

Nächste-Nachbarn-Verfahren zur Reservierung für Einzelschäden



Diplomarbeit

Jens Martin Dittmer

Studiengang Wirtschaftsmathematik

Eingereicht im August 2005

Korrigierte Version (August 2006)

Betreuer:

Prof. Dr. Holger Drees

Schwerpunkt Mathematische Statistik und Stochastische Prozesse

Fachbereich Mathematik der Universität Hamburg

An dieser Stelle möchte ich mich bei Herrn Prof. Dr. Holger Drees ausdrücklich für die engagierte Betreuung bei meiner Diplomarbeit bedanken.

Außerdem gilt mein Dank *AON Rück* - insbesondere Frau Irina Kaiser - für die Bereitstellung der Daten sowie für den Druck und das Binden dieser Arbeit.

Inhaltsverzeichnis

1	Einleitung	5
2	Empirischer Vergleich von Reservierungsverfahren für Einzelschäden	9
2.1	Beurteilungskriterien für Reservierungsverfahren	10
2.2	Das Chain-Ladder-Verfahren für Einzelschäden	14
2.3	Verfahren auf Grundlage <i>eines</i> Vergleichsschadens	15
2.4	Verfahren auf Grundlage der k ähnlichsten Vergleichsschäden	18
3	Genauigkeit des k-nächste-Nachbarn-Schätzers	31
3.1	Normalapproximation des k -nächste-Nachbarn-Schätzfehlers	33
3.1.1	Verwandte Literatur	33
3.1.2	Definitionen und Voraussetzungen	34
3.1.3	Verteilungen der betrachteten Zufallsvariablen	38
3.1.4	Asymptotisches Verhalten von R_n und $G(R_n)$	43
3.1.5	Additive Zerlegung von $m_n(x)$	47
3.1.6	Asymptotische Normalität von $A_n(x)$	67
3.1.7	Stochastische Konvergenz von $B_n(x)$	74
3.2	Optimale Anzahl der Nachbarn und Konfidenzintervalle	77
3.2.1	Optimale Wahl von k_n	77
3.2.2	Konfidenzintervalle für $m(x)$	79
3.3	Berechnung von Konfidenzintervallen für die Beispieldaten	81
3.4	Prognoseintervalle für Y	85
4	Schlussbetrachtung	95
4.1	Fazit	95
4.2	Ausblick	96
	Literaturverzeichnis	97
	Symbolverzeichnis	100

Kapitel 1

Einleitung

Versicherungsunternehmen befinden sich bei Haftpflichtversicherungsverträgen oftmals in der Situation, für Schäden aufkommen zu müssen, die sich bereits einige Jahre oder sogar Jahrzehnte zuvor ereignet haben.

Dem Auftreten solcher *Spätschäden* kann der Umstand zugrunde liegen, dass der betreffende Schaden erst nach langer Zeit festgestellt wird, wie im Fall von Entschädigungsforderungen aufgrund von Produkt- oder Architektenfehlern.¹

Aber auch bei Schadenereignissen, die noch in der Versicherungsperiode gemeldet werden, fallen die zugehörigen Entschädigungen oft erst zu einem späteren Zeitpunkt an.

Besonders im Bereich der Haftpflichtversicherung ist der dem Versicherungsnehmer letztlich zu erstattende Betrag häufig vom Ausgang langwieriger Gerichtsverfahren oder dem Erfolg medizinischer Behandlungen abhängig.²

In einem solchen Fall leistet das Versicherungsunternehmen üblicherweise mittels jährlich variierender Zahlungen, deren Höhe sich aus den im Verlauf der Regulierung jeweils neu gewonnenen Informationen ergibt.

Eine möglichst genaue Prognose von Spätschäden beider Arten stellt nicht nur die Voraussetzung für eine risikogerechte Prämienkalkulation dar, sondern dient insbesondere auch dazu, sowohl die externe Rechnungslegung als auch die interne Erfolgsrechnung realitätsnah gestalten zu können.³

¹Der Versicherungsnehmer kann sich dabei zumeist auf denjenigen Vertrag berufen, der zum Schadeneintrittszeitpunkt gültig gewesen ist. Vgl. hierzu [Mack02, S. 221]. In der Versicherungspraxis werden solche Schäden als „Incurred But Not Reported“ (*IBNR*) bezeichnet.

² Vgl. hierzu [Mack02, S. 221]. Bis der Schaden endgültig reguliert ist, nehmen Experten Schätzungen der zu erwartenden Schäden vor. Die Differenz zwischen tatsächlichen und diesen sogenannten *reservierten* Schäden wird als „Incurred But Not Enough Reserved“ (*IBNER*) bezeichnet. In dieser Arbeit werden jedoch die Expertenschätzungen nicht näher berücksichtigt, sondern stattdessen direkt Schätzungen des eigentlichen (vollen) Schadenbetrages vorgenommen. In diesem Kontext spricht man von Schäden des Typs „Reported But Not Settled“ (*RBNS*).

³Vgl. [Mack02, S. 222].

Nahezu alle üblicherweise verwendeten statistischen Verfahren zur Schätzung der *Schadenreserve* (d.h. der prognostizierten zukünftigen Schadenzahlungen, die aus einem oder mehreren bereits eingetretenen Schäden resultieren) basieren auf einer jährlichen Kumulierung der Daten zum sogenannten *Abwicklungsdreieck*:⁴

Ausgehend davon, dass alle Schäden spätestens nach einer festen Anzahl von I *Abwicklungsjahren* vollständig reguliert sind (oder dass zumindest jeweils nur für die ersten I Abwicklungsjahre prognostiziert werden soll), betrachtet man die Summe S_{ik} über alle beobachteten Beträge, die aus dem k -ten Abwicklungsjahr eines Schadens stammen, der im i -ten *Anfalljahr* eingetreten ist. Meistens handelt es sich hierbei um *Schadenzahlungen*, die im betreffenden Jahr getätigt worden sind.⁵

Definiert man das erste Abwicklungsjahr als das Anfalljahr des Schadens und setzt voraus, dass Beobachtungen aus den vergangenen I Jahren vorliegen, so ergibt sich folgende Dreiecksgestalt der verfügbaren Daten:

$$\begin{array}{ccccccc}
 S_{11} & \cdots & S_{1,I+1-1} & \cdots & S_{1I} & & \\
 \vdots & & \vdots & & & & \\
 S_{i1} & \cdots & S_{i,I+1-i} & & & & (1.1) \\
 \vdots & & & & & & \\
 S_{I1} & & & & & &
 \end{array}$$

Verfahren, die auf diesem *Abwicklungsdreieck* basieren, sind offensichtlich gerade dann besonders effektiv, wenn die zugehörigen versicherten Risiken hinreichend homogen sind.

Bei verhältnismäßig seltenen, aber großen Schäden, die für die Versicherungsunternehmen von besonderem Interesse sind, treten dagegen typischerweise heterogene Risikokollektive auf.

Durch die beschriebene Kumulierung von Einzelschäden im Abwicklungsdreieck gehen in diesem Fall statistisch wertvolle Informationen verloren. In der vorliegenden Arbeit wird diskutiert, inwieweit Verfahren, die auf den Daten *einzelner* Schadenverläufe beruhen, besser zur Vorhersage künftig zu erbringender Leistungen bezüglich bekannter, jedoch nicht vollständig regulierter Schadenfälle⁶ geeignet sind.

[Mack02, S. 312ff.] schlägt in diesem Kontext vor,⁷ die für einen bestimmten Schaden in Zukunft anfallenden Entschädigungen zu schätzen, indem man den (bezüglich des bisherigen Abwicklungsverlaufes) „ähnlichsten“ Schaden aus der beobachteten Vergangenheit heranzieht. Unter der Annahme, dass die Zahlungsverläufe auch in der Folgezeit tendenziell übereinstimmen wer-

⁴Zu den verwendeten Bezeichnungen vgl. [Mack02, S. 223f.].

⁵Sofern man *IBNER*-Schäden (vgl. Fußnote 2) betrachtet, wird S_{ik} als Veränderung der Summe aus bereits bezahlten und (noch) reservierten Schäden aufgefasst.

⁶Es werden also *RBNS*-Schäden betrachtet, vgl. Fußnote 2.

⁷Die Ausrichtung von [Mack02] unterscheidet sich allerdings etwas von der Vorgehensweise in dieser Arbeit. Dort werden auch die Expertenschätzungen für zukünftige Leistungen miteinbezogen, so dass für jeden Schaden jährlich zwei Beträge festgehalten werden.

den, wird der ursprünglich betrachtete Schaden dann entsprechend diesem Vorbild geeignet fortgesetzt.

Anhand konkreter Daten wird in Abschnitt 2.3 untersucht, wie genau dieses Verfahren in Abhängigkeit von der Definition der „Ähnlichkeit“ bisheriger Schadenverläufe sowie von der Art der Fortsetzung prognostiziert.

Um eine Verbesserung gegenüber herkömmlichen Schadenreservierungsverfahren erreichen zu können, ist es notwendig, diesen Ansatz zu verfeinern. Der Idee aus [Mack02, S. 316] folgend, geschieht dies, indem *mehrere* (eine feste Anzahl k) beobachtete Schadenverläufe aus der Vergangenheit herangezogen werden.

Gewichtet nach der Ähnlichkeit zum betrachteten Schaden erhalten diese bei der Prognose zukünftiger Entschädigungen unterschiedlich starken Einfluss.

In Abschnitt 2.4 wird verdeutlicht, dass sich auf diese Art und Weise für eine breite Spanne an Werten für k verhältnismäßig genaue Schätzungen ergeben.

Zur Beurteilung der Schätzgenauigkeit des Verfahrens ist es sinnvoll, diesen Ansatz als nicht-parametrisches Regressionsproblem aufzufassen, bei dem eine eindimensionale Zufallsvariable (beispielsweise der Schaden im nächsten Abwicklungsjahr) durch eine mehrdimensionale Zufallsvariable (den bisherigen Schadenverlauf) erklärt werden soll.

In Anlehnung an [Mack81] wird in Kapitel 3 - unter etwas einschränkenden Annahmen - eine Normalapproximation für den Schätzfehler des eben erläuterten „*k-nächste-Nachbarn*“-Schätzers hergeleitet, mit dem Ziel, Konfidenzintervalle für die Werte der Regressionsfunktion zu berechnen.

Anschließend werden diese Resultate auf Beispieldaten angewendet. Dabei wird auch die Problematik der Herleitung von Prognoseintervallen diskutiert.

Ein kurzer Ausblick auf noch offene Fragestellungen schließt diese Arbeit ab.

Kapitel 2

Empirischer Vergleich von Reservierungsverfahren für Einzelschäden

Anhand eines relativ umfangreichen Datenbestandes aus der Haftpflichtversicherung sollen im Folgenden unterschiedliche Verfahren zur Schätzung von zu *einzelnen Schadenereignissen* gehörigen Spätschäden geeignet verglichen werden.

Aus dem Zeitraum von 1973 bis 2004 liegen konkret Abwicklungsverläufe von insgesamt 1346 Großschäden eines etablierten deutschen Versicherungsunternehmens vor.

Die Daten wurden freundlicherweise von *AON Rück* - in modifizierter Form - zur Verfügung gestellt.

Tendenziell steigt die Anzahl der angefallenen Schäden im betrachteten Zeitverlauf - von unter 10 (vor 1979) bis hin zu 96 (1996) eingetretenen Schadenereignissen.

Die Datensätze haben dabei typischerweise die Gestalt, dass der Großteil der Entschädigungszahlungen in den ersten zwei bzw. drei Abwicklungsjahren anfällt, jedoch bis hin zur vollständigen Regulierung jährliche Zahlungen auftreten, die in Ausnahmefällen auch negativ sein können (im Fall der Rückerstattung überhöhter Entschädigungen).¹

Aufgeführt sind hier sowohl noch nicht abgeschlossene als auch vollständig regulierte Schäden - mit Abwicklungsdauern von bis zu 32 Jahren.

Da die Information über Schadenzuwächse aus vergangenen Jahren genutzt werden sollen, um zukünftige Spätschäden zu prognostizieren, wären Trendeffekte aufgrund variierender Infla-

¹In einigen wenigen Fällen tritt die Situation auf, dass im Jahr, in dem ein Schaden gemeldet wird, noch keine Entschädigungszahlungen geleistet werden. Da (vgl. Fußnote 2 in Kapitel 1) keine Expertenschätzungen berücksichtigt werden, wird hier das Vorgehen gewählt, jenes Kalenderjahr als das *erste* Anfalljahr zu definieren, in dem zum ersten Mal tatsächlich Zahlungen auftreten.

tionsraten störend.

Deswegen werden die vorliegenden Daten diesbezüglich weitgehend bereinigt, indem alle Zahlungen anhand eines Konsumentenindex² in Werte des Jahres 2004 umgerechnet werden.

Abbildung 2.1 gibt einen kurzen Überblick über die Struktur dieser Daten. Neben dem Mittelwert und der Standardabweichung ist auch der geschätzte zeitliche lineare Trend der durchschnittlichen jährlichen Schadenzuwächse für die einzelnen Abwicklungsjahre dargestellt. Letzterer ist der Steigungskoeffizient einer Kleinste-Quadrate-Regression der Zahlungen im entsprechenden Abwicklungsjahr in Abhängigkeit von der Zeit. Dieser ist zunächst für eine breite Spanne an frühen Abwicklungsjahren (2 bis 10) negativ, verhält sich jedoch im späteren Verlauf genau umgekehrt. Dies lässt darauf schließen, dass es zu leichten Strukturveränderungen im Bereich der Zahlungsabwicklungen gekommen ist.³

Da diese Effekte im Folgenden vernachlässigt werden, ist es nicht überraschend, dass die Entschädigungen der ersten Abwicklungsjahre von allen betrachteten Verfahren etwas überschätzt werden.⁴

2.1 Beurteilungskriterien für Reservierungsverfahren

In diesem Abschnitt wird davon ausgegangen, dass das Ziel verfolgt wird, bei vorliegendem bekannten Schadenverlauf die unbekanntes (individuell) zu leistenden Entschädigungen des Folgejahres zu prognostizieren.

Bei anderen Zielsetzungen, z.B. der Ermittlung der Summe aller Schäden, die bis hin zu einem festen Abwicklungsjahr eintreten, lässt sich das hier vorgestellte Beurteilungsverfahren relativ leicht modifizieren (vgl. Abschnitt 2.4).

Eine naheliegende Möglichkeit, die alternativen Schätzverfahren empirisch zu beurteilen, besteht in der Anwendung des sogenannten *Back-Testing*:

Werden Daten aus I Jahren betrachtet, so versetzt man sich nacheinander in die Jahre $p \in \{2, \dots, I-1\}$ und schätzt die Schäden für die $(p+1)$ -te Periode aus den bis zu diesem Zeitpunkt vorliegenden Informationen (für Schäden, die in Jahr 1 angefallen sind, kann so natürlich keine Schätzung vorgenommen werden).

Anschließend vergleicht man dann diese Prognose mit den tatsächlich im jeweiligen Folgejahr aufgetretenen Beträgen.

²Auf die Jahre 1991 bis 2003 wird der (sich auf Gesamtdeutschland beziehende) *Verbraucherpreisindex* angewendet. Für die weiter zurückliegenden Perioden erfolgt die Inflationsbereinigung mit Hilfe des *Preisindex für die Lebenshaltung aller privaten Haushalte* für das frühere Bundesgebiet. Beide Tafeln wurden vom Statistischen Bundesamt (www.destatis.de) veröffentlicht.

³Die Entschädigungszeitpunkte verschieben sich tendenziell etwas in Richtung späterer Abwicklungsjahre. Ein Grund hierfür könnte in der steigenden Dauer von Gerichtsprozessen liegen.

⁴Beim in Abschnitt 2.1 betrachteten Back-Testing betreffen über 70% der Prognosen Schäden aus den ersten zehn Abwicklungsjahren. Dies erklärt, warum die Gesamtsumme aller Vorhersagewerte die Summe der tatsächlichen Schäden in der Regel übersteigt (vgl. Abschnitte 2.3 und 2.4).

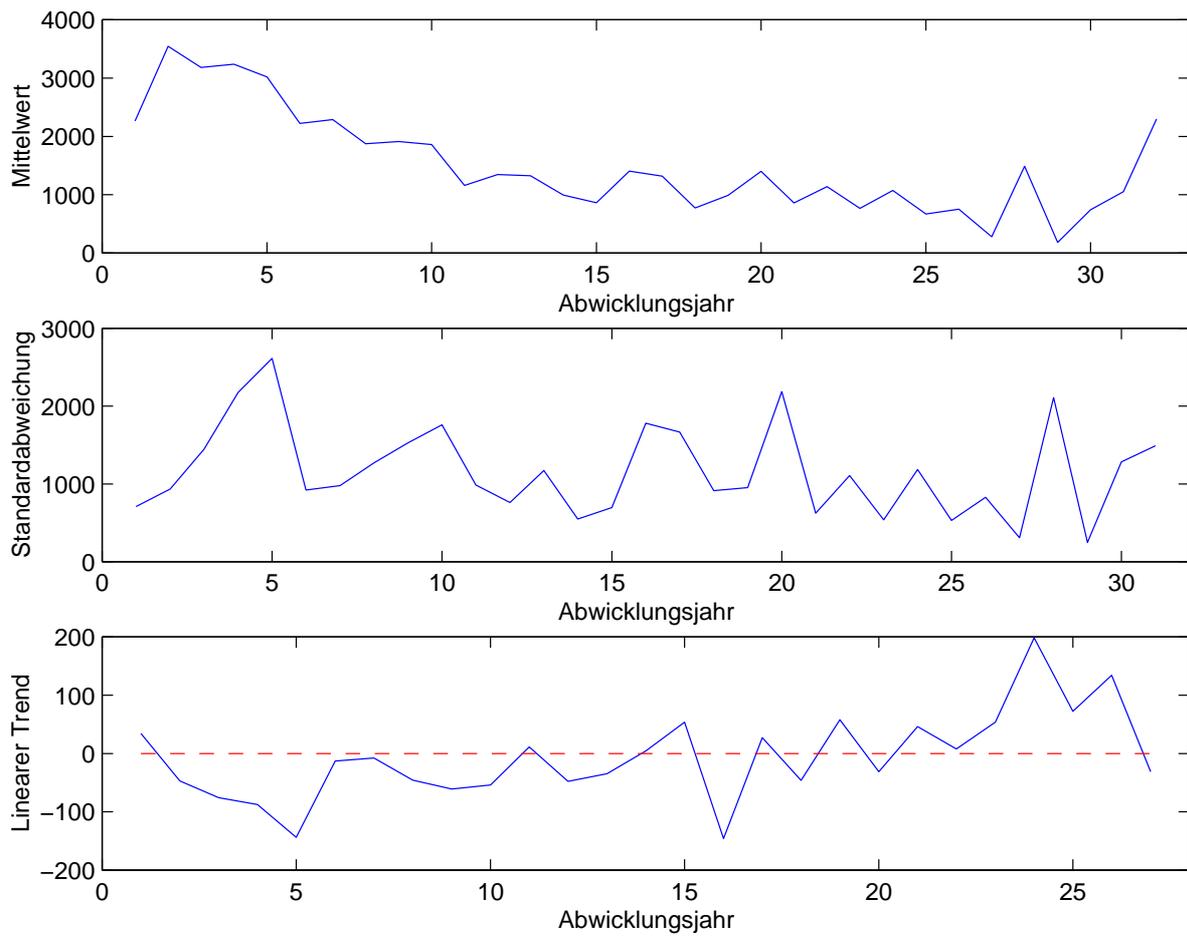


Abbildung 2.1: Durchschnittliche jährliche Schadenzuwächse in den einzelnen Abwicklungsjahren: Verlauf von Mittelwert, Standardabweichung und zeitlichem linearen Trend

Zur Vereinfachung der Notation wird in dieser Arbeit die Schreibweise

$$\underline{n} := \{1, \dots, n\}, \quad n \in \mathbb{N},$$

verwendet.

Man bezeichnet die Prognosewerte für den m -ten unter den n_{ik} Schäden, die dem i -ten Anfalljahr und dem k -ten Abwicklungsjahr zugeordnet sind, als \hat{X}_{ikm} und die (ein Jahr später) tatsächlich beobachteten Werte als X_{ikm} .⁵

Da die Anzahl der beobachteten Kalenderjahre gerade mit der der betrachteten Abwicklungsjahre übereinstimmt, hat man mit den Bezeichnungen aus (1.1) offenbar

$$S_{ik} = \sum_{m=1}^{n_{ik}} X_{ikm}, \quad i, k \in \underline{I}.$$

Entsprechend definiert man

$$\hat{S}_{ik} = \sum_{m=1}^{n_{ik}} \hat{X}_{ikm}, \quad i, k \in \underline{I}.$$

Im Hinblick auf die Beurteilung unterschiedlicher Verfahren muss zunächst festgelegt werden, welche Fragestellung im Rahmen der Schätzung am wichtigsten ist.

Liegt das Hauptaugenmerk auf einer möglichst genauen Vorhersage der Jahreszahlungen jedes einzelnen Schadens, so sollten die Residuen $\hat{X}_{ikm} - X_{ikm}$ betragsmäßig klein sein. Die Summe der quadrierten Residuen über alle Schäden aller Anfall- und Abwicklungsjahre des Zeitraums, in dem das Back-Testing vorgenommen wird,

$$SSR_{\text{Einz}} := \sum_{i=2}^{I-1} \sum_{k=2}^{I-i+1} \sum_{m=1}^{n_{ik}} (\hat{X}_{ikm} - X_{ikm})^2, \quad (2.1)$$

ist also ein adäquates Kriterium für die individuelle Genauigkeit der Prognosen.

Ist man hingegen vorrangig an einer guten Vorhersage der Summe aller Schäden übereinstimmender Anfall- und Abwicklungsjahre (d.h. S_{ik}) interessiert (lässt man also Ausgleichseffekte zwischen den n_{ik} Zahlungen zu), so ist es sinnvoll, entsprechend

$$\begin{aligned} SSR_{\text{Jahr}} &:= \sum_{i=2}^{I-1} \sum_{k=2}^{I-i+1} \left(\sum_{m=1}^{n_{ik}} \hat{X}_{ikm} - \sum_{m=1}^{n_{ik}} X_{ikm} \right)^2 \\ &= \sum_{i=2}^{I-1} \sum_{k=2}^{I-i+1} (\hat{S}_{ik} - S_{ik})^2 \end{aligned} \quad (2.2)$$

⁵Die in diesem Kapitel verwendeten Zufallsvariablen dienen nur zur Erläuterung der jeweils verwendeten Schätzverfahren. Dies geschieht (insbesondere in den Abschnitten 2.3 und 2.4) vor dem Hintergrund, die Notation möglichst eng an der später verwendeten zu halten. Da noch keine theoretischen Resultate hergeleitet werden, kann an dieser Stelle auf eine exakte stochastische Modellierung verzichtet werden, was jedoch in Kapitel 3 nachgeholt wird.

zu betrachten.

Beispielsweise für Jahreserfolgsrechnungen ist es wünschenswert, dass die gesamten Schadenaufwendungen eines bestimmten Kalenderjahres präzise geschätzt werden.

Nummeriert man das erste beobachtete Schadenanfalljahr als *erstes* Kalenderjahr, so erkennt man leicht, dass sich alle Schäden des l -ten Kalenderjahres aus solchen zusammensetzen, die dem i -ten Anfall- sowie dem $(l - i + 1)$ -ten Abwicklungsjahr zugerechnet werden.⁶

In diesem Kontext bietet es sich an, zur Beurteilung eines betrachteten Verfahrens den Wert

$$\begin{aligned} SSR_{\text{KalJ}} &:= \sum_{l=3}^I \left(\sum_{i=2}^{I-1} \sum_{m=1}^{n_{i,l-i+1}} \hat{X}_{i,l-i+1,m} - \sum_{i=2}^{I-1} \sum_{m=1}^{n_{i,l-i+1}} X_{i,l-i+1,m} \right)^2 \\ &= \sum_{l=3}^I \left(\sum_{i=2}^{I-1} \hat{S}_{i,l-i+1} - \sum_{i=2}^{I-1} S_{i,l-i+1} \right)^2 \end{aligned} \quad (2.3)$$

zu berechnen.

Eine Aussage über die Effektivität des Verfahrens liefern auch das *Gesamtresiduum*

$$\begin{aligned} R_{\text{Ges}} &:= \sum_{i=2}^{I-1} \sum_{k=2}^{I-i+1} \sum_{m=1}^{n_{ik}} \hat{X}_{ikm} - \sum_{i=2}^{I-1} \sum_{k=2}^{I-i+1} \sum_{m=1}^{n_{ik}} X_{ikm} \\ &= \sum_{i=2}^{I-1} \sum_{k=2}^{I-i+1} \hat{S}_{ik} - \sum_{i=2}^{I-1} \sum_{k=2}^{I-i+1} S_{ik}, \end{aligned} \quad (2.4)$$

bzw. das relative Gesamtresiduum

$$R_{\text{RelGes}} := \frac{R_{\text{Ges}}}{\sum_{i=2}^{I-1} \sum_{k=2}^{I-i+1} \sum_{m=1}^{n_{ik}} X_{ikm}} = \frac{\sum_{i=2}^{I-1} \sum_{k=2}^{I-i+1} \sum_{m=1}^{n_{ik}} \hat{X}_{ikm}}{\sum_{i=2}^{I-1} \sum_{k=2}^{I-i+1} \sum_{m=1}^{n_{ik}} X_{ikm}} - 1. \quad (2.5)$$

Fallen diese Größen betragsmäßig niedrig aus, so kann man davon ausgehen, dass - langfristig gesehen - die Prognosen und die tatsächlich beobachteten Schäden im Durchschnitt nahe beieinander liegen, dass also keine systematische Über- oder Unterschätzung auftritt.

Neben diesen - auf unterschiedlichen Ebenen - kumulierten (meist quadrierten) Residuen geben spezielle Quantile der empirischen Verteilung der absoluten Residuen $|\hat{X}_{ikm} - X_{ikm}|$ detailliertere Hinweise auf die Genauigkeit der Vorhersage.

Alle Beträge $|\hat{X}_{ikm} - X_{ikm}|$, $i \in \{2, \dots, I-1\}$, $k \in \{2, \dots, I-i+1\}$, $m \in \underline{n_{ik}}$, werden der Größe nach angeordnet; als α -Quantil definiert man dann den $\lceil \frac{\alpha}{n} \rceil$ -ten dieser Werte.

Ein verhältnismäßig kleines 50%-Quantil ist beispielsweise ein Anzeichen dafür, dass zumindest die Hälfte der Daten gut prognostiziert worden ist, während z.B. ein niedriges 95%-Quantil darauf hindeutet, dass nur in sehr wenigen Fällen gravierende Abweichungen vorkommen.

⁶Im Abwicklungsdreieck (1.1) befinden sich Schäden desselben Kalenderjahres gerade auf einer Diagonalen von rechts oben nach links unten.

2.2 Das Chain-Ladder-Verfahren für Einzelschäden

Die in der Praxis meistverwendete Methode zur Reservierung für Spätschäden ist das Chain-Ladder-Verfahren.⁷ Es basiert auf den durch das Abwicklungsdreieck (vgl. (1.1)) gegebenen Schadendaten und dient insofern - in erster Linie - der Prognose entsprechend jährlich zusammengefasster Entschädigungen.

Man betrachtet hier die Schadenstände

$$C_{ik} := \sum_{j=1}^k S_{ij}$$

und unterstellt, dass sich deren multiplikativen Zuwächse $\frac{C_{i,k+1}}{C_{ik}}$ (für $C_{ik} > 0$) für alle Anfalljahre i (bei festgehaltenem Abwicklungsjahr k) ähnlich verhalten.⁸

Aus den verfügbaren Daten $(C_{ik})_{i+k \leq I+1}$ schätzt man daher für jedes Abwicklungsjahr einen Zuwachsfaktor

$$\hat{f}_k := \frac{\sum_{i=1}^{I-k} C_{i,k+1}}{\sum_{i=1}^{I-k} C_{ik}}, \quad k \in \underline{I-1},$$

und prognostiziert zukünftige Schadenstände C_{ik} mit $k \in \underline{I}$ und $i+k > I+1$ dann entsprechend mittels

$$\hat{C}_{ik} := C_{i,I+1-i} \cdot \hat{f}_{I+1-i} \cdot \dots \cdot \hat{f}_{k-1}.$$

Zur Vorhersage von Einzelschäden bietet es sich an, die Faktoren \hat{f}_k ebenfalls auf Grundlage der jährweise kumulierten Zahlungen zu berechnen, diese dann jedoch auf die getrennten Daten anzuwenden.

Ein wesentliches Ziel dieses Kapitels besteht in der Entwicklung eines Schadenreservierungsverfahrens, das sich gegenüber dieser Form des Chain-Ladder-Verfahrens - bezüglich der in Abschnitt 2.1 vorgestellten (empirischen) Beurteilungskriterien - als überlegen erweist.

⁷Vgl. [Mack02, S. 320].

⁸Eine genauere stochastische Modellierung ist in [Mack02, S. 244ff.] zu finden.

2.3 Verfahren auf Grundlage eines Vergleichsschadens

In diesem Abschnitt wird dem Vorschlag von [Mack02, S. 312 ff.] gefolgt, für heterogene Risikokollektive Spätschäden dadurch zu prognostizieren, indem man davon ausgeht, dass sich ein betrachteter Schaden zukünftig tendenziell genauso weiterentwickeln wird, wie dies beim - bezüglich des bisherigen Schadenverlaufes - ähnlichsten Schaden aus den Vorjahren der Fall gewesen ist.

Mit einer empirischen Untersuchung unterschiedlicher Varianten dieses Ansatzes gewinnt man einen Eindruck darüber, welche Gestaltungsformen sich auch im Rahmen der in Abschnitt 2.4 erfolgenden Verallgemeinerung des Verfahrens (mittels Heranziehung von dann $k \in \mathbb{N}$ Vergleichsschäden (*Vorbildern*)) als vorteilhaft erweisen können.

Für einen festgehaltenen Schaden, dessen bisheriger Abwicklungsverlauf (über $p \in \mathbb{N}$ Perioden hinweg) durch den Vektor $x = (x_1, \dots, x_p)^T \in \mathbb{R}^p$ beschrieben wird, sollen die zukünftigen Entschädigungen Y der $(p+1)$ -ten Periode vorhergesagt werden. Aus der Vergangenheit liegen hierfür $n \in \mathbb{N}$ Beobachtungen mit p -jähriger Schadenabwicklung $X_\alpha = (X_{\alpha 1}, \dots, X_{\alpha p})^T$, $\alpha \in \underline{n}$, und zugehörigen Zahlungen Y_α aus der Folgeperiode vor.

Um bewerten zu können, wie nahe bestimmte Abwicklungsverläufe beieinander liegen, muss folglich ein geeignetes Abstandsmaß auf dem \mathbb{R}^p definiert werden.

Misst man den Beträgen aller p Abwicklungsjahre dieselbe Bedeutung zu, kann der euklidische Abstand $d_p(X_\alpha, x)$ zwischen den Schadenverlaufsvektoren X_α und x verwendet werden, für

$$d_p(z_1, z_2) := \|z_1 - z_2\|_2, \quad z_1, z_2 \in \mathbb{R}^p.$$

Eine andere Möglichkeit besteht darin (ähnlich wie beim Chain-Ladder-Verfahren) kumulierte Schadenstände zugrunde zu legen.

Mit der Bezeichnungswise (für beliebiges $z \in \mathbb{R}^p$)

$$z_k^{\text{cum}} := \sum_{j=1}^k z_j, \quad k \in \underline{p},$$

sowie $z^{\text{cum}} := (z_1^{\text{cum}}, \dots, z_p^{\text{cum}})^T$, lassen sich die durch X_α , $\alpha \in \underline{n}$, bzw. x charakterisierten Abwicklungsverläufe alternativ ohne Informationsverlust mittels

$$X_\alpha^{\text{cum}} = (X_{\alpha 1}^{\text{cum}}, \dots, X_{\alpha p}^{\text{cum}})^T$$

bzw.

$$x^{\text{cum}} = (x_1^{\text{cum}}, \dots, x_p^{\text{cum}})^T$$

beschreiben.

Auch deren euklidischer Abstand $d_p^{\text{cum}}(X_\alpha, x)$, für

$$d_p^{\text{cum}}(z_1, z_2) := \|z_1^{\text{cum}} - z_2^{\text{cum}}\|_2, \quad z_1, z_2 \in \mathbb{R}^p,$$

erscheint als sinnvolles Maß für die Ähnlichkeit von X_α zu x .

Gerade bei der Betrachtung kumulierter Daten stellt sich zudem die Frage, ob nicht ein Vergleich lediglich der Schadenstände $X_{\alpha p}^{\text{cum}}$ und x_p^{cum} des letzten bekannten Abwicklungsjahres aussagekräftiger ist.

Daher werden im Folgenden auch die durch

$$d_1(z_1, z_2) := |z_{1p} - z_{2p}|, \quad z_1, z_2 \in \mathbb{R}^p, \quad (2.6)$$

bzw.

$$d_1^{\text{cum}}(z_1, z_2) := |z_{1p}^{\text{cum}} - z_{2p}^{\text{cum}}|, \quad z_1, z_2 \in \mathbb{R}^p, \quad (2.7)$$

definierten Abstandsmaße mit in die Überlegungen einbezogen.

Sei \tilde{X} derjenige Abwicklungsverlauf, dessen Abstand zu x (unter allen X_α , $\alpha \in \underline{n}$) minimal ist,⁹ und \tilde{Y} der zugehörige Betrag aus der $(p+1)$ -ten Periode. Dann wird \tilde{Y} dazu verwendet, den durch x dargestellten Schaden entsprechend fortzusetzen.

Gemäß [Mack02, S. 314] hat man hier die Auswahl zwischen einem multiplikativen und einem additiven Ansatz.

Eine *additive* Fortsetzung des betrachteten Schadens x bedeutet dabei, dass Y mittels

$$\hat{Y}_{(\text{add})} := \tilde{Y}$$

geschätzt wird, also einfach durch den Wert der Schadenzahlungen, die beim nächstgelegenen Vorbild in der dort folgenden Periode aufgetreten sind.

Eine *multiplikative* Fortsetzung hingegen sieht - analog zur Grundidee des Chain-Ladder-Verfahrens - vor, dass die proportionalen Zuwächse des kumulierten Schadenstandes im $(p+1)$ -ten Jahr sich bei jedem Schaden ähnlich verhalten, so dass man Y hier durch

$$\hat{Y}_{(\text{mult})} := x_p^{\text{cum}} \cdot \frac{\tilde{Y}}{\tilde{X}_p^{\text{cum}}}$$

vorhersagt.

Die Begriffe *additiv* und *multiplikativ* ergeben sich also offenbar daraus, dass (mit $\tilde{X}_{p+1}^{\text{cum}} := \tilde{X}_p^{\text{cum}} + \tilde{Y}$) für die geschätzten kumulierten Schadenstände nach $p+1$ Perioden

$$x_p^{\text{cum}} + \hat{Y}_{(\text{add})} = x_p^{\text{cum}} + (\tilde{X}_{p+1}^{\text{cum}} - \tilde{X}_p^{\text{cum}})$$

bzw.

$$x_p^{\text{cum}} + \hat{Y}_{(\text{mult})} = x_p^{\text{cum}} \cdot \frac{\tilde{X}_{p+1}^{\text{cum}}}{\tilde{X}_p^{\text{cum}}}$$

⁹Es wird davon ausgegangen, dass \tilde{X} eindeutig bestimmt ist. Vgl. hierzu Fußnote 13.

gilt.

Das in Abschnitt 2.1 erläuterte Back-Testing wird nun auf die beschriebenen unterschiedlichen Varianten der Vergleichsschaden-Methode angewendet.¹⁰

Ein Vergleich der Ergebnisse (Tabelle 2.1) zeigt, dass die Prognosegenauigkeit des Chain-Ladder-Verfahrens diesen gegenüber klar überlegen ist.

Speziell multiplikative Verfahren schneiden sehr schlecht ab. Dies ist hauptsächlich darauf zurückzuführen, dass bei einigen Schadenverläufen zwar in den früheren Abwicklungsjahren nur geringe Entschädigungen auftreten, später jedoch deutlich höhere Zahlungen beobachtet werden. Relative Zuwächse des kumulativen Schadenstandes fallen dann sehr hoch aus. Additive Verfahren haben naturgemäß weniger mit diesem Problem zu kämpfen.¹¹

Darüberhinaus lassen die Werte aus Tabelle 2.1 die auf kumulierten Daten basierende Quantifizierung der Ähnlichkeit als leicht effizienter erkennen.

Da sich allerdings keine deutliche Überlegenheit der Verwendung von d_p^{cum} gegenüber d_1^{cum} herausstellt, wird die Problematik der Wahl des Abstandes im Rahmen eines allgemeineren Ansatzes (vgl. (2.10)) weiter verfolgt.

Verfahren	Metrik	Fortsetzung	SSR_{Einz}	SSR_{Jahr}	R_{Ges}	R_{RelGes}
Vergleichsschaden	d_p	Additiv	$5,077 \cdot 10^{11}$	$5,719 \cdot 10^{11}$	1377834	0,0501
Vergleichsschaden	d_p	Multiplikativ	$7,219 \cdot 10^{11}$	$8,628 \cdot 10^{11}$	4432152	0,1612
Vergleichsschaden	d_p^{cum}	Additiv	$4,849 \cdot 10^{11}$	$5,476 \cdot 10^{11}$	2192812	0,0798
Vergleichsschaden	d_p^{cum}	Multiplikativ	$5,569 \cdot 10^{11}$	$6,443 \cdot 10^{11}$	3116524	0,1134
Vergleichsschaden	d_1	Additiv	$5,144 \cdot 10^{11}$	$1,142 \cdot 10^{12}$	1453919	0,0529
Vergleichsschaden	d_1	Multiplikativ	$1,300 \cdot 10^{14}$	$1,975 \cdot 10^{14}$	41219723	1,4993
Vergleichsschaden	d_1^{cum}	Additiv	$5,088 \cdot 10^{11}$	$5,407 \cdot 10^{11}$	782133	0,0284
Vergleichsschaden	d_1^{cum}	Multiplikativ	$5,187 \cdot 10^{11}$	$5,606 \cdot 10^{11}$	1120215	0,0407
Chain-Ladder			$3,138 \cdot 10^{11}$	$4,270 \cdot 10^{11}$	1909473	0,0695

Tabelle 2.1: Prognosegenauigkeit der Vergleichsschaden-Methode - im Vergleich mit dem Chain-Ladder-Verfahren

¹⁰Die in dieser Arbeit diskutierten Verfahren sind ausnahmslos in *Microsoft Visual Basic 6.0* (eingebettet in *Microsoft Access 2000*) programmiert worden. Die entsprechenden Datenbanken, der extrahierte Code und genauere Beschreibungen der Programme befinden sich auf der beigelegten CD.

Die zu diesem Abschnitt gehörigen Dateien sind im Ordner *EinNachbar* gespeichert.

¹¹Aus demselben Grund scheidet auch der Versuch, den Schadenverlauf derart zu standardisieren, dass (ab $p = 2$ beobachteten Abwicklungsjahren) statt der Werte x_i , $i \in \underline{p}$, nun x_i/\bar{x} (bzw. $x_i^{\text{cum}}/\bar{x}^{\text{cum}}$ statt x_i^{cum} , $i \in \underline{p}$) betrachtet werden (und analog für $X_{\alpha i}$ bzw. $X_{\alpha i}^{\text{cum}}$, $i \in \underline{p}$, $\alpha \in \underline{n}$), um zu erreichen, dass zwei Schäden, deren Zahlungen bis auf einen festen Faktor identisch verlaufen, der Abstand Null zugewiesen wird. Dabei ist $\bar{x} := \frac{1}{p} \sum_{i=1}^p x_i$ und $\bar{x}^{\text{cum}} := \frac{1}{p} \sum_{i=1}^p x_i^{\text{cum}}$. In diesem Kontext ist dann natürlich konsequenterweise eine multiplikative Fortsetzung des Schadenverlaufes zu wählen.

2.4 Verfahren auf Grundlage der k ähnlichsten Vergleichsschäden

Anhand der Analyse der unterschiedlichen Varianten des auf einem Vergleichsschaden beruhenden Reservierungsverfahrens ist also deutlich geworden, wie diese Methode verallgemeinert werden sollte.

Eine plausible Erklärung für die Unterlegenheit der Ansätze aus Abschnitt 2.3 gegenüber dem Chain-Ladder-Verfahren findet sich im asymptotischen Verhalten der Schätzer. Da für die Schätzung nur ein Vorbild herangezogen wird (also in keinerlei Weise über mehrere Daten gemittelt wird), ist offensichtlich auch bei beliebig wachsender Anzahl verfügbarer beobachteter Schadendaten *nicht* gewährleistet, dass sich die Schätzung entsprechend stabilisiert (beispielsweise um den bedingten Erwartungswert von Y bei gegebenen Schäden x_1, \dots, x_p).

Mit der Einbeziehung mehrerer ähnlicher Vergleichsschäden („Nachbarn“) kann dieser Nachteil behoben werden.

Es bietet sich weiterhin an, den Einfluss der verschiedenen Nachbarn so zu steuern, dass Vorbilder mit geringem Abstand zum betrachteten Schaden die größte Gewichtung erhalten.¹²

Die Resultate aus Abschnitt 2.3 motivieren dabei die Verwendung additiver Fortsetzungen des Schadenverlaufes, d.h. die direkte Prognose der Schäden der Folgeperiode - ohne der Umweg über Zuwachsfaktoren.

Außerdem erscheint es ratsam, für die Berechnung der Abstände kumulierte Daten zu verwenden. Einerseits führt dies bei den Verfahren aus Abschnitt 2.3 zu etwas besseren Ergebnissen, andererseits lassen sich auf einfache Art und Weise Abstandsmaße für Schadenverläufe einführen, die die einzelnen Werte der Abwicklungsjahre unterschiedlich gewichten. Dies kann sinnvoll sein, da in kumulierten Schadenständen natürlich auch Informationen über die Vorperioden mitenthalten sind.

So ist es beispielsweise anschaulich nachvollziehbar, dass die Abstandsmessung d_1^{cum} aus (2.7) der Abstandsmessung d_1 aus (2.6) vorzuziehen ist.

Definiert man für ein festes $a \in [0, \infty)^p$

$$\|u\|_a := \sqrt{\sum_{i=1}^p a_i u_i^2}, \quad u \in \mathbb{R}^p, \quad (2.8)$$

so lassen sich hiermit allgemein Metriken d_a durch

$$d_a(z_1, z_2) := \|z_1 - z_2\|_a, \quad z_1, z_2 \in \mathbb{R}^p, \quad (2.9)$$

festlegen.

¹²[Mack02, S. 316] schlägt einen auf dieser Idee beruhenden Ansatz vor, ohne jedoch nähere theoretische oder praktische Untersuchungen durchzuführen.

Durch die spezielle Auswahl von $a \in [0, \infty)^p$ wird der Einfluss der Schadenstände der einzelnen Abwicklungsjahre auf die Quantifizierung der Ähnlichkeit gesteuert, denn mit den Bezeichnungen aus Abschnitt 2.3 gilt

$$d_a(x_\alpha^{\text{cum}}, x^{\text{cum}}) := \sqrt{\sum_{i=1}^p a_i (X_{\alpha i}^{\text{cum}} - x_i^{\text{cum}})^2}. \quad (2.10)$$

Dies ist offenbar eine Verallgemeinerung der Abstandsmessungen d_p^{cum} bzw. d_1^{cum} , denn mit $a = (1, \dots, 1)^T$ hat man $d_a = d_p^{\text{cum}}$ und für $a = (0, \dots, 0, 1)^T$ ergibt sich $d_a = d_1^{\text{cum}}$.

Bezeichnen $\tilde{X}_1, \dots, \tilde{X}_{k-1}$ (mit zugehörigen $\tilde{Y}_1, \dots, \tilde{Y}_{k-1}$) in analoger Darstellungsweise wie in Abschnitt 2.3 diejenigen Vergleichsschäden unter den $(X_\alpha)_{\alpha \in \underline{n}}$, die bezüglich d_a die geringsten Abstände zu x aufweisen,¹³ so wird nun also der („ k -nächste-Nachbarn“-Schätzer),

$$\hat{Y}_{\text{kNN}} = \sum_{i=1}^{k-1} w_i \tilde{Y}_i \quad (2.11)$$

betrachtet. Dabei seien $(w_i)_{i \in \underline{k-1}}$ positive reelle Zahlen mit

$$\sum_{i=1}^{k-1} w_i = 1.$$

Damit - wie gewünscht - das Gewicht w_i für den Prognosewert \tilde{Y}_i umso größer ist, je kleiner der Abstand von \tilde{X}_i zu x ausfällt, führt man einen sogenannten *Kern* ein, eine integrierbare Funktion

$$K : \mathbb{R}^p \longrightarrow [0, \infty)$$

mit¹⁴

$$K(u) \geq K(v), \quad \text{falls } \|u\| \leq \|v\|.$$

Ist $\sum_{j=1}^k K\left(\frac{\tilde{X}_j - x}{R}\right) > 0$, so definiert man

$$w_i := \frac{K\left(\frac{\tilde{X}_i - x}{R}\right)}{\sum_{j=1}^k K\left(\frac{\tilde{X}_j - x}{R}\right)}, \quad i \in \underline{k-1}, \quad (2.12)$$

¹³Im stochastischen Modell aus Kapitel 3 besitzen die betrachteten Abstände eine stetige Verteilung (vgl. Satz 3.1.2), so dass davon ausgegangen werden kann, dass die Auswahl der $\tilde{X}_1, \dots, \tilde{X}_{k-1}$ stets eindeutig ist. Ansonsten müsste hier eine zusätzliche Entscheidungsregel eingeführt werden.

¹⁴Diese Eigenschaft ist nicht aus technischen Gründen wichtig, sondern resultiert lediglich aus der Überlegung, dass Vergleichsschäden, die in geringerem Abstand zu x liegen, stets stärkeren Einfluss erhalten sollen. Im Rahmen der Herleitung der theoretischen Resultate in Kapitel 3 wird diese Bedingung nicht vorausgesetzt. Stattdessen wird häufig die - hier offenbar unerhebliche - Normierung $\int K(u) du = 1$ gefordert. Vgl. hierzu Fußnote 6 in Kapitel 3.

wobei R den Abstand von x zum (bezüglich d_a) entferntesten der k benachbarten Vergleichsschäden (unter den $(X_\alpha)_{\alpha \in \underline{n}}$) beschreibt,¹⁵

$$R := \inf \left\{ t \geq 0 \mid \sum_{\alpha=1}^n 1_{\{\|X_\alpha - x\| \leq t\}} \geq k \right\}.$$

Oft ist es zweckmäßiger, \hat{Y}_{kNN} als gewichtete Summe *aller* Y_α , $\alpha \in \underline{n}$, darzustellen:

Nach Definition von w_i und R ist es für den Wert \hat{Y}_{kNN} offenbar unerheblich, wie K außerhalb der Menge $\{u \in \mathbb{R}^p \mid \|u\|_a < 1\}$ definiert ist. Setzt man hier $K(u) = 0$ für alle $u \in \mathbb{R}^p$ mit $\|u\|_a \geq 1$, so gilt

$$K\left(\frac{X_\alpha - x}{R}\right) = 0,$$

sofern X_α , $\alpha \in \underline{n}$, nicht zu den $(\tilde{X}_j)_{j \in \underline{k-1}}$ gehört, und daher auch

$$\hat{Y}_{\text{kNN}} = \frac{\sum_{\alpha=1}^n K\left(\frac{X_\alpha - x}{R}\right) Y_\alpha}{\sum_{\alpha=1}^n K\left(\frac{X_\alpha - x}{R}\right)}. \quad (2.13)$$

Für unterschiedliche betrachtete Schadenverläufe x variiert natürlich der Abstand R zum k -ten Vorbild.

Dies charakterisiert den wesentlichen Unterschied zwischen dem k -nächste-Nachbarn-Schätzer \hat{Y}_{kNN} und dem verwandten *Nadaraya-Watson-Kernschätzer*, der in diesem Kontext für ein festes $h > 0$ die Gestalt

$$\hat{Y}_{\text{NW}} := \frac{\sum_{\alpha=1}^n K\left(\frac{X_\alpha - x}{h}\right) Y_\alpha}{\sum_{\alpha=1}^n K\left(\frac{X_\alpha - x}{h}\right)} \quad (2.14)$$

hat.

Im Gegensatz zu R ist die *Bandweite* h (Radius der Kugel, auf der $K(u) \neq 0$ gilt) deterministisch, was die asymptotische Untersuchung von \hat{Y}_{NW} (gegenüber \hat{Y}_{kNN}) etwas vereinfacht.

Stattdessen ist hier jedoch die Anzahl der Nachbarn, über die (gewichtet) gemittelt wird, zufallsabhängig. Für die praktische Anwendung auf die betrachteten heterogenen Schadendaten taucht daher das Problem auf, dass - zumindest wenn K weiterhin ein Kern mit kompaktem Träger ist - \hat{Y}_{NW} nicht definiert ist, sofern sich in einer Kugel mit Radius h um x (bezüglich $\|\cdot\|_a$) kein weiterer Schaden befindet. Wird h dagegen zu groß gewählt, erhalten auch Schäden, deren Abwicklungsverläufe verhältnismäßig weit entfernt von x liegen, eine hohe Gewichtung. So sinkt der Einfluss der ähnlichsten Schäden entsprechend, was tendenziell zu einer ungenaueren Schätzung führt.

¹⁵Dies ist also die Entfernung zum nächsten an x gelegenen Vergleichsschaden, der *nicht* mehr an der Summe (2.11) beteiligt ist. Die Bezeichnung *k-nächste-Nachbarn-Schätzer* rührt somit daher, dass zur Bestimmung von \hat{Y}_{kNN} die k nächstgelegenen Vorbilder ermittelt werden, auch wenn in (2.11) lediglich $k-1$ Summanden auftreten.

Liegen die X_α , $\alpha \in \underline{n}$, sehr ungleichmäßig im \mathbb{R}^p verteilt, müsste man daher h in Abhängigkeit von x festlegen (h wird man sinnvollerweise umso größer wählen, je weniger Werte X_α in einer näheren Umgebung von x liegen, vgl. [Staniswalis89, S. 284]).

Beim k -nächste-Nachbarn-Schätzer \hat{Y}_{kNN} erfolgt hingegen automatisch eine entsprechende Anpassung des Radius R an die Anzahl der nahe an x liegenden Vergleichsschäden. Folglich ist der Schätzer - auch bei einer von x unabhängigen Festlegung von k - stets definiert.

Gerade im Hinblick auf die erforderliche Automatisierung des Verfahrens im Rahmen des Back-Testing erweist sich diese Methode daher als wesentlich praktikabler.

Zur Anwendung des Schätzverfahrens auf die Haftpflicht-Schadendaten wird der p -dimensionale *Epanechnikov-Kern*¹⁶

$$K(u) := \kappa(1 - \|u\|_a^2)1_{[0,1]}(\|u\|_a), \quad \kappa > 0, \quad (2.15)$$

gewählt. Aus (2.12) erkennt man, dass w_i und damit auch \hat{Y}_{kNN} nicht vom Wert $\kappa > 0$ abhängt.

Diese Gewichtsfunktion wird in der Praxis häufig verwendet, auch weil [Gasser/Müller79, S. 41ff.] (für den Fall $p = 1$) im Rahmen der Untersuchung eines verwandten Kernschätzers gewisse Optimalitätseigenschaften des Epanechnikov-Kerns nachgewiesen haben. Vor diesem Hintergrund erscheint die spezielle Wahl von K als nicht unangebracht.

Um die Auswirkungen der in Abbildung 2.1 dargestellten Trendeffekte zu begrenzen, sollen die jeweiligen Vorhersagen zukünftiger Zahlungen nur noch auf den Daten der (maximal) zehn zurückliegenden Kalenderjahre basieren.¹⁷ Weiter zurückliegende Schäden bleiben also unberücksichtigt.¹⁸

Zur Abstandsmessung wird zunächst der euklidische Abstand d_p^{cum} verwendet (d.h. $a = (1, \dots, 1)^T$ in (2.10)).

¹⁶Vgl. [Györfi02, S. 70].

¹⁷Bei Prognosen für in den Kalenderjahren 1974 bis 1982 angefallene Schäden stehen natürlich weiterhin weniger als zehn Jahre zur Verfügung.

¹⁸Natürlich werden dann auch die Abwicklungsfaktoren des Chain-Ladder-Verfahrens nur noch auf Grundlage dieses Zeitraumes berechnet.

Die Anzahl der Nachbarn k wird - der Einfachheit halber - im Back-Testing-Verfahren über alle Abwicklungsjahre hinweg konstant gehalten, sofern $n \geq k$ Vergleichsdatensätze zur Verfügung stehen (ansonsten wird $k = n$ gewählt).

Für die sich so ergebenden Schätzwerte $\hat{Y}_{k,NN}$ werden (für unterschiedliche k) die Beurteilungskriterien SSR_{Einz} , SSR_{Jahr} , SSR_{KalJ} und R_{RelGes} (aus Abschnitt 2.1) berechnet¹⁹ und mit denen des Chain-Ladder-Verfahrens verglichen.

Aus Abbildung 2.2 wird dabei deutlich, dass das k -nächste-Nachbarn-Verfahren für eine breite Spanne an Werten für k eine wesentlich bessere Prognose der zukünftigen Schäden ermöglicht.

Insbesondere erkennt man an den Ausprägungen von SSR_{Jahr} und SSR_{KalJ} , dass diese Methode auch dann vorzuziehen ist, wenn Interesse an der Vorhersage der kumulierten Schäden eines festen Abwicklungs und Anfalljahres (wie ursprünglich beim Chain-Ladder-Verfahren) oder sogar an den Beträgen für das gesamte Kalenderjahr besteht.

Einen weiteren interessanten Aspekt zum Vergleich von Chain-Ladder- und k -nächste-Nachbarn-Verfahren trägt Abbildung 2.3 bei, die die unterschiedlichen Quantile der absoluten Residuen (vgl. Abschnitt 2.1) zeigt.

Dort wird deutlich, dass das Chain-Ladder-Verfahren zwar eine breite Masse der Daten etwas besser prognostiziert, dafür jedoch auch wesentlich häufiger starke Abweichungen von den tatsächlichen Schäden verursacht.

Eine genauere Untersuchung zeigt, dass die betragsmäßig hohen Residuen tendenziell bei größeren Schäden auftreten.²⁰

Es bestätigt sich also die Vermutung, dass das Chain-Ladder-Verfahren eher in homogenen, die k -nächste-Nachbarn-Methode eher in heterogenen (und großschadenlastigen) Risikokollektiven bevorzugt anzuwenden ist.

¹⁹Die Programme (und Erläuterungen) für die in diesem Abschnitt angewendeten Verfahren befinden sich auf der beigelegten CD im Ordner *kNachbarn*. Im Unterordner *Folgejahr* sind die Dateien der zu den Abbildungen 2.2 bis 2.5 gehörigen Verfahren gespeichert. Die später (Tabelle 2.2) diskutierte Prognose von Schadenssummen über mehrere Abwicklungsjahre ist im Unterordner *MehrereJahre* dokumentiert.

²⁰Die durchschnittliche Entschädigungszahlung - berechnet aus allen Schäden aller Anfall- und Abwicklungsjahre - beträgt 2168,5. Mittelt man hingegen nur über diejenigen dieser Schäden, deren Residuen zu den betragsmäßig größten 10% gehören, so ergibt sich ein Wert von 10688,4.

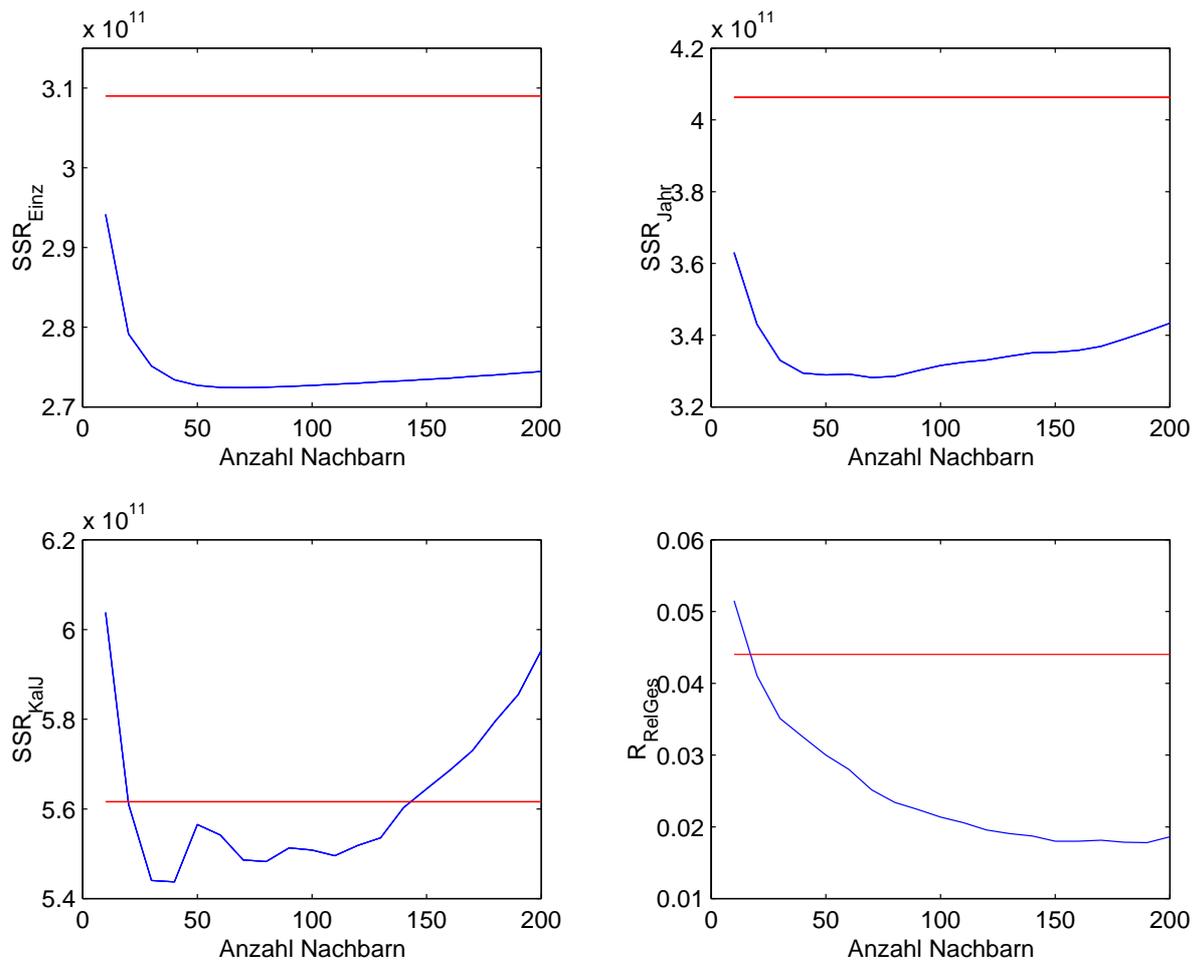


Abbildung 2.2: Beurteilungskriterien für den k -nächste-Nachbarn-Schätzer (blau) im Vergleich mit dem Chain-Ladder-Verfahren (rot).

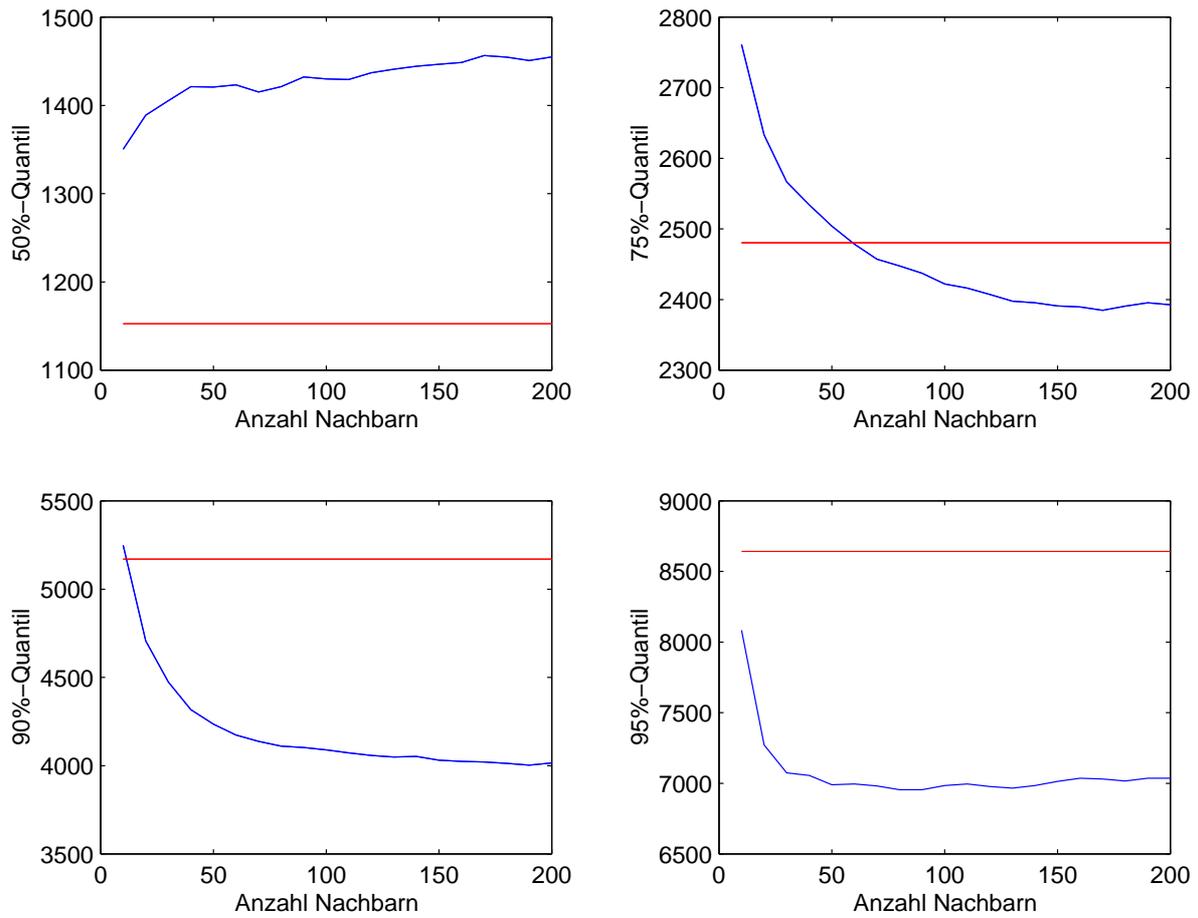


Abbildung 2.3: Quantile der absoluten Residuen beim k -nächste-Nachbarn-Schätzer (blau) und beim Chain-Ladder-Verfahren (rot).

Da die in diesem Abschnitt verwendete Quantifizierung der Ähnlichkeit auf kumulierten Daten (Schadenständen) beruht, beinhalten die Werte der späteren Abwicklungsjahre des Schadenverlaufes auch Informationen über die davorliegenden Perioden.

So stellt sich die Frage, ob mit der Verwendung einer Abstandsmessung d_a (entsprechend (2.9)) mit

$$a_p \geq a_{p-1} \geq \dots \geq a_1 \geq 0$$

die Genauigkeit der Vorhersage noch gesteigert werden kann.

Beispielsweise erreicht man mit der Wahl

$$a_i := 2^{i-1}, \quad i \in \underline{p},$$

dass jedem Schadenstand des Abwicklungsverlaufes - verglichen mit demjenigen der Vorperiode - gerade doppelt soviel Einfluss zugemessen wird.

Unter ansonsten gleichen Gegebenheiten wird das Back-Testing-Verfahren erneut durchgeführt. Die in den Abbildungen 2.4 und 2.5 dargestellten Resultate weisen jedoch keine wesentlichen Veränderungen gegenüber dem zuvor betrachteten Fall auf. Dies legt die Vermutung nahe, dass die Auswahl von a keine erheblichen Auswirkungen auf die Ergebnisse hat.

Eines der Hauptziele der Schadenreserveschätzung besteht natürlich in der Prognose der gesamten Entschädigungszahlungen (über mehrere Abwicklungsjahre hinweg) bis hin zur vollständigen Regulierung des Schadens.

Da das Jahr, in dem der Schaden endgültig abgeschlossen wird, ebenfalls zufällig ist und für hohe Abwicklungsjahre naturgemäß relativ wenige Daten vorliegen, sind hierzu offenbar statistische Zusatzüberlegungen nötig.

In der Praxis ist es in diesem Zusammenhang üblich, die Schäden bis hin zu einem festen Abwicklungsjahr mit Reservierungsverfahren zu prognostizieren und anschließend beispielsweise parametrische Regressionsverfahren zur Schätzung der Schäden darüberhinausgehender Jahre zu verwenden.²¹

²¹Letzteres wird in dieser Arbeit jedoch nicht ausführlicher betrachtet.

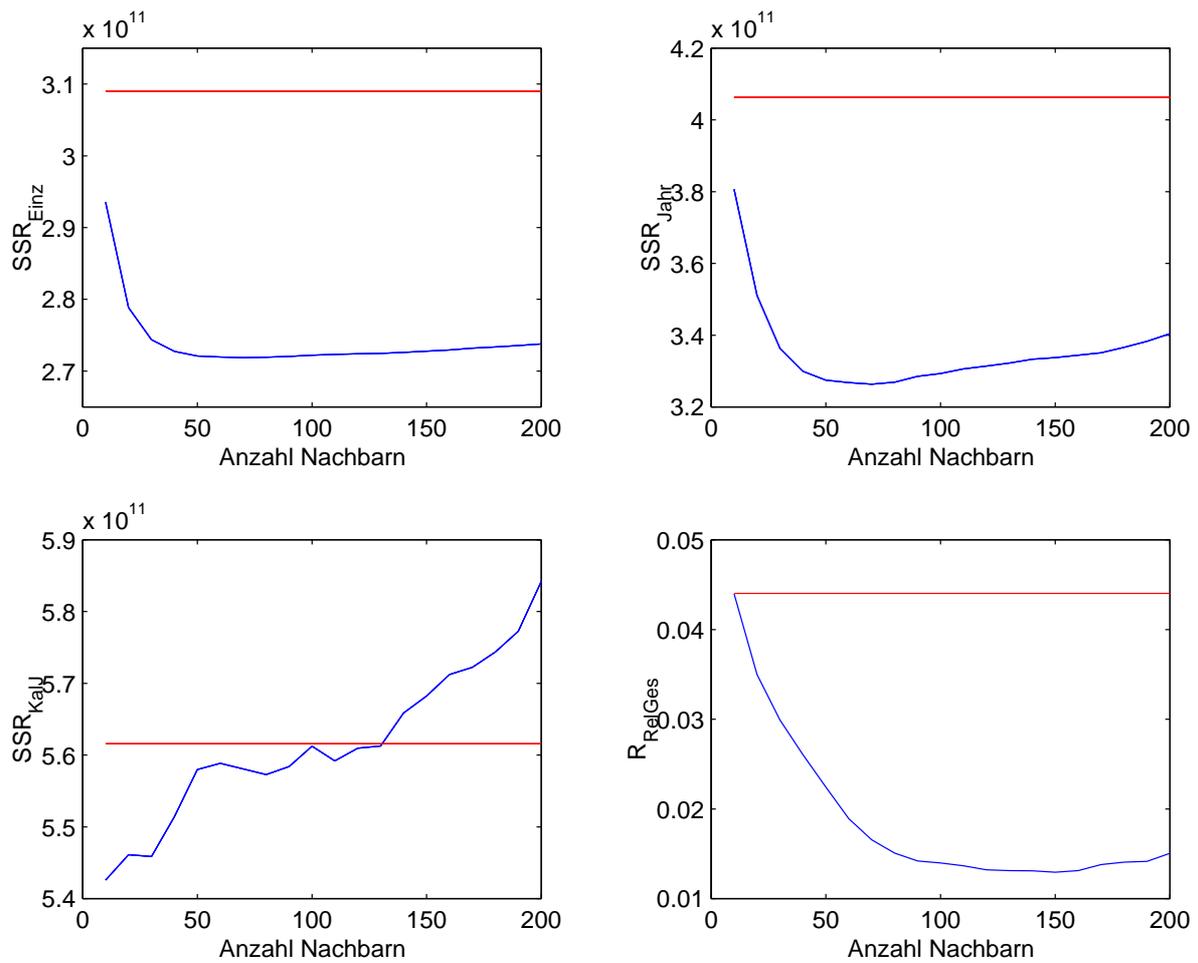


Abbildung 2.4: Beurteilungskriterien für den modifizierten k -nächste-Nachbarn-Schätzer (blau) im Vergleich mit dem Chain-Ladder-Verfahren (rot).

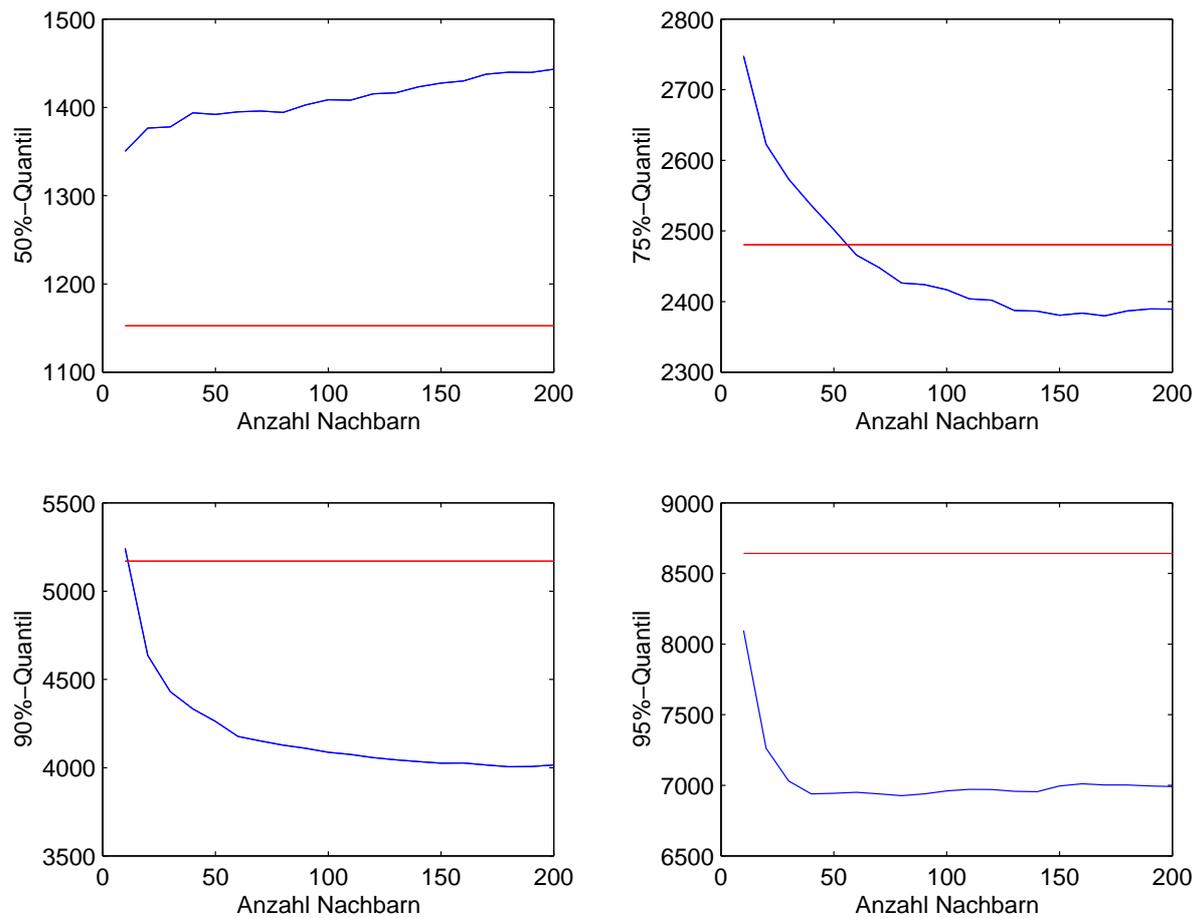


Abbildung 2.5: Quantile der absoluten Residuen beim modifizierten k -nächste-Nachbarn-Schätzer (blau) und beim Chain-Ladder-Verfahren (rot).

Es soll nun untersucht werden, ob die k -nächste-Nachbarn-Methode bei der Schätzung der Schäden bis hin zum L -ten Abwicklungsjahr dem Chain-Ladder-Verfahren ebenfalls überlegen ist.

Erfolgt die entsprechende Prognose nach $p \in \underline{L-1}$ beobachteten Abwicklungsjahren und bezeichnet Y (bzw. Y_α , $\alpha \in \underline{n}$ - in analoger Bezeichnungsweise wie bei der Vorhersage für die Folgeperiode) nun die Summe der zukünftigen Entschädigungen der Perioden $p+1$ bis L , so definiert man den k -nächste-Nachbarn-Schätzer genau wie in (2.13). Die Größen $(X_\alpha)_{\alpha \in \underline{n}}$, x und R müssen hierzu nicht uminterpretiert werden, da die Ermittlung und Gewichtung der Vergleichsschäden völlig identisch abläuft.

Die praktische Durchführung des Verfahrens und insbesondere die Bewertung im Rahmen eines Back-Testing erweisen sich hingegen als problematisch. Einerseits sind beobachtete Schadenverläufe aus Vorjahren nur dann als Vorbilder verwendbar, wenn dort das L -te Abwicklungsjahr bereits erreicht ist. Andererseits muss (für die Beurteilung) auch gewährleistet sein, dass bei den Schäden, für die zukünftige Zahlungen vorhergesagt werden, auch deren tatsächliche Höhe beobachtet wurde.

Liegen wiederum Daten aus insgesamt I Beobachtungsjahren vor, so werden Prognosen also lediglich für solche Zahlungen durchgeführt, die zu in den Jahren L bis $I-L+1$ eingetretenen Schäden gehören.

Im konkreten Fall ist $I = 32$, so dass $L \leq 16$ gewählt werden muss.

Um eine ausreichende Anzahl an Beobachtungen zugrundelegen zu können (auch im Hinblick darauf, dass für die ersten Anfalljahre nur sehr wenige Datensätze vorliegen), erscheint die Wahl eines deutlich geringeren Wertes für L angebracht, beispielsweise $L = 5$ oder $L = 10$.²²

Da bereits bekannt ist, dass das k -nächste-Nachbarn-Verfahren verhältnismäßig genaue Schätzungen der Entschädigungen der Folgeperiode liefert, ist es besonders interessant, zu überprüfen, wie präzise Vorhersagen für längere zukünftige Abwicklungszeiträume ausfallen. Aus diesem Grund sollen nun (mit dem Back-Testing-Verfahren) solche Prognosen analysiert werden, die nach dem ersten Abwicklungsjahr erstellt werden (und daher die Zahlungen der Jahre 2 bis L schätzen).

²²Dabei liegen im Fall $L = 10$ bereits für über 60% der nach dem ersten Abwicklungsjahr durchgeführten Prognosen weniger als 100 Vergleichsschäden vor. Natürlich wird man in der Praxis einen höheren Wert für L verwenden wollen. Dennoch gewinnt man auch bei dem in dieser Arbeit betrachteten (sich aus den Grenzen des Back-Testing ergebenden) Zahlungshorizont einen Eindruck über die Genauigkeit des Verfahrens.

In Anlehnung an (2.1), (2.2) und (2.5) definiert man hier dann (für $L \leq \frac{I+1}{2}$) die Beurteilungskriterien

$$SSR_{\text{Einz}}^* := \sum_{i=L}^{I-L+1} \sum_{m=1}^{n_{ij}} (\hat{X}_{im}^* - X_{im}^*)^2,$$

$$SSR_{\text{Jahr}}^* := \sum_{i=L}^{I-L+1} \left(\sum_{m=1}^{n_{ij}} \hat{X}_{im}^* - \sum_{m=1}^{n_{ij}} X_{im}^* \right)^2$$

sowie

$$R_{\text{RelGes}}^* := \frac{\sum_{i=L}^{I-L+1} \sum_{m=1}^{n_{ij}} \hat{X}_{im}^*}{\sum_{i=L}^{I-L+1} \sum_{m=1}^{n_{ij}} X_{im}^*} - 1.$$

Dabei bezeichnet \hat{X}_{im}^* die nach dem ersten Abwicklungsjahr prognostizierten zukünftigen Entschädigungen (für die Abwicklungsjahre 2 bis L) bezüglich des m -ten unter den n_{ik} Schäden des i -ten Anfalljahres. X_{im}^* stellt die entsprechenden tatsächlich angefallenen Zahlungen dar.

Wird zur Quantifizierung der Ähnlichkeit $\|\cdot\| := |\cdot|$ gewählt, so zeigt ein Vergleich mit dem Chain-Ladder-Verfahren (Tabelle 2.2), dass der k -nächste-Nachbarn-Schätzer auch für mehrjährige Prognosen der zukünftigen Entwicklung einzelner Schäden gut geeignet ist.

Ein ähnliches Bild ergibt sich bei der Vorhersage der jährlichen Summe der Schäden.

	Verfahren	SSR_{Einz}^*	SSR_{Jahr}^*	R_{RelGes}^*
$L = 5$	k -nächste-Nachbarn ($k = 70$)	$1,740 \cdot 10^{11}$	$7,100 \cdot 10^{11}$	0,2154
	Chain-Ladder	$5,902 \cdot 10^{11}$	$1,375 \cdot 10^{12}$	0,2975
$L = 10$	k -nächste-Nachbarn ($k = 30$)	$4,069 \cdot 10^{11}$	$2,311 \cdot 10^{12}$	0,3427
	Chain-Ladder	$1,044 \cdot 10^{12}$	$2,713 \cdot 10^{12}$	0,3523

Tabelle 2.2: Vergleich der Genauigkeit der Prognose der Summe der Zahlungen vom zweiten bis hin zum fünften bzw. zehnten Abwicklungsjahr

Kapitel 3

Genauigkeit des k -nächste-Nachbarn-Schätzers

Im Rahmen einer statistischen Untersuchung des in Abschnitt 2.4 behandelten Verfahrens ist es hilfreich, die Fragestellung der Schätzung von Spätschäden als Regressionsproblem aufzufassen. Mit der Kenntnis des bisherigen Schadenverlaufes über p Abwicklungsjahre hinweg, beschreibbar durch eine \mathbb{R}^p -wertige Zufallsvariable X , soll eine reelle Zufallsvariable Y erklärt werden, die Entschädigungszahlungen während eines bestimmten Zeitraumes charakterisiert. Dabei können die Komponenten von X entweder die Schadenzahlungen der einzelnen Jahre oder aber die jeweiligen Schadenstände (also die kumulierten Zahlungen) darstellen.

Wie in Abschnitt 2.4 ist man in der Regel an den Zahlungen innerhalb eines bestimmten zukünftigen Abwicklungsjahres (z.B. dem folgenden) oder aber innerhalb der nächsten $q \in \mathbb{N}$ Jahre interessiert.¹

Es bietet sich daher an, von einem Modell der Form

$$Y = m(X) + \sigma(X) \cdot \varepsilon \tag{3.1}$$

auszugehen, mit reellwertigen Funktionen m und σ auf dem \mathbb{R}^p , wobei ε eine von X stochastisch unabhängige Zufallsvariable mit $E(\varepsilon) = 0$ und $\text{Var}(\varepsilon) = 1$ darstellt.

Die vorliegenden Daten sind also Ausprägungen von X und Y und sollen dazu verwendet werden, die unbekanntenen Werte $m(x)$ (und gegebenenfalls auch $\sigma(x)$) für gewisse $x \in \mathbb{R}^p$ zu schätzen.

Für P^X -fast alle $x \in \mathbb{R}^p$ gilt offensichtlich gerade

$$m(x) = E(Y|X = x)$$

¹Prinzipiell ließe sich der betrachtete Schätzer aus (3.2) völlig analog auch für ein multivariates Y definieren, etwa um die einzelnen Beträge der Folgejahre in einem gemeinsamen Schritt zu prognostizieren. Ein solcher Ansatz würde jedoch den Rahmen dieser Arbeit sprengen.

und

$$\sigma^2(x) = \text{Var}(Y|X = x).$$

In diesem Kapitel wird nachgewiesen, dass \hat{Y}_{kNN} aus (2.11) dann gerade ein geeigneter Schätzer für den Wert $m(x)$ der Regressionsfunktion ist.

Sei dazu $x \in \mathbb{R}^p$ der Vektor der bisherigen Entschädigungszahlungen eines bestimmten Schadens, dessen zukünftige Entwicklung vorhergesagt werden soll. Weiter nimmt man an, dass die beobachteten Vergleichsschäden (X_i, Y_i) , $i \in \underline{n}$, (wobei wiederum X_i den bisherigen Schadenverlauf und Y_i die entsprechende Fortsetzung verkörpert) stochastisch unabhängig und gemäß $P^{(X,Y)}$ identisch verteilt sind. Für eine Folge natürlicher Zahlen $\{k_n\}_{n \in \mathbb{N}}$, einen geeigneten² Kern K sowie R_n als Abstand von x zum k_n -ten Nachbarn unter den X_i , $i \in \underline{n}$,³ ist

$$m_n(x) := \frac{\sum_{i=1}^n K\left(\frac{X_i - x}{R_n}\right) Y_i}{\sum_{i=1}^n K\left(\frac{X_i - x}{R_n}\right)}, \quad n \in \mathbb{N}, \quad (3.2)$$

der k_n -nächste-Nachbarn-Schätzer für $m(x)$.

Verwendet man zur Abstandsmessung $\|\cdot\|_a$ aus (2.8), so hat man offenbar gerade $m_n(x) = \hat{Y}_{kNN}$ (sofern der selbe Kern K gebraucht wird). Für die theoretischen Resultate muss die Norm jedoch nicht festgelegt werden.

Um die gewünschten Ergebnisse herleiten zu können, sind allerdings zwei Annahmen zu treffen, die nicht völlig im Einklang mit dem Vorgehen in Abschnitt 3.1 stehen.

Noch relativ unproblematisch ist die Forderung, dass X eine Dichte besitzt,⁴ da die Schadenzuwächse der einzelnen Abwicklungsjahre zumeist von Null verschieden sind.⁵

Gravierender ist die Voraussetzung aus (3.5), dass der Kern K im Inneren der Einheitskugel um 0 von Null weg beschränkt ist. Dies hat zur Folge, dass der Epanechnikov-Kern aus (2.15), der beim empirischen Vergleich der Verfahren stets genutzt worden ist, nicht mehr verwendbar ist.

Um die theoretischen Ergebnisse zur Berechnung von Konfidenzintervallen für $m(x)$ in Abschnitt 3.3 auf die Daten anwenden zu können, werden dort alle diejenigen Datensätze nicht zur Prognose herangezogen, in deren bisherigem Schadenverlauf in zumindest einem Abwicklungsjahr keinerlei Zahlungen angefallen sind.

Außerdem wird dort (für feste $\delta > 0$ und $\kappa > 0$) der Kern K definiert als

$$K(u) := \kappa((1 - \|u\|^2) + \delta) 1_{B_1(0)}(u), \quad u \in \mathbb{R}^p, \quad (3.3)$$

²Vgl. Abschnitt 3.1.2.

³Zur exakten Definition von R_n vgl. (3.9).

⁴Vgl. das Vorgehen in Abschnitt 3.3 und die Diskussion in Abschnitt 4.2.

⁵Da diese Restriktion für Y nicht gilt, lassen sich dann jedenfalls solche Schäden betrachten, in deren Abwicklungsverlauf bislang in jedem Jahr echte Zahlungen aufgetreten sind.

wobei die Wahl von $\kappa > 0$ wiederum (wie in (2.15)) keinen Einfluss auf die Schätzwerte hat.⁶

3.1 Normalapproximation des k -nächste-Nachbarn-Schätzfehlers

Im Folgenden soll der k -nächste-Nachbarn-Schätzer im Rahmen eines stochastischen Modells genau formalisiert werden, damit eine Normalapproximation für den Schätzfehler angegeben werden kann. So sind beispielsweise approximative Konfidenzintervalle konstruierbar.

3.1.1 Verwandte Literatur

Der gemäß (3.2) definierte k_n -nächste-Nachbarn-Schätzer wurde von [Collomb79] eingeführt und - unter restriktiven Annahmen - verschiedenen Konsistenzuntersuchungen unterzogen.

[Mack81] gibt asymptotische Ausdrücke für den Bias $Em_n(x) - m(x)$ sowie für die Varianz des Schätzers an und diskutiert die asymptotische Normalität von

$$\sqrt{k_n}(m_n(x) - Em_n(x)).$$

In [Liero89] werden die Resultate von [Collomb79] - in einem allgemeineren Kontext und unter schwächeren Voraussetzungen - (auch mit Angabe von Konvergenzraten) erweitert.

Die Autorin weist darüberhinaus (in [Liero87b]) die asymptotische Normalität des Schätzfehlers $m_n(x) - m(x)$ nach, gebraucht dafür jedoch - im Gegensatz zu [Mack81] - den Ansatz, die zufälligen Größen R_n durch eine deterministische Folge zu approximieren.

Die analog definierten k_n -nächste-Nachbarn-Dichteschätzer (für die Lebesgue-Dichte f von X)⁷

$$f_n(x) := \frac{1}{nR_n^p} \sum_{j=1}^n K\left(\frac{x - X_j}{R_n}\right) \quad (3.4)$$

wurden bereits in [Loftsgaarden/Quesenberry65] vorgeschlagen. [Mack/Rosenblatt79] berechnen hierfür den Bias $Ef_n(x) - f(x)$ und die Varianz von $f_n(x)$, während [Mack80] eine Normalapproximation für $Ef_n(x) - f(x)$ herleitet.⁸

Entsprechende Resultate für den Nadaraya-Watson-Kernschätzer aus (2.14) werden beispielsweise von [Pawlak91] nachgewiesen.

Einen allgemeineren Ansatz verfolgt [Stone77], der Schätzer (für $m(x) = E(Y|X = x)$) der Form

$$\hat{m}(x) := \sum_{i=1}^n w_{ni}(x)Y_i,$$

⁶Die Berechnungen im folgenden Abschnitt vereinfachen sich jedoch etwas, wenn κ so gewählt wird, dass (vgl. (3.7)) $\int K(u)du = 1$ gilt. Außerdem ist diese Voraussetzung notwendig, um $f_n(x)$ aus (3.4) als Dichteschätzer für $f(x)$ - den Funktionswert der Lebesgue-Dichte von X an der Stelle x - ansehen zu können.

⁷Vgl. hierzu Lemma 3.1.11(i) und Satz 3.1.18(i).

⁸Diese beiden Arbeiten sind - im Wesentlichen - Auszüge aus [Mack78].

mit geeigneten Gewichtsfunktionen w_{ni} , $n \in \mathbb{N}$, $i \in \underline{n}$, untersucht und dabei für gewisse $r > 0$ die Konvergenz

$$E|\hat{m}(x) - m(x)|^r \rightarrow 0$$

(für $n \rightarrow \infty$) thematisiert.

Eine gute Übersicht über unterschiedliche nichtparametrische Regressionsverfahren findet sich in [Györfi02].

Das Vorgehen in diesem Abschnitt stützt sich weitgehend auf die Beweistechniken aus [Mack78], [Mack/Rosenblatt79] und [Mack81].

Speziell in letzterer Quelle werden jedoch einige wichtige Aussagen nur unvollständig und auch ungenau nachgewiesen.

Bei den einzelnen Schritten werden hier daher Parallelen und Unterschiede jeweils vermerkt.

3.1.2 Definitionen und Voraussetzungen

Für ein festes $p \in \mathbb{N}$ werden nun also Zufallsvariablen $X_i : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}^p, \mathbb{B}^p)$, $i \in \underline{n}$, und $Y_i : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathbb{B})$, $i \in \underline{n}$, betrachtet, wobei (X_i, Y_i) , $i \in \underline{n}$, unabhängig und identisch verteilt seien, mit $X_i \sim X$ und $Y_i \sim Y$ für alle $i \in \underline{n}$.

Es wird vorausgesetzt, dass es eine stetige Funktion $m : \mathbb{R}^p \rightarrow \mathbb{R}$ gibt mit

$$m(u) = E(Y|X = u) \quad \text{für } P^X\text{-fast alle } u \in \mathbb{R}^p.$$

Dann soll für $x \in \mathbb{R}^p$ der (von $i \in \underline{n}$ unabhängige) Wert $m(x)$ geschätzt werden mittels

$$m_n(x) := \frac{\sum_{j=1}^n K\left(\frac{X_j - x}{R_n}\right) Y_j}{\sum_{j=1}^n K\left(\frac{X_j - x}{R_n}\right)},$$

wobei $K : \mathbb{R}^p \rightarrow \mathbb{R}$ eine Gewichtsfunktion (Kern) mit

$$\exists N_1, N_2 > 0 : \quad N_1 \leq K(u) \leq N_2 \quad \forall u \in \mathbb{R}^p \text{ mit } \|u\| < 1, \quad (3.5)$$

$$K(u) = 0 \quad \forall u \in \mathbb{R}^p \text{ mit } \|u\| \geq 1, \quad (3.6)$$

$$\int K(u) du = 1 \quad \text{sowie} \quad (3.7)$$

$$\int K(u) u_\alpha du = 0 \quad \forall \alpha \in \underline{p} \quad (3.8)$$

darstellt.⁹ Dabei sei $\|\cdot\|$ eine beliebige Norm auf dem \mathbb{R}^p .¹⁰

Als $B_r(x)$ werde im Folgenden die offene Kugel (bezüglich $\|\cdot\|$) um $x \in \mathbb{R}^p$ mit Radius r bezeichnet,

$$B_r(x) := \{u \in \mathbb{R}^p \mid \|u - x\| < r\}.$$

Außerdem sei

$$c := \lambda_p(B_1(0))$$

das Volumen der Einheitskugel im \mathbb{R}^p .

3.1.1 Beispiel. (i) Ist $\|\cdot\|$ die euklidische Norm auf dem \mathbb{R}^p , so gilt

$$c = \frac{\pi^{p/2}}{\Gamma(1 + p/2)}.$$

(ii) Ist allgemeiner $\|\cdot\|$ auf dem \mathbb{R}^p definiert durch

$$\|x\| := \sqrt{\sum_{i=1}^p a_i x_i^2}, \quad a_i \geq 0 \quad \forall i \in \underline{p},$$

so gilt

$$c = \frac{\pi^{p/2}}{\Gamma(1 + p/2)} \prod_{i=1}^p \frac{1}{\sqrt{a_i}}.$$

Beweis. (i) [Königsberger00, S. 282].

(ii) [Königsberger00, S. 291].

□

Die Zufallsvariable $R_n : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathbb{B})$ beschreibe den Abstand von x zu seinem „ k_n -ten Nachbarn“ unter den $X_i, i \in \underline{n}$, sei also die k_n -te Ordnungsstatistik unter den $\{\|X_j - x\|\}_{j \in \underline{n}}$:

$$R_n := \inf \left\{ t \geq 0 \mid \sum_{j=1}^n 1_{\{\|X_j - x\| \leq t\}} \geq k_n \right\}. \quad (3.9)$$

⁹Die Bedingung (3.5) wird in dieser Arbeit - abweichend von [Mack81] - vorausgesetzt. Dies ist jedoch keine Konsequenz der hier getroffenen weniger einschränkenden Annahmen bezüglich der Verteilung von Y bzw. der Wahl von $\|\cdot\|$ (vgl. Fußnoten 11 und 10). (3.5) ist hingegen Grundlage für den Nachweis von Satz 3.1.19(ii) und Lemma 3.1.22; beide Aussagen werden von [Mack81] nicht bewiesen. In der Literatur ist diese Voraussetzung - wenn auch verhältnismäßig restriktiv - nicht unüblich, vgl. etwa [Devroye/Wagner80b, S. 232] und [Pawlak91, S. 312].

¹⁰Im Gegensatz zu den Arbeiten von [Mack81] und [Mack/Rosenblatt79] muss nicht die euklidische Norm zugrundegelegt sein. Dies erlaubt beispielsweise Variationen des Schätzers wie in Abschnitt 2.4.

Dabei sei $\{k_n\}_{n \in \mathbb{N}}$ eine Folge natürlicher Zahlen mit

$$k_n \leq n \quad \forall n \in \mathbb{N}, \quad \lim_{n \rightarrow \infty} k_n = +\infty \quad \text{und} \quad \lim_{n \rightarrow \infty} \frac{k_n}{n} = 0.$$

Weiter sei angenommen, dass X absolutstetig ist,¹¹ mit Lebesgue-Dichte f . Dann sei $x \in \mathbb{R}^p$ fest gewählt mit $f(x) > 0$.

Außerdem werden folgende Glattheitsbedingungen an X und Y gestellt:

- f und m seien beschränkt und in einer Umgebung von x zweimal stetig differenzierbar.
- Für P^X -fast alle $u \in \mathbb{R}^p$ sei $E(|Y|^3|X = u) \leq \bar{M}$ für ein $\bar{M} < \infty$.
- Für eine beschränkte und in x stetige Funktion $l : \mathbb{R}^p \rightarrow \mathbb{R}$ gelte

$$l(u) = E(Y^2|X = u) \quad \text{für } P^X\text{-fast alle } u \in \mathbb{R}^p.$$

Wegen

$$\text{Var}(Y|X) = E(Y^2|X) - E^2(Y|X) \quad P\text{-f.s.}$$

ist $\sigma^2 : \mathbb{R}^p \rightarrow \mathbb{R}$ mit

$$\sigma^2(u) := l(u) - m^2(u), \quad u \in \mathbb{R}^p,$$

beschränkt und stetig in x mit

$$\sigma^2(u) = \text{Var}(Y|X = u) \quad \text{für } P^X\text{-fast alle } u \in \mathbb{R}^p.$$

Dabei sei $\sigma^2(x) > 0$.

Das Ziel dieses Abschnitts 3.1 ist die Angabe einer Normalapproximation für den Schätzfehler $m_n(x) - m(x)$. Dazu erweist es sich als hilfreich, den Schätzer $m_n(x)$ in den Quotienten

$$m_n(x) = \frac{h_n(x)}{f_n(x)} \tag{3.10}$$

zu zerlegen, mit

$$h_n(x) := \frac{1}{nR_n^p} \sum_{j=1}^n K\left(\frac{x - X_j}{R_n}\right) Y_j$$

und

$$f_n(x) := \frac{1}{nR_n^p} \sum_{j=1}^n K\left(\frac{x - X_j}{R_n}\right);$$

¹¹Wiederum im Gegensatz zu den Arbeiten von [Mack81] und [Mack/Rosenblatt79] muss (X, Y) jedoch keine gemeinsame Dichte besitzen. Dies ist für die konkrete Anwendung wichtig, da die Y_j (als Schadenzuwächse interpretiert) Punktmasse in Null aufweisen können sollen.

$f_n(x)$ ist dann gerade der Dichteschätzer für $f(x)$ aus (3.4).

Bei der Untersuchung der Verteilung von $m_n(x)$ treten zwei grundlegende Schwierigkeiten auf. Einerseits handelt es sich hier offenbar um einen Quotienten von Zufallsvariablen, andererseits hängen dabei sowohl der Zähler als auch der Nenner von R_n ab.

Dem ersten Problem begegnet man mit einer approximativen additiven Zerlegung des Quotienten (siehe Abschnitt 3.1.5). Die Abhängigkeit von R_n behandelt man, indem die Zufallsvariablen $(X_i, Y_i)_{i \in \underline{n}}$ zunächst bedingt gegeben R_n analysiert werden.¹²

Man definiert dafür

$$\tilde{X}_i := X_{\chi(i, R_n)}, \quad \tilde{Y}_i := Y_{\chi(i, R_n)}, \quad i \in \underline{k_n - 1},$$

mit

$$\chi(i, r) := \inf \left\{ j \in \underline{n} \mid \sum_{k=1}^j 1_{\{\|X_k - x\| < r\}} = i \right\}, \quad i \in \underline{k_n - 1}, \quad r > 0,$$

sowie

$$\bar{X} := X_{\bar{\chi}}, \quad \bar{Y} := Y_{\bar{\chi}}$$

mit

$$\bar{\chi} := \inf \{ j \in \underline{n} \mid \|X_j - x\| = R_n \}$$

und

$$\ddot{X}_i := X_{\ddot{\chi}(k_n + i, R_n)}, \quad \ddot{Y}_i := Y_{\ddot{\chi}(k_n + i, R_n)}, \quad i \in \underline{n - k_n},$$

mit

$$\ddot{\chi}(\nu, r) := \inf \left\{ j \in \underline{n} \mid \sum_{k=1}^j 1_{\{\|X_k - x\| > r\}} = \nu - k_n \right\}, \quad \nu \in \{k_n + 1, \dots, n\}, \quad r > 0.$$

Da X absolutstetig ist, sind alle X_i , $i \in \underline{n}$, P -fast sicher paarweise verschieden. Dann stellt \bar{X} den k_n -ten Nachbarn von x dar. Die \tilde{X}_i , $i \in \underline{k_n - 1}$, beschreiben die $k_n - 1$ Beobachtungen für X_i , die innerhalb der offenen Kugel mit Radius R_n um x liegen, während die \ddot{X}_i , $i \in \underline{n - k_n}$, die übrigen $n - k_n$ Werte echt außerhalb dieses Bereiches bezeichnen.

¹²Dies entspricht genau der grundsätzlichen Vorgehensweise von [Mack81].

3.1.3 Verteilungen der betrachteten Zufallsvariablen

Unter Verwendung der Theorie über Ordnungsstatistiken wird nun nachgewiesen, dass auch die $(\tilde{X}_i, \tilde{Y}_i)_{i \in \underline{k_n-1}}$ bedingt unabhängig und identisch verteilt gegeben R_n sind. Ebenso wird gezeigt, wie sich $P^{(\tilde{X}_i, \tilde{Y}_i) | R_n=r}$ aus $P^{(X, Y)}$ berechnen lässt:¹³

Sei G die Verteilungsfunktion von $\|x - X\|$. Da G differenzierbar ist (Satz 3.1.2), hat R_n als k_n -te Ordnungsstatistik der unabhängigen und identisch verteilten Stichprobe $\{\|x - X_j\|\}_{j \in n}$ nach [Arnold92, S. 9f.] die Dichte

$$h(r) = n \binom{n-1}{k-1} (G(r))^{k_n-1} (1 - G(r))^{n-k_n} G'(r). \quad (3.11)$$

3.1.2 Satz. G ist auf $(0, \infty)$ stetig differenzierbar.

Beweis. Mit

$$\Lambda : \mathbb{R}^p \setminus \{0\} \longrightarrow (0, \infty) \times \partial B_1(0), \quad x \longmapsto \left(\|x\|, \frac{x}{\|x\|} \right)$$

hat man offenbar die Darstellung

$$\Lambda^{-1}((0, r) \times \partial B_1(0)) = B_r(0) \setminus \{0\}.$$

Bezeichnet $\mathbb{B}(\partial B_1(0))$ die Borel- σ -Algebra auf der Einheitskugel um 0 im \mathbb{R}^p , so ist μ mit

$$\mu(A) := \lambda_p^\Lambda((0, 1) \times A), \quad A \in \mathbb{B}(\partial B_1(0)),$$

ein endliches Maß auf $(\partial B_1(0), \mathbb{B}(\partial B_1(0)))$, denn

$$\mu(\partial B_1(0)) = \lambda_p^\Lambda((0, 1) \times \partial B_1(0)) = \lambda_p(B_1(0)) < \infty.$$

Nun gilt für alle $r > 0$ und alle $D \in \mathbb{B}^p$

$$\lambda_p(r \cdot D) = r^p \lambda_p(D), \quad (3.12)$$

denn für beliebige Rechteckmengen der Form

$$D_{cd} := \times_{i=1}^p [c_i, d_i], \quad c, d \in \mathbb{R}^p,$$

hat man (für beliebige $r > 0$)

$$\lambda_p(r \cdot D_{cd}) = \lambda_p\left(\times_{i=1}^p [r \cdot c_i, r \cdot d_i]\right) = \prod_{i=1}^p (rd_i - rc_i) = r^p \prod_{i=1}^p (d_i - c_i) = r^p \lambda_p(D_{cd})$$

¹³[Mack/Rosenblatt79, S. 3f.] betrachten (im Kontext der Dichteschätzung) nur die $(\tilde{X}_i)_{i \in \underline{k_n-1}}$ bzw. $P^{\tilde{X}_i | R_n=r}$, $i \in \underline{k_n-1}$. Sie motivieren eine Satz 3.1.3 ähnelnde Aussage für die Wahrscheinlichkeitsdichten von $P^{\tilde{X}_i | R_n=r}$ heuristisch. Vgl. auch [Mack81, S. 314].

und $\{D_{cd} \mid c, d \in \mathbb{R}^p\}$ ist ein durchschnitt-stabiler Erzeuger von \mathbb{B}^p .

Dann ergibt sich für alle $r > 0$ und alle $A \in \mathbb{B}(\partial B_1(0))$:

$$\begin{aligned} \lambda_p^\Lambda((0, r) \times A) &= \lambda_p\left(\Lambda^{-1}((0, r) \times A)\right) \\ &= \lambda_p\left\{x \in \mathbb{R}^p \setminus \{0\} \mid \|x\| < r, \frac{x}{\|x\|} \in A\right\} \\ &= \lambda_p\left(r \cdot \left\{x \in \mathbb{R}^p \setminus \{0\} \mid \|x\| < 1, \frac{x}{\|x\|} \in A\right\}\right) \\ &= r^p \lambda_p\left(\Lambda^{-1}((0, 1) \times A)\right) \\ &= r^p \lambda_p^\Lambda((0, 1) \times A) = r^p \mu(A). \end{aligned}$$

Sei H_p ein Maß mit maßdefinierender Funktion $s \mapsto s^p$. Da $A \in \partial B_1(0)$ beliebig ist und $\{(0, r) \mid r > 0\}$ ein durchschnitt-stabiler Erzeuger von $(0, \infty)$ ist, hat man

$$\lambda_p^\Lambda = H_p \otimes \mu.$$

Daher ergibt sich nach dem Transformationssatz und dem Satz von Fubini

$$\begin{aligned} G(r) &= \int_{B_r(x)} f(t) dt \\ &= \int_{B_r(0)} f(t+x) dt \\ &= \int_{B_r(0) \setminus \{0\}} f(t+x) dt \\ &= \int_{\Lambda^{-1}((0,r) \times \partial B_1(0))} f(t+x) \lambda_p(dt) \\ &= \int_{(0,r) \times \partial B_1(0)} f(\Lambda^{-1}(s, \varphi) + x) \lambda_p^\Lambda(ds, d\varphi) \\ &= \int_{(0,r)} \int_{\partial B_1(0)} f(\Lambda^{-1}(s, \varphi) + x) \mu(d\varphi) H_p(ds). \end{aligned}$$

Da H_p die Riemann-Dichte $s \mapsto p \cdot s^{p-1}$ besitzt, folgt wegen

$$\Lambda^{-1}(s, \varphi) = s \cdot \varphi \quad \text{für alle } s > 0, \varphi \in \partial B_1(0),$$

nun

$$G(r) = \int_0^r \int_{\partial B_1(0)} f(s\varphi + x) \mu(d\varphi) p s^{p-1} ds. \quad (3.13)$$

Nachzuweisen bleibt die Stetigkeit von

$$s \mapsto \int_{\partial B_1(0)} f(s\varphi + x) \mu(d\varphi) p s^{p-1} :$$

Sei $\{s_n\}_{n \in \mathbb{N}}$ eine Folge reeller Zahlen mit $\lim_{n \rightarrow \infty} s_n = s_0$. Da f beschränkt ist ($f \leq M_f < \infty$), folgt mit dem Satz von der majorisierten Konvergenz und der Stetigkeit von f

$$\lim_{n \rightarrow \infty} \int_{\partial B_1(0)} f(s_n \varphi + x) \mu(d\varphi) = \int_{\partial B_1(0)} \lim_{n \rightarrow \infty} f(s_n \varphi + x) \mu(d\varphi) = \int_{\partial B_1(0)} f(s_0 \varphi + x) \mu(d\varphi),$$

denn

$$\int_{\partial B_1(0)} f(s_n \varphi + x) \mu(d\varphi) \leq \int_{\partial B_1(0)} M_f \mu(d\varphi) = M_f \cdot \mu(\partial B_1(0)) < \infty.$$

Aus (3.13) erhält man also

$$G'(r) = \int_{\partial B_1(0)} f(r\varphi + x) \mu(d\varphi) p r^{p-1}.$$

□

Wegen (3.6) lässt sich $m_n(x)$ definitionsgemäß (statt durch (X_i, Y_i) , $i \in \underline{n}$) auch mittels der $(\tilde{X}_i, \tilde{Y}_i)$, $i \in \underline{k_n - 1}$, darstellen,

$$m_n(x) = \frac{\sum_{j=1}^{k_n-1} K\left(\frac{\tilde{X}_j - x}{R_n}\right) \tilde{Y}_j}{\sum_{j=1}^{k_n-1} K\left(\frac{\tilde{X}_j - x}{R_n}\right)}. \quad (3.14)$$

Daher werden diese Zufallsvariablen nun genauer untersucht.

3.1.3 Satz. (i) Die $(\tilde{X}_i, \tilde{Y}_i)_{i \in \underline{k_n - 1}}$ sind bedingt unabhängig gegeben R_n .

(ii) Für alle $i \in \underline{k_n - 1}$, $A \in \mathbb{B}^p$ sowie $B \in \mathbb{B}$ gilt (für λ_1 -fast alle $r > 0$)

$$P(\tilde{X}_i \in A, \tilde{Y}_i \in B | R_n = r) = \frac{P\{X_1 \in A \cap B_r(x), Y_1 \in B\}}{G(r)}.$$

Beweis. Beide Aussagen werden gemeinsam bewiesen. Seien $A_i \in \mathbb{B}^p$ und $B_i \in \mathbb{B}$ für alle $i \in \underline{k_n - 1}$; dann gilt

$$\begin{aligned} & P\{\tilde{X}_i \in A_i, \tilde{Y}_i \in B_i \ \forall i \in \underline{k_n - 1}\} \\ &= P\{\tilde{X}_i \in A_i, \tilde{Y}_i \in B_i \ \forall i \in \underline{k_n - 1}, \tilde{X}_j \in \mathbb{R}^p, \tilde{Y}_j \in \mathbb{R} \ \forall j \in \underline{n - k_n}, \bar{X} \in \mathbb{R}^p, \bar{Y} \in \mathbb{R}\} \\ &= \sum_{s \in \underline{n}} \sum_{\substack{T \subset \underline{n} \setminus \{s\} \\ |T| = k_n - 1}} P\left\{(X_{t_i:T}, Y_{t_i:T}) \in (A_i \cap B_{R_n}(x)) \times B_i \ \forall i \in \underline{k_n - 1}, \right. \\ &\quad \left. (X_{t_j:\bar{T}}, Y_{t_j:\bar{T}}) \in (\overline{B_{R_n}(x)})^c \times \mathbb{R} \ \forall j \in \underline{n - k_n}, (X_s, Y_s) \in \partial B_{R_n}(x) \times \mathbb{R}\right\}, \end{aligned}$$

für $\bar{T} := \underline{n} \setminus (T \cup \{s\})$ und $T = \{t_i:T\}_{i \in \underline{k_n - 1}}$, $\bar{T} = \{t_j:\bar{T}\}_{j \in \underline{n - k_n}}$, mit

$$\begin{aligned} t_{1:T} &< t_{2:T} < \dots < t_{(k_n-1):T} && \text{sowie} \\ t_{1:\bar{T}} &< t_{2:\bar{T}} < \dots < t_{(n-k_n):\bar{T}}, \end{aligned}$$

da $\chi(\cdot, r)$ und $\ddot{\chi}(\cdot, r)$ für alle $r > 0$ streng monoton wachsend sind. Damit ergibt sich

$$\begin{aligned}
& P\{\tilde{X}_i \in A_i, \tilde{Y}_i \in B_i \ \forall i \in \underline{k_n - 1}\} \\
&= \sum_{s \in \underline{n}} \sum_{\substack{T \subset \underline{n} \setminus \{s\} \\ |T|=k_n-1}} P\left\{(X_{t_i:T}, Y_{t_i:T}) \in A_i \times B_i, \|X_{t_i:T} - x\| < \|X_s - x\| \ \forall i \in \underline{k_n - 1}, \right. \\
&\quad \left. (X_{t_j:\bar{T}}, Y_{t_j:\bar{T}}) \in \mathbb{R}^p \times \mathbb{R}, \|X_{t_j:\bar{T}} - x\| > \|X_s - x\| \ \forall j \in \underline{n - k_n}, (X_s, Y_s) \in \mathbb{R}^p \times \mathbb{R}\right\} \\
&= \sum_{s \in \underline{n}} \sum_{\substack{T \subset \underline{n} \setminus \{s\} \\ |T|=k_n-1}} \int P\left\{(X_{t_i:T}, Y_{t_i:T}) \in A_i \times B_i, \|X_{t_i:T} - x\| < \|t - x\| \ \forall i \in \underline{k_n - 1}, \right. \\
&\quad \left. \|X_{t_j:\bar{T}} - x\| > \|t - x\| \ \forall j \in \underline{n - k_n}\right\} P^{(X_s, Y_s)}(dt, du),
\end{aligned}$$

denn die $(X_i, Y_i)_{i \in \underline{n}}$ sind stochastisch unabhängig

$$\begin{aligned}
&= \sum_{s \in \underline{n}} \sum_{\substack{T \subset \underline{n} \setminus \{s\} \\ |T|=k_n-1}} \int \prod_{i=1}^{k_n-1} P\{(X_{t_i:T}, Y_{t_i:T}) \in A_i \times B_i, \|X_{t_i:T} - x\| < \|t - x\|\} \\
&\quad \cdot \prod_{j=1}^{n-k_n} P\{\|X_{t_j:\bar{T}} - x\| > \|t - x\|\} P^{(X_s, Y_s)}(dt, du)
\end{aligned}$$

wiederum aufgrund der stochastischen Unabhängigkeit der $(X_i, Y_i)_{i \in \underline{n}}$

$$\begin{aligned}
&= \sum_{s \in \underline{n}} \sum_{\substack{T \subset \underline{n} \setminus \{s\} \\ |T|=k_n-1}} \int \prod_{i=1}^{k_n-1} P\{(X_1, Y_1) \in A_i \times B_i, \|X_1 - x\| < \|t - x\|\} \\
&\quad \cdot \prod_{j=1}^{n-k_n} P\{\|X_1 - x\| > \|t - x\|\} P^{(X, Y)}(dt, du),
\end{aligned}$$

denn $(X_i, Y_i) \sim (X_1, Y_1) \sim (X, Y)$ für alle $i \in \underline{k_n - 1}$.

Da es für festes $s \in \underline{n}$ genau $\binom{n-1}{k_n-1}$ Möglichkeiten gibt, eine $(k_n - 1)$ -elementige Teilmenge T aus der $(n - 1)$ -elementigen Menge $\underline{n} \setminus \{s\}$ auszuwählen, folgt hieraus

$$\begin{aligned}
& P\{\tilde{X}_i \in A_i, \tilde{Y}_i \in V_i \forall i \in \underline{k_n - 1}\} \\
&= n \binom{n-1}{k_n-1} \int \prod_{i=1}^{k_n-1} P\{X_1 \in A_i \cap B_{\|t-x\|}(x), Y_1 \in V_i\} \prod_{j=1}^{n-k_n} P\{X_1 \in (\overline{B_{\|t-x\|}(x)})^c\} \\
&\qquad\qquad\qquad P^{(X,Y)}(dt, du) \\
&= n \binom{n-1}{k_n-1} \int \prod_{i=1}^{k_n-1} P\{X_1 \in A_i \cap B_{\|t-x\|}(x), Y_1 \in V_i\} (1 - G(\|t-x\|))^{n-k_n} \int P^{Y|X=t}(du) \\
&\qquad\qquad\qquad P^X(dt) \\
&= n \binom{n-1}{k_n-1} \int \prod_{i=1}^{k_n-1} P\{X_1 \in A_i \cap B_r(x), Y_1 \in V_i\} (1 - G(r))^{n-k_n} P^{\|X-x\|}(dr) \\
&\quad \text{nach dem Transformationssatz} \\
&= n \binom{n-1}{k_n-1} \int \prod_{i=1}^{k_n-1} P\{X_1 \in A_i \cap B_r(x), Y_1 \in V_i\} (1 - G(r))^{n-k_n} G'(r) dr.
\end{aligned}$$

Wegen

$$P\{\tilde{X}_i \in A_i, \tilde{Y}_i \in V_i \forall i \in \underline{k_n - 1}\} = \int P(\tilde{X}_i \in A_i, \tilde{Y}_i \in V_i \forall i \in \underline{k_n - 1} | R_n = r) h(r) dr$$

erhält man aufgrund der Eindeutigkeit der Lebesgue-Dichte von $P^{(\tilde{X}_i, \tilde{Y}_i)_{i \in \underline{k_n - 1}}}$ also (für λ_1 -fast alle $r > 0$)¹⁴

$$\begin{aligned}
& P(\tilde{X}_i \in A_i, \tilde{Y}_i \in V_i \forall i \in \underline{k_n - 1} | R_n = r) \\
&= \frac{n \binom{n-1}{k_n-1} \prod_{i=1}^{k_n-1} P\{X_1 \in A_i \cap B_r(x), Y_1 \in V_i\} (1 - G(r))^{n-k_n} G'(r)}{h(r)} \\
&= \frac{\prod_{i=1}^{k_n-1} P\{X_1 \in A_i \cap B_r(x), Y_1 \in V_i\}}{(G(r))^{k_n-1}} \tag{3.15}
\end{aligned}$$

¹⁴Vgl. z.B. [Behnen/Neuhaus95, Satz 17.13.].

nach (3.11). Damit folgt für alle $i \in \underline{k_n - 1}$ und für λ_1 -fast alle $r > 0$

$$\begin{aligned}
 & P(\tilde{X}_i \in A_i, \tilde{Y}_i \in V_i | R_n = r) \\
 &= P(\tilde{X}_i \in A_i, \tilde{Y}_i \in V_i, \tilde{X}_j \in \mathbb{R}^p, \tilde{Y}_j \in \mathbb{R} \ \forall j \in \underline{k_n - 1} \setminus \{i\} | R_n = r) \\
 &= \frac{P\{X_1 \in A_i \cap B_r(x), Y_1 \in V_i\} \prod_{j \in \underline{k_n - 1} \setminus \{i\}} P\{X_1 \in B_r(x)\}}{(G(r))^{k_n - 1}} \\
 &= \frac{P\{X_1 \in A_i \cap B_r(x), Y_1 \in V_i\}}{G(r)},
 \end{aligned}$$

denn $P\{X_1 \in B_r(x)\} = G(r)$, also (ii).

Da nun nach (3.15) auch (für λ_1 -fast alle $r > 0$)

$$P(\tilde{X}_i \in A_i, \tilde{Y}_i \in V_i \ \forall i \in \underline{k_n - 1} | R_n = r) = \prod_{i=1}^{k_n - 1} P(\tilde{X}_i \in A_i, \tilde{Y}_i \in V_i | R_n = r)$$

gilt, ist auch (i) bewiesen. \square

3.1.4 Asymptotisches Verhalten von R_n und $G(R_n)$

Für das weitere Vorgehen muss natürlich das stochastische Verhalten von R_n näher untersucht werden. Heuristisch gesehen liegt es nahe, dass der Abstand des k_n -ten Nachbarn zu x mit wachsendem n gegen Null konvergiert.¹⁵

Gemäß (3.14) lässt sich $m_n(x)$ direkt aus den $(\tilde{X}_i, \tilde{Y}_i)$, $i \in \underline{k_n - 1}$, berechnen. Entsprechende Darstellungen gelten auch für $f_n(x)$ und $h_n(x)$. Da diese Schätzer bedingt gegeben R_n analysiert werden sollen,¹⁶ wird vor dem Hintergrund von Satz 3.1.3(ii) deutlich, dass häufig auch die Zufallsvariable

$$T_n := G(R_n)$$

betrachtet werden muss.

An dieser Stelle werden nun einige nützliche technische Resultate nachgewiesen, die das asymptotische Verhalten von R_n und T_n charakterisieren.¹⁷

Da G die Verteilungsfunktion von $\|X - x\|$ ist, gilt $G(\|X - x\|) \sim \mathcal{R}(0, 1)$. R_n ist die k_n -te Ordnungsstatistik aus der Stichprobe der unabhängigen und identisch verteilten Zufallsvariablen $\{\|X_j - x\|\}_{j \in \underline{n}}$. Als Verteilungsfunktion ist G monoton wachsend, $T_n := G(R_n)$ ist also die k_n -te Ordnungsstatistik aus einer Stichprobe von n unabhängigen $\mathcal{R}(0, 1)$ -verteilten Zufallsvariablen.

¹⁵Vgl. Lemma 3.1.4(ii).

¹⁶Vgl. Lemma 3.1.11.

¹⁷Die Beweise orientieren sich dabei eng an [Mack78, S 13f.] (bzw. an [Mack78, S. 36f.] bezüglich Korollar 3.1.7).

Nach [Arnold92, S.15] gilt damit $T_n \sim \text{Beta}(k_n, n - k_n + 1)$, d.h. T_n hat die Lebesgue-Dichte

$$g(y) = n \binom{n-1}{k-1} y^{k_n-1} (1-y)^{n-k_n} \mathbf{1}_{(0,1)}(y), \quad \text{mit}$$

$$ET_n = \frac{k_n}{n+1} \quad \text{und} \quad \text{Var } T_n = \frac{k_n(n+1-k_n)}{(n+1)^2(n+2)}. \quad (3.16)$$

3.1.4 Lemma. (i) Es gilt

$$T_n \xrightarrow{n \rightarrow \infty} 0 \quad \text{n.W.}$$

(ii) Es gilt

$$R_n \xrightarrow{n \rightarrow \infty} 0 \quad \text{n.W.}$$

Beweis. (i) Nach der Markov-Ungleichung gilt für alle $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P\{|T_n| < \varepsilon\} \geq 1 - \lim_{n \rightarrow \infty} \frac{E|T_n|}{\varepsilon} = 1 - \lim_{n \rightarrow \infty} \frac{k_n}{\varepsilon(n+1)} = 1,$$

da $\lim_{n \rightarrow \infty} \frac{k_n}{n} = 0$.

(ii) Wegen $f(x) > 0$ und der Stetigkeit von f gibt es ein $\delta > 0$ mit

$$f(z) > 0 \quad \forall z \in B_\delta(x).$$

Dann ist G an der Stelle 0 wegen

$$G(r) = P\{||X - x|| \leq r\} = P\{X \in B_r(x)\} = \int_{B_r(x)} f(z) dz > 0 \quad \forall r \in (0, \delta)$$

und $G(0) = 0$ streng monoton wachsend und stetig. Damit folgt aus

$$T_n = G(R_n) \xrightarrow{n \rightarrow \infty} 0 \quad \text{n.W.}$$

(nach (i)) auch

$$R_n \xrightarrow{n \rightarrow \infty} 0 \quad \text{n.W.}$$

□

3.1.5 Lemma. Es gilt

$$\frac{k_n}{n} \cdot \frac{1}{T_n} \xrightarrow{n \rightarrow \infty} 1 \quad \text{n.W.}$$

Beweis. Aus der Markov-Ungleichung folgt wegen $ET_n = \frac{k_n}{n+1}$ zunächst

$$\frac{n+1}{k_n} \cdot T_n \xrightarrow{n \rightarrow \infty} 1 \quad \text{n.W.,}$$

denn (nach (3.16))

$$\begin{aligned} \lim_{n \rightarrow \infty} P\left\{\left|T_n \frac{n+1}{k_n} - 1\right| > \varepsilon\right\} &= \lim_{n \rightarrow \infty} P\left\{\left(\frac{n+1}{k_n}\right)^2 \left(T_n - \frac{k_n}{n+1}\right)^2 > \varepsilon^2\right\} \\ &\leq \lim_{n \rightarrow \infty} \left(\frac{n+1}{k_n}\right)^2 \frac{\text{Var } T_n}{\varepsilon^2} \\ &= \lim_{n \rightarrow \infty} \frac{(n+1)^2 k_n (n+1 - k_n)}{k_n^2 (n+1)^2 (n+2) \varepsilon^2} = 0, \end{aligned}$$

da $\lim_{n \rightarrow \infty} \frac{k_n}{n} = 0$ und $\lim_{n \rightarrow \infty} k_n = \infty$. Hieraus folgt

$$\frac{k_n}{n+1} \frac{1}{T_n} \xrightarrow{n \rightarrow \infty} 1 \quad \text{n.W.}$$

und damit

$$\frac{k_n}{n} \cdot \frac{1}{T_n} = \frac{n+1}{n} \cdot \frac{k_n}{n+1} \cdot \frac{1}{T_n} \xrightarrow{n \rightarrow \infty} 1 \quad \text{n.W.}$$

□

3.1.6 Lemma. *Es gilt*

$$\frac{T_n}{cR_n^p} \xrightarrow{n \rightarrow \infty} f(x) \quad \text{n.W.}$$

Beweis. Sei $\varepsilon > 0$ gegeben, dann gibt es aufgrund der Stetigkeit von f ein $r_0(\varepsilon) > 0$ mit

$$|f(x+r) - f(x)| < \varepsilon \quad \forall r \in B_{r_0(\varepsilon)}(0).$$

Entsprechend (3.12) gilt für alle $r > 0$

$$\lambda_p(B_r(x)) = \lambda_p(B_r(0)) = \lambda_p(r \cdot B_1(0)) = r^p \lambda_p(B_1(0)) = cr^p$$

und damit für alle $r \in (0, r_0(\varepsilon))$

$$\begin{aligned} \left| \frac{G(r)}{cr^p} - f(x) \right| &= \left| \frac{P\{\|X-x\| \leq r\}}{cr^p} - f(x) \right| \\ &= \left| \frac{\int_{B_r(x)} f(t) dt - f(x) cr^p}{cr^p} \right| \\ &= \left| \frac{f(x) \lambda_p(B_r(x)) + \int_{B_r(x)} (f(t) - f(x)) dt - f(x) cr^p}{cr^p} \right| \\ &\leq \frac{\int_{B_r(x)} |f(t) - f(x)| dt}{cr^p} \\ &< \frac{\int_{B_r(x)} \varepsilon dt}{cr^p} = \varepsilon. \end{aligned} \tag{3.17}$$

Dann folgt

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{G(R_n)}{cR_n^p} - f(x) \right| < \varepsilon \right\} \geq \lim_{n \rightarrow \infty} P \{ R_n < r_0(\varepsilon) \} = 1,$$

da $R_n \xrightarrow{n \rightarrow \infty} 0$ n.W. nach Lemma 3.1.4(ii), also

$$\frac{T_n}{cR_n^p} \xrightarrow{n \rightarrow \infty} f(x) \quad \text{n.W.}$$

□

Aus dem Beweis zu diesem Lemma erhält man ein weiteres Ergebnis, das insbesondere in Abschnitt 3.1.7 eine entscheidende Rolle spielen wird.

3.1.7 Korollar. Für alle $\lambda > 0$ gilt

$$(G^{-1}(t))^\lambda = (cf(x))^{-\lambda/p} t^{\lambda/p} + o(t^{\lambda/p}) \quad \text{für } t \searrow 0.$$

Beweis. Aus (3.17) folgt für alle $r \geq 0$:

$$t := G(r) = f(x)cr^p + o(r^p) \quad \text{für } r \searrow 0, \quad \text{denn}$$

$$\lim_{r \searrow 0} \frac{|G(r) - f(x)cr^p|}{r^p} = \lim_{r \searrow 0} \left| \frac{G(r)}{r^p} - cf(x) \right| = 0.$$

Da G in einer Umgebung von 0 stetig und streng monoton steigend ist, hat man dort $r = G^{-1}(t)$ und es gilt die Äquivalenz

$$r \searrow 0 \iff t \searrow 0 \tag{3.18}$$

und damit auch

$$\lim_{t \searrow 0} \frac{|t - f(x)c \cdot (G^{-1}(t))^p|}{(G^{-1}(t))^p} = 0.$$

Schließlich erhält man für alle $\lambda > 0$

$$\begin{aligned} & \lim_{t \searrow 0} \left(t(G^{-1}(t))^{-p} - cf(x) \right) = 0 \\ \implies & 1 = \lim_{t \searrow 0} \frac{t(G^{-1}(t))^{-p}}{cf(x)} = \lim_{t \searrow 0} \frac{t^{-\lambda/p} (G^{-1}(t))^\lambda}{(cf(x))^{-\lambda/p}} \\ \implies & \lim_{t \searrow 0} \left| \frac{(G^{-1}(t))^\lambda - (cf(x))^{-\lambda/p} \cdot t^{\lambda/p}}{t^{\lambda/p}} \right| = \lim_{t \searrow 0} \left| t^{-\lambda/p} (G^{-1}(t))^\lambda - (cf(x))^{-\lambda/p} \right| = 0. \end{aligned}$$

□

3.1.5 Additive Zerlegung von $m_n(x)$

An dieser Stelle soll bereits der zentrale Satz dieses Kapitels angegeben werden, dessen Beweis sich über die Abschnitte 3.1.5, 3.1.6 und 3.1.7 erstreckt.

Für eine im Punkt $x \in \mathbb{R}^p$ zweifach stetig differenzierbare Funktion $\psi : \mathbb{R}^p \rightarrow \mathbb{R}$ sei dazu im Folgenden

$$Q(\psi)(x) := \sum_{\alpha=1}^p \sum_{\beta=1}^p \int v_{\alpha} v_{\beta} \frac{\partial^2 \psi(x)}{\partial \alpha \partial \beta} K(v) dv. \quad (3.19)$$

3.1.8 Satz. Sei

$$\eta := \lim_{n \rightarrow \infty} k_n \cdot n^{\frac{-4}{4+p}} < \infty.$$

Dann gilt unter den Voraussetzungen aus Abschnitt 3.1.2 (für $n \rightarrow \infty$):

$$\sqrt{k_n}(m_n(x) - m(x)) \xrightarrow{\mathcal{L}} \mathcal{N}\left(B, c \cdot \sigma^2(x) \int K^2(v) dv\right),$$

mit

$$B := \eta^{1/2+2/p} \frac{Q(mf)(x) - m(x)Q(f)(x)}{2f(x)(cf(x))^{2/p}}. \quad (3.20)$$

Beweisübersicht. Der Beweis wird in zwei Teilen durchgeführt. Mit den Bezeichnungen

$$A_n(x) := \sqrt{k_n} \left(m_n(x) - E(m_n(x) | R_n) \right)$$

und

$$B_n(x) := \sqrt{k_n} \left(E(m_n(x) | R_n) - m(x) \right)$$

gilt offensichtlich

$$A_n(x) + B_n(x) = \sqrt{k_n} (m_n(x) - m(x)).$$

Im ersten Teil des Beweises (Abschnitt 3.1.6) wird - unter Verwendung des zentralen Grenzwertsatzes von Berry-Esséen - die Konvergenz

$$A_n(x) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, c \cdot \sigma^2(x) \int K^2(v) dv\right) \quad \text{für } n \rightarrow \infty$$

gezeigt.¹⁸

Danach wird der Beweis mit dem Nachweis von

$$B_n \xrightarrow{n \rightarrow \infty} B \quad \text{n.W.}$$

¹⁸Das grundsätzliche Vorgehen entspricht dabei der Verfahrensweise in [Mack81, S. 320ff.].

(in Abschnitt 3.1.7) abgeschlossen.¹⁹ □

Da $m_n(x)$ ein Quotient zweier Zufallsvariablen ist, erscheint eine direkte Analyse kaum möglich. Deswegen wählt [Mack81] den Weg, $m_n(x)$ approximativ additiv zu zerlegen und dabei die Differenzen

$$m_n(x) - \frac{E(h_n(x)|R_n)}{E(f_n(x)|R_n)} \quad \text{bzw.} \quad E(m_n(x)|R_n) - \frac{E(h_n(x)|R_n)}{E(f_n(x)|R_n)}$$

näher zu untersuchen, siehe Satz 3.1.19.

Hierfür sind allerdings einige Vorüberlegungen über das stochastische Verhalten spezieller bedingter Momente von $f_n(x)$ und $h_n(x)$ gegeben R_n (siehe Lemma 3.1.11) genauso notwendig wie eine relativ aufwändige Untersuchung von Termen z.B. der Form $\left(f_n(x) - E(f_n(x)|R_n)\right)^2$ (siehe Satz 3.1.18).

Zur Berechnung bedingter Momente auftretender Funktionen von $(\tilde{X}_i, \tilde{Y}_i)_{i \in \underline{k_n-1}}$ eignen sich folgende zwei Lemmas:²⁰

3.1.9 Lemma. *Unter den Annahmen aus Abschnitt 3.1.2 gilt für alle $q \in \{0, 1, 2\}$, $d \in \{1, 2\}$ und für alle $j \in \underline{k_n-1}$:*

$$E\left(\tilde{Y}_j^q K^d\left(\frac{x - \tilde{X}_j}{R_n}\right) \middle| R_n\right) = \frac{R_n^p}{T_n} \int K^d(v) \int y^q P^{(Y|X=x-vR_n)}(dy) f(x - vR_n) dv \quad P\text{-f.s.}$$

Beweis. Seien $q \in \{0, 1, 2\}$ und $d \in \{1, 2\}$ beliebig gewählt. Dann gilt für alle $j \in \underline{k_n-1}$ und für P^{R_n} -fast alle $r > 0$:

$$\begin{aligned} E\left(\tilde{Y}_j^q K^d\left(\frac{x - \tilde{X}_j}{R_n}\right) \middle| R_n = r\right) &= \int y^q K^d\left(\frac{x - u}{r}\right) P^{(\tilde{X}_1, \tilde{Y}_1)|R_n=r}(du, dy) \\ &= \int y^q K^d\left(\frac{x - u}{r}\right) \frac{1_{B_r(x)}(u)}{G(r)} P^{(X,Y)}(du, dy) \quad \text{nach Satz 3.1.3} \\ &= \frac{1}{G(r)} \int \int y^q K^d\left(\frac{x - u}{r}\right) P^{Y|X=u}(dy) f(u) du \\ &\quad \text{nach dem Satz von Fubini und wegen (3.6)} \\ &= \frac{1}{G(r)} \int r^p K^d(v) f(x - v \cdot r) \int y^q P^{Y|X=x-v \cdot r}(dy) dv, \end{aligned}$$

¹⁹[Mack81] weist diese Konvergenz nicht direkt nach. Unter etwas abweichenden Annahmen setzt er sich zum Ziel,

$$\lim_{n \rightarrow \infty} \sqrt{k_n} \left(E(m_n(x)) - m(x) \right) = B$$

und

$$\sqrt{k_n} \left(E(m_n(x)|R_n) - E m_n(x) \right) \xrightarrow{n \rightarrow \infty} 0 \quad \text{n.W.}$$

getrennt zu zeigen.

Einzelne Beweisschritte sind in dieser Arbeit übernommen worden, vgl. Fußnoten 24, 27 und 28.

²⁰Vgl. etwa das Vorgehen in [Mack/Rosenblatt79, S. 10+12].

nach dem Transformationssatz ([Königsberger00, S. 289]), denn für

$$\phi : \mathbb{R}^p \rightarrow \mathbb{R}^p \quad \text{mit} \quad \phi(u) := \frac{x - u}{r}$$

hat die Jacobi-Matrix ϕ' die Gestalt

$$\phi'(u) = \left(\frac{\partial \phi_i(u)}{\partial u_j} \right)_{i,j \in \underline{p}} = -\frac{1}{r} E_p,$$

d.h. $|\det \phi'(u)| = r^{-p}$. Wegen $T_n = G(R_n)$ ist damit die Behauptung bewiesen. \square

3.1.10 Lemma. *Unter den Annahmen aus Abschnitt 3.1.2 gilt für alle $q \in \{0, 1, 2\}$, $d \in \{1, 2\}$ und für alle $j \in \underline{k_n - 1}$:*

$$\frac{T_n}{R_n^p} E \left(\tilde{Y}_j^q K^d \left(\frac{x - \tilde{X}_j}{R_n} \right) \mid R_n \right) \xrightarrow{n \rightarrow \infty} f(x) E(Y^q | X = x) \int K^d(v) dv \quad \text{n.W.}^{21}$$

Beweis. Seien $q \in \{0, 1, 2\}$ und $d \in \{1, 2\}$ beliebig gewählt. Aufgrund von (3.6) gilt offenbar

$$\int K^d(v) dv < \infty. \quad (3.21)$$

Nach Voraussetzung an f , m und l ist (für alle $v \in \mathbb{R}^p$)

$$r \mapsto f(x - vr) \int y^q P^{Y|X=x-vr}(dy)$$

stetig und beschränkt. Da K beschränkt ist, folgt wegen $R_n \xrightarrow{n \rightarrow \infty} 0$ n.W. (Lemma 3.1.4) also (für $n \rightarrow \infty$) nach dem Satz von der majorisierten Konvergenz

$$\begin{aligned} \int K^d(v) f(x - vR_n) \int y^q P^{Y|X=x-vR_n}(dy) dv &\xrightarrow{\text{n.W.}} \int K^d(v) f(x) \int y^q P^{Y|X=x}(dy) dv \\ &= f(x) E(Y^q | X = x) \int K^d(v) dv \end{aligned}$$

und damit nach Lemma 3.1.9 für alle $j \in \underline{k_n - 1}$

$$\begin{aligned} \frac{T_n}{R_n^p} E \left(\tilde{Y}_j^q K^d \left(\frac{x - \tilde{X}_j}{R_n} \right) \mid R_n \right) &= \int K^d(v) f(x - vR_n) \int y^q P^{Y|X=x-vR_n}(dy) dv \\ &\xrightarrow{n \rightarrow \infty} f(x) E(Y^q | X = x) \int K^d(v) dv \quad \text{n.W.} \end{aligned}$$

\square

Nun lässt sich leicht das stochastische Verhalten einiger wichtiger Zufallsvariablen angeben:²²

²¹Dabei wird $E(Y^q | X = x)$ für $q \in \{1, 2\}$ interpretiert als $m(x)$ bzw. $l(x)$.

²²Vgl. [Mack81, S. 320f.].

3.1.11 Lemma. *Unter den Annahmen aus Abschnitt 3.1.2 gilt²³*

(i)

$$E(f_n(x)|R_n) \xrightarrow{n \rightarrow \infty} f(x) \quad \text{n. W.}$$

(ii)

$$k_n \text{Var}(f_n(x)|R_n) \xrightarrow{n \rightarrow \infty} cf^2(x) \int K^2(v)dv - f^2(x) \quad \text{n. W.}$$

(iii)

$$E(h_n(x)|R_n) \xrightarrow{n \rightarrow \infty} m(x)f(x) \quad \text{n. W.}$$

(iv)

$$k_n \text{Cov}(h_n(x), f_n(x)|R_n) \xrightarrow{n \rightarrow \infty} cf^2(x)m(x) \int K^2(v)dv - m(x)f^2(x) \quad \text{n. W.}$$

(v)

$$k_n \text{Var}(h_n(x)|R_n) \xrightarrow{n \rightarrow \infty} cf^2(x)l(x) \int K^2(v)dv - m^2(x)f^2(x) \quad \text{n. W.}$$

Beweis. (i) Da die $(\tilde{X}_i, \tilde{Y}_i)$ gegeben R_n bedingt unabhängig und identisch verteilt sind (gemäß Satz 3.1.3), folgt mit (3.6) und Lemma 3.1.10 (P -fast sicher)

$$\begin{aligned} E(f_n(x)|R_n) &= \frac{1}{nR_n^p} \sum_{j=1}^n E\left(K\left(\frac{x - X_j}{R_n}\right) \mid R_n\right) \\ &= \frac{1}{nR_n^p} \sum_{j=1}^{k_n-1} E\left(K\left(\frac{x - \tilde{X}_j}{R_n}\right) \mid R_n\right) \end{aligned} \quad (3.22)$$

$$\begin{aligned} &= \frac{k_n - 1}{nT_n} \frac{T_n}{R_n^p} E\left(K\left(\frac{x - \tilde{X}_1}{R_n}\right) \mid R_n\right) \\ &\xrightarrow{n \rightarrow \infty} 1 \cdot f(x) \int K(v)dv = f(x) \quad \text{n. W.,} \end{aligned} \quad (3.23)$$

da $\frac{k_n}{nT_n} \xrightarrow{n \rightarrow \infty} 1$ n.W. nach Lemma 3.1.5 und (3.7).

²³Genau genommen reichen an dieser Stelle schwächere Annahmen noch aus. Insbesondere werden hier (sowie für die im Beweis verwendeten Lemmas) die Differenzierbarkeitseigenschaften von f und m nicht benötigt. Daher sind auch (i) und (iii) als eigenständige Resultate mit aufgeführt, die sich ansonsten jeweils als Spezialfall aus Korollar 3.1.27 ergeben würden.

(ii) Es gilt wie in (i) (P -fast sicher)

$$k_n \operatorname{Var}(f_n(x)|R_n) \quad (3.24)$$

$$= \frac{k_n}{n^2 R_n^{2p}} \operatorname{Var}\left(\sum_{j=1}^n K\left(\frac{x - X_j}{R_n}\right) \mid R_n\right)$$

$$= \frac{k_n}{n^2 R_n^{2p}} \operatorname{Var}\left(\sum_{j=1}^{k_n-1} K\left(\frac{x - \tilde{X}_j}{R_n}\right) \mid R_n\right)$$

$$= \frac{k_n(k_n - 1)}{n^2 R_n^{2p}} \operatorname{Var}\left(K\left(\frac{x - \tilde{X}_1}{R_n}\right) \mid R_n\right), \quad (3.25)$$

denn die \tilde{X}_i sind bedingt unabhängig und identisch verteilt gegeben R_n

$$= \frac{k_n(k_n - 1)}{n^2 R_n^{2p} T_n} \frac{T_n}{R_n^p} E\left(K^2\left(\frac{x - \tilde{X}_1}{R_n}\right) \mid R_n\right) - \frac{k_n(k_n - 1)}{n^2 T_n^2} \left(\frac{T_n}{R_n^p} E\left(K\left(\frac{x - \tilde{X}_1}{R_n}\right) \mid R_n\right)\right)^2$$

$$\xrightarrow{n \rightarrow \infty} cf(x) \cdot 1 \cdot f(x) \cdot 1 \cdot \int K^2(v) dv - 1 \cdot f^2(x) \quad \text{n.W.}$$

mit Lemma 3.1.10, denn $\frac{k_n}{nT_n} \xrightarrow{n \rightarrow \infty} 1$ n.W. und $\frac{k_n}{nR_n^p} \xrightarrow{n \rightarrow \infty} cf(x)$ n.W. (gemäß den Lemmas 3.1.5 und 3.1.6).

(iii) Es gilt wie in (i) (P -fast sicher)

$$\begin{aligned} E(h_n(x)|R_n) &= \frac{1}{nR_n^p} E\left(\sum_{j=1}^n K\left(\frac{x - X_j}{R_n}\right) Y_j \mid R_n\right) \\ &= \frac{k_n - 1}{nT_n} \frac{T_n}{R_n^p} E\left(K\left(\frac{x - \tilde{X}_1}{R_n}\right) \tilde{Y}_1 \mid R_n\right) \end{aligned} \quad (3.26)$$

$$\xrightarrow{n \rightarrow \infty} 1 \cdot f(x) E(Y|X = x) \int K(v) dv \quad \text{n.W.}$$

$$= f(x)m(x) \quad (3.27)$$

nach den Lemmas 3.1.5, 3.1.6 und 3.1.10.

(iv) Wegen

$$E(h_n(x)|R_n) = \frac{k_n - 1}{nR_n^p} E\left(K\left(\frac{x - \tilde{X}_1}{R_n}\right) \tilde{Y}_1 \mid R_n\right) \quad (\text{nach (3.26)}), \quad (3.28)$$

$$E(f_n(x)|R_n) = \frac{k_n - 1}{nR_n^p} E\left(K\left(\frac{x - \tilde{X}_1}{R_n}\right) \mid R_n\right) \quad (\text{nach (3.22)}) \quad (3.29)$$

und

$$\begin{aligned} & E(h_n(x)f_n(x)|R_n) \\ &= \frac{1}{n^2 R_n^{2p}} \sum_{i=1}^{k_n-1} \sum_{j=1}^{k_n-1} E\left(K\left(\frac{x - \tilde{X}_i}{R_n}\right) \tilde{Y}_i \cdot K\left(\frac{x - \tilde{X}_j}{R_n}\right) \mid R_n\right) \\ &= \frac{k_n - 1}{n^2 R_n^{2p}} E\left(K^2\left(\frac{x - \tilde{X}_1}{R_n}\right) \tilde{Y}_1 \mid R_n\right) \\ &\quad + \frac{(k_n - 1)(k_n - 2)}{n^2 R_n^{2p}} E\left(K\left(\frac{x - \tilde{X}_1}{R_n}\right) \tilde{Y}_1 \mid R_n\right) E\left(K\left(\frac{x - \tilde{X}_1}{R_n}\right) \mid R_n\right) \quad P\text{-f.s.} \\ & \quad (\text{aufgrund der bedingten Unabhängigkeit der } (\tilde{X}_i, \tilde{Y}_i) \text{ gegebenen } R_n) \end{aligned}$$

hat man insgesamt (P -fast sicher)

$$\begin{aligned} & k_n \text{Cov}(h_n(x), f_n(x)|R_n) \\ &= k_n E(h_n(x)f_n(x)|R_n) - k_n E(h_n(x)|R_n) E(f_n(x)|R_n) \\ &= \frac{k_n(k_n - 1)}{n^2 R_n^{2p}} E\left(K^2\left(\frac{x - \tilde{X}_1}{R_n}\right) \tilde{Y}_1 \mid R_n\right) \\ &\quad + \frac{k_n((k_n - 1)(k_n - 2) - (k_n - 1)^2)}{n^2 R_n^{2p}} E\left(K\left(\frac{x - \tilde{X}_1}{R_n}\right) \tilde{Y}_1 \mid R_n\right) E\left(K\left(\frac{x - \tilde{X}_1}{R_n}\right) \mid R_n\right) \\ &\xrightarrow{n \rightarrow \infty} cf(x) \cdot 1 \cdot f(x)m(x) \int K^2(v)dv - 1 \cdot f(x)m(x)f(x) \quad \text{n.W.} \end{aligned}$$

nach den Lemmas 3.1.5, 3.1.6 und 3.1.10, denn

$$\frac{k_n((k_n - 1)(k_n - 2) - (k_n - 1)^2)}{n^2 R_n^{2p}} = -\frac{k_n(k_n - 1)}{n^2 T_n^2} \frac{T_n^2}{R_n^{2p}}.$$

(v) Wegen (3.28) und (wie in (iv))

$$\begin{aligned} E(h_n^2(x)|R_n) &= \frac{k_n - 1}{n^2 R_n^{2p}} E\left(K^2\left(\frac{x - \tilde{X}_1}{R_n}\right) \tilde{Y}_1^2 \mid R_n\right) \\ &\quad + \frac{(k_n - 1)(k_n - 2)}{n^2 R_n^{2p}} E^2\left(K\left(\frac{x - \tilde{X}_1}{R_n}\right) \tilde{Y}_1 \mid R_n\right) \quad P\text{-f.s.} \end{aligned}$$

hat man insgesamt (P -fast sicher)

$$k_n \operatorname{Var}(h_n(x)|R_n) \quad (3.30)$$

$$\begin{aligned} &= k_n \left(E(h_n^2(x)|R_n) - E^2(h_n(x)|R_n) \right) \\ &= \frac{k_n(k_n - 1)}{n^2 R_n^{2p}} E \left(K^2 \left(\frac{x - \tilde{X}_1}{R_n} \right) \tilde{Y}_1^2 \mid R_n \right) \\ &\quad + \frac{k_n((k_n - 1)(k_n - 2) - (k_n - 1)^2)}{n^2 R_n^{2p}} E^2 \left(K \left(\frac{x - \tilde{X}_1}{R_n} \right) \tilde{Y}_1 \mid R_n \right) \end{aligned} \quad (3.31)$$

$$\xrightarrow{n \rightarrow \infty} c f(x) \cdot 1 \cdot f(x) E(Y^2|X = x) \int K^2(v) dv - 1 \cdot f^2(x) m^2(x) \quad \text{n.W.}$$

$$= c f^2(x) l(x) \int K^2(v) dv - f^2(x) m^2(x),$$

wiederum nach den Lemmas 3.1.5, 3.1.6 und 3.1.10.

□

Entsprechende stochastische Konvergenzen sollen nun auch für die Größen

$$\left(f_n(x) - E(f_n(x)|R_n) \right)^2, \quad \left(h_n(x) - E(h_n(x)|R_n) \right)^2$$

sowie

$$\left(f_n(x) - E(f_n(x)|R_n) \right) \left(h_n(x) - E(h_n(x)|R_n) \right)$$

hergeleitet werden.²⁴

Um dafür die Chebyshev-Ungleichung anwenden zu können, muss das Verhalten von $E \operatorname{Var}(f_n(x)|R_n)$ bzw. $E \operatorname{Var}(h_n(x)|R_n)$ analysiert werden. An der Gestalt von $\operatorname{Var}(f_n(x)|R_n)$ und $\operatorname{Var}(h_n(x)|R_n)$ im Beweis zu Lemma 3.1.11 erkennt man, dass hierfür spezielle Momente von $\frac{1}{R_n^p}$ bzw. $\frac{1}{T_n}$ berechnet werden müssen.²⁵

Da G im Allgemeinen nur innerhalb einer Umgebung von Null streng monoton steigend ist, existiert die Umkehrfunktion G^{-1} nicht zwangsläufig überall. Aus diesem Grund muss die verallgemeinerte Inverse G^{-1} von G betrachtet werden.

²⁴Vgl. Satz 3.1.18. Zusammen mit Lemma 3.1.11 lässt sich dann eine hinreichend genaue Abschätzung für den Ausdruck

$$E(m_n(x)|R_n) - \frac{E(h_n(x)|R_n)}{E(f_n(x)|R_n)}$$

angeben. Dies geschieht in Satz 3.1.19, der von Mack nicht bewiesen wird.

²⁵Die Grundidee für Lemma 3.1.13 stammt aus [Mack/Rosenblatt79, S. 5].

Für beliebige maßdefinierende Funktionen F definiert man dazu

$$F^{-1}(t) := \inf\{x \in \mathbb{R} \mid F(x) \geq t\}, \quad t \in \mathbb{R},$$

mit der Konvention

$$\inf \emptyset := +\infty.$$

Dann ergibt sich folgender Zusammenhang:

3.1.12 Lemma. *Sei F eine maßdefinierende Funktion mit zugehörigem Maß μ_F . Dann gilt*

$$\lambda_1^{F^{-1}} = \mu_F.$$

Beweis. Für alle $t, x \in \mathbb{R}$ gilt $F^{-1}(F(x)) \leq x$ und (da F rechtsseitig stetig ist) $F(F^{-1}(t)) \geq t$. Da F und F^{-1} monoton steigend sind, hat man daher für alle $a, b \in \mathbb{R}$ mit $a \leq b$:

$$\begin{aligned} \lambda_1^{F^{-1}}(a, b] &= \lambda_1\{x \in \mathbb{R} \mid F^{-1}(x) \in (a, b]\} \\ &= \lambda_1\{x \in \mathbb{R} \mid x \in (F(a), F(b))\} + \zeta_1 \lambda_1\{F(a)\} + \zeta_2 \lambda_1\{F(b)\} \\ &\quad \text{für gewisse } \zeta_1, \zeta_2 \in [0, 1] \\ &= \lambda_1(F(a), F(b)] = F(b) - F(a) = \mu_F(a, b]. \end{aligned}$$

□

3.1.13 Lemma. *Für $n \rightarrow \infty$ gilt*

(i)

$$\frac{k_n}{n^2} E\left(\frac{1}{R_n^p T_n}\right) = \frac{cf(x)}{k_n} + o\left(\frac{1}{k_n}\right)$$

(ii)

$$\frac{k_n}{n^2} E\left(\frac{1}{T_n^2}\right) = \frac{1}{k_n} + o\left(\frac{1}{k_n}\right)$$

(iii)

$$\frac{k_n^4}{n^4} E\left(\frac{1}{R_n^{2p} T_n^2}\right) = O(1)$$

(iv)

$$\frac{k_n^4}{n^4} E\left(\frac{1}{T_n^4}\right) = O(1)$$

Beweis. Wegen $T_n = G(R_n)$ folgt, da R_n die Dichte h aus (3.11) besitzt, für alle $i, j \in \mathbb{N}_0$:

$$\begin{aligned}
E\left(\frac{1}{R_n^i T_n^j}\right) &= \int \frac{1}{r^i (G(r))^j} h(r) dr \\
&= n \binom{n-1}{k_n-1} \int_0^\infty r^{-i} (G(r))^{k_n-1-j} (1-G(r))^{n-k_n} G'(r) dr \\
&= n \binom{n-1}{k_n-1} \int_0^\infty r^{-i} (G(r))^{k_n-1-j} (1-G(r))^{n-k_n} G(dr) \\
&= n \binom{n-1}{k_n-1} \int_0^\infty r^{-i} (G(r))^{k_n-1-j} (1-G(r))^{n-k_n} \lambda_1^{G^{-1}}(dr) \\
&\quad \text{nach Lemma 3.1.12} \\
&= n \binom{n-1}{k_n-1} \int_0^1 (G^{-1}(t))^{-i} (G(G^{-1}(t)))^{k_n-1-j} (1-G(G^{-1}(t)))^{n-k_n} dt \\
&= n \binom{n-1}{k_n-1} \int_0^1 (G^{-1}(t))^{-i} t^{k_n-1-j} (1-t)^{n-k_n} dt, \tag{3.32}
\end{aligned}$$

da G stetig ist (und deswegen $G(G^{-1}(t)) = t$ für alle $t \in (0, 1)$).

Im Fall $i > 0$ soll nun der Wert $G^{-1}(t)$ mittels Korollar 3.1.7 dargestellt werden. Für

$$\kappa_\lambda(t) := \left(\frac{t}{cf(x)}\right)^{\lambda/p}, \quad \lambda \in \mathbb{N}_0,$$

hat man dann (für $t \searrow 0$)

$$(G^{-1}(t))^\lambda = \kappa_\lambda(t) + \eta_\lambda(t) \quad \text{mit} \quad \eta_\lambda(t) = o(t^{\lambda/p}) \tag{3.33}$$

und

$$\frac{1}{(G^{-1}(t))^\lambda} = \frac{1}{\kappa_\lambda(t)} - \frac{\eta_\lambda(t)}{\kappa_\lambda(t)(G^{-1}(t))^\lambda}, \quad \text{mit} \tag{3.34}$$

$$\lim_{t \searrow 0} \frac{\eta_\lambda(t)}{(G^{-1}(t))^\lambda} = \lim_{t \searrow 0} \frac{1}{\frac{\kappa_\lambda(t)}{\eta_\lambda(t)} + 1} = 0.$$

Für ein beliebiges festgehaltenes $\lambda \in \mathbb{N}_0$ gibt es also ein $\tau_\lambda \in (0, 1)$ mit $G^{-1}(\tau_\lambda) > 0$ und

$$\left| \frac{\eta_\lambda(t)}{(G^{-1}(t))^\lambda} \right| \leq 1 \quad \text{für alle } t \in (0, \tau_\lambda].$$

Für jedes $t \in (\tau_\lambda, 1)$ ist außerdem

$$\begin{aligned}
\left| \frac{\eta_\lambda(t)}{(G^{-1}(t))^\lambda} \right| &= \left| 1 - \frac{\kappa_\lambda(t)}{(G^{-1}(t))^\lambda} \right| \\
&\leq 1 + \left| \frac{t^{\lambda/p}}{(cf(x))^{\lambda/p} (G^{-1}(t))^\lambda} \right| \\
&\leq 1 + \left| \frac{1}{(cf(x))^{\lambda/p} (G^{-1}(\tau_\lambda))^\lambda} \right| =: \tilde{M}_\lambda(x).
\end{aligned}$$

Es folgt also insgesamt

$$\left| \frac{\eta_\lambda(t)}{(G^{-1}(t))^\lambda} \right| \leq \tilde{M}_\lambda(x) \quad (3.35)$$

für alle $t \in (0, 1)$ und alle $\lambda \in \mathbb{N}_0$.

(i) Nach (3.32) hat man (für o.B.d.A. $k_n \geq 3$)

$$\begin{aligned} \frac{k_n - 1}{n^2} E\left(\frac{1}{R_n^p T_n}\right) &= \frac{k_n - 1}{n} \binom{n-1}{k_n-1} \int_0^1 (G^{-1}(t))^{-p} t^{k_n-2} (1-t)^{n-k_n} dt \\ &= \frac{n-1}{n} \binom{n-2}{k_n-2} \int_0^1 \left(\frac{1}{\kappa_p(t)} - \frac{\eta_p(t)}{\kappa_p(t)(G^{-1}(t))^p} \right) t^{k_n-2} (1-t)^{n-k_n} dt \end{aligned} \quad (3.36)$$

nach (3.34). Wegen $\kappa_p(t) = \frac{t}{cf(x)}$ (für alle $t \in (0, 1)$) ist

$$\begin{aligned} &\frac{n-1}{n} \binom{n-2}{k_n-2} \int_0^1 \frac{1}{\kappa_p(t)} t^{k_n-2} (1-t)^{n-k_n} dt \\ &= cf(x) \frac{n-1}{n(k_n-2)} \int_0^1 \underbrace{(n-2) \binom{n-3}{k_n-3} t^{k_n-3} (1-t)^{n-k_n} dt}_{\text{Dichte von Beta}(k_n-2, n-k_n+1)} \\ &= cf(x) \frac{n-1}{n(k_n-2)} \\ &= \frac{cf(x)}{k_n} + o\left(\frac{1}{k_n}\right) \quad \text{für } n \rightarrow \infty. \end{aligned} \quad (3.37)$$

Entsprechend ist

$$\begin{aligned} &\frac{n-1}{n} \binom{n-2}{k_n-2} \int_0^1 \frac{\eta_p(t)}{\kappa_p(t)(G^{-1}(t))^p} t^{k_n-2} (1-t)^{n-k_n} dt \\ &= \frac{n-1}{n(k_n-2)} \int_0^1 \frac{\eta_p(t)}{(G^{-1}(t))^p} (n-2) \binom{n-3}{k_n-3} t^{k_n-3} (1-t)^{n-k_n} dt \\ &= \frac{n-1}{n(k_n-2)} E\left(\frac{\eta_p(U_n)}{(G^{-1}(U_n))^p}\right), \end{aligned} \quad (3.38)$$

mit $U_n \sim \text{Beta}(k_n - 2, n - k_n + 1)$.

Wegen $EU_n = \frac{k_n - 2}{n - 1}$ folgt genauso wie in Lemma 3.1.4(i)

$$U_n \xrightarrow{n \rightarrow \infty} 0 \quad \text{n.W.};$$

außerdem ist $\frac{\eta_p(t)}{(G^{-1}(t))^p}$ beschränkt (nach 3.35), mit

$$\lim_{t \searrow 0} \frac{\eta_p(t)}{(G^{-1}(t))^p} = 0.$$

Nach dem Satz von der majorisierten Konvergenz gilt damit

$$\lim_{n \rightarrow \infty} E\left(\frac{\eta_p(U_n)}{(G^{-1}(U_n))^p}\right) = 0,$$

nach (3.38) also

$$\frac{n-1}{n} \binom{n-2}{k_n-2} \int_0^1 \frac{\eta_p(t)}{\kappa_p(t)(G^{-1}(t))^p} t^{k_n-2} (1-t)^{n-k_n} dt = o\left(\frac{1}{k_n}\right) \quad \text{für } n \rightarrow \infty.$$

Mit (3.36) und (3.37) folgt dann (für $n \rightarrow \infty$)

$$\frac{k_n}{n^2} E\left(\frac{1}{R_n^p T_n}\right) = \frac{k_n}{k_n-1} \frac{k_n-1}{n^2} E\left(\frac{1}{R_n^p T_n}\right) = \frac{cf(x)}{k_n} + o\left(\frac{1}{k_n}\right).$$

(ii) Nach (3.32) hat man (für o.B.d.A. $k_n \geq 3$)

$$\begin{aligned} \frac{k_n}{n^2} E\left(\frac{1}{T_n^2}\right) &= \frac{k_n}{n} \binom{n-1}{k_n-1} \int_0^1 t^{k_n-3} (1-t)^{n-k_n} dt \\ &= \frac{k_n(n-1)}{n(k_n-1)(k_n-2)} \int_0^1 \underbrace{(n-2) \binom{n-3}{k_n-3} t^{k_n-3} (1-t)^{n-k_n} dt}_{\text{Dichte von Beta}(k_n-2, n-k_n+1)} \\ &= \frac{(n-1)k_n}{n(k_n-1)(k_n-2)} = \frac{1}{k_n} + o\left(\frac{1}{k_n}\right) \quad \text{für } n \rightarrow \infty. \end{aligned}$$

(iii) Nach (3.32) hat man (für o.B.d.A. $k_n \geq 5$)

$$\begin{aligned} \frac{k_n^4}{n^4} E\left(\frac{1}{R_n^{2p} T_n^2}\right) &= \frac{k_n^4}{n^3} \binom{n-1}{k_n-1} \int_0^1 (G^{-1}(t))^{-2p} t^{k_n-3} (1-t)^{n-k_n} dt \\ &= \frac{k_n^4}{n^3} \binom{n-1}{k_n-1} \int_0^1 \left(\frac{1}{\kappa_{2p}(t)} - \frac{\eta_{2p}(t)}{\kappa_{2p}(t)(G^{-1}(t))^{2p}}\right) t^{k_n-3} (1-t)^{n-k_n} dt, \\ &= O(1) \quad \text{für } n \rightarrow \infty, \end{aligned}$$

denn wegen $\kappa_{2p}(t) = (cf(x))^{-2}t^2$ gilt (für $n \rightarrow \infty$)

$$\begin{aligned}
& \frac{k_n^4}{n^3} \binom{n-1}{k_n-1} \int_0^1 \frac{1}{\kappa_{2p}(t)} t^{k_n-3} (1-t)^{n-k_n} dt \\
&= \frac{k_n^4}{n^3} \binom{n-1}{k_n-1} \int_0^1 c^2 f^2(x) t^{k_n-5} (1-t)^{n-k_n} dt \\
&= c^2 f^2(x) \frac{k_n^4 (n-1)(n-2)(n-3)}{n^3 (k_n-1)(k_n-2)(k_n-3)(k_n-4)} \int_0^1 \underbrace{(n-4) \binom{n-5}{k_n-5} t^{k_n-5} (1-t)^{n-k_n}}_{\text{Dichte von Beta}(k_n-4, n-k_n+1)} dt = O(1)
\end{aligned} \tag{3.39}$$

und

$$\begin{aligned}
& \left| \frac{k_n^4}{n^3} \binom{n-1}{k_n-1} \int_0^1 \frac{\eta_{2p}(t)}{\kappa_{2p}(t) (G^{-1}(t))^{2p}} t^{k_n-3} (1-t)^{n-k_n} dt \right| \\
&\leq \frac{k_n^4}{n^3} \binom{n-1}{k_n-1} \int_0^1 \left| \frac{\tilde{M}_{2p}(x) c^2 f^2(x)}{t^2} \right| t^{k_n-3} (1-t)^{n-k_n} dt \quad \text{nach (3.35)} \\
&= \tilde{M}_{2p}(x) c^2 f^2(x) \frac{k_n^4}{n^3} \binom{n-1}{k_n-1} \int_0^1 t^{k_n-5} (1-t)^{n-k_n} dt = O(1) \quad \text{nach (3.39)}.
\end{aligned}$$

(iv) Wiederum nach (3.32) hat man (für o.B.d.A. $k_n \geq 5$)

$$\begin{aligned}
& \frac{k_n^4}{n^4} E\left(\frac{1}{T_n^4}\right) \\
&= \frac{k_n^4}{n^3} \binom{n-1}{k_n-1} \int_0^1 t^{k_n-5} (1-t)^{n-k_n} dt \\
&= \frac{k_n^4 (n-1)(n-2)(n-3)}{n^3 (k_n-1)(k_n-2)(k_n-3)(k_n-4)} \int_0^1 \underbrace{(n-4) \binom{n-5}{k_n-5} t^{k_n-5} (1-t)^{n-k_n}}_{\text{Dichte von Beta}(k_n-4, n-k_n+1)} dt \\
&= O(1) \quad \text{für } n \rightarrow \infty.
\end{aligned}$$

□

Zur Berechnung von $E \operatorname{Var}(f_n(x)|R_n)$ und $E \operatorname{Var}(h_n(x)|R_n)$ werden noch zwei weitere - aufeinander aufbauende - technische Resultate benötigt.²⁶

3.1.14 Lemma. *Seien ψ sowie ψ_j , $j \in \mathbb{N}$, stetige und beschränkte Funktionen. Dann gilt für alle $\alpha \in \mathbb{N}$, $d \in \{1, 2\}$ und $u \in \mathbb{R}^p$:*

(i)

$$\lim_{n \rightarrow \infty} E \left(\int K^d(u) (\psi(x - uR_n) - \psi(x)) du \right) = 0$$

(ii)

$$\lim_{n \rightarrow \infty} E \left(\prod_{j=1}^{\alpha} \int K^d(u) (\psi_j(x - uR_n) - \psi_j(x)) du \right) = 0$$

Beweis. Seien $\alpha \in \mathbb{N}$, $d \in \{1, 2\}$ und $u \in \mathbb{R}^p$ fest.

- (i) Aus der Stetigkeit von ψ folgt $\psi(x - uR_n) \xrightarrow{n \rightarrow \infty} \psi(x)$ n.W. Aufgrund der Beschränktheit von ψ erhält man also nach dem Satz von der majorisierten Konvergenz

$$\lim_{n \rightarrow \infty} E(\psi(x - uR_n) - \psi(x)) = 0 \quad (3.40)$$

und damit (für $n \rightarrow \infty$)

$$\begin{aligned} E \left(\int K^d(u) (\psi(x - uR_n) - \psi(x)) du \right) &= \int K^d(u) E(\psi(x - uR_n) - \psi(x)) du \\ &\quad \text{nach dem Satz von Fubini (wegen (3.40))} \\ &= o(1) \end{aligned}$$

wegen (3.40) und der Beschränktheit von K sowie (3.21) wiederum nach dem Satz von der majorisierten Konvergenz.

- (ii) Dies folgt wegen (für alle $\tilde{\alpha} \in \underline{\alpha}$)

$$\begin{aligned} &\prod_{j=1}^{\tilde{\alpha}} \int K^d(u) (\psi_j(x - uR_n) - \psi_j(x)) du \\ &= \left(\prod_{j=1}^{\tilde{\alpha}-1} \int K^d(u) (\psi_j(x - uR_n) - \psi_j(x)) du \right) \cdot \int K^d(u) (\psi_{\tilde{\alpha}}(x - uR_n) - \psi_{\tilde{\alpha}}(x)) du \end{aligned}$$

sowie der Beschränktheit von ψ_j , $j \in \mathbb{N}$, und K erneut mit dem Satz von der majorisierten Konvergenz induktiv aus (i).

²⁶Das Berechnen dieser beiden Größen erfolgt weitgehend analog zum Vorgehen in [Mack/Rosenblatt79, S. 6f.]. Dort wird der Wert $E \operatorname{Var}(f_n(x)|R_n)$ benötigt, um einen asymptotischen Ausdruck für $\operatorname{Var}(f_n(x))$ zu berechnen. Lemma 3.1.16 entspricht dabei *Proposition 1* in [Mack/Rosenblatt79] (bis auf eine dort überflüssigerweise getroffene Zusatzannahme). Zu Lemma 3.1.17 vgl. [Mack81, S. 317f.].

□

Hieraus folgt

3.1.15 Lemma. *Seien ψ_j , $j \in \mathbb{N}$, stetige und beschränkte Funktionen. Dann gilt für alle $\alpha \in \mathbb{N}$, $d \in \{1, 2\}$ und $u \in \mathbb{R}^p$ (für $n \rightarrow \infty$):*

(i)

$$E\left(\frac{k_n}{n^2} \frac{1}{T_n^2} \left(\prod_{j=1}^{\alpha} \int K^d(u) (\psi_j(x - uR_n) - \psi_j(x)) du\right)\right) = o\left(\frac{1}{k_n}\right)$$

(ii)

$$E\left(\frac{k_n}{n^2} \frac{1}{R_n^p T_n} \left(\prod_{j=1}^{\alpha} \int K^d(u) (\psi_j(x - uR_n) - \psi_j(x)) du\right)\right) = o\left(\frac{1}{k_n}\right)$$

Beweis. (i) Es gilt (für $n \rightarrow \infty$)

$$\begin{aligned} & E\left(\frac{k_n^2}{n^2} \frac{1}{T_n^2} \left(\prod_{j=1}^{\alpha} \int K^d(u) (\psi_j(x - uR_n) - \psi_j(x)) du\right)\right) \\ & \leq \sqrt{E\left(\frac{1}{T_n^4}\right) \frac{k_n^4}{n^4} E\left(\prod_{j=1}^{\alpha} \int K^d(u) (\psi_j(x - uR_n) - \psi_j(x)) du\right)^2} \\ & \quad \text{nach der Ungleichung von Cauchy-Schwarz} \\ & = o(1), \end{aligned}$$

denn

$$\frac{k_n^4}{n^4} E\left(\frac{1}{T_n^4}\right) = O(1) \quad \text{für } n \rightarrow \infty$$

nach Lemma 3.1.13(iv) und

$$\lim_{n \rightarrow \infty} E\left(\prod_{j=1}^{2\alpha} \int K^d(u) (\psi_j(x - uR_n) - \psi_j(x)) du\right) = 0$$

(mit $\psi_{\alpha+j} := \psi_j$ für alle $j \in \underline{\alpha}$) nach Lemma 3.1.14(ii).

(ii) Dies ergibt sich mit Lemma 3.1.14(ii) wegen

$$\frac{k_n^4}{n^4} E\left(\frac{1}{R_n^{2p} T_n^2}\right) = O(1) \quad \text{für } n \rightarrow \infty$$

nach Lemma 3.1.13(iii) völlig analog zu (i).

□

3.1.16 Lemma. Für $n \rightarrow \infty$ gilt

$$E\left(\text{Var}(f_n(x)|R_n)\right) = \frac{cf^2(x)}{k_n} \int K^2(u) du - \frac{f^2(x)}{k_n} + o\left(\frac{1}{k_n}\right)$$

Beweis. Nach (3.24) hat man

$$\begin{aligned} \text{Var}(f_n(x)|R_n) &= \frac{k_n - 1}{n^2 R_n^{2p}} \text{Var}\left(K\left(\frac{x - \tilde{X}_1}{R_n}\right) \mid R_n\right) \\ &= \frac{k_n - 1}{n^2} \frac{1}{R_n^p T_n} \int K^2(u) f(x - uR_n) du - \frac{k_n - 1}{n^2} \frac{1}{T_n^2} \left(\int K(u) f(x - uR_n) du\right)^2. \end{aligned} \quad (3.41)$$

Die beiden Summanden werden nun getrennt behandelt. Es gilt (für $n \rightarrow \infty$)

$$\begin{aligned} &E\left(\frac{k_n - 1}{n^2} \frac{1}{R_n^p T_n} \int K^2(u) f(x - uR_n) du\right) \\ &= \frac{k_n - 1}{n^2} E\left(\frac{1}{R_n^p T_n}\right) f(x) \int K^2(u) du + \frac{k_n - 1}{n^2} E\left(\frac{1}{R_n^p T_n} \int K^2(u) (f(x - uR_n) - f(x)) du\right) \\ &= \frac{cf(x)}{k_n} f(x) \int K^2(u) du + o\left(\frac{1}{k_n}\right) \end{aligned} \quad (3.42)$$

nach Lemma 3.1.13(i) und Lemma 3.1.15(ii), denn f ist stetig und beschränkt. Ebenso hat man wegen

$$\begin{aligned} &\left(\int K(u) f(x - uR_n) du\right)^2 = \left(f(x) \int K(u) du + \int K(u) (f(x - uR_n) - f(x)) du\right)^2 \\ &= f^2(x) + 2f(x) \int K(u) (f(x - uR_n) - f(x)) du + \left(\int K(u) (f(x - uR_n) - f(x)) du\right)^2 \quad (3.43) \\ &\quad (\text{da } \int K(u) du = 1) \end{aligned}$$

dann

$$\begin{aligned} &E\left(\frac{k_n - 1}{n^2} \frac{1}{T_n^2} \left(\int K(u) f(x - uR_n) du\right)^2\right) \\ &= f^2(x) E\left(\frac{k_n - 1}{n^2} \frac{1}{T_n^2}\right) + 2f(x) E\left(\frac{k_n - 1}{n^2} \frac{1}{T_n^2} \int K(u) (f(x - uR_n) - f(x)) du\right) \\ &\quad + E\left(\frac{k_n - 1}{n^2} \frac{1}{T_n^2} \left(\int K(u) (f(x - uR_n) - f(x)) du\right)^2\right) \\ &= \frac{f^2(x)}{k_n} + o\left(\frac{1}{k_n}\right) \quad \text{für } n \rightarrow \infty \end{aligned}$$

nach Lemma 3.1.13(ii) und Lemma 3.1.15(i). Zusammen mit (3.41) und (3.42) ist die Behauptung bewiesen. \square

3.1.17 Lemma. Für $n \rightarrow \infty$ gilt

$$E\left(\text{Var}(h_n(x)|R_n)\right) = \frac{cf^2(x)l(x)}{k_n} \int K^2(u)du - \frac{m^2(x)f^2(x)}{k_n} + o\left(\frac{1}{k_n}\right).$$

Beweis. Nach (3.30) und Lemma 3.1.9 hat man

$$\begin{aligned} & \text{Var}(h_n(x)|R_n) \\ &= \frac{k_n - 1}{n^2 R_n^{2p}} E\left(K^2\left(\frac{x - \tilde{X}_1}{R_n}\right) \tilde{Y}_1^2 \mid R_n\right) - \frac{k_n - 1}{n^2 R_n^{2p}} E^2\left(K\left(\frac{x - \tilde{X}_1}{R_n}\right) \tilde{Y}_1 \mid R_n\right) \\ &= \frac{k_n - 1}{n^2 R_n^{2p}} \frac{R_n}{T_n} \int K^2(u) \int y^2 P^{(Y|X=x-uR_n)}(dy) f(x - uR_n) du \\ & \quad - \frac{k_n - 1}{n^2 R_n^{2p}} \frac{R_n^{2p}}{T_n^2} \left(\int K(u) \int y P^{(Y|X=x-uR_n)}(dy) f(x - uR_n) du \right)^2 \\ &= \frac{k_n - 1}{n^2 R_n^p T_n} \int K^2(u) (lf)(x - uR_n) du - \frac{k_n - 1}{n^2 T_n^2} \left(\int K(u) (mf)(x - uR_n) du \right)^2. \end{aligned} \quad (3.44)$$

Dann ergibt sich (analog zu (3.42))

$$\begin{aligned} & E\left(\frac{k_n - 1}{n^2} \frac{1}{R_n^p T_n} \int K^2(u) (lf)(x - uR_n) du\right) \\ &= \frac{k_n - 1}{n^2} E\left(\frac{1}{R_n^p T_n}\right) l(x) f(x) \int K^2(u) du \\ & \quad + \frac{k_n - 1}{n^2} E\left(\frac{1}{R_n^p T_n} \int K^2(u) ((lf)(x - uR_n) - (lf)(x)) du\right) \end{aligned} \quad (3.45)$$

$$= \frac{cf(x)}{k_n} l(x) f(x) \int K^2(u) du + o\left(\frac{1}{k_n}\right) \quad \text{für } n \rightarrow \infty, \quad (3.46)$$

denn f und l sind stetig und beschränkt. Ebenso hat man wegen (vgl. (3.43))

$$\begin{aligned} & \left(\int K(u) (mf)(x - uR_n) du \right)^2 \\ &= m^2(x) f^2(x) + 2m(x) f(x) \int K(u) ((mf)(x - uR_n) - (mf)(x)) du \\ & \quad + \left(\int K(u) ((mf)(x - uR_n) - (mf)(x)) du \right)^2 \end{aligned}$$

dann (für $n \rightarrow \infty$)

$$\begin{aligned}
& E\left(\frac{k_n - 1}{n^2} \frac{1}{T_n^2} \left(\int K(u)(mf)(x - uR_n) du\right)^2\right) \\
&= m^2(x) f^2(x) E\left(\frac{k_n - 1}{n^2} \frac{1}{T_n^2}\right) \\
&\quad + 2m(x) f(x) E\left(\frac{k_n - 1}{n^2} \frac{1}{T_n^2} \int K(u) ((mf)(x - uR_n) - (mf)(x)) du\right) \\
&\quad + E\left(\frac{k_n - 1}{n^2} \frac{1}{T_n^2} \left(\int K(u) ((mf)(x - uR_n) - (mf)(x)) du\right)^2\right) \\
&= \frac{m^2(x) f^2(x)}{k_n} + o\left(\frac{1}{k_n}\right)
\end{aligned}$$

nach Lemma 3.1.13(ii) und Lemma 3.1.15. Zusammen mit (3.44) und (3.45) ist die Behauptung bewiesen. \square

Nach den ausführlichen Vorbereitungen folgt nun unmittelbar:

3.1.18 Satz. *Unter den Annahmen aus 3.1.2 gilt*

(i)

$$\sqrt{k_n} \left(f_n(x) - E(f_n(x)|R_n) \right)^2 \xrightarrow{n \rightarrow \infty} 0 \quad n.W.$$

(ii)

$$\sqrt{k_n} \left(h_n(x) - E(h_n(x)|R_n) \right)^2 \xrightarrow{n \rightarrow \infty} 0 \quad n.W.$$

(iii)

$$\sqrt{k_n} \left(f_n(x) - E(f_n(x)|R_n) \right) \left(h_n(x) - E(h_n(x)|R_n) \right) \xrightarrow{n \rightarrow \infty} 0 \quad n.W.$$

Beweis. (i) Aus der Markov-Ungleichung folgt für alle $\varepsilon > 0$:

$$\begin{aligned}
& \lim_{n \rightarrow \infty} P\left\{ \sqrt{k_n} \left(f_n(x) - E(f_n(x)|R_n) \right)^2 > \varepsilon \right\} \\
& \leq \lim_{n \rightarrow \infty} \frac{E\left(E\left(\sqrt{k_n} \left(f_n(x) - E(f_n(x)|R_n) \right)^2 | R_n \right) \right)}{\varepsilon} \\
& = \lim_{n \rightarrow \infty} \frac{\sqrt{k_n} E \operatorname{Var}(f_n(x)|R_n)}{\varepsilon} = 0 \quad \text{nach Lemma 3.1.16.}
\end{aligned}$$

(ii) Genauso ergibt sich (nach Lemma 3.1.17) für alle $\varepsilon > 0$:

$$\lim_{n \rightarrow \infty} P\left\{ \sqrt{k_n} \left(h_n(x) - E(h_n(x)|R_n) \right)^2 > \varepsilon \right\} \leq \lim_{n \rightarrow \infty} \frac{\sqrt{k_n} E \operatorname{Var}(h_n(x)|R_n)}{\varepsilon} = 0.$$

(iii) Dies folgt wegen

$$\begin{aligned} & \sqrt{k_n} \left| \left(f_n(x) - E(f_n(x)|R_n) \right) \left(h_n(x) - E(h_n(x)|R_n) \right) \right| \\ &= \sqrt{\sqrt{k_n} \left(f_n(x) - E(f_n(x)|R_n) \right)^2} \sqrt{\sqrt{k_n} \left(h_n(x) - E(h_n(x)|R_n) \right)^2} \end{aligned}$$

sofort aus (i) und (ii). □

Aus der Absolutstetigkeit von X ergibt sich definitionsgemäß $E(f_n(x)|R_n) > 0$ P -f.s. Daher gilt schließlich:²⁷

3.1.19 Satz. Für $a_n := E(h_n(x)|R_n)$ und $b_n := E(f_n(x)|R_n)$ gilt (für $n \rightarrow \infty$)

(i)

$$m_n(x) = \frac{a_n}{b_n} + \frac{h_n(x) - a_n}{b_n} - \frac{a_n(f_n(x) - b_n)}{b_n^2} + o_p\left(\frac{1}{\sqrt{k_n}}\right)$$

(ii)

$$E(m_n(x)|R_n) = \frac{a_n}{b_n} + o_p\left(\frac{1}{\sqrt{k_n}}\right)$$

Beweis. (i) Für reelle Zahlen ξ_1, ξ_2, ξ_3 und ξ_4 mit $\xi_3 \neq 0$ sowie $\xi_3 + \xi_4 \neq 0$ gilt offenbar

$$\frac{\xi_1 + \xi_2}{\xi_3 + \xi_4} = \frac{\xi_1}{\xi_3} + \frac{\xi_2}{\xi_3} - \frac{\xi_1 \xi_4}{\xi_3^2} + \frac{\xi_1 \xi_4^2 - \xi_2 \xi_3 \xi_4}{\xi_3^2(\xi_3 + \xi_4)}. \quad (3.47)$$

Daher folgt

$$\begin{aligned} m_n(x) &= \frac{a_n + (h_n(x) - a_n)}{b_n + (f_n(x) - b_n)} \\ &= \frac{a_n}{b_n} + \frac{h_n(x) - a_n}{b_n} - \frac{a_n(f_n(x) - b_n)}{b_n^2} \\ &\quad + \frac{a_n(f_n(x) - b_n)^2 - b_n(h_n(x) - a_n)(f_n(x) - b_n)}{b_n^2 f_n(x)} \end{aligned} \quad (3.48)$$

und damit wegen

$$\begin{aligned} b_n &= E(f_n(x)|R_n) \xrightarrow{n \rightarrow \infty} f(x) > 0 \text{ n.W.} \quad \text{nach Lemma 3.1.11(i),} \\ a_n &= E(h_n(x)|R_n) \xrightarrow{n \rightarrow \infty} m(x)f(x) \text{ n.W.} \quad \text{nach Lemma 3.1.11(iii),} \\ f_n(x) - b_n &\xrightarrow{n \rightarrow \infty} 0 \text{ n.W.} \quad \text{nach Satz 3.1.18(i)} \end{aligned}$$

²⁷[Mack81, S. 321(47)] gibt lediglich Satz 3.1.19(i) an - ohne vollständigen Beweis (vgl. Fußnote 24). Für Teil (ii) des Satzes ist die Bedingung (3.5) essentiell, für die es in [Mack81] keine Entsprechung gibt. Vgl. hierzu Fußnote 9.

sowie (nach Lemma 3.1.18(i) und (iii))

$$\begin{aligned} & \sqrt{k_n}(f_n(x) - b_n)^2 \xrightarrow{n \rightarrow \infty} 0 \quad \text{n.W.} \\ \text{und} \quad & \sqrt{k_n}(h_n(x) - a_n)(f_n(x) - b_n) \xrightarrow{n \rightarrow \infty} 0 \quad \text{n.W.} \end{aligned}$$

direkt die Behauptung.

(ii) Wegen $a_n = E(h_n(x)|R_n)$ und $b_n = E(f_n(x)|R_n)$ hat man für alle $n \in \mathbb{N}$ mit $b_n > 0$:

$$\begin{aligned} E(m_n(x)|R_n) &= \frac{a_n}{b_n} + 0 - 0 + \frac{a_n}{b_n^2} E\left(\frac{(f_n(x) - b_n)^2}{f_n(x)} \mid R_n\right) \\ &\quad - \frac{1}{b_n} E\left(\frac{(h_n(x) - a_n)(f_n(x) - b_n)}{f_n(x)} \mid R_n\right). \end{aligned} \quad (3.49)$$

Aufgrund der Absolutstetigkeit von X gilt nach (3.5)

$$\sum_{j=1}^n K\left(\frac{x - X_j}{R_n}\right) \in [(k_n - 1)N_1, (k_n - 1)N_2] \quad P\text{-f.s.}$$

sowie

$$E\left(\sum_{j=1}^n K\left(\frac{x - X_j}{R_n}\right)\right) \in [(k_n - 1)N_1, (k_n - 1)N_2] \quad P\text{-f.s.},$$

d.h.

$$1 - \frac{b_n}{f_n(x)} = 1 - \frac{E(\sum_{j=1}^n K(\frac{x - X_j}{R_n}))}{\sum_{j=1}^n K(\frac{x - X_j}{R_n})} \in \left[1 - \frac{N_2}{N_1}, 1 - \frac{N_1}{N_2}\right],$$

also (da $N_1 \leq N_2$)

$$\left|1 - \frac{b_n}{f_n(x)}\right| \leq \frac{N_2}{N_1}.$$

Damit hat man (für $n \rightarrow \infty$)

$$\begin{aligned} & \sqrt{k_n} \frac{a_n}{b_n^2} E\left(\frac{(f_n(x) - b_n)^2}{f_n(x)} \mid R_n\right) \\ &= \frac{a_n}{b_n^2} \sqrt{k_n} E\left(\frac{(f_n(x) - b_n)^2}{b_n} \mid R_n\right) - \frac{a_n}{b_n^2} \sqrt{k_n} E\left(\frac{(f_n(x) - b_n)^2}{b_n} \frac{f_n(x) - b_n}{f_n(x)} \mid R_n\right) \\ &= \frac{a_n}{b_n^3} \sqrt{k_n} \text{Var}(f_n(x)|R_n) - \frac{a_n}{b_n^3} \sqrt{k_n} E\left((f_n(x) - b_n)^2 \left(1 - \frac{b_n}{f_n(x)}\right) \mid R_n\right) = o_p(1), \end{aligned} \quad (3.50)$$

denn

$$\sqrt{k_n - 1} \text{Var}(f_n(x)|R_n) \xrightarrow{n \rightarrow \infty} 0 \quad \text{n.W.} \quad \text{nach Lemma 3.1.11(ii),}$$

$a_n \xrightarrow{n \rightarrow \infty} m(x)f(x)$ n.W., $b_n \xrightarrow{n \rightarrow \infty} f(x)$ n.W. (nach Lemma 3.1.11(iii) und (ii)) und für alle $n > \tilde{N}$ ist

$$\begin{aligned} \left| \sqrt{k_n} E \left((f_n(x) - b_n)^2 \left(1 - \frac{b_n}{f_n(x)}\right) \middle| R_n \right) \right| &\leq \sqrt{k_n} E \left((f_n(x) - b_n)^2 \left|1 - \frac{b_n}{f_n(x)}\right| \middle| R_n \right) \\ &\leq \frac{N_2}{N_1} \sqrt{k_n} E \left((f_n(x) - b_n)^2 \middle| R_n \right) \\ &= o_p(1) \quad \text{für } n \rightarrow \infty. \end{aligned}$$

Ebenso gilt (für $n \rightarrow \infty$)

$$\begin{aligned} &\sqrt{k_n} \frac{1}{b_n} E \left(\frac{(h_n(x) - a_n)(f_n(x) - b_n)}{f_n(x)} \middle| R_n \right) \\ &= \frac{1}{b_n} \sqrt{k_n} E \left(\frac{(h_n(x) - a_n)(f_n(x) - b_n)}{b_n} \middle| R_n \right) \\ &\quad - \frac{1}{b_n} \sqrt{k_n} E \left(\frac{(h_n(x) - a_n)(f_n(x) - b_n)}{b_n} \frac{f_n(x) - b_n}{f_n(x)} \middle| R_n \right) \\ &= \frac{1}{b_n^2} \sqrt{k_n} \text{Cov}(h_n(x), f_n(x) | R_n) \\ &\quad - \frac{1}{b_n^2} \sqrt{k_n} E \left((h_n(x) - a_n)(f_n(x) - b_n) \left(1 - \frac{b_n}{f_n(x)}\right) \middle| R_n \right) = o_p(1), \end{aligned} \quad (3.51)$$

denn

$$\sqrt{k_n} \text{Cov}(h_n(x), f_n(x) | R_n) \xrightarrow{n \rightarrow \infty} 0 \text{ n.W.} \quad \text{nach Lemma 3.1.11(iv)}$$

und für alle $n > \tilde{N}$ ist

$$\begin{aligned} &\left| \sqrt{k_n} E \left((h_n(x) - a_n)(f_n(x) - b_n) \left(1 - \frac{b_n}{f_n(x)}\right) \middle| R_n \right) \right| \\ &\leq \frac{N_2}{N_1} \sqrt{k_n} E \left(|h_n(x) - a_n| \cdot |f_n(x) - b_n| \middle| R_n \right) \\ &\leq \frac{N_2}{N_1} \sqrt{\sqrt{k_n} E \left((h_n(x) - a_n)^2 \middle| R_n \right) \sqrt{k_n} E \left((f_n(x) - b_n)^2 \middle| R_n \right)} \\ &= o_p(1) \quad \text{für } n \rightarrow \infty \end{aligned}$$

nach der Ungleichung von Cauchy-Schwarz sowie Lemma 3.1.11(v) und (ii). Mit (3.49), (3.50) und (3.51) ist die Behauptung bewiesen. \square

3.1.6 Asymptotische Normalität von $A_n(x)$

Mit dem soeben bewiesenen Satz kann man nun relativ einfach zeigen, dass $A_n(x)$ näherungsweise als Summe zentrierter, unabhängiger und identisch verteilter Zufallsvariablen aufgefasst werden kann. Diese Art der Darstellung ist für die Anwendung des zentralen Grenzwertsatzes von Berry-Esséen nötig.

Sei dazu im Folgenden

$$Z_j := \frac{k_n - 1}{nR_n^p} \left(\tilde{Y}_j - \frac{a_n}{b_n^2} \right) K \left(\frac{x - \tilde{X}_j}{R_n} \right), \quad j \in \underline{k_n - 1}, \quad (3.52)$$

mit wieder $a_n := E(h_n(x)|R_n)$, $b_n := E(f_n(x)|R_n)$ für $n \in \mathbb{N}$ und

$$\Phi_n(x) := \frac{1}{\sqrt{k_n - 1}} \sum_{j=1}^{k_n-1} \frac{Z_j}{\sqrt{\text{Var}(Z_j|R_n)}}, \quad n \in \mathbb{N}. \quad (3.53)$$

Ein zentraler Aspekt dieses Abschnittes ist dann der Nachweis von $\Phi_n(x) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$.²⁸ Zunächst werden einige Eigenschaften bestimmter bedingter Momente der Z_j , $j \in \underline{k_n - 1}$, (gegeben R_n) festgehalten.

3.1.20 Lemma. *Es gilt*

(i)

$$E(Z_j|R_n) = 0 \text{ P-f.s. für alle } j \in \underline{k_n - 1}$$

sowie

(ii)

$$A_n(x) = \sqrt{\text{Var}(Z_1|R_n)} \Phi_n(x) + o_p(1) \quad \text{für } n \rightarrow \infty.$$

Beweis. (i) Es gilt für alle $j \in \underline{k_n - 1}$:

$$\begin{aligned} E(Z_j|R_n) &= \frac{k_n - 1}{nR_n^p b_n} E \left(\tilde{Y}_j K \left(\frac{x - \tilde{X}_j}{R_n} \right) \mid R_n \right) - \frac{(k_n - 1)a_n}{nR_n^p b_n^2} E \left(K \left(\frac{x - \tilde{X}_j}{R_n} \right) \mid R_n \right) \\ &= \frac{k_n - 1}{nR_n^p} \left(\frac{E \left(\tilde{Y}_j K \left(\frac{x - \tilde{X}_j}{R_n} \right) \mid R_n \right)}{E(f_n(x)|R_n)} - \frac{E(h_n(x)|R_n) E \left(K \left(\frac{x - \tilde{X}_j}{R_n} \right) \mid R_n \right)}{E^2(f_n(x)|R_n)} \right) \\ &= (k_n - 1) \left(\frac{E \left(\tilde{Y}_j K \left(\frac{x - \tilde{X}_j}{R_n} \right) \mid R_n \right)}{E \left(K \left(\frac{x - \tilde{X}_j}{R_n} \right) \mid R_n \right)} - \frac{E \left(\tilde{Y}_j K \left(\frac{x - \tilde{X}_j}{R_n} \right) \mid R_n \right)}{E \left(K \left(\frac{x - \tilde{X}_j}{R_n} \right) \mid R_n \right)} \right) = 0 \quad \text{P-f.s.} \end{aligned}$$

²⁸Mit der Vorgehensweise in diesem Abschnitt wird der Beweisidee aus [Mack81, S. 321] gefolgt. Die Aussage aus Lemma 3.1.22 wird dort allerdings nicht bewiesen, sondern implizit unterstellt. Für den entsprechenden Nachweis ist jedoch wiederum die Bedingung (3.5) notwendig, die von [Mack81] nicht vorausgesetzt wird, vgl. Fußnote 9.

nach Definition von $f_n(x)$ und $h_n(x)$.

- (ii) Da die $(\tilde{X}_j, \tilde{Y}_j)$, $j \in \underline{k_n - 1}$, gegeben R_n bedingt unabhängig und identisch verteilt sind, gilt dies auch für die Z_j , $j \in \underline{k_n - 1}$, d.h.

$$\text{Var}(Z_j|R_n) = \text{Var}(Z_1|R_n) \quad \forall j \in \underline{k_n - 1}$$

und damit

$$\sqrt{\text{Var}(Z_i|R_n)}\Phi_n(x) = \frac{1}{\sqrt{k_n - 1}} \sum_{j=1}^{k_n-1} Z_j \quad \forall i \in \underline{k_n - 1}.$$

Andererseits hat man gemäß Satz 3.1.19 (für $n \rightarrow \infty$)

$$m_n(x) - E(m_n(x)|R_n) = \frac{h_n(x) - a_n}{b_n} - \frac{a_n(f_n(x) - b_n)}{b_n^2} + o_p\left(\frac{1}{\sqrt{k_n}}\right)$$

und damit (für $n \rightarrow \infty$)

$$\begin{aligned} A_n(x) &= \sqrt{k_n} \left(m_n(x) - E(m_n(x)|R_n) \right) \\ &= \sqrt{k_n} \left(\frac{h_n(x) - a_n}{b_n} - \frac{a_n}{b_n^2} (f_n(x) - b_n) \right) + o_p(1) \\ &= \sqrt{k_n} \left(\frac{h_n(x)}{b_n} - \frac{a_n f_n(x)}{b_n^2} \right) + o_p(1) \\ &= \frac{\sqrt{k_n}}{nR_n^p} \sum_{j=1}^{k_n-1} \left(\frac{K\left(\frac{x - \tilde{X}_j}{R_n}\right) \tilde{Y}_j}{b_n} - \frac{a_n K\left(\frac{x - \tilde{X}_j}{R_n}\right)}{b_n^2} \right) + o_p(1) \\ &= \frac{\sqrt{k_n}}{k_n - 1} \sum_{j=1}^{k_n-1} \frac{k_n - 1}{nR_n^p} \left(\frac{\tilde{Y}_j}{b_n} - \frac{a_n}{b_n^2} \right) K\left(\frac{x - \tilde{X}_j}{R_n}\right) + o_p(1) \\ &= \frac{1}{\sqrt{k_n - 1}} \sum_{j=1}^{k_n-1} Z_j + o_p(1) \\ &= \sqrt{\text{Var}(Z_j|R_n)}\Phi_n(x) + o_p(1). \end{aligned}$$

□

3.1.21 Lemma. Mit Z_j aus (3.52) gilt für alle $j \in \underline{k_n - 1}$:

$$\text{Var}(Z_j|R_n) \xrightarrow{n \rightarrow \infty} c \cdot \sigma^2(x) \int K^2(v) dv \quad n. W.$$

Beweis. Sei $j \in \underline{k_n - 1}$. Wegen $E(Z_j | R_n) = 0$ P -f.s. (Lemma 3.1.20(i)) berechnet man wie in Lemma 3.1.11 (für $n \rightarrow \infty$)

$$\begin{aligned}
 \text{Var}(Z_j | R_n) &= E(Z_j^2 | R_n) \\
 &= \frac{(k_n - 1)^2}{n^2 R_n^{2p}} \frac{1}{b_n^2} E\left(\tilde{Y}_1^2 K^2\left(\frac{x - \tilde{X}_1}{R_n}\right) \mid R_n\right) - \frac{(k_n - 1)^2}{n^2 R_n^{2p}} \frac{2a_n}{b_n^3} E\left(\tilde{Y}_1 K^2\left(\frac{x - \tilde{X}_1}{R_n}\right) \mid R_n\right) \\
 &\quad + \frac{(k_n - 1)^2}{n^2 R_n^{2p}} \frac{a_n^2}{b_n^4} E\left(K^2\left(\frac{x - \tilde{X}_1}{R_n}\right) \mid R_n\right) \\
 &= \frac{(k_n - 1)^2}{n^2 R_n^{2p} b_n^2} \frac{R_n^p}{T_n} l(x) f(x) \int K^2(v) dv - \frac{2(k_n - 1)^2 a_n}{n^2 R_n^{2p} b_n^3} \frac{R_n^p}{T_n} m(x) f(x) \int K^2(v) dv \\
 &\quad + \frac{(k_n - 1)^2 a_n^2}{n^2 R_n^{2p} b_n^4} \frac{R_n^p}{T_n} f(x) \int K^2(v) dv + o_p(1) \quad \text{nach Lemma 3.1.10} \\
 &= \frac{1}{b_n^2} c f^2(x) l(x) \int K^2(v) dv - \frac{2a_n}{b_n^3} c f^2(x) m(x) \int K^2(v) dv + \frac{a_n^2}{b_n^4} c f^2(x) \int K^2(v) dv + o_p(1) \\
 &\quad \text{für } n \rightarrow \infty, \text{ denn } \frac{k_n - 1}{n R_n^p} \xrightarrow{n \rightarrow \infty} c f(x) \text{ n.W. und } \frac{k_n - 1}{n T_n} \xrightarrow{n \rightarrow \infty} 1 \text{ n.W.} \\
 &\quad \text{(Lemmas 3.1.5 und 3.1.6)} \\
 &\xrightarrow{n \rightarrow \infty} c l(x) \int K^2(v) dv - 2c m^2(x) \int K^2(v) dv + c m^2(x) \int K^2(v) dv \quad \text{n.W.}, \\
 &\quad \text{denn } a_n = E(h_n(x) | R_n) \xrightarrow{n \rightarrow \infty} m(x) f(x) \text{ n.W. und } b_n = E(f_n(x) | R_n) \xrightarrow{n \rightarrow \infty} f(x) \text{ n.W.} \\
 &\quad \text{(Lemma 3.1.11(iii) und (i))} \\
 &= c(l(x) - m^2(x)) \int K^2(v) dv \\
 &= c\sigma^2(x) \int K^2(v) dv.
 \end{aligned}$$

□

3.1.22 Lemma. Sei Z_1 wie in (3.52). Dann gibt es ein $\Delta > 0$ mit

$$E(|Z_1|^3 | R_n = r) \leq \Delta$$

für alle $n \in \mathbb{N}$ und für P^{R_n} -fast alle $r > 0$.

Beweis. Sei $n \in \mathbb{N}$ fest gewählt. Nach (3.28) und (3.29) hat man

$$a_n = E(h_n(x) | R_n) = \frac{k_n - 1}{n R_n^p} E\left(K\left(\frac{x - \tilde{X}_1}{R_n}\right) \tilde{Y}_1 \mid R_n\right)$$

sowie

$$b_n = E(f_n(x)|R_n) = \frac{k_n - 1}{nR_n^p} E\left(K\left(\frac{x - \tilde{X}_1}{R_n}\right) \mid R_n\right)$$

und damit

$$\frac{a_n}{b_n} = \frac{E(K(\frac{x - \tilde{X}_1}{R_n})\tilde{Y}_1 \mid R_n)}{E(K(\frac{x - \tilde{X}_1}{R_n}) \mid R_n)}.$$

Daher lässt sich Z_1 darstellen als

$$\begin{aligned} Z_1 &= \frac{k_n - 1}{nR_n^p} \cdot \frac{1}{b_n} \left(\tilde{Y}_1 - \frac{a_n}{b_n}\right) K\left(\frac{x - \tilde{X}_1}{R_n}\right) \\ &= \left(\tilde{Y}_1 - \frac{a_n}{b_n}\right) \frac{K\left(\frac{x - \tilde{X}_1}{R_n}\right)}{E(K(\frac{x - \tilde{X}_1}{R_n}) \mid R_n)}. \end{aligned}$$

Für alle reellen Zahlen $a, b \geq 0$ gilt

$$(a + b)^3 \leq (2 \max(a, b))^3 = 8 \max(a^3, b^3) \leq 8(a^3 + b^3)$$

und daher - für P^{R_n} -fast alle $r > 0$

$$\begin{aligned} E(|Z_1|^3 | R_n = r) &= E\left(\left|\tilde{Y}_1 - \frac{a_n}{b_n}\right|^3 \underbrace{\frac{K^3\left(\frac{x - \tilde{X}_1}{R_n}\right)}{E^3(K(\frac{x - \tilde{X}_1}{R_n}) | R_n)}}_{\leq (N_2/N_1)^3} \mid R_n = r\right) \\ &\leq \left(\frac{N_2}{N_1}\right)^3 E\left(\left(|\tilde{Y}_1| + \left|\frac{a_n}{b_n}\right|\right)^3 \mid R_n = r\right) \\ &\leq 8\left(\frac{N_2}{N_1}\right)^3 E\left(|\tilde{Y}_1|^3 + \left|\frac{a_n}{b_n}\right|^3 \mid R_n = r\right). \end{aligned} \quad (3.54)$$

Dabei ist für P^{R_n} -fast alle $r > 0$

$$\begin{aligned} E\left(\left|\frac{a_n}{b_n}\right|^3 \mid R_n = r\right) &= E\left(\frac{\left|E(K(\frac{x - \tilde{X}_1}{R_n})\tilde{Y}_1 | R_n)\right|^3}{E^3(K(\frac{x - \tilde{X}_1}{R_n}) | R_n)} \mid R_n = r\right) \\ &= \frac{\left|E(K(\frac{x - \tilde{X}_1}{R_n})\tilde{Y}_1 \mid R_n = r)\right|^3}{E^3(K(\frac{x - \tilde{X}_1}{R_n}) \mid R_n = r)} \\ &\leq \left(\frac{N_2}{N_1}\right)^3 |E(\tilde{Y}_1 | R_n = r)|^3 \\ &\leq \left(\frac{N_2}{N_1}\right)^3 E(|\tilde{Y}_1|^3 | R_n = r) \end{aligned} \quad (3.55)$$

nach der Jensenschen Ungleichung für bedingte Erwartungswerte²⁹, denn $\nu \mapsto |\nu|^3$ ist konvex. Nach Voraussetzung aus Abschnitt 3.1.2 gilt $E(|Y_j|^3|X = u) \leq \bar{M}$, $j \in \underline{k_n - 1}$, für P^X -fast alle $u \in \mathbb{R}^p$ und damit (für P^{R_n} -fast alle $r > 0$)

$$\begin{aligned}
E(|\tilde{Y}_j|^3|R_n = r) &= \int |y|^3 P^{\tilde{Y}_j|R_n=r}(dy) \\
&= \int |y|^3 P^{(\tilde{X}_j, \tilde{Y}_j)|R_n=r}(du, dy) \\
&= \frac{1}{G(r)} \int 1_{B_r(x)}(u) |y|^3 P^{(X,Y)}(du, dy) \quad \text{nach Satz 3.1.3(ii)} \\
&= \frac{1}{G(r)} \int 1_{B_r(x)}(u) \int |y|^3 P^{Y|X=u}(dy) P^X(du) \\
&= \frac{1}{G(r)} \int_{B_r(x)} E(|Y|^3|X = u) f(u) du \\
&\leq \bar{M} \frac{1}{G(r)} \int_{B_r(x)} f(u) du \\
&= \bar{M} \frac{1}{G(r)} \underbrace{P\{|X - x| < r\}}_{=G(r)} = \bar{M}.
\end{aligned}$$

Daher folgt aus (3.54) und (3.55) nun

$$E(|Z_1|^3|R_n = r) \leq 8 \left(\frac{N_2}{N_1} \right)^3 \bar{M} \left(1 + \left(\frac{N_2}{N_1} \right)^3 \right) =: \Delta$$

für P^{R_n} -fast alle $r > 0$. □

3.1.23 Satz. *Unter den Voraussetzungen aus 3.1.2 gilt*

$$A_n(x) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, c\sigma^2(x) \int K^2(v) dv\right).$$

Wie bereits angedeutet, wird dieser Satz mit Hilfe des zentralen Grenzwertsatzes von Berry-Esséen bewiesen.

3.1.24 Satz. *(Berry-Esséen) Sei $(\Theta_i)_{i \in \mathbb{N}}$ eine Folge unabhängiger, zentrierter Zufallsvariablen über (Ω, \mathcal{A}, P) mit $\sigma_i^2 := \text{Var}(\Theta_i) \in (0, \infty)$. Sei $\Lambda_n := \sum_{i=1}^n \Theta_i$, F_n die Verteilungsfunktion von $\frac{\Lambda_n}{\sqrt{\text{Var} \Lambda_n}}$ und $s_n := \sqrt{\text{Var} \Lambda_n} = \sqrt{\sum_{i=1}^n \sigma_i^2}$. Weiter sei Φ die Verteilungsfunktion von $\mathcal{N}(0, 1)$. Dann gilt für alle $n \in \mathbb{N}$:*

$$\sup_{x \in \mathbb{R}} |F_n(x) - \Phi(x)| \leq \frac{6}{s_n^3} \sum_{i=1}^n E|\Theta_i|^3.$$

Beweis. [Gänssler77, S. 165.] □

²⁹Vgl. z.B. [Gänssler77, S. 202].

3.1.25 Bemerkung. Wird der Satz von Berry-Esséen nur für ein festgehaltenes $n \in \mathbb{N}$ benutzt, so ist es natürlich auch möglich, ihn auf Verteilungen P^{Θ_i} von Zufallsvariablen Θ_i anzuwenden, die ihrerseits auch von n abhängen.

Beweis von Satz 3.1.23. Seien Z_i , $i \in \underline{k_n - 1}$, und $\Phi_n(x)$ wie in (3.52) bzw. (3.53). Mit Bemerkung 3.1.25 lässt sich der zentrale Grenzwertsatz von Berry-Esséen - bei festgehaltenen $r > 0$ und $n \in \mathbb{N}$ - auf die Verteilungen $P^{Z_i|R_n=r}$, $i \in \underline{k_n - 1}$, anwenden. Betrachtet man nämlich unabhängige und identisch verteilte Zufallsvariablen $U_i \sim P^{Z_i|R_n=r}$, $i \in \underline{k_n - 1}$, so gilt offensichtlich

$$P^{\Phi_n(x)|R_n=r} = \mathcal{L}\left(\frac{\sum_{j=1}^{k_n-1} U_j}{\sqrt{k_n-1}\sqrt{\text{Var} U_j}}\right) = \mathcal{L}\left(\frac{\sum_{j=1}^{k_n-1} U_j}{\sqrt{\text{Var}(\sum_{j=1}^{k_n-1} U_j)}}\right)$$

und Satz 3.1.24 ist verwendbar. Bezeichnet $F_n(\cdot|r)$ die Verteilungsfunktion von $P^{\Phi_n(x)|R_n=r}$, d.h.

$$F_n(t|r) = P(\Phi_n(x) \leq t | R_n = r),$$

und Φ die Verteilungsfunktion der Standardnormalverteilung, so folgt also (für P^{R_n} -fast alle $r > 0$)

$$\begin{aligned} |F_n(t|r) - \Phi(t)| &\leq \frac{6 \sum_{j=1}^{k_n-1} E|U_j|^3}{(\text{Var}(\sum_{j=1}^{k_n-1} U_j))^{3/2}} \\ &= \frac{6(k_n-1)E(|Z_1|^3|R_n=r)}{(k_n-1)^{3/2}(\text{Var}(Z_1|R_n=r))^{3/2}} = 6 \frac{E(|Z_1|^3|R_n=r)}{\sqrt{k_n-1}(\text{Var}(Z_1|R_n=r))^{3/2}}. \end{aligned} \quad (3.56)$$

Gemäß Lemma 3.1.22 gibt es ein $\Delta > 0$ mit

$$E(|Z_1|^3|R_n=r) \leq \Delta$$

für P^{R_n} -fast alle $r > 0$. Nach (3.56) erhält man also

$$|F_n(t|r) - \Phi(t)| \leq 6 \frac{\Delta}{\sqrt{k_n-1}(\text{Var}(Z_1|R_n=r))^{3/2}} \quad P^{R_n}\text{-fast überall.}$$

Für $\varepsilon := \frac{c}{2}\sigma^2(x)\int K^2(v)dv$ gilt dann

$$\begin{aligned}
& |P(\Phi_n(x) \leq t) - \Phi(t)| \\
&= \left| \int P(\Phi_n(x) \leq t | R_n = r) h(r) dr - \Phi(t) \right| \\
&= \left| \int (P(\Phi_n(x) \leq t | R_n = r) - \Phi(t)) h(r) dr \right| \\
&\leq \int |F_n(t|r) - \Phi(t)| h(r) dr \\
&= \int |F_n(t|r) - \Phi(t)| 1_{\{\text{Var}(Z_j | R_n=r) > \varepsilon\}} h(r) dr + \int |F_n(t|r) - \Phi(t)| 1_{\{\text{Var}(Z_j | R_n=r) \leq \varepsilon\}} h(r) dr \\
&\leq \frac{6\Delta}{\sqrt{k_n - 1}} \int \left| \frac{1}{(\text{Var}(Z_1 | R_n = r))^{3/2}} \right| 1_{\{\text{Var}(Z_j | R_n=r) > \varepsilon\}} h(r) dr + P\{\text{Var}(Z_j | R_n) \leq \varepsilon\} \\
&\leq \frac{6\Delta}{\sqrt{k_n - 1}\varepsilon^{3/2}} P\{\text{Var}(Z_j | R_n) > \varepsilon\} + P\{\text{Var}(Z_j | R_n) \leq \varepsilon\},
\end{aligned}$$

d.h.

$$\lim_{n \rightarrow \infty} |P(\Phi_n(x) \leq t) - \Phi(t)| = 0,$$

da

$$\text{Var}(Z_j | R_n) \xrightarrow{n \rightarrow \infty} c\sigma^2(x) \int K^2(v)dv > \varepsilon \quad \text{n.W.}$$

(nach Lemma 3.1.21) und $\sigma^2(x) > 0$ laut Voraussetzung. Es gilt also $\Phi_n(x) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$.

Wegen

$$A_n(x) = \sqrt{\text{Var}(Z_j | R_n)} \Phi_n(x) + o_p(1) \quad \text{für } n \rightarrow \infty$$

(nach Lemma 3.1.20(ii)) und wiederum nach Lemma 3.1.21 folgt dann mit dem Lemma von Slutsky

$$A_n(x) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, c\sigma^2(x) \int K^2(v)dv\right).$$

□

3.1.7 Stochastische Konvergenz von $B_n(x)$

Für den Beweis von Satz 3.1.8 wird neben Satz 3.1.23 auch noch die stochastische Konvergenz von

$$B_n(x) = \sqrt{k_n} \left(E(m_n(x)|R_n) - Em_n(x) \right)$$

benötigt.

Daher liegt es nahe, $E(m_n(x)|R_n)$ wiederum gemäß Satz 3.1.19(ii) näherungsweise als Quotient $\frac{E(h_n(x)|R_n)}{E(f_n(x)|R_n)}$ darzustellen und hier die Konvergenz getrennt für Zähler und Nenner zu untersuchen. Entscheidend hierfür ist folgendes Resultat:³⁰

3.1.26 Lemma. *Sei $\psi : \mathbb{R}^p \rightarrow \mathbb{R}$ beschränkt und in einer Umgebung von x zweimal stetig differenzierbar. Mit $Q(\psi)(x)$ aus (3.19) gilt dann für $n \rightarrow \infty$*

$$\int K(u) (\psi(x - uR_n) - \psi(x)) du = \frac{1}{2} Q(\psi)(x) \cdot R_n^2 + o_p(R_n^2).$$

Beweis. Sei $\{r_n\}_{n \in \mathbb{N}}$ eine Folge positiver reeller Zahlen mit $\lim_{n \rightarrow \infty} r_n = 0$. Mittels einer Taylor-Entwicklung um x erhält man für $n \rightarrow \infty$

$$\psi(x - u \cdot r_n) = \psi(x) - \sum_{\alpha=1}^p \frac{\partial \psi(x)}{\partial x_\alpha} u_\alpha r_n + \sum_{\alpha=1}^p \sum_{\beta=1}^p \frac{\partial^2 \psi(x)}{\partial x_\alpha \partial x_\beta} \frac{u_\alpha u_\beta}{2} r_n^2 + o(\|u\|^2 r_n^2)$$

und damit - aufgrund der Beschränktheit von ψ und K - nach dem Satz von der majorisierten Konvergenz

$$\begin{aligned} & \int K(u) (\psi(x - ur_n) - \psi(x)) du \\ &= -r_n \sum_{\alpha=1}^p \frac{\partial \psi(x)}{\partial x_\alpha} \underbrace{\int K(u) u_\alpha du}_{=0 \text{ nach (3.8)}} + \frac{1}{2} r_n^2 \sum_{\alpha=1}^p \sum_{\beta=1}^p \frac{\partial^2 \psi(x)}{\partial x_\alpha \partial x_\beta} \int K(u) u_\alpha u_\beta du + o(r_n^2) \\ &= \frac{1}{2} Q(\psi)(x) \cdot r_n^2 + o(r_n^2) \quad \text{für } n \rightarrow \infty. \end{aligned}$$

Da K und ψ beschränkt sind, folgt mit der Stetigkeit von ψ aus dem Satz von der majorisierten Konvergenz, dass

$$r \mapsto \int K(u) (\psi(x - ur) - \psi(x)) du$$

stetig ist. Dann ergibt sich also auch (für $n \rightarrow \infty$)

$$\int K(u) (\psi(x - uR_n) - \psi(x)) du = \frac{1}{2} Q(\psi)(x) R_n^2 + o_p(R_n^2).$$

□

³⁰Vgl. etwa [Mack81, S. 317 (30)].

3.1.27 Korollar. Für $n \rightarrow \infty$ gilt

(i)

$$E(f_n(x)|R_n) = \frac{k_n - 1}{nT_n} f(x) + \frac{Q(f)(x)}{2(cf(x))^{2/p}} \left(\frac{k_n}{n}\right)^{2/p} + o_p\left(\left(\frac{k_n}{n}\right)^{2/p}\right)$$

(ii)

$$E(h_n(x)|R_n) = \frac{k_n - 1}{nT_n} m(x)f(x) + \frac{Q(mf)(x)}{2(cf(x))^{2/p}} \left(\frac{k_n}{n}\right)^{2/p} + o_p\left(\left(\frac{k_n}{n}\right)^{2/p}\right)$$

Beweis. (i) Nach Lemma 3.1.26 (denn f ist in x zweimal stetig differenzierbar) hat man wie in Lemma 3.1.11(i) und Lemma 3.1.9 (für $n \rightarrow \infty$)

$$\begin{aligned} E(f_n(x)|R_n) &= \frac{k_n - 1}{nR_n^p} E\left(K\left(\frac{x - \tilde{X}_1}{R_n}\right) \mid R_n\right) \\ &= \frac{k_n - 1}{nT_n} \int K(u) f(x - uR_n) du \\ &= \frac{k_n - 1}{nT_n} f(x) + \underbrace{\frac{k_n - 1}{nT_n} \int K(u) (f(x - uR_n) - f(x)) du}_{\xrightarrow{n \rightarrow \infty} 1 \text{ n.W.}} \\ &= \frac{k_n - 1}{nT_n} f(x) + (1 + o_p(1)) \left(\frac{1}{2} Q(f)(x) R_n^2 + o_p(R_n^2)\right) \\ &= \frac{k_n - 1}{nT_n} f(x) + \frac{1}{2} Q(f)(x) R_n^2 + o_p(R_n^2) \\ &= \frac{k_n - 1}{nT_n} f(x) + \frac{1}{2} Q(f)(x) \cdot (cf(x))^{-2/p} \left(\frac{k_n}{n}\right)^{2/p} + o_p\left(\left(\frac{k_n}{n}\right)^{2/p}\right), \end{aligned}$$

denn nach den Lemmas 3.1.5 und 3.1.6 gilt

$$cf(x) R_n^p \frac{n}{k_n - 1} = \underbrace{\frac{R_n^p cf(x)}{T_n}}_{\xrightarrow{n \rightarrow \infty} 1} \underbrace{\frac{nT_n}{k_n - 1}}_{\xrightarrow{n \rightarrow \infty} 1} \xrightarrow{n \rightarrow \infty} 1 \quad \text{n.W.}$$

und damit (für $n \rightarrow \infty$)

$$R_n^2 = \underbrace{\left(cf(x) R_n^p \frac{n}{k_n - 1}\right)^{2/p}}_{\xrightarrow{n \rightarrow \infty} 1 \text{ n.W.}} \left(cf(x) \frac{n}{k_n - 1}\right)^{-2/p} = (cf(x))^{-2/p} \left(\frac{k_n}{n}\right)^{2/p} (1 + o_p(1)).$$

(ii) Genauso ergibt sich (mit der zweifachen stetigen Differenzierbarkeit von $m \cdot f$) aus Lemma

3.1.26 (wie in Lemma 3.1.11(iii) und Lemma 3.1.9) für $n \rightarrow \infty$

$$\begin{aligned}
E(h_n(x)|R_n) &= \frac{k_n - 1}{nR_n^p} E\left(K\left(\frac{x - \tilde{X}_1}{R_n}\right) \tilde{Y}_1 \mid R_n\right) \\
&= \frac{k_n - 1}{nT_n} \int K(u)(mf)(x - uR_n) du \\
&= \frac{k_n - 1}{nT_n} m(x)f(x) + \underbrace{\frac{k_n - 1}{nT_n} \int K(u)((mf)(x - uR_n) - (mf)(x)) du}_{\xrightarrow{n \rightarrow \infty} 1 \text{ n.W.}} \\
&= \frac{k_n - 1}{nT_n} m(x)f(x) + \frac{1}{2} Q(mf)(x) R_n^2 + o_p(R_n^2) \\
&= \frac{k_n - 1}{nT_n} m(x)f(x) + \frac{1}{2} Q(mf)(x) \cdot (cf(x))^{-2/p} \left(\frac{k_n}{n}\right)^{2/p} + o_p\left(\left(\frac{k_n}{n}\right)^{2/p}\right).
\end{aligned}$$

□

3.1.28 Satz. *Unter den Annahmen aus 3.1.2 gilt (für $n \rightarrow \infty$)*

$$E(m_n(x)|R_n) = m(x) + \frac{Q(mf)(x) - m(x)Q(f)(x)}{2f(x)(cf(x))^{2/p}} \left(\frac{k_n}{n}\right)^{2/p} + o_p\left(\left(\frac{k_n}{n}\right)^{2/p}\right) + o_p\left(\frac{1}{\sqrt{k_n}}\right).$$

Beweis. Nach Satz 3.1.19(ii) gilt zunächst

$$E(m_n(x)|R_n) = \frac{E(h_n(x)|R_n)}{E(f_n(x)|R_n)} + o_p\left(\frac{1}{\sqrt{k_n}}\right) \quad \text{für } n \rightarrow \infty.$$

Aus Korollar 3.1.27 ergibt sich außerdem (für $n \rightarrow \infty$)

$$\begin{aligned}
&\frac{E(h_n(x)|R_n)}{E(f_n(x)|R_n)} - m(x) \\
&= \frac{E(h_n(x)|R_n) - m(x)f(x)\frac{k_n-1}{nT_n}}{E(f_n(x)|R_n)} - \frac{m(x)E(f_n(x)|R_n) - m(x)f(x)\frac{k_n-1}{nT_n}}{E(f_n(x)|R_n)} \\
&= \frac{\frac{Q(mf)(x)}{2(cf(x))^{2/p}} \left(\frac{k_n}{n}\right)^{2/p} + o_p\left(\left(\frac{k_n}{n}\right)^{2/p}\right)}{f(x)(1 + o_p(1))} - \frac{\frac{m(x)Q(f)(x)}{2(cf(x))^{2/p}} \left(\frac{k_n}{n}\right)^{2/p} + o_p\left(\left(\frac{k_n}{n}\right)^{2/p}\right)}{f(x)(1 + o_p(1))} \\
&= \frac{Q(mf)(x) - m(x)Q(f)(x)}{2f(x)(cf(x))^{2/p}} \left(\frac{k_n}{n}\right)^{2/p} + o_p\left(\left(\frac{k_n}{n}\right)^{2/p}\right).
\end{aligned}$$

Aus

$$E(m_n(x)|R_n) = m(x) + \left(\frac{E(h_n(x)|R_n)}{E(f_n(x)|R_n)} - m(x)\right) + \left(E(m_n(x)|R_n) - \frac{E(h_n(x)|R_n)}{E(f_n(x)|R_n)}\right)$$

folgt daher die Behauptung. □

Beweis zu Satz 3.1.8. Die Behauptung folgt nun wegen

$$\sqrt{k_n}(m_n(x) - m(x)) = A_n(x) + B_n(x)$$

mit dem Lemma von Slutsky unmittelbar aus den Sätzen 3.1.23 und 3.1.28, denn es gilt

$$\begin{aligned} \lim_{n \rightarrow \infty} \left(\frac{k_n}{n}\right)^{2/p} \sqrt{k_n} &= \lim_{n \rightarrow \infty} \left(k_n \cdot n^{-\frac{2/p}{2/p+1/2}}\right)^{2/p+1/2} \\ &= \lim_{n \rightarrow \infty} \left(k_n \cdot n^{-4/(4+p)}\right)^{2/p+1/2} \\ &= \eta^{2/p+1/2}. \end{aligned} \tag{3.57}$$

□

3.2 Optimale Anzahl der Nachbarn und Konfidenzintervalle

3.2.1 Optimale Wahl von k_n

Die Folge $\{k_n\}_{n \in \mathbb{N}}$ ist durch die Anforderungen aus 3.1.2 natürlich noch nicht eindeutig festgelegt. In diesem Abschnitt soll ermittelt werden, wie $\{k_n\}_{n \in \mathbb{N}}$ asymptotisch optimal gewählt werden kann, wenn die Zielsetzung darin besteht, dass (für $n \rightarrow \infty$) der Term $m_n(x) - m(x)$ schnellstmöglich gegen Null konvergiert.

Aus den Sätzen 3.1.23 und 3.1.28 erkennt man, dass

$$m_n(x) - E(m_n(x)|R_n) = \frac{A_n}{\sqrt{k_n}}$$

umso langsamer konvergiert, je kleiner k_n (relativ zu n) gewählt wird, während sich

$$E(m_n(x)|R_n) - m(x) = O\left(\frac{k_n}{n}\right) \quad \text{für } n \rightarrow \infty$$

genau umgekehrt verhält.

Die naheliegende Vermutung, dass $|m_n(x) - m(x)|$ asymptotisch minimal wird, wenn $\sqrt{k_n} \frac{k_n}{n}$ (für $n \rightarrow \infty$) gegen eine echt positive reelle Zahl konvergiert,³¹ bestätigt sich mit dem folgenden Satz.

Dabei sei (für eine Folge reellwertiger Zufallsvariablen $\{V_n\}_{n \in \mathbb{N}}$)

$$V_n \xrightarrow{n \rightarrow \infty} \infty \quad \text{n.W.}$$

definiert durch

$$\forall N > 0 : \quad \lim_{n \rightarrow \infty} P\{V_n > N\} = 1.$$

³¹Vgl. [Mack81, S. 320].

3.2.1 Satz. Seien die Annahmen aus Abschnitt 3.1.2 erfüllt.

(i) Im Fall $\lim_{n \rightarrow \infty} k_n n^{\frac{-4}{4+p}} > 0$ gilt

$$n^{\frac{2}{4+p}} (m_n(x) - m(x)) = O_p(1) \quad \text{für } n \rightarrow \infty.$$

(ii) Im Fall $k_n = o(n^{\frac{4}{4+p}})$ gilt

$$n^{\frac{2}{4+p}} |m_n(x) - m(x)| \xrightarrow{n \rightarrow \infty} \infty \quad n. W.$$

(iii) Im Fall $n^{\frac{4}{4+p}} = o(k_n)$ gilt

$$n^{\frac{2}{4+p}} |m_n(x) - m(x)| \xrightarrow{n \rightarrow \infty} \infty \quad n. W.$$

Beweis. (i) Aus $\eta := \lim_{n \rightarrow \infty} k_n \cdot n^{\frac{-4}{4+p}} > 0$ folgt (für $n \rightarrow \infty$)

$$\sqrt{\eta} \cdot n^{\frac{2}{4+p}} = \sqrt{k_n} (1 + o(1)) \quad \text{sowie} \quad \sqrt{k_n} = O\left(\left(\frac{k_n}{n}\right)^{-2/p}\right),$$

denn wie in (3.57) hat man

$$\lim_{n \rightarrow \infty} \sqrt{k_n} \left(\frac{k_n}{n}\right)^{2/p} \eta^{2/p+1/2}.$$

Damit gilt (nach den Sätzen 3.1.23 und 3.1.28)

$$\begin{aligned} & n^{\frac{2}{4+p}} (m_n(x) - m(x)) \\ &= \underbrace{\sqrt{k_n} (m_n(x) - E(m_n(x)|R_n)) (1 + o(1)) \frac{1}{\sqrt{\eta}}}_{\xrightarrow{\mathcal{L}} \mathcal{N}(0, c \cdot \sigma^2(x) \int K^2(v) dv \frac{1}{\eta})} \\ & \quad + \sqrt{k_n} (E(m_n(x)|R_n) - m(x)) (1 + o(1)) \frac{1}{\sqrt{\eta}} \quad P\text{-f.s.} \\ &= O_p(1) + O\left(\left(\frac{k_n}{n}\right)^{-2/p}\right) O_p\left(\left(\frac{k_n}{n}\right)^{2/p}\right) = O_p(1) \end{aligned}$$

für $n \rightarrow \infty$.

(ii) Aus $k_n = o(n^{\frac{4}{4+p}})$ folgt (für $n \rightarrow \infty$)

$$n^{\frac{2}{4+p}} \left(\frac{k_n}{n}\right)^{2/p} = n^{\frac{2}{4+p}} \cdot o\left(\left(n^{\frac{4}{4+p}-1}\right)^{2/p}\right) = n^{\frac{2}{4+p}} \cdot o(n^{\frac{-2}{4+p}}) = o(1)$$

und damit (P -f.s.)

$$\begin{aligned}
 n^{\frac{2}{4+p}} |m_n(x) - m(x)| &= \left| n^{\frac{2}{4+p}} \left(m_n(x) - E(m_n(x)|R_n) \right) + n^{\frac{2}{4+p}} \left(E(m_n(x)|R_n) - m(x) \right) \right| \\
 &= \left| \underbrace{\frac{n^{\frac{2}{4+p}}}{\sqrt{k_n}}}_{\xrightarrow{n \rightarrow \infty} \infty} \underbrace{\sqrt{k_n} \left(m_n(x) - E(m_n(x)|R_n) \right)}_{\xrightarrow{\mathcal{L}} \mathcal{N}(0, c \cdot \sigma^2(x) \int K^2(v) dv)} + \underbrace{n^{\frac{2}{4+p}} O_p \left(\frac{k_n}{n} \right)^{2/p}}_{=o_p(1)} \right| \\
 &\xrightarrow{n \rightarrow \infty} \infty \quad \text{n.W.},
 \end{aligned}$$

denn $\sigma^2(x) > 0$.

(iii) Aus $n^{\frac{4}{4+p}} = o(k_n)$ folgt $n^{\frac{2}{4+p}} = o(\sqrt{k_n})$ und damit (P -f.s.)

$$\begin{aligned}
 n^{\frac{2}{4+p}} |m_n(x) - m(x)| &= \left| \underbrace{n^{\frac{2}{4+p}} \left(m_n(x) - E(m_n(x)|R_n) \right)}_{=o_p(1)} + \underbrace{n^{\frac{2}{4+p}} \left(E(m_n(x)|R_n) - m(x) \right)}_{|\dots| \xrightarrow{n \rightarrow \infty} \infty \text{ n.W.}} \right| \\
 &\xrightarrow{n \rightarrow \infty} \infty \quad \text{n.W.},
 \end{aligned}$$

denn

$$n^{\frac{2}{4+p}} \left(\frac{k_n}{n} \right)^{2/p} = n^{\frac{2}{4+p} - \frac{2}{p}} \cdot k_n^{2/p} = \left(n^{\frac{p}{4+p} - 1} \cdot k_n \right)^{2/p} = \left(\frac{k_n}{n^{\frac{4}{4+p}}} \right)^{2/p} \xrightarrow{n \rightarrow \infty} \infty.$$

□

3.2.2 Konfidenzintervalle für $m(x)$

Aus Satz 3.1.8 lassen sich nun approximative Konfidenzintervalle für $m(x)$ einfach berechnen. Man hat

$$\sqrt{k_n} (m_n(x) - m(x)) \xrightarrow{\mathcal{L}} \mathcal{N} \left(B, c \cdot \sigma^2(x) \int K^2(v) dv \right), \quad n \rightarrow \infty, \quad (3.58)$$

für

$$B = \eta^{1/2+2/p} \frac{Q(mf)(x) - m(x)Q(f)(x)}{2f(x)(cf(x))^{2/p}},$$

mit

$$\eta = \lim_{n \rightarrow \infty} k_n \cdot n^{\frac{-4}{4+p}} < \infty.$$

Es gilt also (für alle $t \in \mathbb{R}$)

$$P \left\{ \frac{\sqrt{k_n} (m_n(x) - m(x)) - B}{\sqrt{c \sigma^2(x) \int K^2(v) dv}} \leq t \right\} \xrightarrow{n \rightarrow \infty} \Phi(t). \quad (3.59)$$

Bezeichnet $z_{1-\alpha/2}$ das $(1 - \alpha/2)$ -Quantil der Standardnormalverteilung, so ergibt sich wegen der Äquivalenz

$$\begin{aligned} & \left| \frac{\sqrt{k_n}m_n(x) - \sqrt{k_n}m(x) - B}{\sqrt{c \cdot \sigma^2(x) \int K^2(v) dv}} \right| \leq z_{1-\alpha/2} \\ \Leftrightarrow & \sqrt{k_n}m(x) \in \left[\sqrt{k_n}m_n(x) - B - z_{1-\alpha/2} \sqrt{c \cdot \sigma^2(x) \int K^2(v) dv}, \right. \\ & \left. \sqrt{k_n}m_n(x) - B + z_{1-\alpha/2} \sqrt{c \cdot \sigma^2(x) \int K^2(v) dv} \right] \\ \Leftrightarrow & m(x) \in \left[m_n(x) - \frac{B}{\sqrt{k_n}} - \tau_n(x), m_n(x) - \frac{B}{\sqrt{k_n}} + \tau_n(x) \right] \end{aligned}$$

für

$$\tau_n(x) := z_{1-\alpha/2} \sqrt{\frac{c}{k_n} \sigma^2(x) \int K^2(v) dv}$$

also

$$P \left\{ m(x) \in \left[m_n(x) - \frac{B}{\sqrt{k_n}} - \tau_n(x), m_n(x) - \frac{B}{\sqrt{k_n}} + \tau_n(x) \right] \right\} \xrightarrow{n \rightarrow \infty} 1 - \alpha. \quad (3.60)$$

3.2.2 Bemerkung. Die Werte $\int K^2(u) du$ und c in (3.58) werden durch die Wahl des Kernes K und der Norm $\|\cdot\|$ festgelegt. $\sigma^2(x)$ und gegebenenfalls B (sofern $\lim_{n \rightarrow \infty} k_n n^{\frac{-4}{4+p}} > 0$) müssen hingegen ihrerseits aus den Daten geschätzt werden.

Sind allerdings \hat{B} und $\hat{\sigma}^2(x)$ konsistente Schätzer für B und $\sigma^2(x)$, so ist für

$$\hat{\tau}_n(x) := z_{1-\alpha/2} \sqrt{\frac{c}{k_n} \hat{\sigma}^2(x) \int K^2(v) dv}$$

auch

$$\left[m_n(x) - \frac{\hat{B}}{\sqrt{k_n}} - \hat{\tau}_n(x), m_n(x) - \frac{\hat{B}}{\sqrt{k_n}} + \hat{\tau}_n(x) \right]$$

(approximatives) Konfidenzintervall zum Niveau $1 - \alpha$. Denn dann gilt (für $n \rightarrow \infty$)

$$\frac{\sqrt{k_n}(m_n(x) - m(x)) - \hat{B}}{\sqrt{c \hat{\sigma}^2(x) \int K^2(v) dv}} \xrightarrow{\text{n.W.}} \frac{\sqrt{k_n}(m_n(x) - m(x)) - B}{\sqrt{c \sigma^2(x) \int K^2(v) dv}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1),$$

d.h. für alle $t \in \mathbb{R}$ erhält man (vgl. (3.59))

$$P \left\{ \frac{\sqrt{k_n}(m_n(x) - m(x)) - \hat{B}}{\sqrt{c \hat{\sigma}^2(x) \int K^2(v) dv}} \leq t \right\} \xrightarrow{n \rightarrow \infty} \Phi(t).$$

3.3 Berechnung von Konfidenzintervallen für die Beispieldaten

Um einen Eindruck über die Schätzgenauigkeit des betrachteten k_n -nächste-Nachbarn-Verfahrens zur Prognose von Spätschäden zu erhalten, sollen konkret - basierend wiederum auf dem Datenbestand aus Kapitel 2 - approximative Konfidenzintervalle berechnet werden.

Beispielsweise interessiert man sich für die Vorhersage der Entschädigungszahlungen Y des *fünften* Abwicklungsjahres, wobei die ersten vier Beträge bekannt seien.³² Um in Abschnitt 3.4 die Prognosen mit den tatsächlich eingetretenen Schäden vergleichen zu können, wird die Schätzung für die 1998 eingetretenen Schäden auf Grundlage des Informationsstandes nach Ablauf des Jahres 2001 durchgeführt.

In Abhängigkeit von $x \in \mathbb{R}^4$ (einem festgehaltenen vierjährigen Abwicklungsverlauf) wird dann die Regressionsfunktion $m(x) = E(Y|X = x)$ geschätzt mittels

$$m_n(x) = \frac{\sum_{i=1}^n K\left(\frac{X_i - x}{R_n}\right) Y_i}{\sum_{i=1}^n K\left(\frac{X_i - x}{R_n}\right)}.$$

Entsprechend der Erläuterung in der Einleitung zu Kapitel 3 werden zur konkreten Berechnung nur solche Schäden betrachtet, bei denen in allen fünf bisherigen Abwicklungsjahren von Null verschiedene Zahlungen angefallen sind (hier dann $n = 536$).

Außerdem wird (um die Normalapproximation aus Abschnitt 3.1 und damit das Konfidenzintervall aus 3.60 bzw. Bemerkung 3.2.2 nutzen zu können), der Kern K mit

$$K(u) := \kappa((1 - \|u\|^2) + \delta) 1_{B_1(0)}(u), \quad u \in \mathbb{R}^p, \quad (3.61)$$

(für ein $\delta > 0$) angewendet (vgl. (3.3)), wobei $\kappa > 0$ wegen (3.7) so festgelegt wird, dass $\int K(u) du = 1$ erreicht wird.

Es wird davon ausgegangen, dass die übrigen Voraussetzungen aus Abschnitt 3.1.2 erfüllt sind. Damit ist Satz 3.1.8 anwendbar; zur Vereinfachung werde angenommen, dass dort ein etwaiges $B \neq 0$ vernachlässigbar klein ist. Hierfür ist k_n kleiner als gemäß Satz 3.2.1(i) zu wählen (vgl. Unterabschnitt 3.2.1).

Beim Anblick der Ergebnisse aus Abschnitt 2.4 erscheint es angebracht, den Fall $k_n = 30$ zu betrachten.³³

Gemäß Bemerkung 3.2.2 muss dann zur Berechnung von Konfidenzintervallen für $m(x)$ lediglich noch $\sigma^2(x)$ konsistent geschätzt werden.

³²Die Wahl gerade des fünften Abwicklungsjahres begründet sich lediglich darin, dass sichergestellt werden soll, dass eine ausreichend hohe Anzahl an Datensätzen zugrundegelegt werden kann und der bisherige Schadenverlauf dennoch mehrere Jahre umfasst.

³³Eine solche globale Wahl der Anzahl an Nachbarn ist vertretbar, da in Abschnitt 2.4 erkennbar geworden ist, dass das k_n -nächste-Nachbarn-Verfahren für eine breite Spanne an Werten für k_n gute Prognosen liefert. Das eigentlich optimale k_n hängt aber natürlich auch vom konkreten $x \in \mathbb{R}^4$ ab.

Die Theorie aus Abschnitt 3.1 lässt sich natürlich (mittels Ersetzung von Y durch Y^2) unmittelbar auf die Betrachtung von

$$l(x) = E(Y^2|X = x)$$

übertragen, so dass

$$l_n(x) := \frac{\sum_{i=1}^n K\left(\frac{X_i - x}{R_n}\right) Y_i^2}{\sum_{i=1}^n K\left(\frac{X_i - x}{R_n}\right)}$$

ein konsistenter Schätzer für $l(x)$ ist. Wie aus den Beweisen zu Lemma 3.1.11 und Satz 3.1.18 deutlich wird, genügt hierfür die zusätzliche Voraussetzung, dass die Funktion

$$u \mapsto E(Y^4|X = u)$$

beschränkt und in x stetig ist.

Wegen

$$\sigma^2(x) = l(x) - m^2(x) = |l(x) - m^2(x)|$$

gilt dann auch

$$\sigma_n^2(x) := |l_n(x) - m_n^2(x)| \xrightarrow{n \rightarrow \infty} \sigma^2(x) \quad \text{n.W.} \quad (3.62)$$

für alle $x \in \mathbb{R}^p$ mit $\sigma^2(x) > 0$. Letzteres soll im Folgenden stets vorausgesetzt werden.

Schließlich wird noch der Wert $\int K^2(u) du$ benötigt.

3.3.1 Lemma. Für eine Funktion $h : \mathbb{R} \rightarrow \mathbb{R}$ gilt (für alle $p \in \mathbb{N}$)

$$\int_{B_1(0)} h(\|u\|) du = \int_0^1 h(t) c p t^{p-1} dt,$$

sofern das rechte Integral definiert ist, wobei $B_1(0)$ die Einheitskugel um 0 im \mathbb{R}^p bezüglich einer beliebigen Norm $\|\cdot\|$ bezeichnet, sowie $c = \lambda_p(B_1(0))$.

Beweis. Sei $p \in \mathbb{N}$ fest gewählt. Für beliebiges $t > 0$ gilt (wie in (3.12))

$$\lambda_p^{\|\cdot\|}(0, t] = \lambda_p\{u \in \mathbb{R}^p \mid \|u\| \in (0, t]\} = \lambda_p(B_t(0)) = t^p \cdot c,$$

d.h. $\lambda_p^{\|\cdot\|}$ besitzt die Lebesgue-Dichte $t \mapsto c p t^{p-1}$. Damit ergibt sich nach dem Transformationsatz

$$\begin{aligned} \int_{B_1(0)} h(\|u\|) du &= \int h(\|u\|) 1_{[0,1]}(\|u\|) \lambda_p(du) \\ &= \int h(t) 1_{[0,1]}(t) \lambda_p^{\|\cdot\|}(dt) \\ &= \int_0^1 h(t) c p t^{p-1} dt. \end{aligned}$$

□

Mit den Bezeichnungen aus (3.3) muss κ für festes $\delta > 0$ (entsprechend (3.7)) so festgelegt werden, dass $\int K(u)du = 1$ erfüllt ist. Wegen (für alle $p \in \mathbb{N}$)

$$\begin{aligned}\int K(u)du &= \kappa \int_{B_1(0)} ((1 - \|u\|^2) + \delta) du \\ &= \kappa \int_0^1 (1 + \delta - t^2) c p t^{p-1} dt \quad \text{nach Lemma 3.3.1} \\ &= \kappa c p \left(\frac{1 + \delta}{p} t^p - \frac{1}{p+2} t^{p+2} \right) \Big|_0^1 \\ &= \kappa c \left(\delta + \frac{2}{p+2} \right)\end{aligned}$$

hat man also

$$\kappa = \frac{1}{c} \left(\delta + \frac{2}{p+2} \right)^{-1}. \quad (3.63)$$

Entsprechend ergibt sich (ebenfalls nach Lemma 3.3.1)

$$\begin{aligned}\int K^2(u)du &= \kappa^2 \int_{B_1(0)} (1 + \delta - \|u\|^2)^2 du \\ &= \kappa^2 \int_0^1 (1 + \delta - t^2)^2 c p t^{p-1} dt \\ &= \kappa^2 c p \left(\frac{(1 + \delta)^2}{p} t^p - \frac{2(1 + \delta)}{p+2} t^{p+2} + \frac{1}{p+4} t^{p+4} \right) \Big|_0^1 \\ &= \kappa^2 c \left(\delta^2 + \frac{4\delta}{p+2} + \frac{8}{(p+2)(p+4)} \right).\end{aligned}$$

Sei nun beispielsweise $\delta = 0,05$. Verwendet man zur Abstandsmessung die euklidische Norm auf dem \mathbb{R}^4 , so gilt nach Beispiel 3.1.1(i)

$$c = \frac{\pi^2}{\Gamma(3)} \approx 4,935$$

und damit (gemäß (3.63))

$$\kappa \approx 0,5286$$

sowie

$$c \int K^2(u)du \approx 1,3781.$$

In Abbildung 3.1 sind die Breiten

$$2 \cdot z_{0,975} \cdot \sqrt{\frac{c}{k_n} \int K^2(u)du \cdot \sigma_n^2(x)}$$

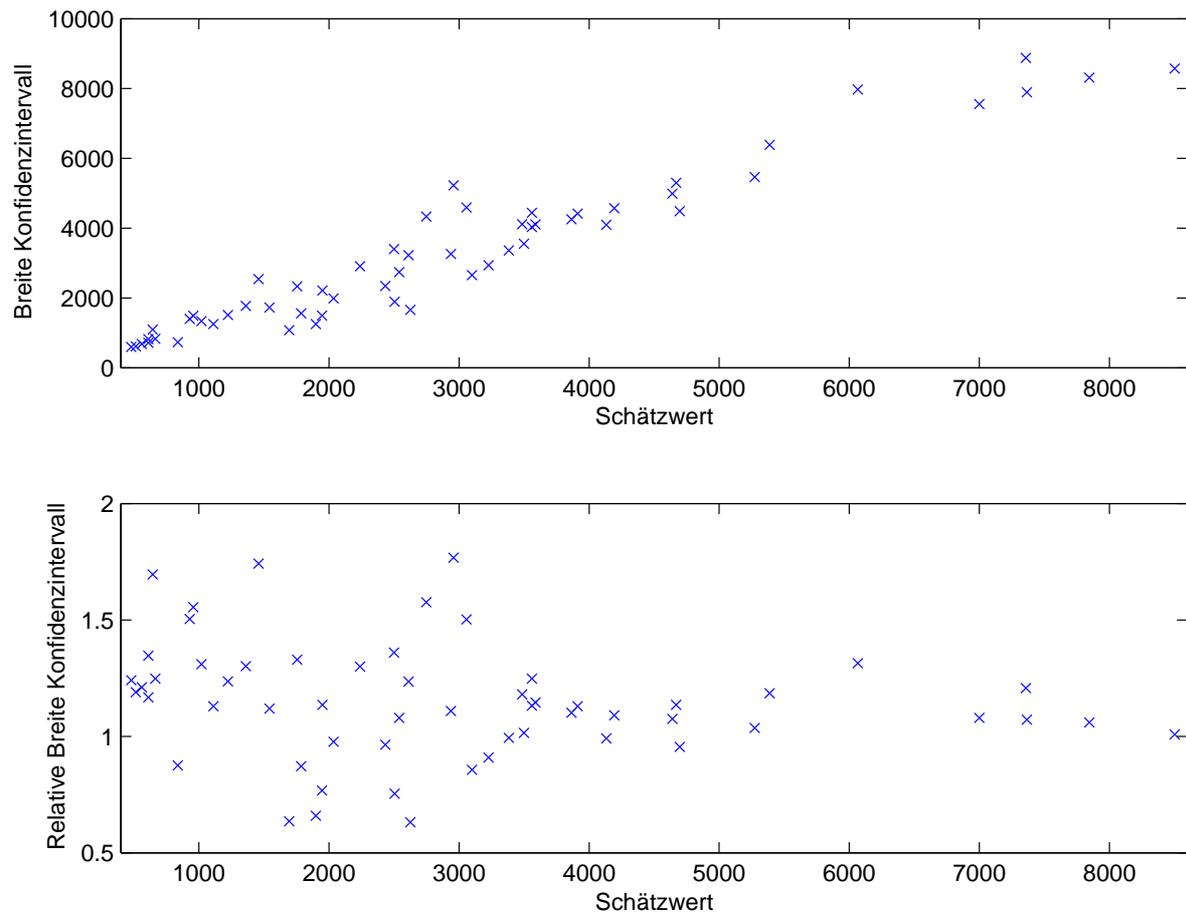


Abbildung 3.1: Absolute und relative Breiten der approximativen Konfidenzintervalle für $m(x)$ in Abhängigkeit von $m_n(x)$

der sich ergebenden (um $m_n(x)$ zentrierten) asymptotischen Konfidenzintervalle (zum Niveau 0,95) für $m(x)$ gegen die zugehörigen Werte des Schätzers $m_n(x)$ abgetragen - absolut und relativ zu $m_n(x)$.³⁴

Die relative Breite der Konfidenzintervalle scheint sich also bei wachsendem Schätzwert $m_n(x)$ nicht systematisch zu verändern, durchschnittlich beträgt sie dabei näherungsweise 1,148.

3.4 Prognoseintervalle für Y

Im Abschnitt 3.1 ist - ohne genaue Festlegung der Verteilung $P^{(X,Y)}$ - auf Grundlage des Modells aus (3.1) für festes $x \in \mathbb{R}^p$ die asymptotische Normalität von $m_n(x) - m(x)$ gezeigt worden. Darauf basierend sind im Abschnitt 3.2 Konfidenzintervalle für $m(x)$ hergeleitet worden.

Noch interessanter ist oftmals die Fragestellung, wie genau Y durch $m_n(x)$ vorhergesagt wird. Aus (3.1) erhält man als Darstellung für den *Prognosefehler*

$$m_n(x) - Y = (m_n(x) - m(x)) - \sigma(x) \cdot \varepsilon, \tag{3.64}$$

mit $E(\varepsilon) = 0$ und $\text{Var}(\varepsilon) = 1$. Zur Berechnung von Prognoseintervallen für Y muss also die Verteilung von ε bestimmt werden.

Geht man davon aus, dass Y und $(X_i, Y_i)_{i \in \underline{n}}$ stochastisch unabhängig sind, so sind auch $m_n(x)$ und ε stochastisch unabhängig. Dann kann $m_n(x) - Y$ also als Faltung der Zufallsvariablen $m_n(x) - m(x)$ und $-\sigma(x) \cdot \varepsilon$ angesehen werden.

Im einfachsten Fall ist ε selbst normalverteilt, also $\varepsilon \sim \mathcal{N}(0, 1)$. Sei $\{k_n\}_{n \in \mathbb{N}}$ so gewählt, dass Satz 3.1.8 anwendbar ist, mit

$$\sqrt{k_n}(m_n(x) - m(x)) \xrightarrow{\mathcal{L}} \mathcal{N}\left(B, \sigma^2(x) \int K^2(u) du\right)$$

(für $B \geq 0$ entsprechend (3.20)). Anhand der Zerlegung

$$\begin{aligned} \sqrt{k_n}(m_n(x) - Y) &= \underbrace{\sqrt{k_n}(m_n(x) - m(x))}_{\xrightarrow{\mathcal{L}} \mathcal{N}\left(B, \sigma^2(x) \int K^2(u) du\right)} - \underbrace{\sqrt{k_n}\sigma(x)\varepsilon}_{\sim \mathcal{N}\left(0, k_n\sigma^2(x)\right)} \end{aligned}$$

erkennt man dann, dass sich die Verteilungsfunktion von $\sqrt{k_n}(m_n(x) - Y)$ für wachsendes n wie diejenige von

$$\mathcal{N}\left(B, \sigma^2(x) \left(k_n + \int K^2(u) du\right)\right) \tag{3.65}$$

verhält. Genauer gilt

³⁴Die entsprechende Berechnung ist auf der beigelegten CD im Ordner *Beispiel* dokumentiert. Dieser umfasst einerseits eine *Access*-Datenbank zur Durchführung der k_n -nächste-Nachbarn-Verfahren (für $m_n(x)$ und $l_n(x)$) sowie andererseits ein *Microsoft Excel 2000*-Arbeitsblatt (jeweils gleichnamig) zur zusammenfassenden Darstellung der Konfidenzintervalle - und auch der standardisierten Residuen im Kontext von Abbildung 3.2 im folgenden Abschnitt.

3.4.1 Satz. Sei $(C_n)_{n \in \mathbb{N}}$ eine Folge reeller Zahlen mit $\lim_{n \rightarrow \infty} C_n = 1$. Außerdem sei

$$\rho_n(x) := z_{1-\alpha/2} \sigma(x) C_n, \quad n \in \mathbb{N},$$

wobei $z_{1-\alpha/2}$ wiederum das $(1 - \alpha/2)$ -Quantil der Standardnormalverteilung bezeichnet, sowie

$$I_n(x) := \left[m_n(x) - \frac{B}{\sqrt{k_n}} - \rho_n(x), m_n(x) - \frac{B}{\sqrt{k_n}} + \rho_n(x) \right], \quad n \in \mathbb{N}. \quad (3.66)$$

(i) Es gilt

$$\frac{m_n(x) - Y - \frac{B}{\sqrt{k_n}}}{\sigma(x) C_n} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1), \quad n \rightarrow \infty, \quad (3.67)$$

und

$$\lim_{n \rightarrow \infty} P\{Y \in I_n(x)\} = 1 - \alpha. \quad (3.68)$$

(ii) Für $\gamma := \sqrt{c \int K^2(v) dv}$ und

$$W_n := \frac{\sqrt{k_n}(m_n(x) - m(x)) - B}{\sigma(x) \cdot \gamma}$$

gelte

$$\lim_{n \rightarrow \infty} E W_n^2 = 1 \quad \text{und} \quad \lim_{n \rightarrow \infty} \frac{1}{\sqrt{k_n}} E |W_n|^3 = 0. \quad (3.69)$$

Außerdem existiere

$$\mu := \lim_{n \rightarrow \infty} 2k_n(C_n - 1) \quad (3.70)$$

in \mathbb{R} .³⁶ Dann gilt

$$\lim_{n \rightarrow \infty} k_n (P\{Y \in I_n(x)\} - (1 - \alpha)) = \phi(z_{1-\alpha/2}) z_{1-\alpha/2} (\mu - \gamma^2),$$

wobei ϕ die Dichte der Standardnormalverteilung bezeichnet.

3.4.2 Bemerkung. (i) Aus Satz 3.4.1(i) folgt genau wie in Bemerkung 3.2.2, dass auch

$$\left[m_n(x) - \frac{\hat{B}}{\sqrt{k_n}} - \hat{\rho}_n(x), m_n(x) - \frac{\hat{B}}{\sqrt{k_n}} + \hat{\rho}_n(x) \right]$$

für

$$\hat{\rho}_n(x) := z_{1-\alpha/2} \hat{\sigma}(x) C_n$$

(approximatives) Prognoseintervall für Y zum Niveau $1 - \alpha$ ist, sofern \hat{B} und $\hat{\sigma}(x)$ konsistente Schätzer für B und $\sigma(x)$ sind.

³⁵Vgl. Bemerkung 3.4.2(ii).

³⁶Im Fall $k_n |C_n - 1| \rightarrow \infty$ konvergiert $k_n (P\{Y \in I_n(x)\} - (1 - \alpha))$ nicht, vgl. (3.72).

- (ii) Aufgrund von Satz 3.1.8 gilt $W_n \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$ für $n \rightarrow \infty$. Für die Gültigkeit der beiden Konvergenzaussagen in (3.69) sind - im Vergleich zu Abschnitt 3.1 - zusätzliche Annahmen insbesondere an die Verteilung von X nötig. Man beachte etwa das Vorgehen in [Mack81, S. 313f.], der sich zum Ziel setzt, das asymptotische Verhalten von $Em_n(x)$ und $\text{Var } m_n(x)$ zu untersuchen.

Beweis von Satz 3.4.1. (i) Mit Satz 3.1.8 folgt (für $n \rightarrow \infty$)

$$\frac{m_n(x) - Y - \frac{B}{\sqrt{k_n}}}{\sigma(x)C_n} = \underbrace{\frac{m_n(x) - m(x) - \frac{B}{\sqrt{k_n}}}{\sigma(x)C_n}}_{\substack{n \rightarrow \infty \\ \rightarrow 0 \quad \text{n.W.}}} - \underbrace{\frac{\varepsilon}{C_n}}_{\rightarrow \mathcal{N}(0,1)} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

unmittelbar aus dem Lemma von Slutsky. Wegen der Äquivalenz

$$\left| \frac{m_n(x) - Y - \frac{B}{\sqrt{k_n}}}{\sigma(x)C_n} \right| \leq z_{1-\alpha/2}$$

$$\iff m_n(x) - \frac{B}{\sqrt{k_n}} - \rho_n(x) \leq Y \leq m_n(x) - \frac{B}{\sqrt{k_n}} + \rho_n(x)$$

ergibt sich hieraus direkt (3.68).

- (ii) Aufgrund der Unabhängigkeit von W_n und ε hat man wegen (3.1) und $\varepsilon \sim \mathcal{N}(0,1)$ zunächst

$$\begin{aligned} & P\{Y \in I_n(x)\} - (1 - \alpha) \\ &= P\left\{ \frac{m_n(x) - m(x) - \frac{B}{\sqrt{k_n}} - \rho_n(x)}{\sigma(x)} \leq \varepsilon \leq \frac{m_n(x) - m(x) - \frac{B}{\sqrt{k_n}} + \rho_n(x)}{\sigma(x)} \right\} - (1 - \alpha) \\ &= P\left\{ \frac{\gamma}{\sqrt{k_n}} W_n - z_{1-\alpha/2} C_n \leq \varepsilon \leq \frac{\gamma}{\sqrt{k_n}} W_n + z_{1-\alpha/2} C_n \right\} - (1 - \alpha) \\ &= \int \left(\Phi\left(\frac{\gamma}{\sqrt{k_n}} w + z_{1-\alpha/2} C_n\right) - \Phi\left(\frac{\gamma}{\sqrt{k_n}} w - z_{1-\alpha/2} C_n\right) - (1 - \alpha) \right) P^{W_n}(dw). \quad (3.71) \end{aligned}$$

Für die Dichte $\phi(x) = \exp(-x^2/2)$ der Standardnormalverteilung gilt für alle $x \in \mathbb{R}$

$$\phi(x) = \phi(-x) \quad \text{sowie} \quad \phi'(x) = -x\phi(x)$$

und damit

$$\phi'(-x) = x\phi(-x) = -\phi'(x).$$

Daher liefern die Taylor-Entwicklungen von $\Phi\left(\frac{\gamma}{\sqrt{k_n}}w \pm z_{1-\alpha/2}C_n\right)$ um $\pm z_{1-\alpha/2}$ dann

$$\begin{aligned}
& \Phi\left(\frac{\gamma}{\sqrt{k_n}}w + z_{1-\alpha/2}C_n\right) - \Phi\left(\frac{\gamma}{\sqrt{k_n}}w - z_{1-\alpha/2}C_n\right) - (1 - \alpha) \\
&= \underbrace{\Phi(z_{1-\alpha/2})}_{=1-\alpha/2} + \phi(z_{1-\alpha/2})\left(\frac{\gamma}{\sqrt{k_n}}w + z_{1-\alpha/2}(C_n - 1)\right) \\
&\quad + \frac{1}{2}\phi'(z_{1-\alpha/2})\left(\frac{\gamma}{\sqrt{k_n}}w + z_{1-\alpha/2}(C_n - 1)\right)^2 + T_{3,1}(w, n) \\
&\quad - \underbrace{\Phi(-z_{1-\alpha/2})}_{=\alpha/2} - \phi(-z_{1-\alpha/2})\left(\frac{\gamma}{\sqrt{k_n}}w + z_{1-\alpha/2}(1 - C_n)\right) \\
&\quad - \frac{1}{2}\phi'(-z_{1-\alpha/2})\left(\frac{\gamma}{\sqrt{k_n}}w + z_{1-\alpha/2}(1 - C_n)\right)^2 - T_{3,2}(w, n) - (1 - \alpha) \\
&= 2\phi(z_{1-\alpha/2})z_{1-\alpha/2}(C_n - 1) + \frac{1}{2}\phi'(z_{1-\alpha/2})\left(2\frac{\gamma^2}{k_n}w^2 + 2z_{1-\alpha/2}^2(C_n - 1)^2\right) \\
&\quad + T_{3,1}(w, n) - T_{3,2}(w, n) \\
&= \phi(z_{1-\alpha/2})z_{1-\alpha/2}\left(2(C_n - 1) - \frac{\gamma^2}{k_n}w^2 - (C_n - 1)^2z_{1-\alpha/2}^2\right) + T_{3,1}(w, n) - T_{3,2}(w, n),
\end{aligned} \tag{3.72}$$

mit

$$|T_{3,i}(w, n)| \leq \frac{1}{6} \sup_{\xi \in \mathbb{R}} |\phi''(\xi)| \left(\frac{\gamma}{\sqrt{k_n}}|w| + z_{1-\alpha/2}|C_n - 1| \right)^3 =: H_n(w)$$

für alle $w \in \mathbb{R}$, $n \in \mathbb{N}$, $i \in \{1, 2\}$. Dabei ist $|\phi''|$ wegen

$$\lim_{x \rightarrow \pm\infty} \phi''(x) = \lim_{x \rightarrow \pm\infty} (-1 + x^2)e^{-x^2/2} = 0$$

durch ein $M^* < \infty$ beschränkt. Mit

$$(a + b)^3 \leq 8(a^3 + b^3) \quad \text{für alle } a, b \geq 0$$

hat man für alle $n \in \mathbb{N}$

$$\begin{aligned}
k_n H_n(W_n) &\leq k_n \frac{M^*}{6} \left(\frac{\gamma}{\sqrt{k_n}}|W_n| + z_{1-\alpha/2}|C_n - 1| \right)^3 \\
&\leq \frac{4M^*}{3} \left(\frac{\gamma^3}{\sqrt{k_n}}|W_n|^3 + k_n z_{1-\alpha/2}^3 |C_n - 1|^3 \right).
\end{aligned}$$

Wegen (3.69) und (3.70) ist also

$$\lim_{n \rightarrow \infty} k_n E H_n(W_n) = 0$$

und damit

$$\lim_{n \rightarrow \infty} k_n ET_{3,i}(W_n, n) = 0, \quad i \in \{1, 2\}.$$

Nach (3.71) und (3.72) gilt (mit (3.69)) also insgesamt

$$\begin{aligned} & \lim_{n \rightarrow \infty} k_n (P\{Y \in I_n(x)\} - (1 - \alpha)) \\ &= \phi(z_{1-\alpha/2}) z_{1-\alpha/2} (\mu - \gamma^2 \lim_{n \rightarrow \infty} EW_n^2 - 0) \\ & \quad + \lim_{n \rightarrow \infty} k_n ET_{3,1}(W_n, n) - \lim_{n \rightarrow \infty} k_n ET_{3,2}(W_n, n) \\ &= \phi(z_{1-\alpha/2}) z_{1-\alpha/2} (\mu - \gamma^2). \end{aligned}$$

□

Approximative Prognoseintervalle zum Niveau $1 - \alpha$ können nach Satz 3.4.1(i) also entsprechend (3.66) festgelegt werden, wobei die optimale Konvergenzgeschwindigkeit der Überdeckungswahrscheinlichkeit (gemäß Teil (ii) des Satzes) nur für Folgen $(C_n)_{n \in \mathbb{N}}$ mit $\mu = \gamma^2$ erreicht wird. Insbesondere kann hier

$$C_n = \sqrt{1 + \frac{\gamma^2}{k_n}} = \sqrt{1 + \frac{c}{k_n} \int K^2(v) dv}$$

gewählt werden (vgl. (3.65)), denn

$$\lim_{n \rightarrow \infty} 2k_n \left(\sqrt{1 + \frac{\gamma^2}{k_n}} - 1 \right) = \lim_{n \rightarrow \infty} 2k_n \frac{1 + \frac{\gamma^2}{k_n} - 1}{\sqrt{1 + \frac{\gamma^2}{k_n}} + 1} = \gamma^2.$$

Um nun zu überprüfen, ob ε als standardnormalverteilt angenommen werden kann, bietet es sich (wegen $\varepsilon = \frac{Y - m(x)}{\sigma(x)}$ für $\sigma(x) > 0$) an, sobald Y bekannt ist, das Verhalten der standardisierten Residuen

$$\frac{Y - m_n(x)}{\sqrt{\sigma_n^2(x)}}$$

(mit $\sigma_n^2(x) = |l_n(x) - m_n^2(x)|$, wie in (3.62)) zu untersuchen.

Im Beispiel aus Abschnitt 3.3 zeigt jedoch bereits ein Histogramm der Daten - in Verbindung mit einem Vergleich der Quantile von empirischer Verteilungsfunktion und Standardnormalverteilung (Abbildung 3.2), dass dort keine Normalverteilung der Residuen vorliegt, sondern eine rechtsschiefe Verteilung. Dies ist naheliegend, da Y eine Schadenszufallsvariable ist und dabei sogar Großschäden beschreibt, so dass stärkere Abweichungen vom Erwartungswert nur in einer Richtung auftreten.

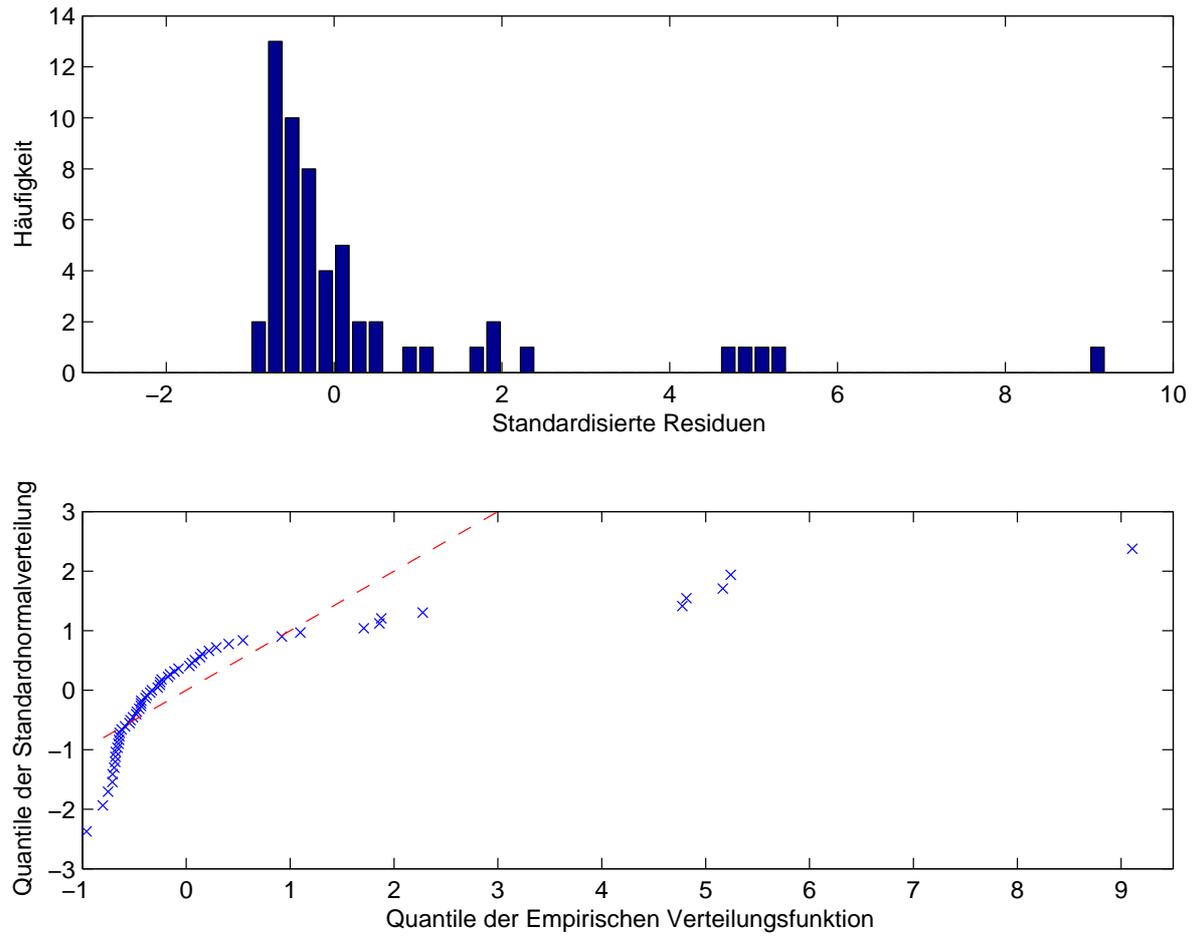


Abbildung 3.2: Standardisierte Residuen im Beispiel aus Abschnitt 3.3: Histogramm und Vergleich der Quantile von empirischer Verteilungsfunktion und Standardnormalverteilung (Q-Q-Plot)

Im allgemeinen Fall ist die analytische Berechnung von Prognoseintervallen nicht möglich.³⁷

Es lässt sich jedoch feststellen, dass Prognoseintervalle für Y meistens wesentlich breiter als Konfidenzintervalle für $m(x)$ sind. Denn in der Zerlegung (3.64) gilt $\text{Var}(\sigma(x) \cdot \varepsilon) = \sigma^2(x)$, während wegen Satz 3.1.8 im Regelfall $\text{Var}(m_n(x) - m(x)) \approx \sigma^2(x) \frac{c}{k_n} \int K^2(u) du$ deutlich geringer ausfällt. So gilt mit K aus (3.61) und der euklidischen Norm $\|\cdot\|$ (für beliebige $p \in \mathbb{N}$ und $\delta > 0$) stets $c \int K^2(u) du \leq 2$, so dass hier der Einfluss von $m_n(x) - m(x)$ und k_n bei wachsender Anzahl an Nachbarn schnell klein wird.

In Abbildung 3.3 werden die Prognosefehler $m_n(x) - Y$ der behandelten Daten mit den Werten $\sigma_n(x) := \sqrt{\overline{\sigma_n^2(x)}}$ verglichen. Um einen groben Eindruck über die entsprechenden Größenordnungen zu erhalten, sind neben den Konfidenzintervallen für $m(x)$ (blau) auch die Prognoseintervalle (rot) dargestellt, die sich ergeben würden, wenn ε doch normalverteilt wäre. Bezüglich ihrer Breite übertreffen letztere dabei die jeweiligen Konfidenzintervalle um den Faktor 4,77.

Fasst man allerdings die Schätzungen für mehrere verhältnismäßig ähnliche Einzelschäden zusammen, so relativiert sich dieser Unterschied. Während in (3.64) der Schätzfehler $m_n(x) - m(x)$ bei der Betrachtung der unterschiedlichen Schäden tendenziell nur wenig variiert (für $x \approx x'$ sind die zugehörigen Schätzungen $m_n(x)$ und $m_n(x')$ in der Regel positiv korreliert), sind hier die *Prognoseanteile* $Y - m(x) = \sigma(x) \cdot \varepsilon$ untereinander stochastisch unabhängig.

Im Extremfall weisen alle diese Schäden einen identischen bisherigen Abwicklungsverlauf x auf. Da dann natürlich auch alle Vorhersagen $m_n(x)$ übereinstimmen, bleibt der durchschnittliche Schätzfehler auch bei einer wachsenden Anzahl zusammengefasster Schäden unverändert, wohingegen sich der durchschnittliche Prognoseanteil gemäß dem schwachen Gesetz der großen Zahlen verringert.

Ähnlich stellt sich die Situation bei der Untersuchung der relativen Standardabweichung des Prognosefehlers dar. Zur Verdeutlichung betrachte man $q \in \mathbb{N}$ Schäden mit Abwicklungsverlauf x und zukünftigen Entschädigungen

$$Y^{(i)} = m(x) + \sigma(x) \cdot \varepsilon^{(i)}, \quad i \in \underline{q}.$$

³⁷Die approximative Verteilung von $Y - m_n(x)$ müsste mit geeigneten numerischen Verfahren aus der Faltung in (3.64) berechnet werden.

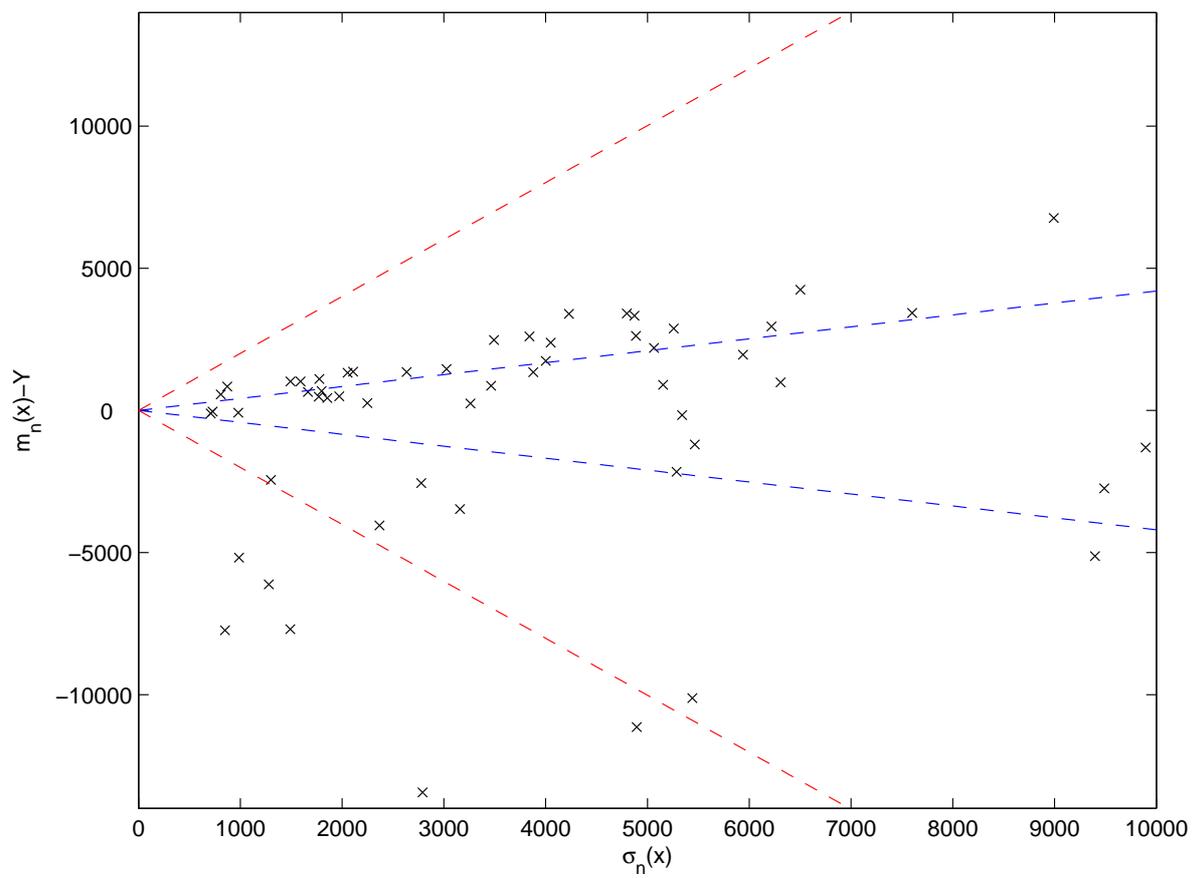


Abbildung 3.3: Vergleich von Vorhersagefehler und Schätzwert für $\sigma(x)$ (blau: Konfidenz-, rot: Prognoseintervall)

Mit der stochastischen Unabhängigkeit der $Y^{(i)}$, $i \in \underline{q}$, sowie von $Y^{(i)}$ und $(X_j, Y_j)_{j \in \underline{n}}$ (für alle $i \in \underline{q}$) erhält man wegen $\text{Var}(\varepsilon^{(i)}) = 1$

$$\begin{aligned}
 \frac{1}{q} \sqrt{\text{Var}\left(\sum_{i=1}^q (m_n(x) - Y^{(i)})\right)} &= \frac{1}{q} \sqrt{\text{Var}\left(\sum_{i=1}^q (m_n(x) - m(x)) - \sum_{i=1}^q \sigma(x)\varepsilon^{(i)}\right)} \\
 &= \frac{1}{q} \sqrt{\text{Var}(q \cdot (m_n(x) - m(x))) + \text{Var}\left(\sigma(x) \sum_{i=1}^q \varepsilon^{(i)}\right)} \\
 &= \frac{1}{q} \sqrt{q^2 \text{Var}(m_n(x) - m(x)) + \sigma^2(x)q \text{Var}(\varepsilon^{(1)})} \\
 &= \sqrt{\text{Var}(m_n(x)) + \frac{\sigma^2(x)}{q}}.
 \end{aligned}$$

Für eine wachsende Anzahl zusammengefasster Schäden steigt also der relative Anteil der Streuung des Schätzers $m_n(x)$ an der Streuung von $m_n(x) - Y$.

Kapitel 4

Schlussbetrachtung

4.1 Fazit

In dieser Arbeit ist diskutiert worden, wie sich Spätschäden anhand beobachteter Daten aus vergangenen Perioden präzise prognostizieren lassen.

Üblicherweise werden die zukünftigen Entschädigungen in der versicherungsmathematischen Praxis auf Basis nach Anfall- und Abwicklungsjahr gegliederter, jedoch ansonsten kumulierter Daten vorhergesagt.

Liegen die Beobachtungen auch getrennt für jeden einzelnen Schaden vor, so lässt sich diese Zusatzinformation nutzen, um bessere Schätzungen zu ermöglichen.

Hier ist der Ansatz verfolgt worden, für gemeldete, jedoch nicht vollständig regulierte Schäden zukünftige Leistungen auf der Grundlage ähnlicher Schadenverläufe aus den Vorjahren zu prognostizieren.

Genauere Prognosen als mit dem Chain-Ladder-Verfahren werden erreicht, indem hierfür die k nächstgelegenen Vorbilder - abstandsgewichtet - verwendet werden.

Fasst man die Vorhersage von Spätschäden in Abhängigkeit vom bisherigen Abwicklungsverlauf als Regressionsproblem auf, lassen sich diese k -nächste-Nachbarn-Schätzer als Schätzer für die Regressionsfunktion interpretieren.

Durch den Nachweis der asymptotischen Normalität des Schätzfehlers ist es ermöglicht worden, Konfidenzintervalle für die erwarteten zukünftigen Entschädigungszahlungen zu berechnen.

4.2 Ausblick

Nach Abschluss dieser Arbeit ergeben sich nun einige weitere interessante Fragestellungen.

Um das diskutierte k -nächste-Nachbarn-Verfahren zur Prognose von Einzelschäden in der Praxis leichter anwenden zu können, ist eine automatisierte Festlegung der Anzahl an Nachbarn (k) hilfreich.

Entsprechend Satz 3.2.1 ist bekannt, wie sich k asymptotisch (bei steigender Anzahl von Vergleichsschäden) verhalten sollte. Daher wird man k in Abhängigkeit von der Datenanzahl wählen. Denkbar ist aber auch, dass k bei der Prognose von $E(Y|X = x)$ außerdem umso größer gewählt wird, je stärker die Ausprägungen von Y lokal (unter den (X, Y) mit $X \approx x$) schwanken.

Für den Nachweis der asymptotischen Normalität des Schätzfehlers für $E(Y|X = x)$ wird die Voraussetzung benötigt, dass die erklärende Variable X eine Lebesgue-Dichte besitzt.

Dies führt dazu, dass im Kontext der Reserveschätzung nur für solche Schäden Aussagen über die Genauigkeit der Prognose gemacht werden können, deren bisheriger Abwicklungsverlauf durch ausnahmslos von Null verschiedenen Schäden charakterisiert ist.¹

In der Realität kann jedoch durchaus der Fall eintreten, dass in manchen Abwicklungsjahren eines Schadenverlaufes keinerlei Zahlungen anfallen, ohne dass dieser als abgeschlossen angesehen werden kann.

Es ist also wünschenswert, die Annahme an X zumindest insofern abschwächen zu können, dass die Schadenzuwächse in den einzelnen Abwicklungsjahren Punktmasse in Null aufweisen dürfen.

Dies führt allerdings schon allein deswegen zu Schwierigkeiten, weil die Eindeutigkeit der Auswahl der k Vorbilder dann nicht mehr gewährleistet ist.²

Im Abschnitt 2.4 ist verdeutlicht worden, dass mit dem k -nächste-Nachbarn-Verfahren sowohl Schäden der Folgeperiode als auch Zahlungen bis hin zu einem festen Abwicklungsjahr vergleichsweise präzise vorhergesagt werden können.

Führt man diese Prognose für alle einzelnen Abwicklungsjahre getrennt durch, so erhält man eine Schätzung für den gesamten zukünftigen Schadenverlauf.

Zu diesem Zweck ließe sich alternativ auch direkt eine - die restliche Abwicklung in den Jahren $p + 1$ bis I beschreibende - \mathbb{R}^{I-p} -wertige zu erklärende Zufallsvariable Y einführen, die völlig analog zum reellen Fall anhand von k Vergleichsschäden prognostiziert werden kann. Eine Untersuchung der Prognosegenauigkeit dürfte sich hier jedoch als aufwändig erweisen.

¹Vgl. das Vorgehen in Abschnitt 3.3.

²Da (in Satz 3.1.8) die asymptotische Grenzverteilung des Schätzfehlers für $E(Y|X = x)$ nicht mehr von der Dichte von X abhängt, besteht grundsätzlich Hoffnung, eine entsprechende Aussage - unter geeigneten Zusatzannahmen - auch im allgemeineren Fall herleiten zu können.

Literaturverzeichnis

- [Arnold92] Arnold, B. C., N. Balakrishnan und H. N. Nagaraja (1992): *A First Course in Order Statistics*. Wiley Series in Probability and Mathematical Statistics, New York.
- [Barndorff-Nielsen/Cox89] Barndorff-Nielsen, O. E. und D. R. Cox (1989): *Asymptotic Techniques for Use in Statistics*. Chapman and Hall, London.
- [Bauer74] Bauer, H. (1974): *Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie*. 2. Auflage; de Gruyter, Berlin.
- [Behnen/Neuhaus95] Behnen, K. und G. Neuhaus (1995): *Grundkurs Stochastik*. 3. Auflage; Teubner, Stuttgart.
- [Carroll/Härdle88] Carroll, R. J. und W. Härdle (1988): *Symmetrized Nearest Neighbor Regression Estimates*. In: *Statistics & Probability Letters* 7, S. 315-318.
- [Cheng95] Cheng, P. E. (1995): *A note on Strong Convergence Rates in Nonparametric Regression*. In: *Statistics & Probability Letters* 24(4), S. 357-364.
- [Collomb79] Collomb, G. (1979): *Estimation de la Régression par la Méthode des k Points les plus Proches avec Noyau: Quelques Propriétés de Convergence Ponctuelle*. In: *Lecture Notes in Mathematics* 821, S. 159-175.
- [Collomb81] Collomb, G. (1981): *Estimation Non-paramétrique de la Régression: Revue Bibliographique*. In: *International Statistical Review* 49, S. 75-93.
- [Devroye81] Devroye, L. P. (1981): *On the almost everywhere Convergence of Nonparametric Regression Function Estimation*. In: *The Annals of Statistics* 9, S. 1310-1319.
- [Devroye/Wagner80a] Devroye, L. P. und T. J. Wagner (1980): *On the L_1 Convergence of Kernel Estimates of Regression Functions with Applications in Discrimination*. In: *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 51, S. 15-25.
- [Devroye/Wagner80b] Devroye, L. P. und T. J. Wagner (1980): *Distribution-free Consistency Results in Nonparametric Discrimination and Regression Function Estimation*. In: *The Annals of Statistics* 8, S. 231-239.

- [Gänssler77] Gänssler, P. und W. Stute (1977): *Wahrscheinlichkeitstheorie*. Springer, Berlin.
- [Gasser/Müller79] Gasser, T. und H.-G. Müller (1979): *Kernel Estimation of Regression Functions*. In: *Lecture Notes in Mathematics* 757, S. 23-68.
- [Györfi02] Györfi, L., M. Kohler, A. Krzyzak und H. Walk (2002): *A Distribution-free Theory of Nonparametric Regression*. Springer, New York.
- [Härdle93] Härdle, W. (1993): *Applied Nonparametric Regression*. Econometric Society Monographs, 19; Cambridge University Press.
- [Hoeffding63] Hoeffding, W. (1963): *Probability Inequalities for Sums of Bounded Random Variables*. In: *Journal of the American Statistical Association* 58, S. 13-30.
- [Königsberger00] Königsberger, K. (2000): *Analysis 2*. 3. Auflage; Springer, Berlin.
- [Liero87a] Liero, H. (1987): *Limit Theorems for a k_n -Nearest Neighbour Estimate of the Regression Function*. Preprint-Serie Karl-Weierstrass-Institut für Mathematik der Akademie der Wissenschaften der DDR, 25/87.
- [Liero87b] Liero, H. (1987): *On the Asymptotic Behaviour of a k_n -Nearest Neighbour Estimate of the Regression Function*. In: *Seminarbericht, Sektion Mathematik der Humboldt-Universität Berlin* 89, S. 184-196.
- [Liero89] Liero, H. (1989): *Strong Uniform Consistency of Nonparametric Regression Function Estimates*. In: *Probability Theory and Related Fields* 82(4), S. 587-614.
- [Loftsgaarden/Quesenberry65] Loftsgaarden, D. O. und C. P. Quesenberry (1965): *A Nonparametric Estimate of a Multivariate Density Function* In: *The Annals of Mathematical Statistics* 36, S. 1049-1051.
- [Mack02] Mack, T. (2002): *Schadenversicherungsmathematik*. 2. Auflage; Verlag Versicherungswirtschaft, Karlsruhe.
- [Mack78] Mack, Y.-P. (1978): *k-Nearest-Neighbor Estimation*. Dissertation, University of California, San Diego.
- [Mack80] Mack, Y.-P. (1980): *Asymptotic Normality of Multivariate k-NN Density Estimates* In: *Sankhyā* 42, Series A, S. 53-63.
- [Mack81] Mack, Y.-P. (1981): *Local Properties of k-NN Regression Estimates* In: *SIAM Journal on Algebraic and Discrete Methods* 2(3), S. 311-323.
- [Mack/Rosenblatt79] Mack, Y.P. und M. Rosenblatt (1979): *Multivariate k-Nearest Neighbor Density Estimates* In: *Journal of Multivariate Analysis* 9, S. 1-15.

- [Pawlak91] Pawlak, M. (1991): *On the almost everywhere Properties of the Kernel Regression Estimate*. In: Annals of the Institute of Statistical Mathematics 43(2), S. 311-326.
- [Staniswalis89] Staniswalis, J.G. (1989): *Local Bandwidth Selection for Kernel Estimates*. In: Journal of the American Statistical Association 84(405), S. 284-288.
- [Stone77] Stone, C. (1977): *Consistent Nonparametric Regression*. In: The Annals of Statistics 5(4), S. 595-645.
- [Stute84] Stute, W. (1984): *Asymptotic Normality of Nearest Neighbor Regression Function Estimation*. In: The Annals of Statistics 12, S. 917-926.

Symbolverzeichnis

\emptyset	Leere Menge
1_A	Indikatorfunktion bezüglich einer Menge A
Beta(a, b)	Beta-Verteilung mit Parametern a und b
\mathbb{B}	Borel- σ -Algebra bezüglich \mathbb{R}
$\mathbb{B}(\partial B_1(0))$	Borel- σ -Algebra auf der Einheitskugel um 0
\mathbb{B}^p	Borel- σ -Algebra bezüglich \mathbb{R}^p
$B_r(x)$	Offene Kugel mit Radius r um x
∂A	Rand einer Menge A
\overline{A}	Abschluss einer Menge A
A^c	Komplement einer Menge A
E_n	$n \times n$ -Einheitsmatrix
H^T	Transponierte einer Matrix H
Γ	Gammafunktion
$\mathcal{L}(X)$	Verteilung einer Zufallsvariablen X
λ_p	Lebesgue-Maß über \mathbb{R}^p
\underline{n}	Menge der Zahlen $\{1, 2, \dots, n\}$
\mathbb{N}	Menge der natürlichen Zahlen $\{1, 2, \dots\}$
\mathbb{N}_0	$\mathbb{N} \cup 0$
$\mathcal{N}(a, \sigma^2)$	Normalverteilung mit Erwartungswert a und Varianz σ^2
P -f.s.	P -fast sicher
P -f.ü.	P -fast überall
Φ	Verteilungsfunktion der Standardnormalverteilung
ϕ	Lebesgue-Dichte der Standardnormalverteilung
\mathbb{R}	Menge der reellen Zahlen
$\mathcal{R}(0, 1)$	Rechteckverteilung über $(0, 1)$
$X_n \rightarrow X$ n.W.	Konvergenz von X_n gegen X nach Wahrscheinlichkeit
$X_n \xrightarrow{\mathcal{L}} X$	Konvergenz von X_n gegen X nach Verteilung
$z_{1-\alpha}$	$(1 - \alpha)$ -Quantil der Standardnormalverteilung
$ B $	Anzahl der Elemente einer Menge B
$\ \cdot\ _2$	Euklidische Norm