der Beweis nicht konstruktiv, es wird das Auswahlaxiom gebraucht. Nun ist ein notwendiges (Existenz der Abbildung Φ) und ein hinreichendes Kriterium ($(\|\cdot\|_{L_1(\mathbb{R})}, d)$-Stetigkeit von Φ) gefunden. Angestrebt wird eine äquivalente Formulierung. Anhand eines Beispiels wird gezeigt, dass die Abbildung Φ durchaus als nicht stetige Abbildung existieren kann. Satz 3.4 zeigt weiterhin, dass die $(\|\cdot\|_{L_1(\mathbb{R})}, d)$-Stetigkeitsforderung zur Stetigkeit bis auf abzählbar viele $h \in \mathcal{H}$ abgeschwächt und dennoch die robust konsistente Schätzbarkeit bewahrt werden kann. Aber auch die pure Existenz der Abbildung Φ kann als notwendige Bedingung noch verschärft werden. In Satz 3.5 wird die stetigen Approximierbarkeit von Φ, d.h. es existiert eine Folge $(\Phi_m)_{m \in \mathbb{N}}$ von Abbildungen $\mathcal{H}$ nach $\{x \in X \mid \|x\|_X \leq C\}$, die gleichmäßig $(L_1, \|\cdot\|_X)$ -stetig sind und die Abbildung Φ punktweise approximieren, als notwendige Bedingung nachgewiesen. Dabei ist anzumerken, dass nun kein allgemeiner metrischer Raum X , sondern ein reflexiver, separabler Banachraum mit Norm $\|\cdot\|_X$ betrachtet wird, da Erwartungswertbildung über den Raum X im Beweis benötigt wird. Ein nachfolgendes Beispiel zeigt dann unter Verwendung des Satzes von Baire aus der topologischen Funktionalanalysis, dass es ein Schätzproblem gibt, bei dem die Abbildung Φ zwar existiert, aber es dennoch keinen robust d^k -konsistenten Schätzer gibt, auch wenn dieses Beispiel sehr konstruiert erscheinen mag. Im Abschnitt 3.4 wird auf quasi-gleichmäßig konsistent schätzbare Dichteklassen $\mathcal{H}$ spezialisiert, d.h.

$$E_h \|\hat{h}_n - h\|_{L_1} \leq D(h) \cdot o_n \quad , \quad \forall\, h \in \mathcal{H}.$$

Für solche Schätzprobleme kann man ein äquivalentes Kriterium der robust d^k -konsistenten Schätzbarkeit ermitteln (Satz 3.6). Dieser sagt aus, dass eine robust X^k -konsistente Schätzerfolge $(\hat{f}_n)_n$ genau dann existiert, wenn die Abbildung Φ gleichmäßig X -stetig approximierbar durch eine Folge $(\Phi_m)_{m\in\mathbb{N}}$ ist, welche noch eine Ungleichung erfüllen, die eine Verschärfung der Stetigkeit der Φ approximierenden Abbildungen darstellt, und die ständige Voraussetzung der gleichgradigen $\|\cdot\|_X$ -Beschränktheit von $\mathcal{F}$ erfüllt ist.

In **Kapitel 4** soll nun die gleichmäßig robust d^k -konsistente Schätzbarkeit beleuchtet werden. Diese ist eine stärkere Forderung als die robuste d^k -Konsistenz. Daher ist auch hier die Existenz der in Kapitel 3 definierten Abbildung Φ notwendige Bedingung. Satz 4.1 liefert die Aussage, dass sogar die gleichmäßige $(L_1(\mathbb{R}), d)$ -Stetigkeit von Φ notwendiges Kriterium ist. Daraus folgt auch, dass die $L_1(\mathbb{R})$ -Konvergenz in $\mathcal{F}$ die d -Konvergenz impliziert (Korollar 4.1). In Abschnitt 4.2 wird ein hinreichendes Kriterium der gleichmäßigen robusten d^k -Konsistenz ermittelt (Satz 4.2), die Voraussetzungen sind eine abgeschwächte Variante der gleichmäßigen b -konsistenten Schätzbarkeit der verrauschten Dichte h bezüglich einer Metrik b und die gleichmäßige (b, d) -Stetigkeit der Abbildung Φ. Nachdem Satz 4.3 diese Abschwächung gleichmäßiger b -Konsistenz für $b = \|\cdot\|_{L_1(\mathbb{R})}$ auch als notwendige Bedingung ausweist, kann der Äquivalenzsatz Satz 4.4 als Hauptsatz über gleichmäßig robuste L_1-Konsistenz mit den Bedingungen aus Satz 4.2 für $b = \|\cdot\|_{L_1(\mathbb{R})}$ formuliert werden. Auf $L_2(\mathbb{R})$ lassen sich diese Ergebnisse nicht ohne Einschränkung übertragen. Daher betrachtet man nun $L_2(\mathbb{R})$-Dichteklassen $\mathcal{H}$ mit gleichgradig kompaktem Träger. Für solche Dichteklassen wird die $L_1(\mathbb{R})$-Konvergenz von der $L_2(\mathbb{R})$-Konvergenz dominiert. Somit kann Satz 4.4 für diese Dichteklassen $\mathcal{H}$ bezüglich der $L_2(\mathbb{R})$-Metrik formuliert werden. Dies geschieht in Satz 4.5.

In **Kapitel 5** sollen diese theoretischen Ergebnisse nun angewandt werden, um (gleichmäßig) robust $L_2(\mathbb{R})^2$-konsistente Schätzer zu berechnen. Die Sätze in den Kapiteln 3 und 4 liefern (gleichmäßig) robust konsistente Schätzer, die nur gemäß dem Auswahlaxiom existieren und die zudem die Abbildung Φ benutzen. Im allgemeinen Fall ist Φ nicht in jedem Nicht-Überlappungsfall auch kalkulierbar, da Φ von den Dichteklassen $\mathcal{F}$ und $\mathcal{G}$ abhängt, von denen nur die Inklusion in der Menge aller Dichten gefordert wird. Daher betrachte ich in diesem Kapitel spezielle Dichteklassen. Im Abschnitt 5.1 stellt sich das Schätzproblem

$$\mathcal{F} := \{f \; L_2(\mathbb{R})\text{-Dichte} \mid \tfrac{1}{t^2} \geq |\psi_f(t)| \geq \tfrac{1}{2t^2}\,,\ \forall t \text{ mit } |t| \geq t_0 \geq 1\}$$

$$\mathcal{G} := \{g_L, g_N\},$$

wobei g_L die Laplace-Dichte und g_N die Standardnormalverteilungsdichte bezeichne. t_0 sei fest und bekannt. Hier ist die untere Schranke an das Abklingverhalten der Fouriertransformierten $|\psi_f(t)| \geq \frac{1}{2t^2}$ besonders wichtig. Auf diese Bedingung kann nicht gänzlich verzichtet werden, da sonst nach den Ergebnissen aus Kapitel 3 eine robust konsistente Schätzung unmöglich ist. Die Abbildung Φ setzt sich hier als Komposition der beiden inversen Faltungsoperatoren $B_{g_L}^{-1}$ und $B_{g_N}^{-1}$ zusammen, die jeweils auf eine Teilmenge einer zweielementigen Partition von $\mathcal{H}$ wirken. Die Idee der Schätzung beruht darauf, dass aufgrund des empirischen Datensatzes $Y_1, \ldots, Y_n$ getestet werden kann, ob g_L oder g_N die Fehlerdichte ist. Auf diese Weise kann ein und derselbe Datensatz zweimal verwendet werden, weswegen ich diese Methode als doppelte Ausnutzung der Empirik bezeichnet habe. Der Schätzer lautet

$$\hat{f}_n := \left(B^{\dagger}_{\hat{g}_{T,n}} B_{\hat{g}_{T,n}}\right)^{-1} B^{\dagger}_{\hat{g}_{T,n}} \hat{h}_{\omega_n}$$

mit der Bandbreitenfolgenwahl $(\omega_n)_n = (\frac{1}{2}\sqrt{\ln n})_n$ und

$$\hat{g}_{T,n}(Y_1, \dots, Y_n) := \begin{cases} g_L & , \text{ falls } \left| \frac{1}{n} \sum\limits_{j=1}^{n} e^{iTY_j} \right| \geq \frac{1}{2}(O(T) + U(T)) \\ \\ g_N & , \text{ falls } \left| \frac{1}{n} \sum\limits_{j=1}^{n} e^{iTY_j} \right| < \frac{1}{2}(O(T) + U(T)). \end{cases}$$

Als obere Schranke für den MISE des Schätzers $\hat{f}_n$ erhält man in Satz 5.1

$$\sup_{g \in \mathcal{G}} \sup_{f \in \mathcal{F}} E_{f,g} \|\hat{f}_n - f\|^2_{L_2(\mathbb{R})} \leq Const. \cdot (\ln n)^{-3/2},$$

die übliche logarithmische Konvergenzrate bei Verrauschung mit normalverteilten Fehlern. Dieses Verfahren lässt sich auch auf den Fall übertragen, dass die konkurrierenden Fehlerdichten zwei Normalverteilungsdichten mit verschiedenen Varianzen, aber identischem Erwartungswert sind. Im Abschnitt 5.2 wird nun ein Schätzproblem betrachtet, bei dem nicht nur zwei, sondern ein Kontinuum von Dichten als Fehlerdichte g zur Auswahl steht. Das Problem lautet:

$$\mathcal{F} := \{f \; L_2(\mathbb{R})\text{-Dichte} \mid C_2|t|^{-\beta} \geq |\psi_f(t)| \geq C_1|t|^{-\beta}, \forall t \text{ mit } |t| \geq T > 0\}$$

$$\mathcal{G} := \{\mathcal{N}(\mu, \sigma^2) \mid 0 < \sigma^2 \leq \sigma_0^2\}.$$

Dabei seien $C_2 > C_1 > 0$, $\beta > 1$, $T > 0$, μ, $\sigma_0^2 > 0$ bekannt, aber beliebig bis auf die Tatsache, dass C_1, C_2, T so gewählt seien, dass $\mathcal{F}$ nicht leer ist. Auch hier kann ein Schätzer bestimmt werden, und zwar

$$\hat{f}_n(x) := \frac{1}{2\pi} \int_{-\omega_n}^{\omega_n} \exp(-it(x+\mu) + \frac{1}{2}\hat{\sigma_n^2} t^2) \, \frac{1}{n} \sum_{j=1}^{n} \exp(itY_j) \, dt$$

mit der Bandbreitenfolgenwahl $(\omega_n)_{n \in \mathbb{N}} = (u_1 \cdot (\ln n)^{1/2})_{n \in \mathbb{N}}$ und

$$\hat{\sigma_n^2} := \max\Big(-\frac{2}{t_n^2} \cdot \ln\left(\frac{\hat{\varphi}_n(t_n)}{C_1 t_n^{-\beta}}\right), 0\Big),$$

wobei $(t_n)_{n\in\mathbb{N}} = (u_2 \cdot (\ln n)^{1/2})_{n\in\mathbb{N}}$ mit positiven Konstanten u_1, u_2, für welche gilt $u_1^2 + u_2^2 < \frac{1}{\sigma_0^2}$. Als Konvergenzrate des MISE erhält man nach Satz 5.2 ab einem gewissen $N \in \mathbb{N}$

$$\sup_{g\in\mathcal{G}} \sup_{f\in\mathcal{F}} E_{f,g} \|\hat{f}_n - f\|^2_{L_2(\mathbb{R})} \leq Const. \cdot \begin{cases} (\ln n)^{(1/2)-\beta} & , \text{ falls } \beta < 5/2 \\ (\ln n)^{-2} \cdot \ln(\ln(n)) & , \text{ falls } \beta = 5/2 \\ (\ln n)^{-2} & , \text{ falls } \beta > 5/2 \end{cases}$$

für alle $n \geq N$. Dieser Schätzer wurde auch implementiert und in einem Diagramm anhand Simulationen veranschaulicht. Im Abschnitt 5.3 wird schließlich gezeigt, dass für eine technische Bedingung an T und C_2 diese Konvergenzraten tatsächlich optimal für $\beta \neq 5/2$ sind. Die in Satz 5.3 gefundene untere Schranke ist für $\beta = 5/2$ von der Ordnung $(\ln n)^{-2}$, unterscheidet sich also geringfügig von der oberen Schranke aus Satz 5.2 (um Multiplikation eines iterierten Logarithmus). Damit stimmen für $1 < \beta < 5/2$ die Konvergenzraten mit denen des gewöhnlichen Dekonvolutionsschätzers bei bekannter Fehlerdichte überein (siehe *Fan* (1993)), doch stagnieren bezüglich des Exponenten des Logarithmus für $\beta > 5/2$ bei $(\ln n)^{-2}$.

Anhang C

Abstract

The basic problem in deconvolution estimation is the estimation of a probability density f based on contamined observations $Y_1, \ldots, Y_n$. Mathematically spoken, identically distributed random variables $X_1, \ldots, X_n$ with density f and also identically distributed random variables $\epsilon_1, \ldots, \epsilon_n$ with density g which represent the error or the contamination are given. Furthermore, $X_1, \epsilon_1, \ldots, X_n, \epsilon_n$ are independent. The random variables $Y_1, \ldots, Y_n$ that are defined by $Y_j = X_j + \epsilon_j$ for all $j \in \{1, \ldots, n\}$ are observed. So the density of Y_j equals $h = f * g$, i.e. the convolution of the densities f and g. Now it is our aim to find an estimator $\hat{f}_n$ of the density f based on the observations $Y_1, \ldots, Y_n$. Deterministic stipulations of nonparametric character like conditions refering to the asymptotic behaviour of the Fourier-transform for the densities f and g are made. These can be expressed by the definition of the density classes $\mathcal{F}$ and $\mathcal{G}$ which are subsets of the set of all densities.
In the classical approach of deconvolution estimation, the error density g is supposed to be known like in *Stefanski and Carroll* (1990) or *Devroye* (1989). However, this assumption is not always realistic. The papers of *Efromovich* (1997) and *Neumann* (1997) deal with the situation of an error density which is not exactly known. In both papers the error density is estimated from additional independent observations that come directly from the error distribution.
In this dissertation with the translated title "Robustness Properties of Deconvolution Density Estimators Relating to Misspecification of the Error Density" , the case that the error density is neither exactly known nor can be estimated from additional data is regarded for the first time. The informa-

tions, which the density f should be estimated from, only consists of the observations $Y_1, \dots, Y_n$ and the membership of f in $\mathcal{F}$ and of g in $\mathcal{G}$. A property of asymptotic quality of an estimator $\hat{f}_n$ in this situation is the robustly d^k-consistent estimability for a metrical space (X, d) ($k > 0$), for which $\mathcal{F} \subseteq X \cap \mathcal{D}$ is postulated. This property is defined by

$$E_{f,g}\, d(\hat{f}_n(Y_1, \dots, Y_n), f)^k \quad \underset{n\to\infty}{\longrightarrow} 0 \qquad \forall f \in \mathcal{F}, \forall g \in \mathcal{G}.$$

A stronger property of an estimator $\hat{f}_n$ is the uniformly robustly d^k-consistent estimability which is defined by

$$\sup_{g\in\mathcal{G}} \sup_{f\in\mathcal{F}} E_{f,g}\, d(\hat{f}_n(Y_1, \dots, Y_n), f)^k \quad \underset{n\to\infty}{\longrightarrow} 0.$$

Both definitions are based on the condition that there is no a-priori weight of the error densities but all error densities in $\mathcal{G}$ are equal before the empirical data become known .

In **chapter 2** the common deconvolution estimator is introduced. This estimator was developped for the situation of an exactly known g and it uses g in its construction. For the M(ean) I(ntegrated) S(quared) E(rror), $E\|\hat{f}_n - f\|^2_{L_2(\mathbb{R})}$, the asymptotic effects of a misspecification of the error density are investigated. A false error density $\tilde{g}$ instead of the real error density g is used for constructing the deconvolution estimator. Therefore, there is the condition for both $\tilde{g}$ and g

$$|\psi_g(t)| \neq 0 \quad \forall t \in \mathbb{R} \text{ and } |\psi_g(t)| \geq D|t|^{-\eta} \quad \forall t \text{ with } |t| \geq T$$

for smooth error densities or for supersmooth error densities

$$|\psi_g(t)| \neq 0 \quad \forall t \in \mathbb{R} \text{ and } |\psi_g(t)| \geq B|t|^{\eta} \exp(-\delta|t|^{\xi}) \quad \forall t \text{ with } |t| \geq T.$$

These stipulations are the defining properties of the density classes $\mathcal{G}$ in two separate cases. Furthermore, the fact that all densities in $\mathcal{F}$, $\mathcal{G}$ are square integrable is required. Here, $L_2(\mathbb{R})$ is the metric space X. The intersection of $L_2(\mathbb{R})$ and the set of all densities $\mathcal{D}$ is called $\mathcal{D}_2$. Theorem 2.1 says that the estimator

$$\hat{f}_{\omega_n} := (B_{\tilde{g}}^{\dagger} B_{\tilde{g}})^{-1} B_{\tilde{g}}^{t} R_{\omega_n} \hat{h}_{\omega_n}$$

has the asymptotic behaviour

$$\sup_{f \in \mathcal{F}} E_{f,g} \, \| \hat{f}_{\omega_n} - f \|^2_{L_2(\mathbb{R})} \stackrel{n \to \infty}{\longrightarrow} d_{\mathcal{F}}(g, \tilde{g}),$$

with the distance $d_{\mathcal{F}}(\cdot,\cdot)$ defined by

$$d_{\mathcal{F}}(g, \tilde{g}) := \frac{1}{2\pi} \sup_{f \in \mathcal{F}} \int_{\mathbb{R}} \left| \frac{\psi_g(t)}{\psi_{\tilde{g}}(t)} - 1 \right|^2 |\psi_f(t)|^2 dt.$$

In Theorem 2.1 the bandwidth sequence $(\omega_n)_n$ can be chosen arbitrarily as long as the estimator remains L_2-consistent in the case of correct specification of the error density ($\tilde{g} = g$), i.e. the MISE has to tend to zero for n tending to infinity. Even if a deterioration of the rate of convergence in the case of correct specification is accepted while the pure consistence is kept, one is not able to change something about the convergence of the MISE to $d_{\mathcal{F}}(g, \tilde{g})$. The distance $d_{\mathcal{F}}$ is illustrated in section 2.2. It is proved there that $d_{\mathcal{F}}$ is positive definite under very weak conditions and this property makes the existence of a uniformly robustly L_2-consistent density estimator which is construced by a suitable choice of the error density impossible. Section 2.3 deals with the question if a change of the kernel (and not only the bandwidth sequence) can improve the asymptotic quality of the estimator that possesses the shape

$$\hat{f}_n(x) := \frac{1}{2\pi} \int \exp(-itx) L_n(t) \frac{1}{n} \sum_{j=1}^{n} e^{itY_j} / \xi(t) \, dt.$$

Therefore, ξ is a function $\mathbb{R} \to \mathbb{C}$, whose infimum on each compact interval $[-R, R]$ is positive, and it does not have to be the Fourier-transform of a density. Such a misspecification that leads out of the set $\mathcal{G}$ may be able to be avoided theoretically, but if numerical effects are considered in a simulation for example it cannot completely be avoided. Theorem 2.2 makes all attempts for achieving a better asymptotic quality in this way fail: Under some technical conditions resulting from the postulation of convergence in the case of the correct selection of the error density, one receives that the estimator sequence $(\sup\limits_{f\in\mathcal{F}} E_{f,g}\,\|\hat{f}_n - f\|^2_{L_2(\mathbb{R})})_{n\in\mathbb{N}}$ possesses no accumulation point that is smaller than $\frac{1}{2\pi}\sup\limits_{f\in\mathcal{F}} \int |\psi_f(t)|^2 \left|\frac{\psi_g(t)}{\xi(t)} - 1\right|^2 dt$. Under additional conditions related to the kernel, the MISE converges or diverges, respectively, to $\frac{1}{2\pi}\sup\limits_{f\in\mathcal{F}} \int |\psi_f(t)|^2 \left|\frac{\psi_g(t)}{\xi(t)} - 1\right|^2 dt$. The distance $d_{\mathcal{F}}(g,\tilde{g})$ can also be infinity. In section 2.4 this fact is illustrated with a common example: The density class $\mathcal{F}$ is defined by a stipulation related to the asymptotic behaviour of the Fourier-transform. The Laplace-density $f_L(t) = \exp(-|t|)/2$ is in this set. Only the normal distribution density and the Laplace-density are possible as error density. If the normal distribution density is the real density and the Laplace-density is used, the MISE will converge to a finite positive number. In the reverse case, $d_{\mathcal{F}}$ is infinity and hence the MISE diverges to infinity. The more observations one has got the worse the estimator's risk is. Viewed under the aspects of robustness properties, the Laplace-deconvolution with the Laplace density should be prefered in comparison to the Gauss-deconvolution with normal densities. A similar problem occurs if two normal densities with different variances compete as error density. In this case, the variance has to be selected rather too large than too small in order to avoid an explosion of the MISE. However, if the set $\mathcal{G}$ consists of two normal densities with the same variance but different means then the MISE will converge to a finite positive number in every case. The estimator can be modified so that an explosion of the MISE cannot occur any more by building an upper bound for the estimator. This requires the set $\mathcal{F}$ to be uniformly bounded in the L_2-norm (theorem 2.3) but this is valid in most of the types of $\mathcal{F}$ that are considered. But this result is not satisfying because in nonparametric density estimation the density ought to be estimated very precisely.

Hence, in **chapter 3** the question is investigated whether there is a robustly d^k-consistent estimator for a proper choice of $\mathcal{F}$ and $\mathcal{G}$ at all. The estimator may possibly be of an entirely different shape. In section 3.1 the condition of non-overlappingness of $\mathcal{F}$ and $\mathcal{G}$ is introduced

$$\exists f,\tilde{f}\in\mathcal{F}\,\exists g,\tilde{g}\in\mathcal{G}\;:\; f\neq\tilde{f}\;\wedge\; f*g=\tilde{f}*\tilde{g}$$

as a necessary condition for the existence of a robustly d^k-consistent estimator for any metric d and any $k > 0$. In section 3.2 examples for overlapping density classes $\mathcal{F}$ and $\mathcal{G}$ are regarded. Among those, there are often used density classes like in *Fan* (1993), *Neumann* (1997), *Hesse and Meister* (2001). For these classes it is shown that even the fact that $\mathcal{G}$ possesses more than one element, which is equivalent to the not exact knowledge of the error distribution, is enough for the impossibility of the existence of a robustly d^k-consistent estimator. In section 3.3 the investigation for necessary and sufficient conditions for the existence of such an estimator continues. Firstly, it is shown that the condition of non-overlappingness implies the existence of a function Φ from $\mathcal{H}$ to $\mathcal{F}$ with $\mathcal{H} = \mathcal{F} * \mathcal{G}$ which is exactly defined there and which replaces the inverse convolution operator in the case of a known error density. However, in general, this function Φ is no linear operator. If the contamined and directly observed density $h \in \mathcal{H}$ can consistently be estimated related to an arbitrary metric b and the function Φ is (b, d)-continuous related to the metric d, in which the density f should be estimated, then there is a robustly d^k-consistent estimator, so says theorem 3.3. According to the equivalence theorem of *Devroye*, it is always possible to estimate a directly observed density consistently if $b = \| \cdot \|_{L_1(\mathbb{R})}$. So one can derive corollary 3.1 which says that a robustly d^k-consistent estimator exists if Φ is $(\|\cdot\|_{L_1(\mathbb{R})}, d)$-continuous. The proof is not constructive as Zorn's Lemma is used. Now a necessary and an sufficient condition for the robustly d^k-consistent estimability has been found and we aim for an equivalent condition. By an example, it is shown that Φ can exist as a non continuous function. Theorem 3.4 proves that the continuity of Φ can be changed to the weaker stipulation of the continuity of Φ outside of a countable subset of $\mathcal{H}$ and still the robustly d^k-consistent estimability holds. But on the other hand, the mere existence of Φ can also be changed to a stronger necessary condition. In theorem 3.5 the continuous approachability of Φ is proved to be a necessary stipulation, i.e. there is a sequence $(\Phi_m)_{m \in \mathbb{N}}$ of functions from $\mathcal{H}$ to $\{x \in X \mid \|x\|_X \leq C\}$, which are uniformly (but not uniformly related to the whole density class!) $(L_1, \| \cdot \|_X)$-continuous and which approach Φ pointwise. Therefore, (X, d) is no longer an arbitrary metric space but a separable reflexive Banach space because in the proof the expectation is taken over X. In the sequel an example shows that there is an estimation problem where the function Φ exists but there is no robustly d^k-consistent estimator. In section 3.4 the conditions are specialized for so called almost-uniformly d^k-consistently estimatable density classes $\mathcal{H}$, i.e. classes $\mathcal{H}$ for which

$$E_h \, \|\hat{h}_n - h\|_{L_1} \;\leq\; D(h) \cdot o_n \quad , \quad \forall\, h \in \mathcal{H}.$$

For such estimation problems, an equivalent condition of the robustly d^k-consistent estimability is obtained (theorem 3.6). This says that a robustly X^k-consistent estimator sequence $(\hat{f}_n)_n$ exists if and only if Φ is uniformly X-continuous approachable by a sequence $(\Phi_m)_{m\in\mathbb{N}}$ which even fulfills an inequality which is stronger than the uniformly X-continuous approachability, and the permanent stipulation of the $\|\cdot\|_X$-upper bound of $\mathcal{F}$ is fulfilled.

In **chapter 4** the uniformly robustly d^k-consistent estimability shall be investigated. This is a stronger condition than the robustly d^k-consistent estimability. Hence, the existence of the function Φ which is defined in chapter 3 is necessary for the existence of a uniformly robustly d^k-consistent estimator. Theorem 4.1 says that the uniformly robustly d^k-consistent estimability even requires the uniform $(L_1(\mathbb{R}), d)$-continuity of Φ. It follows from this that the $L_1(\mathbb{R})$-convergence in $\mathcal{F}$ implies the d-convergence (corollary 4.1). In section 4.2 a suuficient condition of the uniformly robust d^k-consistence is obtained (theorem 4.2), the stipulations are a weaker version of the uniformly b-consistent estimatabiliy of the directly observed and contamined density h related to the metric b and the uniform (b, d)-continuity of the function Φ. After theorem 4.3 says that this weaker version of b-consistence for $b = \|\cdot\|_{L_1(\mathbb{R})}$ is also a necessary condition, theorem 4.4 as a central theorem about the uniformly robust L_1-consistency can be proved. It contains an equivalent condition for this kind of estimability. These results cannot be transformed to $L_2(\mathbb{R})$-estimation without further restrictions. Therefore one regards $L_2(\mathbb{R})$-density classes $\mathcal{H}$ with a uniformly bounded support. For such density classes the $L_1(\mathbb{R})$-convergence is dominated by the $L_2(\mathbb{R})$-convergence. This is why theorem 4.4 can be formulated for those density classes $\mathcal{H}$ related to the $L_2(\mathbb{R})$-metric. This is accomplished in theorem 4.5.

In **chapter 5** these theoretical results are applied to calculate (uniformly) robustly $L_2(\mathbb{R})^2$-consistent estimators. The theorems in chapters 3 und 4 prove the existence of (uniformly) robustly d^k-estimators in a non constructive way by using Zorn's Lemma. Furthermore, in the general case, in which $\mathcal{F}$ and $\mathcal{G}$ only have to be subsets of $\mathcal{D}$, the function Φ is too complicated to be calculated. We consider a special example, namely

$$\mathcal{F} := \{f\ L_2(\mathbb{R})\text{-density} \mid \tfrac{1}{t^2} \geq |\psi_f(t)| \geq \tfrac{1}{2t^2}\,,\ \forall t \text{ with } |t| \geq t_0 \geq 1\}$$

$$\mathcal{G} := \{g_L, g_N\},$$

with the Laplace-density g_L and the normal density g_N. Here t_0 is fixed and known. In this case the lower bound of the Fourier-transform of the densities in $\mathcal{F}$ is very important. This condition cannot totally be eliminated, otherwise it can be shown by the results of chapter 3 that robustly $L_2(\mathbb{R})^2$-consistent estimation cannot be accomplished. The function Φ from chapter 3 is the composition of two inverse convolution operators that operate on two different sets of a partition of $\mathcal{H}$ that only consists of these two sets. The idea of the estimation procedure is to construct a test based on the empirical data $Y_1, \ldots, Y_n$ if the error density is g_N or g_L. So the given sample of independent data can be used twice, that is why I call this method double use of the empirical data. The estimator is

$$\hat{f}_n := \left(B^{\dagger}_{\hat{g}_{T,n}} B_{\hat{g}_{T,n}}\right)^{-1} B^{\dagger}_{\hat{g}_{T,n}} \hat{h}_{\omega_n}$$

with the bandwidth sequence $(\omega_n)_n = (\frac{1}{2}\sqrt{\ln n})_n$ and

$$\hat{g}_{T,n}(Y_1, \ldots, Y_n) := \begin{cases} g_L & ,\ \text{if } \left|\frac{1}{n}\sum\limits_{j=1}^{n} e^{iTY_j}\right| \geq \frac{1}{2}(O(T) + U(T)) \\[2ex] g_N & ,\ \text{if } \left|\frac{1}{n}\sum\limits_{j=1}^{n} e^{iTY_j}\right| < \frac{1}{2}(O(T) + U(T)). \end{cases}$$

As an upper bound for the MISE of this estimator one gets according to theorem 5.1

$$\sup_{g \in \mathcal{G}} \sup_{f \in \mathcal{F}} E_{f,g} \|\hat{f}_n - f\|^2_{L_2(\mathbb{R})} \leq Const. \cdot (\ln n)^{-3/2},$$

the usual slow rate in the situation of contamination with normal distributions. This procedure can also be transformed to the situation that $\mathcal{G}$ consists of two competing normal densities with different variances but identical means. In section 5.2 we arrive at an estimation problem, in which uncountably many error distributions are possible. The function classes are

$$\mathcal{F} := \{f \; L_2(\mathbb{R})\text{-density} \mid C_2|t|^{-\beta} \geq |\psi_f(t)| \geq C_1|t|^{-\beta}, \forall t \text{ with } |t| \geq T > 0\}$$

$$\mathcal{G} := \{\mathcal{N}(\mu, \sigma^2) \mid 0 < \sigma^2 \leq \sigma_0^2\}$$

where $C_2 > C_1 > 0$, $\beta > 1$, $T > 0$, μ, $\sigma_0^2 > 0$ are known but arbitrary as long as C_1, C_2, T are selected so that $\mathcal{F}$ is not empty. Also in this case an estimator can be defined by

$$\hat{f}_n(x) := \frac{1}{2\pi} \int_{-\omega_n}^{\omega_n} \exp(-it(x+\mu) + \frac{1}{2}\hat{\sigma_n^2} t^2) \, \frac{1}{n} \sum_{j=1}^{n} \exp(itY_j) \, dt$$

with bandwidth sequence $(\omega_n)_{n\in\mathbb{N}} = (u_1 \cdot (\ln n)^{1/2})_{n\in\mathbb{N}}$ and

$$\hat{\sigma_n^2} := \max\Big(-\frac{2}{t_n^2} \cdot \ln\left(\frac{\hat{\varphi}_n(t_n)}{C_1 t_n^{-\beta}}\right), 0\Big),$$

with $(t_n)_{n\in\mathbb{N}} = (u_2 \cdot (\ln n)^{1/2})_{n\in\mathbb{N}}$ with positive constants u_1, u_2, for which $u_1^2 + u_2^2 < \frac{1}{\sigma_0^2}$ is valid. As upper bound of the MISE one obtains according to theorem 5.2

$$\sup_{g\in\mathcal{G}} \sup_{f\in\mathcal{F}} E_{f,g}\|\hat{f}_n - f\|^2_{L_2(\mathbb{R})} \leq Const. \cdot \begin{cases} (\ln n)^{(1/2)-\beta} & , \text{ if } \beta < 5/2 \\ (\ln n)^{-2} \cdot \ln(\ln(n)) & , \text{ if } \beta = 5/2 \\ (\ln n)^{-2} & , \text{ if } \beta > 5/2 \end{cases}$$

if n is larger than a certain $N \in \mathbb{N}$. This estimator can also be simulated and the result is shown in a graphical box. In section 5.3 it is proved that under a technical condition for T and C_2 these rates of convergence are really optimal for $\beta \neq 5/2$. For $\beta = 5/2$ the rate differs from the upper bound of the estimator $\hat{f}_n$ by the multiplication of an iterated logarithm. For $\beta < 5/2$ the upper bound $(\ln n)^{(1/2)-\beta}$ corresponds to the one obtained by estimating the density with known error density. For $\beta > 5/2$ the upper bound remains of order $(\ln n)^{-2}$, so here we lose asymptotic quality caused by not knowing the error distribution.

Anhang D

Literaturverzeichnis

[1] Devroye, L.. *Consistent Deconvolution in density estimation.* Canadian Journal of Statistics, 17, 235-239, (1989).

[2] Efromovich, S.. *Density estimation for the case of supersmooth Measurement Error.* J. o. the American Statistical Association, 92, 526-535, (1997).

[3] Fan, J.. *On the optimal rates of convergence for nonparametric deconvolution problems.* Ann. Statist., 19, 1257-1272, (1991a).

[4] Fan, J.. *Adaptively local one-dimensional subproblems with application to a deconvolution problem.* Ann. Statist., 21, 600-610, (1993).

[5] Hesse, C.H.. *Data-Driven Deconvolution.* J. o. Nonparametric Statistics, 10, 343-373, (1999).

[6] Hesse, C.H. und Meister, A.. *Optimal Iterative Density Deconvolution.* Report, Math. Inst. A., Universität Stuttgart , (2001).

[7] Meister, A.. *Iterative Dichte-Dekonvolution aufgrund verrauschter Beobachtungen.* Diplomarbeit, Math. Inst. A , Universität Stuttgart , (2001).

[8] Neumann, M.H.. *On the effect of estimating the error density in nonparametric deconvolution.* J. o. Nonparametric Statistics, 7, 307-330, (1997).

[9] Stefanski, L.A. und Carroll, R.J.. *Deconvoluting kernel density estimators.* Statistics, 21, 169-184, (1990).

Bücher

[10] DEVROYE, L. und GYÖRFI, L.. *The Nonparametric Density Estimation. L_1 View*. New York: Wiley, 1985

[11] DURRETT, R. *Probability: theory and examples*. Duxberry Press, 2. ed., 1996

[12] HEUSER, H.. *Funktionalanalysis - Theorie und Anwendungen*. B.G. Teubner Stuttgart, 3. Auflage, 1992.

[13] LEDOUX, M. und TALAGRAND, M.. *Probability in Banach Spaces*. Springer-Verlag, 1991.

[14] SCHMITZ, N.. *Vorlesungen über Wahrscheinlichkeitstheorie*. Teubner Studienbücher Mathematik, 1996.

[15] WERNER, D.. *Funktionalanalysis*. Springer-Verlag, 2. Auflage, 1997.

Anhang E

Lebenslauf

Herkunft :

Name:	Alexander Meister
Geburtsdatum:	2. Oktober 1976
Geburtsort:	Filderstadt
Familienstand:	ledig
Staatsangehörigkeit:	deutsch

Werdegang :

1982-1986	Besuch der Grundschule in Esslingen a. N.
1986-1995	Besuch des Mörike-Gymnasiums Esslingen
1995	Allgemeine Hochschulreife
1995-2001	Studium der Mathematik (Diplomstudiengang) mit Nebenfach Physik an der Universität Stuttgart
1.4.1996-31.1.1997	Grundwehrdienst in Ellwangen an der Jagst, Beurlaubung vom Studium im SS 96 und WS 96/97
9.5.2001	Diplom in Mathematik, Diplomarbeit „*Iterative Dichte-Dekonvolution aufgrund verrauschter Beobachtungen*“, Betreuer: Prof. Dr. C.H. Hesse
seit 1.10.2001	Promotionsstudium (Mathematik) an der Universität Stuttgart, Betreuer: Prof. Dr. C.H. Hesse wissenschaftlicher Angestellter an der Universität Hohenheim