

Generalisierte Kernschätzung aus stochastischen Prozessen mit Anwendung in der Risikoanalyse

Der Fakultät Statistik
an der Technischen Universität Dortmund
zur
Erlangung des Grades
eines Doktors der Naturwissenschaften
eingereichte

D i s s e r t a t i o n

vorgelegt von

MSc Datenwiss. Vladyslav Ponyatovskyy

aus Kiev/Ukraine

Betreut durch PD Dr. Rafael Weißbach

Tag der Einreichung:

Tag der mündlichen Prüfung:

Hiermit versichere ich, die vorliegende Doktorarbeit unter der Betreuung von PD Dr. Rafael Weißbach nur mit den angegebenen Hilfsmitteln selbstständig angefertigt zu haben.

Dortmund, den 06.08.09

Danksagung

An dieser Stelle möchte ich allen danken, die mir bei dieser Arbeit geholfen haben. Mein primärer Dank gilt Herrn PD Dr. Rafael Weißbach für die geeignete Themenstellung und die inspirierende Betreuung.

Ich bedanke mich auch bei meinen Kollegen am Institut für Wirtschafts- und Sozialstatistik und persönlich beim Herrn Prof. Dr. Krämer für die Hilfsbereitschaft und eine freundliche Arbeitsatmosphäre.

Die Arbeit wäre ohne die finanzielle Unterstützung der Deutschen Forschungsgemeinschaft im Rahmen des Graduiertenkollegs „Statistische Modellbildung“ nicht möglich gewesen, besonders danke ich hierbei Frau Eva Brune.

Für die Begeisterung danke ich außerordentlich meinen Eltern und für ihre Unterstützung meiner Freundin Tatiana Vlasenco.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Kreditportfolios	3
1.2	Struktur der Arbeit	8
2	Verallgemeinerte Funktionalschätzer	11
2.1	Generalisierter Kernschätzer	11
2.1.1	Voraussetzungen	17
2.1.2	Konvergenz des Kernschätzers	19
2.1.3	Optimale Bandbreite	22
2.2	Lokal konsistenter Schätzer	24
2.3	Global konsistenter Schätzer	33
3	Schätzung aus stochastischen Prozessen	39
3.1	Empirische Wahrscheinlichkeitsfunktion	39
3.2	Empirische Dichtefunktion	41
3.3	Überlebensprozess	44
3.3.1	Modell	44
3.3.2	Empirische kumulative Hazardrate	46
3.3.3	Annahmen	48
3.3.4	Kernschätzer für Hazardrate	49
3.4	Markov-Prozess	55
3.4.1	Modell	55

3.4.2	Kumulative gewichtete Migrationsintensität	60
3.4.3	Beobachtungsschranken	62
3.4.4	Annahmen	63
3.4.5	Kernschätzer für Migrationsintensitäten	65
4	Simulationen stochastischer Prozesse	73
4.1	Überlebensprozess	74
4.2	Markov-Prozess	78
4.2.1	Testprozess aus exakten Migrationsintensitäten	78
4.2.2	Simulationsalgorithmus	86
4.2.3	Simulation eines Markov-Prozesses	91
5	Anwendung in der Risikoanalyse	95
5.1	Kreditrisikoaufschlag	96
5.2	Digitale Konkursversicherung	101
5.3	Beispiel einer Risikobewertung	104
6	Diskussion	111
	Literatur	116
A	Theoretische Grundlagen	121
A.1	Limes superior Formulierung der Konvergenz	121
A.2	Verteilung des absoluten Mittelwertes	122

Kapitel 1

Einleitung

Schon in den alten Zeiten hat man bemerkt, dass der Erfolg ohne Risiko undenkbar ist. Während der ganzen menschlichen Geschichte haben sich Tausende dem Schicksal überlassen und sind rasch nach oben geklettert... oder genauso schnell untergegangen. „Wer wagt, gewinnt!“ war seit Jahrhunderten das Motto von Abenteurern aller Art. Aber was, wenn es um die Verantwortlichkeit geht? Was, wenn nicht mit dem eigenen Leben, sondern mit dem Wohl der anderen gespielt wird? Desto aktueller klingt diese Frage im Finanzgeschäft, bei dem die Ersparnisse und Lebenswerke von Millionen aufs Spiel gesetzt werden. Die unzähligen Finanzkrisen zeigen, wie hoch das Risiko dabei ist, das Spiel mit dem Geschick zu verlieren und noch einmal den Untergang der Welt zu provozieren.

Das Wort *Risiko* verdankt seinen Ursprung den italienischen Seeleuten im Hochmittelalter, die auf eine ähnliche Weise das Umsteuern der Klippe bezeichnet haben. Die wachsende Komplexität der technischen und finanziellen Systeme in den späteren Jahrhunderten ließ allerdings keine manuelle Risikosteuerung mehr zu und oft wurde alles nur dem Zufall überlassen. Das 20. Jahrhundert zeigte deutlich, dass die Hoffnung keine gute Strategie ist. Die unzähligen Katastrophen, sowie wirtschaftliche und politische Krisen regten die Entwicklung von Methoden zur Risikosteuerung an. Die ersten

Schritte zum modernen Risikoverständnis hat dabei einer der bedeutendsten Wirtschaftswissenschaftler des 20. Jahrhunderts, Frank Knight (1885 - 1972), gemacht. In einer Reihe der Veröffentlichungen hat er das Risiko und die Ungewissheit definiert. Dabei erläuterte er in der Zeit der Weltwirtschaftskrise der 1930er Jahre die Idee, dass die historischen Daten nicht unbedingt auf die Zukunft hinweisen sollen.

Der bekannte niederländische Künstler Vincent van Gogh sagte einmal: „Was wäre das Leben, hätten wir nicht den Mut, etwas zu riskieren?“ und hat so eine andere Seite vom Risiko erläutert, die das Leben bunter und interessanter, dafür aber unsicherer macht. Genau das Spiel mit dem Schicksal finden so viele Menschen interessant, egal ob es etwas bringt oder nicht. So ist es nicht verwunderlich, dass die weitere Entwicklung des mathematischen Konzeptes der Unsicherheit im Rahmen der Spieltheorie erfolgte. Dabei leisteten die Mathematiker John von Neumann (1903 - 1957) und Oskar Morgenstern (1902 - 1977) einen großen Beitrag zur Untersuchung von Gesellschaftsspielen. Eine Ähnlichkeit von Spiel- und Finanzprozessen ermöglichte die Anwendung von Ansätzen aus der Spieltheorie auch zu effizienten Entscheidungen in der Wirtschaft. Allgemein haben die Strategien des Risikomanagements von der Risikovermeidung in den 1970er Jahren zur Risikosteuerung und risikoorientierten Performance in den 1990er Jahren evolutioniert. Trotzdem sind zur Zeit noch gar nicht alle Aspekte der Risikosteuerung untersucht. Wie weit wir noch von der risikofreien Wirtschaft stehen, zeigt uns die aktuelle Finanzkrise vom Jahre 2008. Die rasche Entwicklung von Rechenkapazitäten und mathematischen Methoden im 21. Jahrhundert lässt uns allerdings auf die Verminderung von Risiken der modernen Zivilisation und auf eine bessere Zukunft für die Menschheit hoffen.

Neben den mathematischen Verfahren zur Einschätzung von Risiken spielt auch die rechtliche Basis eine große Rolle in der Kontrolle von Gefahren wie menschliche Fehler, Ignoranz und Gier. Eine solche gesetzliche Grundlage bilden in Deutschland die Eigenkapitalvorschriften *Basel II*. Unter anderem

schaft diese Regelung einen Anreiz für die Banken, möglichst fortschrittliche Methoden zur Kreditausfallrisiken zu verwenden. Ein Beitrag zu einer präziseren Bewertung von Ausfallrisiken wird auch in dieser Arbeit geleistet, wobei die Verwendung von stochastischen Prozessen zur Risikobewertung in Kreditportfolios betrachtet wird.

1.1 Kreditportfolios

Wie bereits erwähnt, ist auf dem Finanzmarkt eine bessere Abschätzung von Kreditausfallrisiken gefragt. Zu den wichtigsten Finanzinstrumenten der Risikobewertung gehören Kreditratings, d. h. die Bewertungen der Kreditwürdigkeit von Emittenten. In der Regel stehen hinter den Ratingsklassen die Ausfallwahrscheinlichkeiten, so dass eine „gute“ Ratingsklasse auf eine kleine Ausfallwahrscheinlichkeit hinweist und umgekehrt. Die Kreditratings werden üblich intern durch die finanziellen Institutionen oder extern durch die Ratingagenturen (*Fitch*, *Standard & Poor's*, *Moody's*) erstellt. Aufgrund der unbeständigen Situation auf dem Finanzmarkt werden die Ratings im Laufe der Zeit immer wieder aktualisiert, so dass die Ratingsverläufe wie in Abbildung 1.1 entstehen.

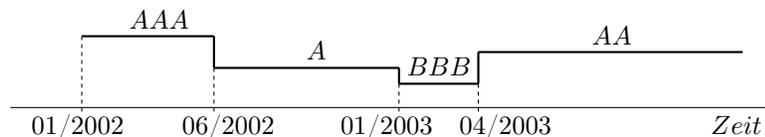


Abbildung 1.1: Ratingmigrationen eines Emittenten.

Die erste intuitive Idee wäre nun, die Anleihen nur bei Emittenten in den besten Ratingsklassen zu kaufen, was in der Realität nicht unbedingt der Fall ist. Trotz der scheinbaren Einfachheit bevorzugen finanzielle Institutionen eher risikobehaftete Anleihen, da diese wesentlich mehr Erlös bringen können. Die dabei entstehenden Risiken werden nicht mehr einzeln, son-

dern in Rahmen der so genannten Kreditportfolios (siehe Abbildung 1.2) gesteuert. Unter den Kreditportfolios kann man sowohl die Gesamtheit aller Kredite eines Instituts, als auch nur eine nach bestimmten Kriterien ausgewählte Teilmenge verstehen. Neben der Situation auf dem Finanzmarkt ist das Gesamtrisiko eines Kreditportfolios, auch sehr stark von dem Auswahlverfahren geprägt. Darüber hinaus treten die Emittenten normalerweise unabhängig voneinander und zu verschiedenen Zeitpunkten ins Portfolio ein und erschaffen somit das einzigartige Risikomuster für jedes Kreditportfolio. Bei der Risikoanalyse von Kreditportfolios wird das Ziel gesetzt, unter Beachtung von Ausfallrisiken den erwarteten Gewinn des gesamten Portfolios zu maximieren oder zumindest auf das Niveau vom risikofreien Portfolio zu bringen.

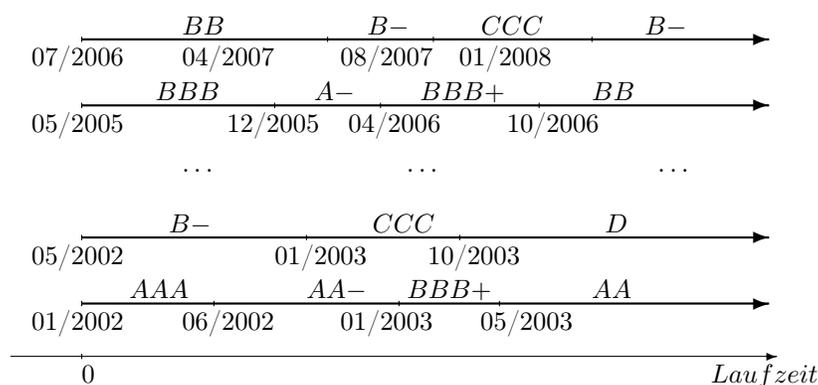


Abbildung 1.2: Ratingsverläufe von Emittenten in einem Kreditportfolio.

Als ein Maß zur Risikobewertung betrachten wir in dieser Arbeit den Kreditrisikoaufschlag (engl. Credit Spread). Dabei handelt es sich von einem erhöhten Zinssatz für die risikobehafteten Anleihen im Vergleich zu den risikofreien. Zur Berechnung vom Credit Spread werden in der Praxis hauptsächlich drei Ansätze verwendet. Bei der ersten Methode aus Merton (1974) wird versucht, den Firmenwert nach dem Black-Scholes-Ansatz (Black und Scholes, 1973) zu modellieren. Der Credit Spread wird dabei aus dem

Vergleich der ausstehenden Zahlungen mit dem modellierten Firmenwert am Fälligkeitstag berechnet. Leider setzt das Merton-Modell die Firmenwerte als bekannt voraus, was in der Wirklichkeit nicht der Fall ist und die Realisierbarkeit dieses Modells begrenzt. Dieser Nachteil wird z.T. in den Ansätzen kommerzieller Anbieter wie **CreditRisk**, **KMV** u. a. behoben. Trotzdem lassen wir im weiteren auch diese Methoden außer Acht.

Die dritte Klasse von Methoden zur Risikobewertung stellen die so genannten intensitätsbasierten Modelle dar, bei denen der Fokus auf dem Zeitpunkt des Defaults liegt und der Insolvenzprozess exogen modelliert wird. Ein gängiger Ansatz in diesem Fall ist, einen stochastischen Prozess mit dem finiten Zustandsraum den Ratingmigrationen von Emittenten zu unterstellen. Dabei werden die Informationen über Kreditnehmer in der Form von Kreditratings zur Risikobewertung verwendet. Es sollte noch betont werden, dass die Verwendung von Ausfallwahrscheinlichkeiten bzw. Kreditratings zur Ermittlung von Risiken heutzutage sehr populär ist. Auch nach den *Basel II*-Vorschriften ist das Mindestkapital zur Eigenmittelunterlegung durch die IRB-Formel (engl. **I**nternal **R**ating **B**ased Approach) zu berechnen.

Das Gesamtrisiko eines Portfolios kann bei dem letzten Ansatz anhand der endlichen Menge von funktionalen Charakteristiken des unterstellten stochastischen Prozesses ausgedrückt werden. Als funktionale Charakteristiken verstehen wir in dieser Arbeit eine Menge von Funktionen, die bestimmte Eigenschaften eines stochastischen Prozesses festlegen. Für einen Überlebensprozess kann z. B. die Hazardrate als funktionale Charakteristik bezeichnet werden. Für einen Markov-Prozess sind es die Migrationsintensitäten. In beiden Fällen wird die zeitabhängige Neigung zum Übergang in den nächsten Zustand durch die funktionalen Charakteristiken festgelegt. Die Berechnung vom Credit Spread wird dementsprechend auf die Schätzung dieser Charakteristiken zurückgeführt. Um die Schätzung zu ermöglichen, werden die Ratingsverläufe von einzelnen Emittenten im Rahmen eines Portfolios als stochastisch unabhängige Realisationen des unterstellten Prozesses betrach-

tet. Die Unabhängigkeitsannahme kann auf Grund des zufälligen Charakters des Eintritts von Emittenten ins Portfolio getroffen werden.

Um einen Insolvenzprozess zu modellieren, wird im einfachsten Fall den aggregierten Ratingsverläufen ein Überlebensprozess unterstellt. Dieser Prozess wird häufig eingesetzt, um das Ausscheiden der Objekte aus der statistischen Erfassung im Laufe der Zeit zu beschreiben. Es wird dabei angenommen, dass jeder Emittent sich zunächst in einem nichtinsolventen Zustand befindet. Der weitere Prozessverlauf - wie in Abbildung 1.3 - entsteht, indem der Emittent im Zeitpunkt τ in die Insolvenz wechselt. Durch die Zufallsvariable τ wird dabei der Zeitpunkt des Ausfalls bezeichnet.

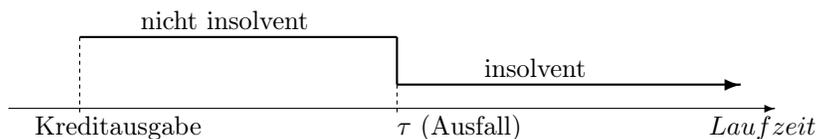


Abbildung 1.3: Verlauf des Überlebensprozesses.

Dieser Ansatz (in der Literatur als Hazardratenmodell genannt) wurde erstmals von Jarrow und Turnbull (1995) vorgeschlagen und später von Duffie und Singleton (1999) erweitert. Seinen Namen verdankt das Hazardratenmodell der Hazardrate

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} P(\tau \in [t, t + \Delta t] \mid \tau \geq t),$$

die als ein stetiges Analogon zu den momentanen Ausfallwahrscheinlichkeiten zur Definition der Überlebensprozesse eingesetzt wird.

Die Schätzung der Hazardrate kann sowohl parametrisch durch eine Verteilungsannahme als auch nichtparametrisch erfolgen. So wird z. B. in Madan und Unal (2000) die Hazardrate als ein Produkt der konstanten Intensität des Ausfallprozesses und der Verteilung der Verluste modelliert. Durch die Verteilungsannahmen für den Ausfallprozess und die Verluste kommt ein zweiparametrisches Modell zustande. Leider können häufig keine Annahmen über

den Verlauf der Hazardrate getroffen werden und nichtparametrische Verfahren zur Funktionsschätzung müssen zum Einsatz kommen. Eine Vielzahl davon wurde z. B. in einer Monographie von Rao (1983) präsentiert.

Zu den nichtparametrischen funktionalen Schätzern gehören u. a. auch die Kernschätzer, deren Verallgemeinerung in dieser Arbeit betrachtet wird. Unter geringen Modellannahmen ermöglicht diese Klasse von Schätzern eine konsistente und stetige Schätzung einer Funktion $\psi(\cdot)$, falls ein konsistenter Schätzer für die kumulative Funktion $\Psi(x) = \int_{-\infty}^x \psi(u)du$ vorliegt.

Die Kernschätzer wurden erstmals in Parzen (1962) zur Dichteschätzung eingesetzt und wurden in den folgenden Jahren von vielen Autoren weiterentwickelt und auf andere Funktionen wie die Hazardrate erweitert. Besonders hilfreich ist der Kernschätzer für die Hazardrate bei unvollständigen Daten, da der konsistente Schätzer für die kumulative Hazardrate $\Lambda(x) = \int_{-\infty}^x \lambda(u)du$ in der Regel problemlos zu ermitteln ist. Der Fall von rechtszensierten Beobachtungen wurde z. B. in Schäfer (1985) betrachtet, wobei die Konvergenzrate des Kernschätzers geliefert wurde. Die Kernschätzung der Hazardrate bei einer weiteren Art der Unvollständigkeit - der Linkstrunkierung - wird in dieser Arbeit präsentiert.

Der wesentliche Nachteil des Hazardratenmodells liegt allerdings darin, dass der Credit Spread unabhängig vom Ausgangsrating berechnet wird. In Wirklichkeit besteht auf dem Finanzmarkt ein deutlicher Zusammenhang zwischen Kreditratings und dem Zinssatz für die Anleihen. Dieser Nachteil kann behoben werden, indem den Ratingsverläufen ein Markov-Prozess $\{X_t \in \mathbb{S}\}_{t \in [0, \infty)}$ unterstellt wird. Das Ratingssystem stellt dabei den Zustandsraum \mathbb{S} des Markov-Prozesses dar. Als funktionale Charakteristiken des Markov-Prozesses gelten die Migrationsintensitäten:

$$q_{ih}(t) = \lim_{\Delta t \rightarrow 0} \frac{P(X_{t+\Delta t} = h \mid X_t = i) - P(X_t = h \mid X_t = i)}{\Delta t}, \quad \forall i, h \in \mathbb{S}.$$

Dieser Ansatz wurde erstmals in Jarrow *et al.* (1997) als eine Erweiterung des Jarrow-Turnbull-Modells präsentiert. Der Einfachheit halber setzen

die Autoren einen homogenen Markov-Prozess voraus, bei dem die Migrationsintensitäten zeitkonstant bleiben. Eine Studie von Weißbach *et al.* (2009) zeigte allerdings, dass diese Annahme nicht immer gerechtfertigt ist. Dabei wurden Ratingsverläufe in einem Teilportfolio der *WestLB* auf das homogene Verhalten überprüft. Der Test auf die Homogenität aus Weißbach und Dette (2007) zeigte, dass nur einzelne externe Ratings einem homogenen Markov-Prozess folgen. Der Verlauf von internen Ratings ist dagegen inhomogen, so dass ein funktionaler Schätzer für die Migrationsintensitäten gefragt wird. Da so ein Schätzer in der Literatur kaum zu finden ist, wird er in dieser Arbeit entwickelt.

Als eine Grundlage benutzen wir dabei den generalisierten Ansatz zur Kernschätzung aus Weißbach (2006). Aus dem Konvergenzbeweis zu diesem generalisierten Schätzer folgt die Konsistenz für die früheren Anwendungen des Kernschätzers. Als eine Erweiterung dieses Ansatzes wird die Konstruktion des generalisierten Kernschätzers in dieser Arbeit präzisiert. Der vollständige generalisierte Ansatz zur Kernschätzung ermöglicht somit eine schnelle und einfache Konstruktion von neuen konsistenten nichtparametrischen funktionalen Schätzern mit der bekannten Konvergenzrate, was auf dem Beispiel von Migrationsintensitäten gezeigt wird.

1.2 Struktur der Arbeit

Die Arbeit gliedert sich in vier Bereiche. In Kapitel 2 wird zuerst auf historische Grundlagen der Kernschätzung eingegangen und eine präzisierte Konstruktion des Kernschätzers aus Weißbach (2006) gezeigt. Die Anwendungsbeispiele des generalisierten Kernschätzers werden dann in Kapitel 3 präsentiert. Dabei werden konsistente Schätzer für eine Reihe von stochastischen Prozessen nur durch die Überprüfung von Regularitätsbedingungen hergeleitet. Die Korrektheit der vorgestellten Schätzer wird für die Überlebens- und Markov-Prozesse anhand von Simulationsstudien in Kapitel 4 noch einmal

bestätigt. In Kapitel 5 wird eine Anwendung der Kernschätzung in der Risikoanalyse erläutert. Der existierende Rating-basierte Ansatz aus Jarrow und Turnbull (1995) zur Berechnung des Credit Spreads wird dabei mit Hilfe von inhomogenen Markov-Prozessen erweitert.

Den Schwerpunkt der Arbeit stellen Abschnitte 2.2 und 2.3 dar, in denen die Konstruktion des generalisierten Kernschätzers, sowie Konvergenzbeweise beschrieben werden. Auf eine mögliche Form der variablen Bandbreite wird dabei in Abschnitt 2.1.3 eingegangen.

Die Anwendung der hergeleiteten Konstruktion wird mit Hilfe von insgesamt drei Beispielen gezeigt. In Abschnitt 3.2 wird Konvergenz des wohl bekannten Kerndichteschätzers besprochen. Die Überlebensprozesse mit links-trunkierten und rechtszensierten Beobachtungen betrachten wir in Abschnitt 3.3, wobei die Konvergenz des Kernschätzers für die Hazardrate gezeigt wird. Das letzte Beispiel aus Abschnitt 3.4 bezieht sich auf die inhomogenen Markov-Prozesse. Die verallgemeinerte Konstruktion des Kernschätzers hilft in diesem Fall einen neuen funktionalen Schätzer für Migrationsintensitäten herzuleiten.

Die Simulationsstudien bestehen aus zwei Unterabschnitten. In Abschnitt 4.1 werden Überlebensprozesse betrachtet, wobei Realisationen eines Überlebensprozesses mit unvollständigen Daten erzeugt werden. Aus den simulierten Daten wird anschließend die Hazardrate geschätzt und mit der theoretischen grafisch verglichen.

Für Markov-Prozesse wird eine Simulationsstudie in Abschnitt 4.2 präsentiert. Um eine exakte Simulation zu ermöglichen, wird in Abschnitt 4.2.1 zuerst eine analytische Beschreibung des Markov-Prozesses mit angegebenen Migrationsintensitäten erläutert. Wie Realisationen eines Markov-Prozess aus dieser Beschreibung erzeugt werden, wird in Abschnitt 4.2.2 gezeigt. Zum Schluss in Abschnitt 4.2.3 erfolgt ein grafischer Vergleich von geschätzten Migrationsintensitäten mit den theoretischen.

Da eine Erweiterung des Rating-basierten Ansatzes im Hintergrund der

aktuellen Arbeit steht, wird in Abschnitt 5.1 eine Anwendung von inhomogenen Markov-Prozessen auf die Risikobewertung in Kreditportfolien vorgestellt. Dabei wird die Kernschätzung zur Berechnung des minimalen risikodeckenden Credit Spreads verwendet. Zusätzlich wird in Abschnitt 5.2 gezeigt, wie eine Risikoprämie bei der digitalen Konkursversicherung (engl. Digital Default Put) unter der Markov-Annahme berechnet wird. Ein Beispiel der Risikobewertung in einem Kreditportfolio ist zum Schluss in Abschnitt 5.3 zu sehen.

Kapitel 2

Die verallgemeinerten Funktionschätzer

2.1 Der generalisierte Kernschätzer

In diesem Kapitel wird auf die Grundideen des generalisierten Kernschätzers eingegangen. Dabei werden einige historischen Aspekte, sowie das Konvergenzverhalten dieses Schätzers betrachtet. Zum Schluss werden einige Richtlinien zur praktischen Anwendung des Kernschätzers vorgeschlagen.

Stochastische Prozesse werden nichtparametrisch durch funktionale Charakteristika spezifiziert, z. B. durch die Überlebensfunktion oder kumulative Hazardrate in der Verweildaueranalyse oder durch die Migrationsintensitäten in der Analyse von Markov-Prozessen. Wir bezeichnen eine funktionale Charakteristik dieser Art durch $\Psi(x)$, $x \in \mathbb{R}$. Häufig ist allerdings auch ein stetiger und konsistenter funktionaler Schätzer für die erste Ableitung $\psi(x) = d\Psi(x)/dx$, $x \in \mathbb{R}$ gefragt. Während konsistente Schätzer für $\Psi(x)$ in der Regel keine Probleme bereiten, waren die gefragten Schätzer für die ersten Ableitungen offensichtlich bis vor kurzem ein Stein des Anstoßes. Seit kurzem bietet der so genannte generalisierte Kernschätzer einen flexiblen Ansatz zur Schätzung von $\psi(x)$ dar, wobei die Verallgemeinerung des wohl

bekanntem Kerndichteschätzer gemeint ist. Aus diesem Grund ist die historische Entwicklung des Kerndichteschätzers von großer Bedeutung bei der Beschreibung des verallgemeinerten Ansatzes.

Wie erwähnt, geht die Konstruktion des Kernschätzers auf die Dichteschätzung zurück. Das Histogramm ist als ein intuitiver nichtparametrischer Dichteschätzer längst bekannt. Allerdings weist das Histogramm einige Nachteile auf, die eine „gute“ Schätzung erschweren können. So ist es z. B. der Fall, wenn die theoretische Dichte stetig ist. Das Histogramm ist dagegen weder stetig, noch konsistent bezüglich der theoretischen Dichtefunktion. Die Idee eines stetigen und konsistenten Dichteschätzers wurde erstmals 1956 von Murray Rosenblatt erläutert (Rosenblatt, 1956). Die endgültige Konstruktion des Kerndichteschätzers ist allerdings etwas später im Jahre 1962 in einem Artikel von Emanuel Parzen erschienen (Parzen, 1962). Als Kernfunktion (engl. kernel function) versteht man dabei einen Baustein in der Konstruktion des Dichteschätzers von Rosenblatt. Parzen zeigte in seinem Artikel die Konsistenz des neu entwickelten Schätzers, präziserte die Voraussetzungen für Kernfunktionen und stellte eine Reihe von möglichen Funktionen dieser Art vor. Wegen der Stetigkeitsannahme an die Kernfunktionen ist dabei auch der Kerndichteschätzer stetig, wodurch die bereits erwähnten Mängel des Histogramms behoben werden.

Angenommen, die u.i.v. Zufallsvariablen X_1, \dots, X_n mit der Verteilungsfunktion $F_X(\cdot)$ liegen vor. Bekanntlich kann $F_X(x)$ konsistent und erwartungstreu nichtparametrisch als $F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i \leq x\}}$ geschätzt werden. Wie kann aber die Dichtefunktion $f_X(x) = dF_X(x)/dx$ geschätzt werden? Es ist möglich, einen intuitiven Schätzer für $f_X(\cdot)$ als

$$f_n(x) = \frac{F_n(x + R_n) - F_n(x - R_n)}{2R_n} \quad (2.1)$$

mit $n \rightarrow \infty$, $R_n > 0$ und $R_n \rightarrow 0$ darzustellen. Dabei bleibt allerdings unklar, wie schnell R_n gegen Null streben soll. Die Konsistenz von $f_n(\cdot)$ bleibt ebenso ungewiss. Um diese Fragen zu untersuchen, stellen wir den intuitiven

Dichteschätzer (2.1) in einer allgemeiner Form wie folgt dar:

$$f_n(x) = \int_{\mathbb{R}} \frac{1}{R_n} \text{Kr} \left(\frac{x-y}{R_n} \right) dF_n(y), \quad (2.2)$$

mit $\text{Kr}(u) = 1/2$ für $|u| \leq 1$ und 0 sonst. Parzen (1962) beweist, dass auch andere Funktionen $\text{Kr}(\cdot)$ zu einem konsistenten Dichteschätzer führen. Falls die Konsistenz bzgl. $f_X(x)$ und eine beschränkte Varianz für $f_n(x)$ vorausgesetzt werden, folgen Eigenschaften von $\text{Kr}(\cdot)$:

- $\int_{\mathbb{R}} \text{Kr}(u) du = 1,$
- $\int_{\mathbb{R}} u \text{Kr}(u) du = 0,$
- $\int_{\mathbb{R}} \text{Kr}^2(u) du < \infty.$

Anschaulich passiert folgendes: die Wahrscheinlichkeitsmasse $1/n$ sind nicht in den Punkten X_i , $i = 1, \dots, n$ konzentriert, sondern in den näheren Umgebungen von diesen Punkten „verschmiert“. Dies kann realisiert werden, indem ein unabhängiges Y_i jedem X_i aufaddiert wird. Die Zufallsvariable Y_i steht für eine zufällige Schwankung um Null mit einer Dichte $f_Y(\cdot) = \frac{1}{R_n} \text{Kr}(\frac{\cdot}{R_n})$, wobei $R_n \in \mathbb{R}^+$ ein Skalierungsparameter und $\text{Kr}(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$ eine Funktion sind. Offensichtlich ist $f_Y(\cdot)$ eine Dichtefunktion, falls $\int_{\mathbb{R}} \text{Kr}(u) du = 1$ gilt. Die Bedingung $\int_{\mathbb{R}} u \text{Kr}(u) du = 0$ entspricht dabei $E(Y_i) = 0$. Die Dichte der „verschmierten“ Punkte $Z_i := X_i + Y_i$ wird wie folgt berechnet:

$$f_Z(z) = \int_{\mathbb{R}} f_Y(z-y) f_X(y) dy = \int_{\mathbb{R}} \frac{1}{R_n} \text{Kr} \left(\frac{z-y}{R_n} \right) dF_X(y).$$

Der Schätzer für $f_Z(z)$ bietet sich folgenderweise als Lebesgue-Stieltjes-Integral, wobei Dichte von Y_i bereits bekannt ist:

$$\hat{f}_Z(z) := \int_{\mathbb{R}} \frac{1}{R_n} \text{Kr} \left(\frac{z-y}{R_n} \right) dF_n(y) \quad (2.3)$$

Wenn die Varianz von Y_i wegen $R_n \rightarrow 0$ gegen 0 geht, stellt (2.3) den konsistenten Dichteschätzer für X_i dar. Daher ist dieser Schätzer äquivalent zu (2.2).

Wir bezeichnen nun die Funktionen der Art $\text{Kr}(\cdot)$ als Kernfunktionen, R_n als Bandbreite und die Formel (2.2) als Kerndichteschätzer. Seit der Entwicklung des Kerndichteschätzers wurde eine Reihe von Kernfunktionen entwickelt, deren Einfluss auf die Güte der Schätzung von vielen Autoren wie z. B. Gasser *et al.* (1985) oder Wand und Jones (1995) untersucht. Ohne in die Diskussion einzutreten, welche Kernfunktion besser ist, listen wir in Tabelle 2.1 die gebräuchlichsten auf. Der Epanechnikov-Kern stammt dabei aus Epanechnikov (1969).

Bezeichnung	Kernfunktion
Gaussian	$\text{Kr}(x) := \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}x^2)$
Epanechnikov	$\text{Kr}(x) := \frac{3}{4}(1 - x^2)\mathbf{1}_{\{ x \leq 1\}}$
Cosine	$\text{Kr}(x) := \frac{\pi}{4} \cos\left(\frac{\pi}{2}x\right) \mathbf{1}_{\{ x \leq 1\}}$
Triangle	$\text{Kr}(x) := (1 - x)\mathbf{1}_{\{ x \leq 1\}}$
Uniform	$\text{Kr}(x) := \frac{1}{2}\mathbf{1}_{\{ x \leq 1\}}$
Quartic	$\text{Kr}(x) := \frac{15}{16}(1 - x^2)^2\mathbf{1}_{\{ x \leq 1\}}$
Triweight	$\text{Kr}(x) := \frac{35}{32}(1 - x^2)^3\mathbf{1}_{\{ x \leq 1\}}$
Triangle	$\text{Kr}(x) := (1 - x)\mathbf{1}_{\{ x \leq 1\}}$

Tabelle 2.1: Die gebräuchlichsten Kernfunktionen.

Eine weitere Entwicklung des Kerndichteschätzers beruht auf der Tatsache, dass die Konzentration von Messpunkten in verschiedenen Regionen des

Wertebereichs auch unterschiedlich ist. Die aktuelle Bandbreite R_n hängt zwar vom Datenumfang n ab, bleibt aber sowohl für alle Beobachtungen X_1, \dots, X_n als auch für verschiedene Bereiche von $F_n(\cdot)$ fix. So besteht die Gefahr, dass der Schätzer „überglättet“ und in den steilen Stellen eine Verzerrung verursacht, falls zu viele Messungen zur Schätzung herangezogen werden. In den Intervallen mit einer kleinen Dichte liegen dagegen zu wenige Beobachtungen vor, deren Ausprägungen den Schätzer in diesem Fall sehr stark beeinflussen. Die Fluktuationen des Schätzers sind dann sehr groß. Eine flexible Bandbreite $R_n(x \mid X_1, \dots, X_n)$, $x \in \mathbb{R}$ wäre die Lösung dieses Problems.

Eine intuitive Idee zur Realisierung der flexiblen Bandbreite ist, nicht die Bandbreite, sondern die Beobachtungsanzahl innerhalb der Bandbreite konstant zu halten. Die weiteren Konsistenzbeweise basieren im wesentlichen auf dem Vorschlag aus Moore und Yackel (1977), eine stochastische Bandbreite $R_n(x)$ als den Abstand von x zu seinem k -nächsten Nachbarn unter X_1, \dots, X_n zu wählen. Eine solche variable Bandbreite wurde Nächste-Nachbar-Bandbreite genannt und kann formal folgendermaßen ausgedrückt werden:

$$R_n(x) := \inf\{r > 0 : |F_n(x - r/2) - F_n(x + r/2)| \geq p_n\}, \quad (2.4)$$

wobei $p_n = k/n$. Dementsprechend wird der Kerndichteschätzer

$$f_n(x) = \int_{\mathbb{R}} \frac{1}{R_n(y)} \text{Kr} \left(\frac{x - y}{R_n(y)} \right) dF_n(y) \quad (2.5)$$

als Nächste-Nachbar-Schätzer bezeichnet.

Der Nächste-Nachbar-Ansatz wird in Schäfer (1986a) durch die Unterstellung (p_n) einer positiven Nullfolge mit einer beschränkten Konvergenzgeschwindigkeit erweitert. Die Verallgemeinerung der Nächste-Nachbar-Methode erfolgt allerdings erst in Weißbach (2006), wo ein konsistenter Kernschätzer für eine allgemeine Funktion $\psi(x) = d\Psi(x)/dx$ präsentiert wird. Die Konsi-

stanz wird dabei unter den Voraussetzungen für $\Psi(\cdot)$ und $\psi(\cdot)$ aus Abschnitt 2.1.1 gezeigt. Die Grundidee ist dabei, auf der Basis von existierenden Kernschätzern den Konvergenzbeweis für die neuen Schätzer zu vereinfachen.

Darüber hinaus wird in der Nächste-Nachbar-Bandbreite (2.4) die empirische Verteilungsfunktion durch einen Schätzer $\tilde{\Psi}_n(\cdot)$ einer Funktion $\tilde{\Psi}(\cdot)$ ersetzt, so dass sich die folgende Bandbreite ergibt:

$$R_n(t) := \inf\{r > 0 : |\tilde{\Psi}_n(t - r/2) - \tilde{\Psi}_n(t + r/2)| \geq p_n\}. \quad (2.6)$$

Die Funktion $\tilde{\Psi}(\cdot)$ mit ähnlichen Eigenschaften wie $\Psi(\cdot)$ bezeichnen wir im folgenden als Glätter. Die Einführung einer neuen Funktion zur Bestimmung der Bandbreite verleiht der Nächste-Nachbar-Methode eine gewisse Flexibilität. Wird z. B. $\tilde{\Psi}(\cdot)$ als eine lineare Funktion

$$\tilde{\Psi}(\cdot) = \tilde{\Psi}_n(\cdot) = \tilde{c} \cdot \text{id}(\cdot) + \tilde{d} \quad (2.7)$$

definiert, so ergibt sich die konstante Bandbreite $R_n = p_n/|\tilde{c}|$, deren Herleitung in Weißbach (2006) zu finden ist. Durch $\text{id}(\cdot)$ wird dabei die Identität bezeichnet. Zusammen mit der Auflistung der notwendigen Voraussetzungen zeigt Weißbach in demselben Artikel die Konvergenzrate für den generalisierten Kernschätzer. Der generalisierte Kernschätzer

$$\psi_n(x) := \int_{\mathbb{R}} \frac{1}{R_n(t)} \text{Kr} \left(\frac{x-t}{R_n(t)} \right) d\Psi_n(t) \quad (2.8)$$

entstand aus der einfachen Überlegung, die Idee des Kerndichteschätzers auch auf die anderen Funktionen zu übertragen. Wie bereits erwähnt, verstehen wir als generalisierten Kernschätzer einen Schätzer für $\psi(x) = d\Psi(x)/dx$, wobei $\Psi(\cdot)$ durch $\Psi_n(\cdot)$ konsistent geschätzt wird.

In Abschnitt 2.1.2 wird unter den Voraussetzungen aus Abschnitt 2.1.1 gezeigt, dass $\psi_n(\cdot)$ auf einem Intervall $[A, B]$ mit der Rate $(\sqrt{\log(n)/(np_n)} + p_n)$ gegen $\psi(\cdot)$ konvergiert.

2.1.1 Voraussetzungen

Damit der Kernschätzer (2.8) auf einem Intervall $[A, B]$ gegen die wahre Funktion $\psi(\cdot)$ mit der Rate $(\sqrt{\log(n)/(np_n)} + p_n)$ konvergiert, müssen die Kernfunktion, der Glätter und die Funktionen $\Psi(\cdot)$ und $\tilde{\Psi}(\cdot)$ gewisse Bedingungen erfüllen. Die im weiteren aufgelisteten Voraussetzungen stammen ursprünglich aus Weißbach (2006) und stellen eine notwendige Bedingung zur oben erwähnten Konvergenzgeschwindigkeit dar. Da der generalisierte Kernschätzer eine Verallgemeinerung des Kerndichteschätzers darstellt, werden einige Voraussetzungen für diesen wie z. B. die Rechtsstetigkeit übernommen.

(K1) Für die Funktionen $\Psi : \mathbb{R} \rightarrow [0, \infty)$ und $\tilde{\Psi} : \mathbb{R} \rightarrow [0, \infty)$ wird die Existenz von solchen $0 < m \leq M < \infty$ und $0 < \tilde{m} \leq \tilde{M} < \infty$ angenommen, dass

$$\begin{aligned} m|x - y| &\leq |\Psi(x) - \Psi(y)| \leq M|x - y| \\ \tilde{m}|x - y| &\leq |\tilde{\Psi}(x) - \tilde{\Psi}(y)| \leq \tilde{M}|x - y| \end{aligned}$$

für alle $x, y \in [A, B] \subset \mathbb{R}(A < B)$ gelten.

Dadurch wird die Existenz von ersten Ableitungen $\psi(\cdot)$ und $\tilde{\psi}(\cdot)$ mit $m \leq |\psi(x)| \leq M$ und $\tilde{m} \leq |\tilde{\psi}(x)| \leq \tilde{M}$ für alle $x \in [A, B]$ gewährleistet und strenge Monotonie von $\Psi(\cdot)$ und $\tilde{\Psi}(\cdot)$ auf $[A, B]$ vorausgesetzt.

(K2) $\psi(\cdot)$ und $\tilde{\psi}(\cdot)$ sind Lipschitz-stetig auf $[A, B]$ mit Konstanten L_ψ und $L_{\tilde{\psi}}$.

Da der Kernschätzer stetig ist, ist er auch nur zur Schätzung von stetigen Funktionen geeignet.

(K3) $\Psi_n(\cdot)$ und $\tilde{\Psi}_n(\cdot)$ sind rechtsstetig und monoton mit

$$P \left\{ \limsup_{n \rightarrow \infty} \frac{\sup_{J \subseteq [A, B], \Psi(J) \leq p_n} |\Psi_n(J) - \Psi(J)|}{\sqrt{\log(n)p_n/n}} = D \right\} = 1$$

$$P \left\{ \limsup_{n \rightarrow \infty} \frac{\sup_{J \subseteq [A, B], \tilde{\Psi}(J) \leq p_n} |\tilde{\Psi}_n(J) - \tilde{\Psi}(J)|}{\sqrt{\log(n)p_n/n}} = \tilde{D} \right\} = 1$$

für ein $0 < D < \infty$ und ein $0 < \tilde{D} < \infty$, wobei der theoretische und der empirische Maß jeweils als $\Psi(J) := \int_J \psi(t)dt = \Psi(b) - \Psi(a)$ und $\Psi_n(J) := \Psi_n(b) - \Psi_n(a)$ definiert werden und (p_n) eine Nullfolge aus der Voraussetzung (K5) ist. Die Definition der Maße bzgl. des Glätters ist dieselbe.

Eine mögliche Konstruktion der Schätzer $\Psi_n(\cdot)$ und $\tilde{\Psi}_n(\cdot)$ wird in Kapitel 2.2 näher erläutert. Diese basiert auf einer Verallgemeinerung der empirischen Verteilungsfunktion, sowie der empirischen kumulativen Hazardrate. Die Konvergenzbeweise für die oben genannten Schätzer stellen den methodischen Kern der aktuellen Arbeit dar.

(K4) Die Funktionen $\psi(\cdot)$ und $\tilde{\psi}(\cdot)$ sind streng positiv auf $[A, B]$.

Ursprünglich wurde der Kerndichteschätzer für positive Dichtefunktionen entwickelt. Diese Eigenschaft wurde auch im generalisierten Schätzer übernommen.

(K5) Für eine Folge (p_n) wird $p_n > 0$, $p_n \rightarrow 0$ und $np_n/\log(n) \rightarrow \infty$ angenommen.

(K6) Die Kernfunktion $\text{Kr}(\cdot)$ ist stetig und es gilt $\text{Var}(\text{Kr}) < \infty$.

(K7) Die Kernfunktion $\text{Kr}(\cdot)$ ist Lipschitz-stetig (mit einer Konstante L_{Kr}) und es gilt $\sup(\text{Kr}) < \infty$.

Es muss beachtet werden, dass Annahme (K7) für viele bekannte Kernfunktionen gilt, außer der Gaus'schen Kernfunktion. Im Bezug auf die Minimierung des asymptotischen integrierten mittleren quadratischen Fehlers haben Wand und Jones (1995) herausgestellt, dass die Epanen-chikovs Kernfunktion (siehe Tabelle 2.1) am effizientesten ist.

2.1.2 Konvergenz des Kernschätzers

Die fast sichere Konvergenz des Kerndichteschätzers mit der konstanten Bandbreite wurde erstmals von Nadaraya (1965) untersucht. Es wurde gezeigt, dass $\sup_{x \in \mathbb{R}} |f_n(x) - f(x)|$ in Wahrscheinlichkeit 1 gegen 0 konvergiert, wobei $f_n(\cdot)$ der Kerndichteschätzer (2.1) ist. Erst 1984 haben Luc Devroye und Clark S. Penrod ein allgemeines Konsistenzergebnis für die variablen Bandbreiten geliefert (Devroye und Penrod, 1984). Obwohl diese Autoren die Nächste-Nachbar-Bandbreite erwähnt haben, unterscheidet sich der im Artikel vorgeschlagene Kerndichteschätzer von dem Nächste-Nachbar-Schätzer (2.5).

Die weiteren Untersuchungen bezogen sich auf die Übertragung der Idee der Kerndichteschätzung auf andere Schätzprobleme. So haben Földes *et al.* (1981) und Failing (1983) unabhängig voneinander den Kernschätzer auf zensierte Daten angewendet, wobei die empirische Verteilungsfunktion durch den Kaplan-Meier-Schätzer ersetzt wurde. Der Beweis der starken gleichmäßigen Konvergenz wurde unter Benutzung der globalen Konvergenz des Kaplan-Meier-Schätzers geliefert. Gleichzeitig haben verschiedene Autoren (Ramlau-Hansen (1983), Yandell (1983), Tanner und Wong (1983)) unabhängig voneinander Kernschätzer mit fester Bandbreite für die Hazardrate entwickelt. Im Jahre 1985 zeigte Helmut Schäfer die gleichmäßige fast sichere Konvergenz für die Kernschätzung der Hazardrate mit der variablen Bandbreite, wobei der Nächste-Nachbar-Schätzer entsprechend modifiziert wurde (Schäfer, 1985). Die ausführlichen Beweise dazu finden sich in Schäfer (1986a). Die Tatsache, dass der Kernschätzer nicht nur für die konsistente Dichte-

schätzung anwendbar ist, führte zur Verallgemeinerung dieses Ansatzes. Auf der Basis der Arbeiten von Schäfer zeigte Rafael Weißbach 2006 die gleichmäßige fast sichere Konvergenz für den generalisierten Kernschätzer (2.8) (Weißbach, 2006). Dabei wurde u. a. die Konvergenzrate des generalisierten Kernschätzers geliefert.

Weiter betrachten wir den Konvergenzsatz für den verallgemeinerten Kernschätzer, sowie ein paar Hinweise zur Wahl des Glätters.

Satz 2.1.1. *Seien die Voraussetzungen (K1) bis (K7) erfüllt.*

Dann existiert eine Konstante $0 \leq D_0 \leq \max\{D_1 + D_2, D_3\}$, so dass

$$P \left\{ \limsup_{n \rightarrow \infty} \frac{\sup_{x \in [A, B]} |\psi_n(x) - \psi(x)|}{\sqrt{\log(n)/(np_n) + p_n}} = D_0 \right\} = 1$$

wobei

$$D_1 := 2\tilde{D}M\tilde{M}^2\tilde{m}^{-2}[\text{sup}(\text{Kr}) + L_{K_r}\tilde{M}\tilde{m}^{-1}]$$

$$D_2 := D\tilde{M}M^{1/2}\text{Var}(\text{Kr})\tilde{m}^{-1/2}$$

$$D_3 := 2\tilde{M}^3ML_{K_r}L_{\tilde{\psi}}\tilde{m}^{-5} + 2\text{sup}(\text{Kr})L_{\tilde{\psi}}\tilde{M}^2M\tilde{m}^{-4} + L_{\psi}\tilde{m}^{-1}$$

gelten.

Beweis. Der Beweis findet sich in Weißbach (2006) □

Im einfachsten Fall kann die Funktion $\Psi(\cdot)$ als Glätter gewählt werden. Da diese alle Voraussetzungen an $\tilde{\Psi}(\cdot)$ erfüllen sollte, fallen dabei zusätzliche Konvergenzbeweise für den Glätter weg. Die Konstanten D_1 , D_2 und D_3 werden dann wie folgt vereinfacht:

$$D_1 := 2DM^3m^{-2}[\text{sup}(\text{Kr}) + L_{K_r}Mm^{-1}]$$

$$D_2 := DM^{3/2}\text{Var}(\text{Kr})m^{-1/2}$$

$$D_3 := 2M^4L_{K_r}L_{\psi}m^{-5} + 2\text{sup}(\text{Kr})L_{\psi}M^3m^{-4} + L_{\psi}m^{-1}$$

Eine andere Möglichkeit besteht in der Darstellung des Glätters durch eine deterministische lineare Funktion (2.7), wobei die Bandbreite $R_n(x)$ für

alle $x \in \mathbb{R}$ bei festen n konstant ist. Wegen $\tilde{\psi}_n(\cdot) = \tilde{c}$ und der Annahme (K4) gilt auch $\tilde{c} > 0$ auf $[A, B]$. Der Satz 2.1.1 kann weiter wesentlich vereinfacht werden.

Korollar 2.1.1. *Seien die Voraussetzungen (K1) bis (K6) erfüllt und es gelte $\tilde{\Psi}(\cdot) = \tilde{\Psi}_n(\cdot) = \tilde{c} \cdot \text{id}(\cdot) + \tilde{d}$.*

Dann existiert eine Konstante $0 \leq D_0 \leq \max\{D_1, D_2\}$, so dass

$$P \left\{ \limsup_{n \rightarrow \infty} \frac{\sup_{x \in [A, B]} |\psi_n(x) - \psi(x)|}{\sqrt{\log(n)/(np_n)} + p_n} = D_0 \right\} = 1 \quad (2.9)$$

wobei

$$D_1 := DM^{1/2} \tilde{c}^{1/2} \text{Var}(\text{Kr})$$

$$D_2 := L_\psi \tilde{c}^{-1}$$

gelten.

Beweis. Offensichtlich gelten folgende Aussagen bei beliebigen n :

- $\tilde{m} = \tilde{M} = \tilde{c}$
- $L_{\tilde{\psi}} = 0$
- $\tilde{D} = 0$

Die Behauptung folgt dann direkt aus dem Satz 2.1.1. □

Obwohl der Konvergenzsatz 2.1.1 ein allgemeines Ergebnis für eine breite Klasse von Kernschätzern liefert, basiert die Konvergenzgeschwindigkeit $\mathcal{O}(\sqrt{\log(n)/(np_n)} + p_n)$ im Wesentlichen auf der lokalen Konvergenz des Schätzers $\Psi_n(\cdot)$. Dabei wird die lokale Konvergenzrate $\mathcal{O}(\sqrt{\log(n)p_n/n})$ vorausgesetzt. Abgesehen von einigen Spezialfällen ist allerdings eine allgemeine Konstruktion des Schätzers $\Psi_n(\cdot)$ in der Literatur nicht beschrieben. Daher wird in dieser Arbeit angestrebt, diese Lücke zu schließen und eine generalisierte Konstruktion des Schätzers mit der gefragten lokalen Konvergenzrate

herzuleiten. Die generalisierte Darstellung lässt dann neue Kernschätzer mit den bekannten Konvergenzraten für die funktionalen Charakteristiken verschiedener stochastischer Prozesse herleiten.

2.1.3 Optimale Bandbreite

Zur praktischen Anwendung des Kernschätzers fehlen noch Kriterien zur Wahl einer optimalen Bandbreite. Offensichtlich hängt die Nächste-Nachbar-Bandbreite (2.6) von der Wahl der Folge (p_n) und des Glätters $\tilde{\Psi}(\cdot)$ ab. Da die möglichen Glätter bereits in Abschnitt 2.1.2 besprochen wurden, konzentrieren wir uns weiter auf (p_n) .

Zuerst stellen wir (p_n) als Produkt einer Konstante \tilde{p} und einer Nullfolge (\tilde{p}_n) dar. Obwohl die Optimalitätskriterien für (\tilde{p}_n) auf die Konvergenzgeschwindigkeit der Folge $(\sqrt{\log(n)/(np_n)} + p_n)$ zurückzuführen sind, untersuchen wir der Einfachheit halber in dieser Arbeit nur den Fall, wenn (p_n) bzw. (\tilde{p}_n) so schnell wie möglich gegen Null strebt. Die Geschwindigkeit dieser Folge ist allerdings durch die Bedingung $np_n/\log n \rightarrow \infty$ beschränkt, so dass (\tilde{p}_n) über $(\log n/n)$ liegen muss. Als eine mögliche Wahl von (\tilde{p}_n) passt dann $(\log n/n)^a$ mit $0 < a < 1$. Die Voraussetzung (K5) ist in diesem Fall erfüllt.

Wie aus Abschnitt 2.1 bekannt, bestand ursprünglich die Idee der Nächste-Nachbar-Methode in der Festlegung einer konstanten Anzahl der Sprünge von $\Psi_n(\cdot)$ innerhalb der Bandbreite. Durch diese Methode sollte die Bandbreite besser an die Daten angepasst werden. Seit der Einführung des Glätters $\tilde{\Psi}(\cdot)$ wird die Bandbreite allerdings nicht mehr durch die Anzahl der Sprünge von $\Psi_n(\cdot)$, sondern von $\tilde{\Psi}_n(\cdot)$ bestimmt. Diese Tatsache führt dazu, dass die gewählten Bandbreiten viel zu klein oder viel zu groß für $\Psi_n(\cdot)$ sein können. Zur Herleitung der Konstante \tilde{p} wird daher die kleinste Anzahl der Sprünge (k_{min}) von $\Psi_n(\cdot)$ auf $[t - R_n(t)/2, t + R_n(t)/2]$ für alle $t \in [A, B]$ festgelegt. Der Einfachheit halber fordern wir für $R_n(\cdot)$, dass zumindest k_{min} größte Zuwächse von $\Psi_n(\cdot)$ auf $[\Psi_n(t - R_n(t)/2), \Psi_n(t + R_n(t)/2)]$ ab einem Datenumfang

\tilde{n} passen. Dies impliziert die Bedingung:

$$k_{min} \cdot \sup_{s \in [A, B]} (d\Psi_n(s)) \leq |\Psi_n(t + R_n(t)/2) - \Psi_n(t - R_n(t)/2)| \quad (2.10)$$

für alle $n \geq \tilde{n} \in \mathbb{N}$ und alle $t \in [A, B]$, wobei $d\Psi_n(s) := \lim_{\Delta s \rightarrow 0} [\Psi_n(s + \Delta s) - \Psi_n(s)]$ der Zuwachs von $\Psi_n(\cdot)$ in s ist.

Im Falle $\tilde{\Psi}(\cdot) = \Psi(\cdot)$ verwandelt sich die Bedingung (2.10) in die folgende:

$$k_{min} \cdot \sup_{s \in [A, B]} (d\Psi_n(s)) \leq |\tilde{\Psi}_n(t + R_n(t)/2) - \tilde{\Psi}_n(t - R_n(t)/2)|,$$

was $p_n \geq k_{min} \cdot \sup_{s \in [A, B]} (d\Psi_n(s))$ impliziert. Die Konstante \tilde{p} wird dabei als $\tilde{p}^{(\tilde{n})} = k_{min} \cdot \sup_{s \in [A, B]} (d\Psi_{\tilde{n}}(s)) (\log \tilde{n}/\tilde{n})^{-a}$ für alle $n \geq \tilde{n} \in \mathbb{N}$ abgeschätzt, falls $\sup_{s \in [A, B]} (d\Psi_{\tilde{n}}(s)) (\log \tilde{n}/\tilde{n})^{-a}$ bei wachsenden \tilde{n} monoton fällt. D. h. die Konstante $\tilde{p}^{(\tilde{n})}$ wird so gewählt, dass die minimale Anzahl der Sprünge innerhalb der Bandbreite erst ab einem Datenumfang \tilde{n} gewährleistet wird. Dabei wird \tilde{n} nur einmal festgelegt und bleibt weiterhin konstant. Die Folge (p_n) ist somit bei $\tilde{\Psi}(\cdot) = \Psi(\cdot)$ für alle $n \geq \tilde{n} \in \mathbb{N}$ wie folgt spezifiziert:

$$p_n^{(\tilde{n})} = k_{min} \cdot \sup_{s \in [A, B]} (d\Psi_{\tilde{n}}(s)) (\log \tilde{n}/\tilde{n})^{-a} (\log n/n)^a. \quad (2.11)$$

Die Folge (2.11) erfüllt die Voraussetzung (K5) erst ab dem Datenumfang \tilde{n} , was durch das hochgestellte (\tilde{n}) gekennzeichnet wird und auch für die weiteren Bezeichnungen gilt. Den Grundgedanken der oben beschriebenen Vorgehensweise kann man wie folgt ausdrücken: Falls eine Stichprobe vom Umfang \tilde{n} vorliegt, so interessiert es uns mehr, ob der auf der Basis dieser Stichprobe gebaute Schätzer weiterhin gegen die theoretische Funktion konvergiert. Für die kleineren Datenumfänge ist die Erfüllung der Konvergenzeigenschaft und somit auch der Voraussetzung (K5) im Prinzip nicht nötig.

Falls die lineare Funktion (2.7) als Glätter gewählt wird, muss neben der Folge (p_n) auch das lineare Parameter \tilde{c} bestimmt werden. Bekanntlich ergibt

sich beim linearen Glätter die konstante Bandbreite $R_n = p_n/\tilde{c}$. Im Unterschied zur variablen Bandbreite fordern wir hier, dass nur beim Datenumfang m die Mindestanzahl der Sprünge k_{min} in der Bandbreite garantiert wird. Die Konstante \tilde{c} kann dann wie folgt abgeschätzt werden:

$$\tilde{c}^{(\tilde{n})} = \frac{p_{\tilde{n}}^{(\tilde{n})}}{\sup_{s \in [A, B]} R_{\tilde{n}}(s)},$$

wobei $\sup_{s \in [A, B]} R_{\tilde{n}}(s)$ die maximale variable Bandbreite auf $[A, B]$ beim Datenumfang \tilde{n} und k_{min} die Mindestanzahl der Sprünge ist. Die Bandbreite für alle $n > \tilde{n} \in \mathbb{N}$ wird folgenderweise berechnet:

$$R_n^{(\tilde{n})} = \frac{\sup_{s \in [A, B]} R_{\tilde{n}}(s)}{p_{\tilde{n}}^{(\tilde{n})}} \cdot p_n^{(\tilde{n})}.$$

Zu den Vorteilen der variablen Bandbreite gehört u. a. ihre Flexibilität, so dass ihre Länge ständig an die Daten angepasst wird. Bei den gleichen Bedingungen ist die konstante Bandbreite i. d. R. viel größer als die variable, wobei der Schätzer leicht überglättet werden kann.

Allgemein hängt die Wahl des Glätters sehr stark von der Form der zu schätzenden Funktion ab. Die Optimalitätskriterien für die Glätter behandeln wir allerdings in dieser Arbeit nicht.

2.2 Lokal konsistenter Schätzer

Wie im Abschnitt 2.1.2 erläutert, sind Schätzer mit der lokalen Konvergenzrate $\mathcal{O}(\sqrt{\log(n)p_n/n})$ nur für einige Spezialfälle bekannt, wie die empirische Verteilungsfunktion aus Stute (1982) oder die kumulierte Hazardrate bei rechtszensierten Beobachtungen aus Schäfer (1986b). Als Ziel dieser Arbeit setzen wir die Herleitung einer allgemeinen Konstruktion des Schätzers $\Psi_n(\cdot)$ mit der gefragten Konvergenzrate.

Sei $(\mathbf{S}_i)_{i=1, \dots, n}$ eine Folge von u.i.v. Zufallsvektoren $\mathbf{S}_i : \Omega \rightarrow \mathbb{R}^d$. Die

Funktion $G : \mathbb{R} \rightarrow \mathbb{R}_0^+$ sei stetig und die Funktion $G_n : \mathbb{R} \times (\mathbb{R}^d)^n \rightarrow \mathbb{R}_0^+$, $(x, s_1, \dots, s_n) \mapsto G_n(x)(s_1, \dots, s_n)$ sei symmetrisch für jedes feste $x \in \mathbb{R}$ und beliebige $s_1, \dots, s_n \in \mathbb{R}^d$. Mit der Symmetrie wird hier die Invarianz bzgl. der Permutationen der Argumente s_i gemeint. Weiter benutzen wir die vereinfachte Bezeichnung $G_n(x, \omega)$ für $G_n(x)(\mathbf{S}_1(\omega), \dots, \mathbf{S}_n(\omega))$.

Zusätzlich betrachten wir eine Abbildung $\Delta : \mathbb{R} \times \mathbb{R}^d \rightarrow \mathbb{R}_0^+$, $(x, s) \mapsto \Delta^x(s)$ mit der vereinfachten Bezeichnung $\Delta_i^x(\omega)$ für $\Delta^x(\mathbf{S}_i(\omega))$. Der Schätzer für $\Psi(\cdot)$ kann wie folgt konstruiert werden:

$$\Psi_n(x) := \frac{1}{n} \sum_{i=1}^n \frac{\mathbf{1}_{\{S_i^1 \leq x\}} \cdot \Delta_i^x}{G_n(S_i^1)}, \quad (2.12)$$

wobei S_i^1 das erste Element aus dem Vektor \mathbf{S}_i ist. Wie $\Psi(\cdot)$ und \mathbf{S}_i zusammenhängen, sehen wir in den Beispielen von Kapitel 3 und in der folgenden Voraussetzung (L6).

Die Formel (2.12) stellt eine Generalisierung des Schätzers für die kumulative Hazardrate aus Schäfer (1986a) und Andersen *et al.* (1993) dar. Die Beobachtungsvektoren $(X_i, \delta_i)_{i=1, \dots, n}$ (X_i und δ_i stehen jeweils für die Rechtszensierungszeit und das Zensierungsindikator) werden dabei durch die Zufallsvektoren $(\mathbf{S}_i)_{i=1, \dots, n}$ verallgemeinert. Die Einführung von Zufallsvektoren $(\mathbf{S}_i)_{i=1, \dots, n}$ ermöglicht es, beliebig viele stochastisch abhängige Beobachtungen eines Objekts in die Schätzung einzubeziehen. So kann z. B. der Schätzer für die kumulative Hazardrate durch die Linkstrunkierungszeit L_i erweitert werden, ohne die Konvergenz des Schätzers erneut zu beweisen (siehe Kapitel 3.3).

Die weiteren Verallgemeinerungen resultieren, indem die Zählprozesse $dN(\cdot)$ und $Y(\cdot)$ im Schätzer (3.6) jeweils durch Δ_i^x und $G_n(\cdot)$ ersetzt werden.

Im weiteren wird gezeigt, dass die obere Darstellung des Schätzers für $\Psi(\cdot)$ die Voraussetzung (K3) erfüllt und in der Konstruktion des Kernschätzers für $\psi(x) = d\Psi(x)/dx$ eingesetzt werden kann. Für die Konvergenz des

generalisierten Kernschätzers (2.8) muss bekanntlich $\Psi_n(\cdot)$ gegen $\Psi(\cdot)$ mit der lokalen Rate $\mathcal{O}(\sqrt{\log(n)p_n/n})$ konvergieren.

Die lokale Konsistenz von (2.12) braucht einige Annahmen für die zu schätzende Funktion, Beobachtungen und die Konvergenzrate von $G_n(\cdot)$ bzgl. $G(\cdot)$.

(L1) Angenommen, $[A, B]$ ist ein Intervall mit $A \in \mathbb{R}$, $B \in \mathbb{R}$ und $A < B$.

Die Konvergenz wird nur für Intervalle mit der Wahrscheinlichkeitsmasse großer Null gezeigt. Wie diese Intervalle gewählt werden, ist in Kapitel 3 beschrieben.

(L2) Wir nehmen an, dass die Funktion $\Psi : \mathbb{R} \rightarrow \mathbb{R}_0^+$ stetig, positiv und streng monoton wachsend auf $[A, B]$ ist.

Diese Voraussetzung berücksichtigt, dass der Kernschätzer nur für streng positive und stetige Funktionen existiert.

(L3) Wir nehmen an, dass Δ_i^x , $i = 1, \dots, n$ für alle $x \in [A, B]$ nach oben durch $\Delta_{max} < \infty$ beschränkt ist.

Damit die lokale Konvergenz erreichbar ist, müssen alle Summanden von (2.12) nach oben beschränkt sein.

(L4) Für eine Konstante $M := \sup_{x \in [A, B]} [G(x)]^{-1} < \infty$ wird angenommen, dass

$$[\mathbf{1}_{\{x \leq a\}} \Delta_i^a - \mathbf{1}_{\{x \leq b\}} \Delta_i^b] [G(x)]^{-1} < 2\Delta_{max} M, \quad i = 1, \dots, n$$

für beliebige $a, b \in [A, B]$ und $x \in \mathbb{R}$ gilt.

Bei dieser Voraussetzung wird gemeint, dass die angeführte Ungleichung nicht nur bei $x \in [A, B]$, sondern für alle $x \in \mathbb{R}$ gilt. Dies wird erreicht, indem $G(\cdot) := 1$ bei $\Delta_i^x \neq const$ oder $\Delta_i^x := const$ bei funktionswertigen $G(\cdot)$ gesetzt werden.

(L5) Zusätzlich wird angenommen, dass

$$[\mathbf{1}_{\{x \leq a\}} \Delta_i^a - \mathbf{1}_{\{x \leq b\}} \Delta_i^b][G_n(x) - G(x)] = 0, \quad i = 1, \dots, n$$

für alle $x \in [a, b] \not\subseteq [A, B]$ gilt.

Hier wird gemeint, dass entweder $G_n(\cdot) = G(\cdot)$ oder $\Delta_i^x = \text{const}$ auf jeden Fall gelten muss.

(L6) Weiterhin wird der Erwartungswert $E\left(\frac{\mathbf{1}_{\{S_i^1 \leq x\}} \cdot \Delta_i^x}{G(S_i^1)}\right) = \Psi(x)$ für jedes feste $x \in [A, B]$ vorausgesetzt.

Diese Voraussetzung legt den Zusammenhang zwischen Zufallsvektoren S_i und der Funktion $\Psi(\cdot)$ fest. Anders gesagt, müssen Δ_i^x und $G(\cdot)$ so gewählt werden, dass diese Momentbedingung erfüllt ist. Darüber hinaus muss $G(\cdot)$ einen konsistenten Schätzer besitzen, der die folgenden Eigenschaften besitzt.

Es kann durchaus passieren, dass nicht alle S_i 's beobachtbar sind. So können z. B. die Vektoren (L_i, X_i, δ_i) nur bei $L_i \leq X_i$ beobachtet werden, wobei L_i und X_i jeweils die Links- und Rechtszensierungszeiten sind und durch δ_i das Zensierungsindikator bezeichnet wird. Der Erwartungswert (L6) verwandelt sich im Falle der bedingten Beobachtungen in die bedingte Erwartung. In dem o.a. Beispiel wird der Schätzer (2.12) aus den Vektoren (L_i, X_i, δ_i) konstruiert. Der resultierende Erwartungswert (L6) muss dabei auf $L_i \leq X_i$ bedingt werden.

(L7) Wir nehmen an, dass eine Konstante $0 \leq D \leq D_G$ existiert, so dass

$$P \left\{ \limsup_{n \rightarrow \infty} \frac{\sup_{x \in [A, B]} |G_n(x) - G(x)|}{\sqrt{\log(n)/n}} = D \right\} = 1$$

Damit die Konvergenz in Kapitel 2.2 beweisbar ist, muss $G_n(x)$ gegen

$G(x)$ mit der Rate $\mathcal{O}(\sqrt{\log(n)/n})$ konvergieren. Dadurch wird auch der Zusammenhang zwischen $G_n(\cdot)$ und $G(\cdot)$ festgelegt.

(L8) Das Intervall $J := [a, b] \subseteq [A, B]$ and $a \leq b$ wird eingeführt.

Die lokale Konvergenz des Schätzers (2.12) wird gezeigt, indem die Hilfsfunktion $\Psi_n^*(x) := \frac{1}{n} \sum_{i=1}^n \frac{\mathbf{1}_{\{S_i^1 \leq x\}} \cdot \Delta_i^x}{G(S_i^1)}$ und das Hilfsmaß $\Psi_n^*(J) := \Psi_n^*(b) - \Psi_n^*(a)$ zusätzlich eingeführt werden.

Die ursprüngliche Differenz $|\Psi_n(J) - \Psi(J)|$ wird dann in zwei zerlegt, nämlich $|\Psi_n^*(J) - \Psi(J)|$ und $|\Psi_n(J) - \Psi_n^*(J)|$, und der weitere Beweis folgt in drei Schritten.

Zunächst werden Schranken für die erste Differenz gezeigt, indem die Ungleichung aus Abschnitt A.2 und Borel-Cantelli Lemma verwendet werden.

Lemma 2.2.1. *Die Voraussetzungen (L1) bis (L8) seien erfüllt.*

Für alle $p > 0$, $\varepsilon > 0$, $n \in \mathbb{N}_{>0}$ und jedes feste $J \subseteq [A, B]$ mit $\Psi(J) \leq p$ gilt dann

$$P(|\Psi_n^*(J) - \Psi(J)| > \varepsilon) < 2 \exp\left(\frac{-n\varepsilon^2}{2(2\Delta_{max}M + \Psi(B))(p + \varepsilon)}\right).$$

Beweis. Wir betrachten die folgende Differenz:

$$\Psi_n^*(J) - \Psi(J) = \frac{1}{n} \sum_{i=1}^n \left(\frac{\mathbf{1}_{\{S_i^1 \leq b\}} \cdot \Delta_i^b}{G(S_i^1)} - \frac{\mathbf{1}_{\{S_i^1 \leq a\}} \cdot \Delta_i^a}{G(S_i^1)} - \Psi(J) \right). \quad (2.13)$$

Für jedes feste $J \subseteq [A, B]$ wird die Differenz (2.13) als das arithmetische Mittel aus n unabhängigen beschränkten Zufallsvariablen ausgeschrieben, die identisch wie

$$R_J := \frac{\mathbf{1}_{\{S_i^1 \leq b\}} \cdot \Delta_i^b}{G(S_i^1)} - \frac{\mathbf{1}_{\{S_i^1 \leq a\}} \cdot \Delta_i^a}{G(S_i^1)} - \Psi(J)$$

verteilt sind.

Der Erwartungswert, die Varianz und Schranken für R_J werden bei festen $J \subseteq [A, B]$ mit $\Psi(J) \leq p$ berechnet.

Der Erwartungswert von R_J folgt aus Annahme (L6):

$$E(R_J) = E\left(\frac{\mathbf{1}_{\{S_i^1 \leq b\}} \cdot \Delta_i^b}{G(S_i^1)}\right) - E\left(\frac{\mathbf{1}_{\{S_i^1 \leq a\}} \cdot \Delta_i^a}{G(S_i^1)}\right) - \Psi(b) + \Psi(a) = 0. \quad (2.14)$$

Die Schranken für $|R_J|$ auf $[A, B]$ bekommen wir aus Annahme (L4) wie folgt:

$$\begin{aligned} |R_J| &= \left| \frac{\mathbf{1}_{\{S_i^1 \leq b\}} \cdot \Delta_i^b}{G(S_i^1)} - \frac{\mathbf{1}_{\{S_i^1 \leq a\}} \cdot \Delta_i^a}{G(S_i^1)} - \Psi(J) \right| \\ &< 2\Delta_{max}M + \Psi(B) - \Psi(A) < 2\Delta_{max}M + \Psi(B) =: g. \end{aligned} \quad (2.15)$$

Die Varianz von R_J ergibt sich aus dem Erwartungswert (2.14) und Schranken (2.15) folgendermaßen:

$$\begin{aligned} \sigma_J^2 := \text{Var}(R_J) &= E \left[\left(\frac{\mathbf{1}_{\{S_i^1 \leq b\}} \cdot \Delta_i^b}{G(S_i^1)} - \frac{\mathbf{1}_{\{S_i^1 \leq a\}} \cdot \Delta_i^a}{G(S_i^1)} - \Psi(J) \right)^2 \right] \\ &< 2\Delta_{max}M \cdot E \left(\frac{\mathbf{1}_{\{S_i^1 \leq b\}} \cdot \Delta_i^b}{G(S_i^1)} - \frac{\mathbf{1}_{\{S_i^1 \leq a\}} \cdot \Delta_i^a}{G(S_i^1)} \right) \\ &= 2\Delta_{max}M \cdot \Psi(J) < g \cdot p. \end{aligned} \quad (2.16)$$

Wegen (2.13), (2.14), (2.15) und (2.16) gilt dann

$$P(|\Psi_n^*(J) - \Psi(J)| > \varepsilon) < 2 \exp\left(\frac{-n\varepsilon^2}{2g(p + \varepsilon)}\right)$$

nach dem Satz A.2.1 für jedes feste Intervall $J \subseteq [A, B]$ mit $\Psi(J) \leq p$. \square

Lemma 2.2.2. *Die Voraussetzungen (L1) bis (L8) seien erfüllt und die Folge (p_n) erfüllt die Voraussetzung (K5).*

Dann existiert eine Konstante $C > \sqrt{2(2\Delta_{max}M + \Psi(B))}$, so dass

$$\sup_{J \subseteq [A, B], \Psi(J) \leq p_n} |\Psi_n^*(J) - \Psi(J)| \leq C \sqrt{\log(n)p_n/n} \text{ für } n \text{ groß genug f.s.}$$

Beweis. Mit p_n und $\varepsilon_n := C \sqrt{\log(n)p_n/n}$, $C > 0$, $n > 1$ an der Stelle von p und ε bekommen wir folgende rechte Seite der exponentiellen Abschätzung

aus Lemma 2.2.1

$$< 2 \cdot \exp\left(-\log(n) \frac{C^2}{2g} \frac{p_n}{(p_n + \varepsilon_n)}\right) = 2n^{-\frac{C^2}{2g} \frac{p_n}{(p_n + \varepsilon_n)}} =: A_n,$$

wobei $g = 2\Delta_{\max}M + \Psi(B)$ gilt.

Die Folge (A_n) ist dann summierbar ab einem großen $n < \infty$, wenn

$$\beta_n := \frac{C^2}{2g} \frac{p_n}{(p_n + \varepsilon_n)} > 1$$

für große n gilt.

Aus $\varepsilon_n/p_n = C\sqrt{\log(n)/(np_n)}$ und Annahme (K5) folgen $\varepsilon_n/p_n \rightarrow 0$ und $p_n/(p_n + \varepsilon_n) \rightarrow 1$ für große n . Die Bedingung $\beta_n > 1$ kann mit $C^2/2g > 1$ bzw. $C > \sqrt{2g}$ erreicht werden, wobei die Folge (A_n) summierbar ab einem großen $n < \infty$ ist.

Aus der Summierbarkeit von (A_n) folgt dann:

$$\exists C > \sqrt{2g} \exists m < \infty, m \in \mathbb{N} : \sum_{n=m}^{\infty} P(|\Psi_n(J) - \Psi(J)| > \varepsilon_n) < \sum_{n=m}^{\infty} A_n < \infty$$

für beliebige $J \in [A, B]$ mit $\Psi(J) \leq p_n$ und offensichtlich auch:

$$\exists m < \infty, m \in \mathbb{N} : \sum_{n=1}^m P(|\Psi_n(J) - \Psi(J)| > \varepsilon_n) \leq m < \infty.$$

Wegen Summierbarkeit von $P(|\Psi_n(J) - \Psi(J)| > \varepsilon_n)$ gilt bei $C > \sqrt{2g}$ nach Borel-Cantelli Lemma:

$$P\left(\limsup_{n \rightarrow \infty} |\Psi_n(J) - \Psi(J)| > \varepsilon_n\right) = 0,$$

d.h. $|\Psi_n(J) - \Psi(J)|$ überschreitet ε_n für fast alle n nicht, wobei

$$|\Psi_n(J) - \Psi(J)| \leq C\sqrt{\log(n)p_n/n} \tag{2.17}$$

für alle $J \in [A, B]$ mit $\Psi(J) \leq p_n$ bei großen n fast sicher gilt. \square

Im zweiten Schritt werden Schranken für die zweite Differenz gezeigt, indem die Ungleichung aus Lemma 2.2.1 und die Annahme (L7) zum Einsatz kommen.

Lemma 2.2.3. *Die Voraussetzungen (L1) bis (L8) seien erfüllt und die Folge (p_n) erfüllt die Voraussetzung (K5).*

Dann existiert eine Konstante $C > D_G \cdot M$, so dass

$$\sup_{J \subseteq [A, B], \Psi(J) \leq p_n} |\Psi_n(J) - \Psi_n^*(J)| \leq C \cdot p_n \sqrt{\log(n)/n} \text{ für } n \text{ groß genug f.s.}$$

Beweis. Sei $C'_1 > D_G$. Nach Annahme (L7) und Lemma A.1.1 gelten

$$G_n(x) - G(x) \leq |G_n(x) - G(x)| \leq \sup_{x \in [A, B]} |G_n(x) - G(x)| \leq C'_1 \cdot \sqrt{\log(n)/n}$$

und

$$G_n(x) \geq G(x) - C'_1 \sqrt{\log(n)/n} \geq \inf_{t \in [A, B]} G(t) - C'_1 \cdot \sqrt{\log(n)/n}.$$

fast sicher für große n und alle $x \in [A, B]$.

Wegen Annahme (L7) und $\sqrt{\log(n)/n} \rightarrow 0$ sind die Ungleichungen

$$\inf_{t \in [A, B]} G(t) - C'_1 \cdot \sqrt{\log(n)/n} > 0$$

und

$$\frac{1}{G_n(x)} \leq \frac{1}{\inf_{t \in [A, B]} G(t) - C'_1 \cdot \sqrt{\log(n)/n}}$$

auf $[A, B]$ für große n erfüllt. Es gilt für $x \in [A, B]$ und große n fast sicher

$$\frac{|G_n(x) - G(x)|}{G_n(x)} \leq \frac{C'_1 \cdot \sqrt{\log(n)/n}}{\inf_{t \in [A, B]} G(t) - C'_1 \cdot \sqrt{\log(n)/n}}.$$

Sei $C'_2 > \sqrt{2 \cdot (2\Delta_{max}M + \Psi(B))}$. Es gilt dann nach Lemma 2.2.2 für $J \subseteq [A, B]$ mit $\Psi(J) \leq p_n$ und für große n fast sicher

$$\Psi_n^*(J) - \Psi(J) \leq \sup_{J \subseteq [A, B], \Psi(J) \leq p_n} |\Psi_n^*(J) - \Psi(J)| \leq C'_2 \sqrt{\log(n)p_n/n}$$

und folglich

$$\Psi_n^*(J) \leq \Psi(J) + C'_2 \sqrt{\log(n)p_n/n} \leq p_n + C'_2 \sqrt{\log(n)p_n/n}.$$

Aus der Annahme (L5) bekommen wir folgende Abschätzung der Differenz $|\Psi_n(J) - \Psi_n^*(J)|$ für jedes $J \subseteq [A, B]$ mit $\Psi(J) \leq p_n$ und bei großen n :

$$\begin{aligned} |\Psi_n(J) - \Psi_n^*(J)| &= \frac{1}{n} \sum_{i=1}^n \left| \frac{1}{G_n(S_i^1)} - \frac{1}{G(S_i^1)} \right| (\mathbf{1}_{\{S_i^1 \leq b\}} \cdot \Delta_i^b - \mathbf{1}_{\{S_i^1 \leq a\}} \cdot \Delta_i^a) \\ &= \frac{1}{n} \sum_{i=1}^n \left| \frac{G(S_i^1) - G_n(S_i^1)}{G_n(S_i^1)} \right| \frac{\mathbf{1}_{\{S_i^1 \leq b\}} \cdot \Delta_i^b - \mathbf{1}_{\{S_i^1 \leq a\}} \cdot \Delta_i^a}{G(S_i^1)} \\ &\leq \frac{C'_1 \sqrt{\log(n)/n} \cdot \Psi_n^*(J)}{\inf_{t \in [A, B]} G(t) - C'_1 \sqrt{\log(n)/n}} \leq \frac{C'_1 \sqrt{\log(n)/n} \cdot (p_n + C'_2 \sqrt{\log(n)p_n/n})}{\inf_{t \in [A, B]} G(t) - C'_1 \sqrt{\log(n)/n}}. \end{aligned}$$

Es gilt $p_n + C'_2 \sqrt{\log(n)p_n/n} = p_n [1 + C'_2 \sqrt{\log(n)/(p_n n)}]$, wobei der Summand $C'_2 \sqrt{\log(n)/(p_n n)}$ wegen Annahme (K5) bei großen n vernachlässigt werden kann. Auch $\sqrt{\log(n)/n}$ im Nenner kann bei großen n vernachlässigt werden. Wir bekommen also für alle $J \subseteq [A, B]$ mit $\Psi(J) \leq p_n$ und für große n fast sichere Ungleichung

$$|\Psi_n(J) - \Psi_n^*(J)| \leq C'_1 \cdot M \cdot p_n \sqrt{\log(n)/n}$$

Für ein $C > D_G \cdot M$ gilt dann

$$\sup_{J \subseteq [A, B], \Psi(J) \leq p_n} |\Psi_n(J) - \Psi_n^*(J)| \leq C \cdot p_n \sqrt{\log(n)/n}$$

fast sicher für große n . □

Zum Schluss werden Schranken für die gesamte Differenz $|\Psi_n(J) - \Psi(J)|$ auf der Basis von Lemmata 2.2.2 und 2.2.3 konstruiert.

Korollar 2.2.1. *Die Voraussetzungen (L1) bis (L8) seien erfüllt und die Folge (p_n) erfüllt die Voraussetzung (K5).*

Dann existiert eine Konstante

$$D \leq 2(\sqrt{2 \cdot (2\Delta_{\max} M + \Psi(B))} + D_G \cdot M)$$

so dass

$$P \left\{ \limsup_{n \rightarrow \infty} \frac{\sup_{J \subseteq [A, B], \Psi(J) \leq p_n} |\Psi_n(J) - \Psi(J)|}{\sqrt{\log(n)p_n/n}} = D \right\} = 1.$$

Beweis. Zur Behandlung der lokalen Abweichung

$$\sup_{J \subseteq [A, B], \Psi(J) \leq p_n} |\Psi_n(J) - \Psi(J)|$$

schieben wir das Hilfsmaß $\Psi_n^*(J)$ ein:

$$\begin{aligned} & \sup_{J \subseteq [A, B], \Psi(J) \leq p_n} |\Psi_n(J) - \Psi(J)| \\ & \leq \sup_{J \subseteq [A, B], \Psi(J) \leq p_n} |\Psi_n(J) - \Psi_n^*(J)| + \sup_{J \subseteq [A, B], \Psi(J) \leq p_n} |\Psi_n^*(J) - \Psi(J)| \end{aligned}$$

Wegen Annahme (K5) strebt $\frac{p_n \sqrt{\log(n)/n}}{\sqrt{\log(n)p_n/n}} = \sqrt{p_n}$ gegen Null, d. h. $p_n \sqrt{\log(n)/n} \leq \sqrt{\log(n)p_n/n}$ gilt bei großen n .

Dann folgt aus Lemmata 2.2.2 und 2.2.3 die Existenz einer Konstante

$$C > \sqrt{2 \cdot (2\Delta_{max}M + \Psi(B))} + D_G \cdot M,$$

so dass

$$\begin{aligned} \sup_{J \subseteq [A, B], \Psi(J) \leq p_n} |\Psi_n(J) - \Psi(J)| & \leq C(\sqrt{\log(n)p_n/n} + p_n \sqrt{\log(n)/n}) \\ & \leq 2C\sqrt{\log(n)p_n/n} \end{aligned}$$

fast sicher für große n gilt. Wegen der Symmetrie von $\Psi_n(J)$ folgt die Behauptung direkt aus Lemma A.1.1. \square

2.3 Global konsistenter Schätzer

Aus der Annahme (L7) folgt die Notwendigkeit, einen Schätzer $G_n(\cdot)$ für $G(\cdot)$ mit der globalen Konvergenzrate $\mathcal{O}(\sqrt{\log(n)/n})$ zu konstruieren.

Sei $(\mathbf{S}_i)_{i=1, \dots, n}$ eine Folge von u.i.v. Zufallsvektoren aus Kapitel 2.2. Wir definieren die Abbildung $H : \mathbb{R}^d \rightarrow \mathcal{P}(\mathbb{R})$ mit der vereinfachten Darstellung

$H_i(\omega)$ für $H(\mathbf{S}_i(\omega))$.

Eine mögliche Konstruktion von $G_n(\cdot)$ kann aus der Generalisierung der empirischen Verteilungsfunktion analog zum Schätzer (2.12) wie folgt hergeleitet werden:

$$G_n(x) := \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{x \in H_i\}} \cdot \Gamma_i^x, \quad (2.18)$$

wobei die Definition von Γ_i^x identisch zu Δ_i^x aus Abschnitt 2.2 ist. Wie $G(\cdot)$ und \mathbf{S}_i zusammenhängen, sehen wir in den Beispielen von Kapitel 3 und in der folgenden Voraussetzung (G4).

Ähnlich zum lokal konsistenten Schätzer (2.12) braucht die globale Konvergenz von (2.18) einige Annahmen an die zu schätzende Funktion und an die beobachtbaren Variablen.

(G1) Angenommen, $[A, B]$ ist ein Intervall mit $A \in \mathbb{R}$, $B \in \mathbb{R}$ und $A < B$.

Analog zu (L1).

(G2) Wir nehmen an, dass Γ_i^x , $i = 1, \dots, n$ für alle $x \in [A, B]$ von rechts durch $\Gamma_{max} < \infty$ beschränkt ist.

Analog zu (L3).

(G3) Die Funktion $G : \mathbb{R} \rightarrow \mathbb{R}_0^+$ wird als nichtnegativ mit $G(x) \leq G_{max}$ auf $[A, B]$ angenommen.

Vor allem wird hier eine Konstruktion des Schätzers für die positive Funktion $G(\cdot)$ aus Abschnitt 2.2 gezeigt. Das heißt aber nicht, dass der Schätzer (2.18) nur im Zusammenhang mit dem Schätzer (2.12) verwendet werden kann. So wird z. B. die Konvergenz der empirischen Verteilungsfunktion in Abschnitt 3.2 durch den Konvergenzbeweis für (2.18) gezeigt.

(G4) Weiterhin wird der Erwartungswert $E(\mathbf{1}_{\{x \in H_i\}} \cdot \Gamma_i^x) = G(x)$ für jedes feste $x \in [A, B]$ angenommen. Analog zu (L6) legt diese Voraussetzung

den Zusammenhang zwischen Zufallsvektoren S_i und der Funktion $G(\cdot)$ fest. Vor allem betrifft es den Aufbau von Γ_i^x .

Falls nicht alle S_i 's beobachtbar sind, so gelten die Überlegungen aus der Voraussetzung (L6). In diesem Fall wird die Erwartung (G4) zur bedingten Erwartung.

Die Voraussetzungen (G1) bis (G4) sind dabei viel leichter zu überprüfen als die Annahme (L7).

Der folgende Beweis ist ähnlich zu diesem aus Abschnitt 2.2 und verläuft in zwei Schritten.

Im ersten Schritt wird die Ungleichung aus Abschnitt A.2 eingesetzt.

Lemma 2.3.1. *Die Voraussetzungen (G1) bis (G4) seien erfüllt.*

Für alle $\varepsilon \in \mathbb{R}_{>0}$, $n \in \mathbb{N}_{>0}$ und jedes feste $x \in [A, B]$ gilt dann

$$P(|G_n(x) - G(x)| > \varepsilon) < 2 \exp\left(\frac{-n\varepsilon^2}{2(G_{max} + \Gamma_{max})(G_{max} + \varepsilon)}\right).$$

Beweis. Der Beweis verläuft analog zum Beweis von Lemma 2.2.1. Für jedes feste $x \in [A, B]$ ist

$$G_n(x) - G(x) = \frac{1}{n} \sum_{i=1}^n (\mathbf{1}_{\{x \in H_i\}} \cdot \Gamma_i^x - G(x)) \quad (2.19)$$

das arithmetische Mittel aus n unabhängigen beschränkten Zufallsvariablen, die identisch wie

$$R_x := \mathbf{1}_{\{x \in H_i\}} \cdot \Gamma_i^x - G(x)$$

verteilt sind.

Wie üblich werden der Erwartungswert, die Varianz und Schranken für R_x bei festen $x \in [A, B]$ berechnet.

Der Erwartungswert von R_x folgt aus Annahme (G4).

$$\begin{aligned} E(R_x) &= E(\mathbf{1}_{\{x \in H_i\}} \cdot \Gamma_i^x - G(x)) \\ &= E(\mathbf{1}_{\{x \in H_i\}} \cdot \Gamma_i^x) - G(x) = 0 \end{aligned} \quad (2.20)$$

Die Schranken für $|R_x|$ auf $[A, B]$ werden aus Annahme (G3) wie folgt berechnet:

$$|R_x| = |\mathbf{1}_{\{x \in H_i\}} \cdot \Gamma_i^x - G(x)| < \Gamma_{max} + G_{max} =: g \quad (2.21)$$

Dann gilt für die Varianz von R_x folgende Ungleichung:

$$\begin{aligned} \sigma_x^2 &:= \text{Var}(R_x) = E \left[(\mathbf{1}_{\{x \in H_i\}} \cdot \Gamma_i^x - G(x))^2 \right] \\ &= E (\mathbf{1}_{\{x \in H_i\}} \cdot (\Gamma_i^x)^2) - (G(x))^2 \leq E (\mathbf{1}_{\{x \in H_i\}} \cdot (\Gamma_i^x)^2) \\ &\leq \Gamma_{max} \cdot E (\mathbf{1}_{\{x \in H_i\}} \cdot \Gamma_i^x) \\ &= \Gamma_{max} \cdot G(x) < g \cdot G_{max} \end{aligned} \quad (2.22)$$

Wegen (2.19), (2.20), (2.21) und (2.22) gilt

$$P(|G_n(x) - G(x)| > \varepsilon) < 2 \exp \left(\frac{-n\varepsilon^2}{2(\sigma^2 + g\varepsilon/3)} \right) < 2 \exp \left(\frac{-n\varepsilon^2}{2g(G_{max} + \varepsilon)} \right)$$

nach dem Satz A.2.1 für jedes feste $x \in [A, B]$. \square

Das Borel-Cantelli Lemma wird verwendet, um die resultierende Konvergenz zu zeigen.

Korollar 2.3.1. *Die Voraussetzungen (G1) bis (G4) seien erfüllt.*

Dann existiert eine Konstante

$$D \leq \sqrt{2G_{max} \cdot (\Gamma_{max} + G_{max})},$$

so dass

$$P \left\{ \limsup_{n \rightarrow \infty} \frac{\sup_{x \in [A, B]} |G_n(x) - G(x)|}{\sqrt{\log(n)/n}} = D \right\} = 1.$$

Beweis. Wie im Beweis zu Lemma 2.2.2 bekommen wir mit $\varepsilon_n := C' \sqrt{G_{max} \cdot \sqrt{\log(n)/n}}$, $C' > 0$, $n > 1$ an der Stelle von ε folgende rechte Seite der exponentiellen Abschätzung aus Lemma 2.3.1

$$< 2 \cdot \exp \left(-\log(n) \frac{(C')^2}{2g} \frac{G_{max}}{(G_{max} + \varepsilon_n)} \right) = 2n^{-\frac{(C')^2}{2g} \frac{G_{max}}{(G_{max} + \varepsilon_n)}} =: A_n,$$

wobei $g = \Gamma_{max} + G_{max}$ gilt.

Die Folge (A_n) ist dann summierbar ab einem großen $n < \infty$, wenn

$$\beta_n := \frac{(C')^2}{2g} \frac{G_{max}}{(G_{max} + \varepsilon_n)} > 1$$

für große n gilt.

Wegen $G_{max}/(G_{max} + \varepsilon_n) < 1$ kann $\beta_n > 1$ nur bei $(C')^2/2g > 1$ (Wahl von C' !) gelten. Aus $\log(n)/n \rightarrow 0$ folgt dabei $G_{max}/(G_{max} + \varepsilon_n) \rightarrow 1$ und $\beta_n > 1$ gilt bei großen n . Die Folge (A_n) ist somit summierbar ab einem großen $n < \infty$ nur bei $C' > \sqrt{2g}$.

Ähnlich wie bei Lemma 2.2.2 wird durch Borel-Cantelli Lemma gezeigt, dass

$$\sup_{x \in [A, B]} |G_n(x) - G(x)| \leq C \sqrt{\log(n)/n}$$

für $C > \sqrt{2gG_{max}}$ bei großen n fast sicher gilt. Wegen der Symmetrie von $G_n(\cdot)$ folgt die Behauptung direkt aus Lemma A.1.1. \square

Kapitel 3

Nichtparametrische Schätzung aus stochastischen Prozessen

In diesem Kapitel zeigen wir, wie die generalisierte Konstruktion des Kernschätzers auf die bekannten stochastischen Prozesse angewendet werden kann. Beginnend mit dem Kerndichteschätzer wird weiter der Fall der linkstrunkierten und rechtszensierten Daten betrachtet. Auf der Basis der generalisierten Darstellung wird dabei die Konvergenz des Kernschätzers für die Hazardrate gezeigt. Darüber hinaus wird gezeigt, wie der Ansatz der Kernschätzung für die Migrationsintensitäten bei einem Markov-Prozess eingesetzt werden kann.

3.1 Empirische Wahrscheinlichkeitsfunktion

In diesem Abschnitt wird der generalisierte Ansatz zur konsistenten Schätzung von zeitabhängigen positiven Funktionen zwischen 0 und 1 vorgestellt. Die Konvergenzrate des resultierenden Schätzers wird in weiteren Kapiteln zur Vereinfachung des Beweises eingesetzt.

Sei $(\mathbf{S}_i)_{i=1,\dots,n}$ eine Folge von u.i.v. Zufallsvektoren $\mathbf{S} : \Omega \rightarrow \mathbb{R}^d$ aus Abschnitt 2.2, wobei $S_i^d : \Omega \rightarrow \{s_1, \dots, s_k\} \subset \mathbb{R}$ für $i = 1, \dots, n$ angenommen

wird. S_i^d ist die die d -te Zufallsvariable aus \mathbf{S}_i . Wir betrachten zusätzlich eine Zufallsmenge H_i wie in Abschnitt 2.3.

Wir bezeichnen $G(x) := P(x \in H_i, S_i^d = s_{const})$ als Wahrscheinlichkeitsfunktion, die gemäß der Annahme (G3) durch $G_{max} := 1$ beschränkt ist. Dabei kommt der Index i in der Definition von $G(\cdot)$ vor, weil die Zufallsmenge H_i und nicht die Funktion $G(\cdot)$ vom Zufallsvektor \mathbf{S}_i abhängt und alle \mathbf{S}_i 's u.i.d. verteilt sind. Von daher sind die Wahrscheinlichkeiten $P(x \in H_i, S_i^d = s_{const})$ für jedes $i = 1, \dots, n$ gleich.

Seien $G(x)$ eine nichtnegative Funktion auf $[A, B]$ und

$$G_n(x) := \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{\{x \in H_j, S_j^d = s_{const}\}}$$

ein Schätzer für $G(x)$, wobei $s_{const} \in \{s_1, \dots, s_k\} \subset \mathbb{R}$ eine Konstante ist.

Der Term $S_i^d = s_{const}$ wird in der Wahrscheinlichkeitsfunktion gebraucht, um zusätzliche von x unabhängige Bedingungen zu modellieren. So wird z. B. die Bedingung $\delta_i = 1$ bei der Verweildaueranalyse in Kapitel 3.3 durch diesen Term ausgedrückt. Falls eine solche Bedingung nicht nötig ist, kann einfach eine Zufallsvariable $S_j^d : \Omega \rightarrow \{s_{const} \in \mathbb{R}\}$ zusammen mit der Nebenbedingung $S_j^d = s_{const}$ definiert werden. S_j^d besitzt dabei eine einelementige Wertemenge.

Lemma 3.1.1. *Seien $(\mathbf{1}_{\{x \in H_i\}}, S_i^d)_{i=1, \dots, n}$ eine Folge unabhängiger und identisch verteilter Paare von Zufallsvariablen für jedes feste $x \in [A, B]$.*

Dann existiert eine Konstante $D \leq 2$, so dass

$$P \left\{ \limsup_{n \rightarrow \infty} \frac{\sup_{x \in [A, B]} |G_n(x) - G(x)|}{\sqrt{\log(n)/n}} = D \right\} = 1,$$

wobei $[A, B] \subseteq \mathbb{R}$ ein beliebiges Intervall ist.

Beweis. Zunächst werden die Zufallsvariablen $\Gamma_i^x = \mathbf{1}_{\{S_i^d = s_{const}\}} \leq 1 =: \Gamma_{max}$, $i = 1, \dots, n$ für jedes feste $x \in [A, B]$ eingeführt. Der Schätzer $G_n(x)$

kann dann als

$$G_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{x \in H_i\}} \cdot \Gamma_i^x$$

ausgedrückt werden und folgende Erwartungswerte ergeben sich für feste x :

$$\begin{aligned} E(\mathbf{1}_{\{x \in H_i\}} \cdot \Gamma_i^x) &= E(\mathbf{1}_{\{x \in H_i\}} \cdot \mathbf{1}_{\{S_i^d = s_{const}\}}) \\ &= E(\mathbf{1}_{\{x \in H_i, S_i^d = s_{const}\}}) = P(x \in H_i, S_i^d = s_{const}) = G(x). \end{aligned}$$

Offensichtlich sind alle Voraussetzungen (G1) bis (G4) erfüllt und die Behauptung folgt aus Korollar 2.3.1. \square

Wenn die S_i 's nur unter einer Bedingung beobachtet werden können, gilt Lemma 3.1.1 auch für die bedingte Wahrscheinlichkeitsfunktion unter Verwendung der bedingten Erwartung im Beweis.

3.2 Empirische Dichtefunktion

Der Kerndichteschätzer ist in der Literatur gut bekannt und stellt, historisch gesehen, den ersten Meilenstein für die weitere Entwicklung der Kernschätzung dar. Abgesehen von vielen Veröffentlichungen zu dem Thema zeigen wir in diesem Abschnitt, wie der generalisierte Ansatz zur Schätzung von Dichtefunktionen eingesetzt werden kann.

Sei S_1, \dots, S_n eine Folge von u.i.v. Zufallsvariablen mit auf $[A, B]$ streng positiven und Lipschitz-stetigen Verteilungs- und Dichtefunktionen $F(x)$ bzw. $f(x)$. Zuerst untersuchen wir die Konvergenzeigenschaften der empirischen Verteilungsfunktion $F_n(x) := \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{S_i \leq x\}}$.

Satz 3.2.1. *Seien $[A, B] \subseteq \mathbb{R}$ ein beliebiges Intervall und S_1, \dots, S_n eine Folge von unabhängigen und identisch verteilten Zufallsvariablen mit der Verteilungsfunktion $F(x) := P(S_i \leq x)$.*

Dann existiert eine Konstante $D \leq 2$, so dass

$$P \left\{ \limsup_{n \rightarrow \infty} \frac{\sup_{x \in [A, B]} |F_n(x) - F(x)|}{\sqrt{\log(n)/n}} = D \right\} = 1.$$

Beweis. Wir definieren u.i.v. Zufallsvektoren $(S_i, S_C)_{i=1, \dots, n}$, wobei $S_C : \Omega \rightarrow \{s_{const} \in \mathbb{R}\}$ eine Zufallsvariable mit der einelementigen Wertemenge ist. Zusätzlich werden die Abbildung $H(x_1, x_2) := \{y : x_1 \leq y\}$ und die Zufallsmenge $H_i = \{y \in \mathbb{R} : S_i \leq y\}$ eingeführt. Analog zu Abschnitt 3.1 definieren wir eine Wahrscheinlichkeitsfunktion $G(x) := P(x \in H_i, S_C = s_{const})$ und ihren Schätzer $G_n(x) := \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{\{x \in H_j, S_C = s_{const}\}}$.

Die Behauptung folgt aus $F(x) = G(x)$ und $F_n(x) = G_n(x)$ und Lemma 3.1.1. \square

Neben der globalen Konvergenz von $F_n(x)$ kann auch die lokale gezeigt werden, d. h. die Konvergenz des empirischen Maßes $F_n(J) := F_n(b) - F_n(a)$ gegen das theoretische $F(J) := F(b) - F(a)$ für die Intervalle $J := [a, b] \subseteq [A, B] \subseteq \mathbb{R}$.

Satz 3.2.2. Seien $[A, B] \subseteq \mathbb{R}$ ein beliebiges Intervall und S_1, \dots, S_n eine Folge von u.i.v. Zufallsvariablen mit der Lipschitz-stetigen, positiven und streng monotonen auf $[A, B]$ Verteilungsfunktion $F(x) := P(S_i \leq x)$. Die Folge (p_n) erfülle die Voraussetzung (K5).

Dann existiert eine Konstante $D \leq 2\sqrt{6}$, so dass

$$P \left\{ \limsup_{n \rightarrow \infty} \frac{\sup_{J \subseteq [A, B], F(J) \leq p_n} |F_n(J) - F(J)|}{\sqrt{\log(n)p_n/n}} = D \right\} = 1.$$

Beweis. Folgende Definitionen werden eingeführt:

- Δ_i^x werden als Konstanten $\Delta_i^x := 1 = \Delta_{max}$, $i = 1, \dots, n$ für jedes feste $x \in [A, B]$ definiert.
- Funktionen $\Psi(x) := F(x)$, $G(x) := 1$ und $G_n(x) := 1$ werden definiert.

- Eine Konstante $D_G := 0$ wird eingeführt.

Nun werden die Annahmen (L1) bis (L8) überprüft.

Für alle $n > 0$, $x \in [A, B]$ und $C_G > D_G$ gilt $|G_n(x) - G(x)| = 0 \leq C_G \sqrt{\log(n)/n}$ und Annahme (L7) wird erfüllt.

Für jedes feste $x \in [A, B]$ bekommen wir dann Folgendes

$$E \left(\frac{\mathbf{1}_{\{S_i \leq x\}} \cdot \Delta_i^x}{G(S_i)} \right) = E(\mathbf{1}_{\{S_i \leq x\}}) = \Psi(x).$$

Die Annahme (L5) ist erfüllt, weil Δ_i^x , $G(\cdot)$ und $G_n(\cdot)$ konstant sind. Wegen konstanten Δ_i^x bekommen wir ebenso die Annahme (L4) für beliebige $a, b \in [A, B]$ und $x \in \mathbb{R}$ wie folgt:

$$[\mathbf{1}_{\{x \leq a\}} \Delta_i^a - \mathbf{1}_{\{x \leq b\}} \Delta_i^b][G(x)]^{-1} = \mathbf{1}_{\{a \leq x \leq b\}} \Delta_i^b < 2\Delta_{max} M = 2\Delta_{max}.$$

Die Voraussetzungen (L1) bis (L8) sind erfüllt und die Behauptung folgt aus Korollar 2.2.1. \square

Auf $[A, B]$ nehmen wir nun an, dass die Verteilungsfunktion $F(\cdot)$ Lipschitz-stetig ist und eine positive und Lipschitz-stetige Dichtefunktion $f_{min} \leq f(\cdot) \leq f_{max}$ mit Konstante L_f besitzt. Der Kernschätzer für $f(\cdot)$ kann aus der generalisierten Form (2.8) wie folgt hergeleitet werden:

$$f_n(x) := \sum_{i=1}^n \frac{1}{R_n(S_i)} \text{Kr} \left(\frac{x - S_i}{R_n(S_i)} \right), \quad (3.1)$$

wobei $R_n(\cdot)$ die Nächste-Nachbar-Bandbreite ist.

Da der Schätzer $F_n(\cdot)$ eine Konvergenzrate von $O(\sqrt{\log(n)p_n/n})$ aufweist und die Annahmen (K1)-(K4) für $F(\cdot)$ und $f(\cdot)$ erfüllt sind, kann die Konvergenz von $f_n(\cdot)$ für verschiedene Glätter $\tilde{\Psi}(\cdot)$ gezeigt werden. Es existiert dann eine Konstante $D_0 < \infty$, so dass

$$P \left\{ \limsup_{n \rightarrow \infty} \frac{\sup_{x \in [A, B]} |f_n(x) - f(x)|}{\sqrt{\log(n)/(np_n)} + p_n} = D_0 \right\} = 1.$$

Nach dem Korollar 2.1.1 gilt $D_0 \leq \max\{D_1, D_2\}$ für den linearen Glätter (2.7) und eine konstante Bandbreite $R_n = p_n/\tilde{c}$, $\tilde{c} > 0$, wobei

$$\begin{aligned} D_1 &= 24^{1/2} \tilde{c}^{1/2} \text{Var}(\text{Kr}) f_{max}^{1/2}, \\ D_2 &= L_f \tilde{c}^{-1} \end{aligned}$$

und die Kernfunktion $\text{Kr}(\cdot)$ die Voraussetzung (K6) erfüllt.

Für die variable Bandbreite kann $\tilde{\Psi}(\cdot) = F(\cdot)$ gewählt werden, da die Voraussetzungen für den Glätter identisch zu diesen von $F(\cdot)$ sind. Es gilt dann $D_0 \leq \max\{D_1 + D_2, D_3\}$, wobei

$$\begin{aligned} D_1 &= 96^{1/2} f_{min}^{-2} f_{max}^3 [\text{sup}(\text{Kr}) + L_{Kr} f_{min}^{-1} f_{max}], \\ D_2 &= 24^{1/2} \text{Var}(\text{Kr}) f_{min}^{-1/2} f_{max}^{3/2}, \\ D_3 &= 2L_{Kr} L_f f_{min}^{-5} f_{max}^4 + 2 \text{sup}(\text{Kr}) L_f f_{min}^{-4} f_{max}^3 + L_f f_{min}^{-1} \end{aligned}$$

und die Kernfunktion $\text{Kr}(\cdot)$ die Voraussetzungen (K6) und (K7) erfüllt.

3.3 Überlebensprozess

3.3.1 Modell

Die Überlebensanalyse beschäftigt sich mit der Untersuchung der Lebensdauer von gleichartigen Objekten, bei denen der Übergang in einen bestimmten Anfangszustand festgestellt wurde. Bei jedem Objekt wechselt dann der Anfangszustand nach einer zufälligen Zeitperiode T zum Endzustand.

Wegen der begrenzten Beobachtungsdauer ist allerdings die vollständige Lebensdauer nicht bei allen Objekten zu bestimmen. So können manche Objekte noch vor dem Wechsel zum Endzustand nicht beobachtbar werden. Es ist ebenfalls nicht bekannt, wie viele Objekte vor dem Anfang der Beobachtung zum Endzustand gewechselt haben.

Den Informationsmangel zu beheben, helfen zusätzliche Informationen,

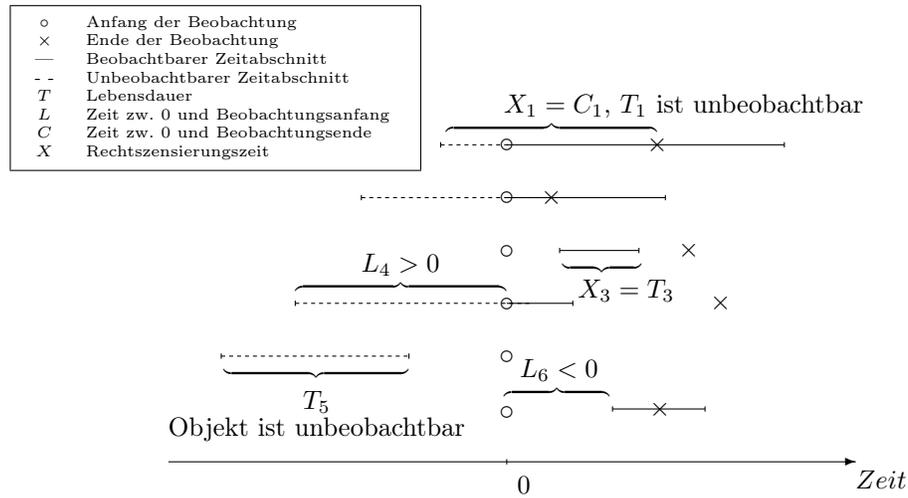


Abbildung 3.1: Überlebensprozess und zeitliche Rahmen der Beobachtung in realer Zeit.

wie Zeitabschnitte zwischen dem Entstehen des Objektes und dem Anfang der Beobachtung L (auch Linkstrunkierungszeit genannt) bzw. dem Ende der Beobachtung C . Die Linkstrunkierungszeit L wird dabei nur dann positiv, wenn die Beobachtung erst nach dem Entstehen des Objektes anfängt. Der Verlauf des oben beschriebenen Prozesses wird anschaulich auf Abbildung 3.1 dargestellt.

Angenommen, in einem Zeitraum befinden sich insgesamt N Objekte $i = 1, \dots, N$ im Anfangszustand. Die Anzahl der beobachtbaren Objekte n kann allerdings kleiner als N sein, da die Objekte mit $L_i > T_i$ noch vor dem Beobachtungsanfang den Zustand wechseln und in die Studie nicht kommen. Auch bei den beobachtbaren Objekten sind Lebenszeiten nicht immer vollständig zu beobachten. Aus diesem Grund wird die Lebensdauer T_i durch die Rechtszensierungszeit $X_i = \min(T_i, C_i)$ und den Zensierungsindikator $\delta_i = \mathbf{1}_{\{X_i=T_i\}} = \mathbf{1}_{\{T_i \leq C_i\}}$ ersetzt.

3.3.2 Empirische kumulative Hazardrate

Zur Charakterisierung des Überlebensprozesses wird die so genannte Hazardfunktion eingeführt. Dabei handelt es sich von der momentanen Neigung zum Zustandswechsel. Die Definition der Hazardfunktion $\lambda(t)$ erfolgt z. B. in Andersen *et al.* (1993) durch die momentane Wahrscheinlichkeit des Zustandswechsels in t für alle Objekte, die sich in t noch im Anfangszustand befinden:

$$\lambda(t)dt = P(t \leq T < t + dt \mid t \leq T) = \frac{dF^T(t)}{1 - F^T(t)} \quad (3.2)$$

Die Aufsummierung dieser Wahrscheinlichkeiten liefert die kumulative Hazardrate $\Lambda(t)$:

$$\begin{aligned} \Lambda(t) &:= \int_0^t \lambda(s)ds = \int_0^t \frac{dF^T(s)}{1 - F^T(s)} \\ &= \int_0^t \frac{dF^{X^*}(s)}{G(s)} = -\log[1 - F^T(t)], \end{aligned} \quad (3.3)$$

wobei

$$\begin{aligned} F^{X^*}(x) &:= P(X_i \leq x, \delta_i = 1 \mid L_i \leq X_i) \\ &= \alpha^{-1} \int_0^x F^L(t)(1 - F^C(t))dF^T(t) \end{aligned} \quad (3.4)$$

und

$$\begin{aligned} G(x) &= P(L_i \leq x \leq X_i \mid L_i \leq X_i) \\ &= \alpha^{-1} F^L(x)(1 - F^T(x))(1 - F^C(x)) \end{aligned} \quad (3.5)$$

mit $\alpha := P(L_i \leq X_i)$.

Da die kumulative Hazardrate auch aus mangelnden Daten gut geschätzt werden kann, hat sie eine große Bedeutung in der Überlebensanalyse. Auch die Hazardfunktion $\lambda(t)$ und die Verteilungsfunktion $F^T(x)$ können mit Hilfe der kumulativen Hazardrate geschätzt werden.

Die Schätzung der kumulativen Hazardrate erfolgt aus dem Zählprozess $N_t := \#\{i : X_i \leq t, \delta_i = 1\}$ in Abbildung 3.2.

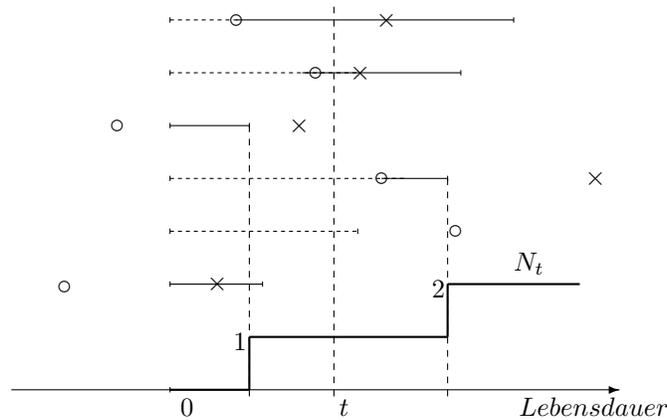


Abbildung 3.2: Zählprozess N_t .

In Andersen *et al.* (1993) wird gezeigt, dass der Schätzer für die kumulative Hazardrate als

$$\hat{\Lambda}_n(x) := \int_0^x \frac{\mathbf{1}_{\{Y(s) > 0\}}}{Y(s)} dN(s) \quad (3.6)$$

ausgedrückt werden kann. Die Anzahl der unter dem Risiko stehenden Objekte zum Zeitpunkt t wird durch $Y(t)$ bezeichnet. In Hougaard (2000, S. 72) wird dabei empfohlen bei linkstrunkierten und rechtszensierten Daten alle beobachtbaren Objekte im Anfangszustand als Objekte unter dem Risiko in t zu interpretieren:

$$Y(t) := \#\{i : L_i \leq t \leq X_i\}.$$

Alternativ wird die kumulative Hazardrate durch die Einsetzung der folgenden erwartungstreuen und konsistenten Schätzer für $F^{X^*}(\cdot)$ und $G(\cdot)$ (sie-

he Lemmata 3.3.1 und 3.3.2) in Cao *et al.* (2005) gezeigt:

$$F_n^{X^*}(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i \leq x, \delta_i=1\}} \quad (3.7)$$

und

$$G_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{L_i \leq x \leq X_i\}}. \quad (3.8)$$

Sowohl der Ansatz aus Andersen *et al.* (1993) als auch aus Cao *et al.* (2005) ergeben die folgende kumulative Hazardrate:

$$\hat{\Lambda}_n(t) := \sum_{i=1}^n \frac{\mathbf{1}_{\{X_i \leq t, \delta_i=1\}}}{nG_n(X_i)} = \sum_{i: X_{(i)} \leq t} \frac{\delta_i}{\#\{j : L_j \leq X_{(i)} \leq X_j\}} \quad (3.9)$$

wobei $X_{(i)}$ eine geordnete Stichprobe ist.

Offenbar kann der Schätzer (3.9) verwendet werden, auch wenn nur rechtszensierte Daten vorliegen. In diesem Fall sind alle Linkstrunkierungszeiten L_i kleiner Null. Für X_i gilt dagegen immer $X_i \geq 0$, die Bedingung $L_j \leq X_{(i)}$ wird dann überflüssig und der Schätzer (3.9) verwandelt sich in den Nelson-Aalen-Schätzer aus z. B. Andersen *et al.* (1993).

3.3.3 Annahmen

Damit die Konvergenz des Schätzers (3.9) gezeigt werden kann, müssen folgende Annahmen getroffen werden.

(Z1) $(T_i \in \mathbb{R}_0^+)_{i \in \mathbb{N}}$, $(C_i \in \mathbb{R}_0^+)_{i \in \mathbb{N}}$ und $(L_i \in \mathbb{R})_{i \in \mathbb{N}}$ sind unabhängige Folgen unabhängiger, je identisch verteilter Zufallsvariablen

(Z2) Durch X_i bzw. δ_i werden die Zufallsvariablen.

$$X_i := \min(T_i, C_i)$$

und

$$\delta_i := \mathbf{1}_{\{T_i \leq C_i\}}$$

bezeichnet.

(Z3) F^T, F^C, F^L bzw. F^X bezeichnen die durch

$$F^T(x) := P(T_i \leq x)$$

$$F^C(x) := P(C_i \leq x)$$

$$F^L(x) := P(L_i \leq x)$$

$$F^X(x) := P(X_i \leq x)$$

definierten stetigen und streng monotonen Verteilungsfunktionen.

(Z4) $A > 0$ ist eine Konstante, für die gilt $F^L(A) > 0$

(Z5) $B > A$ ist eine Konstante, für die gilt $F^X(B) < 1$

(Z6) J ist ein Intervall mit $J := [a, b] \subseteq [A, B]$ und $a \leq b$

(Z7) $\frac{0}{0} = 0$ wird vereinbart

3.3.4 Kernschätzer für Hazardrate

Zur Herleitung des Schätzers für die Hazardrate muss zunächst die Konvergenzrate der empirischen kumulativen Hazardrate ermittelt werden. Vor allem zeigen wir, dass der Schätzer (3.9) tatsächlich aus zwei konsistenten Schätzern konstruiert wird.

Lemma 3.3.1. *Die Annahmen (Z1) bis (Z7) seien erfüllt und eine Folge der Beobachtungen $(L_i, X_i, \delta_i)_{i=1, \dots, n}$ mit $L_i \leq X_i$ liegt vor.*

Dann existiert eine Konstante $D \leq 2$, so dass

$$P \left\{ \limsup_{n \rightarrow \infty} \frac{\sup_{x \in [A, B]} |F_n^{X^*}(x) - F^{X^*}(x)|}{\sqrt{\log(n)/n}} = D \right\} = 1$$

gilt.

Beweis. Um die Funktion $F^{X^*}(\cdot)$ an die Form der Wahrscheinlichkeitsfunktion aus Abschnitt 3.1 anzupassen, setzen wir zufällige Mengen $H_i := \{y \in \mathbb{R} : X_i \leq y\}$, Zufallsvariablen $S_i = \delta_i$, $i = 1, \dots, n$ und eine Konstante $s_{const} = 1$. Dann ergeben sich:

$$F^{X^*}(x) = P(X_i \leq x, \delta_i = 1 \mid L_i \leq X_i) = P(x \in H_i, S_i = s_{const} \mid L_i \leq X_i)$$

und

$$F_n^{X^*}(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i \leq x, \delta_i = 1\}} = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{x \in H_i, S_i = s_{const}\}}$$

Das Lemma 3.1.1 liefert die gewünschte Konvergenzrate. Beachte, dass $F^{X^*}(\cdot)$ die bedingte Wahrscheinlichkeitsfunktion ist. \square

Besonders wichtig ist die Konvergenzgeschwindigkeit des Schätzers für $G(x)$, da die lokale Rate des Schätzers (3.9) dadurch geprägt wird.

Lemma 3.3.2. Die Annahmen (Z1) bis (Z7) seien erfüllt und eine Folge der Beobachtungen $(L_i, X_i)_{i=1, \dots, n}$ mit $L_i \leq X_i$ liegt vor.

Dann existiert eine Konstante $D \leq 2$, so dass

$$P \left\{ \limsup_{n \rightarrow \infty} \frac{\sup_{x \in [A, B]} |G_n(x) - G(x)|}{\sqrt{\log(n)/n}} = D \right\} = 1$$

gilt.

Beweis. Analog zum Lemma 3.3.1 werden zufällige Mengen $H_i := \{y \in \mathbb{R} : L_i \leq y \leq X_i\}$ und Zufallsvariablen $S_i : \Omega \rightarrow \{s_{const} \in \mathbb{R}\}$, $i = 1, \dots, n$ gesetzt.

Wir bekommen dann:

$$\begin{aligned} G(x) &= P(L_i \leq x \leq X_i \mid L_i \leq X_i) \\ &= P(L_i \leq x \leq X_i, S_i = s_{const} \mid L_i \leq X_i) \\ &= P(x \in H_i, S_i = s_{const} \mid L_i \leq X_i) \end{aligned}$$

und

$$\begin{aligned} G_n(x) &= \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{L_i \leq x \leq X_i\}} \\ &= \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{L_i \leq x \leq X_i, S_i = s_{const}\}} \\ &= \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{x \in H_i, S_i = s_{const}\}} \end{aligned}$$

Schließlich folgt die gewünschte Konvergenzrate aus Lemma 3.1.1. Beachte, dass $G(\cdot)$ die bedingte Wahrscheinlichkeitsfunktion ist. \square

Somit ist der Schätzer (3.9) konsistent und seine lokale Konvergenzrate lässt sich bestimmen. Dabei handelt es sich um die Konvergenz des empirischen Maßes $\hat{\Lambda}_n(J) := \hat{\Lambda}_n(b) - \hat{\Lambda}_n(a)$ gegen das theoretische $\Lambda(J) := \int_J \lambda(s) ds = \int_a^b \lambda(s) ds = \Lambda(b) - \Lambda(a)$ mit $J \subseteq [A, B]$.

Satz 3.3.1. *Die Annahmen (Z1) bis (Z7) seien erfüllt, die Folge (p_n) erfülle die Voraussetzung (K5) und eine Folge von Beobachtungen $(L_i, X_i, \delta_i)_{i=1, \dots, n}$ mit $L_i \leq X_i$ liegt vor.*

Dann existiert eine Konstante $D \leq 2(\sqrt{2 \cdot (2M + \Lambda(B))} + 2M)$, so dass

$$P \left\{ \limsup_{n \rightarrow \infty} \frac{\sup_{J \subseteq [A, B], \Lambda(J) \leq p_n} |\hat{\Lambda}_n(J) - \Lambda(J)|}{\sqrt{\log(n)p_n/n}} = D \right\} = 1$$

wobei $M := \sup_{x \in [A, B]} [P(L_i \leq x \leq X_i \mid L_i \leq X_i)]^{-1}$.

Beweis. Wie üblich überprüfen wir die Annahmen (L1) bis (L8) für die lokale Konvergenz.

- Die Zufallsvariablen $\Delta_i^x = \mathbf{1}_{\{\delta_i=1\}} \leq 1 =: \Delta_{max}$, $i = 1, \dots, n$ werden für alle $x \in [A, B]$ eingeführt,
- $\Psi(x) := \Lambda(x)$,
- $G(x) := P(L_i \leq x \leq X_i \mid L_i \leq X_i) > 0$ auf $[A, B]$,
- $G_n(x) := \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{L_i \leq x \leq X_i\}}$,
- $D_G := 2$.

Gemäß dem Lemma 3.3.2 gilt $|G_n(x) - G(x)| \leq C \sqrt{\log(n)/n}$ f.s. für große n , $x \in \mathbb{R}$ und $C > D_G$.

Nun wird der Ausdruck (3.5) betrachtet, der wegen Stetigkeit von $F^L(x)$ als

$$\begin{aligned} G(x) &= \alpha^{-1} F^L(x)(1 - F^T(x))(1 - F^C(x)) \\ &= \int_{-\infty}^x \alpha^{-1}(1 - F^T(x))(1 - F^C(x)) f^L(s) ds \end{aligned}$$

geschrieben werden kann.

Da $(1 - F^T(x))(1 - F^C(x)) \leq 1$, $f^L(x) < \infty$ und $\alpha^{-1} < \infty$ gelten, existiert ein $L_{Lipschitz}^G < \infty$, so dass $\alpha^{-1}(1 - F^T(x))(1 - F^C(x)) f^L(s) \leq \alpha^{-1} f^L(s) \leq L_{Lipschitz}^G$ gilt.

Die Lipschitz-Stetigkeit von $G(x)$ ergibt sich dann für alle $x \in [A, B]$ wie folgt:

$$\begin{aligned} |G(x) - G(y)| &= \left| \int_{-\infty}^x \alpha^{-1}(1 - F^T(x))(1 - F^C(x)) f^L(s) ds \right. \\ &\quad \left. - \int_{-\infty}^y \alpha^{-1}(1 - F^T(y))(1 - F^C(y)) f^L(s) ds \right| \\ &\leq L_{Lipschitz}^G \left| \int_{-\infty}^x ds - \int_{-\infty}^y ds \right| = L_{Lipschitz}^G |x - y|. \end{aligned}$$

Wegen der Definitionen von $\lambda(x)$ und $\Lambda(x)$ ist $\Psi(x)$ offensichtlich monoton wachsend. Die Lipschitz-Stetigkeit von $\Lambda(x)$ auf $[A, B]$ folgt daraus, dass $G(x)$ Lipschitz-stetig und positiv ist. Da $\alpha^{-1} < \infty$, $F^L(s)(1 - F^C(s)) \leq 1$, $f^T(s) <$

∞ und $G(s) > 0$ für alle $s \in [A, B]$ gelten, existiert auch ein $L_{Lipschitz} < \infty$, so dass $\alpha^{-1}F^L(s)(1 - F^C(s))f^T(s)/G(s) \leq L_{Lipschitz}$ auf $[A, B]$. Analog zu $G(x)$ wird gezeigt, dass $\Lambda(x)$ Lipschitz-stetig für alle $x \in [A, B]$ ist.

Die Annahmen (L4) und (L5) gelten wegen der konstanten Δ_i^x .

Es bleibt also nur noch die Annahme (L6) zu überprüfen.

$$\begin{aligned} E\left(\frac{\mathbf{1}_{\{X_i \leq x\}} \cdot \Delta_i^x}{G(X_i)} \mid L_i \leq X_i\right) &= E\left(\frac{\mathbf{1}_{\{X_i \leq x, \delta_i = 1\}}}{G(X_i)} \mid L_i \leq X_i\right) \\ &= \sum_{\delta_1=0}^1 \int_{-\infty}^{\infty} \frac{\mathbf{1}_{\{x_1 \leq x, \delta_1 = 1\}}}{G(x_1)} dF^{X, \delta}(x_1, \delta_1) = \int_{-\infty}^x \frac{dF^{X, \delta}(x_1, 1)}{G(x_1)} \end{aligned}$$

Durch $F^{X, \delta}(x, y)$ wird die gemeinsame Verteilungsfunktion von (X, δ) bedingt $L \leq X$ bezeichnet.

Für die Intervalle $J := [a, b] \subseteq [A, B]$ kann die Ableitung $dF^{X, \delta}(x_1, 1)$ aus der Wahrscheinlichkeit $P(X_i \in J, \delta_i = 1 \mid L_i \leq X_i)$ berechnet werden.

$$\begin{aligned} P(X_i \in J, \delta_i = 1 \mid L_i \leq X_i) &= \alpha^{-1}P(X_i \in J, \delta_i = 1, L_i \leq X_i) \\ &= \alpha^{-1}[P(T_i \in J, T_i \leq C_i, L_i \leq T_i, T_i \leq C_i) \\ &\quad + P(C_i \in J, T_i \leq C_i, L_i \leq C_i, C_i < T_i)] = \alpha^{-1}P(T_i \in J, L_i \leq T_i \leq C_i). \end{aligned} \quad (3.10)$$

Die Wahrscheinlichkeiten $P(X_i \in J, \delta_i = 1 \mid L_i \leq X_i)$ und $P(T_i \in J, L_i \leq T_i \leq C_i)$ können dabei über Integrale weiter ausgeschrieben werden. Wir bekommen dann

$$\begin{aligned} P(X_i \in J, \delta_i = 1 \mid L_i \leq X_i) &= E(\mathbf{1}_{\{X_i \in J, \delta_i = 1\}} \mid L_i \leq X_i) \\ &= \sum_{\delta_1=0}^1 \int_{-\infty}^{\infty} \mathbf{1}_{\{x_1 \in J, \delta_1 = 1\}} dF^{X, \delta}(x_1, \delta_1) = \int_{x_1 \in J} dF^{X, \delta}(x_1, 1) \end{aligned} \quad (3.11)$$

und

$$\begin{aligned} \alpha^{-1}P(T_i \in J, L_i \leq T_i \leq C_i) &= \alpha^{-1}E(\mathbf{1}_{\{T_i \in J, L_i \leq T_i \leq C_i\}}) \\ &= \int_{t \in \mathbb{R}} \int_{c \in \mathbb{R}} \int_{l \in \mathbb{R}} \alpha^{-1} \mathbf{1}_{\{t \in J\}} \mathbf{1}_{\{l \leq t\}} \mathbf{1}_{\{t \leq c\}} dF^T(t) dF^C(c) dF^L(l) \\ &= \int_{t \in J} \alpha^{-1} F^L(t) (1 - F^C(t)) dF^T(t). \end{aligned} \quad (3.12)$$

Bei einelementigen Intervallen $J = [j, j]$, $j \in [A, B]$ liefern die Ausdrücke (3.10), (3.11) und (3.12) die Ableitung $dF^{X,\delta}(x_1, 1)$ wie folgt:

$$\begin{aligned} \int_{x_1 \in J} dF^{X,\delta}(x_1, 1) &= \int_{t \in J} \alpha^{-1} F^L(t)(1 - F^C(t)) dF^T(t) \\ &= dF^{X,\delta}(j, 1) = \alpha^{-1} F^L(j)(1 - F^C(j)) dF^T(j). \end{aligned} \quad (3.13)$$

Durch die Einsetzung von (3.13) kann die Annahme (L6) leicht folgenderweise hergeleitet werden:

$$\begin{aligned} E \left(\frac{\mathbf{1}_{\{X_i \leq x\}} \cdot \Delta_i^x}{G(X_i)} \mid L_i \leq X_i \right) &= \int_{-\infty}^x \frac{dF^{X,\delta}(x_1, 1)}{G(x_1)} \\ &= \int_{-\infty}^x \frac{\alpha^{-1} F^L(x_1)(1 - F^C(x_1)) dF^T(x_1)}{G(x_1)} \\ &= \int_{-\infty}^x \frac{\alpha^{-1} F^L(x_1)(1 - F^C(x_1)) dF^T(x_1)}{\alpha^{-1} F^L(x_1)(1 - F^C(x_1))(1 - F^T(x_1))} = \int_{-\infty}^x \frac{dF^T(x_1)}{1 - F^T(x_1)} = \Lambda(x) = \Psi(x). \end{aligned}$$

Wir bezeichnen weiter $\Psi_n(x) := \Lambda_n(x)$ und $\Psi_n(J) := \hat{\Lambda}_n(J)$.

Die Voraussetzungen (L1) bis (L8) sind dann erfüllt und aus Korollar 2.2.1 folgt die Existenz einer Konstante

$$D \leq 2(\sqrt{2 \cdot (2M + \Lambda(B))} + 2M),$$

so dass

$$P \left\{ \limsup_{n \rightarrow \infty} \frac{\sup_{J \subseteq [A, B], \Lambda(J) \leq p_n} |\hat{\Lambda}_n(J) - \Lambda(J)|}{\sqrt{\log(n) p_n / n}} = D \right\} = 1.$$

□

Gemäß der Formel (2.8) schreiben wir nun den Kernschätzer für die Hazardrate wie folgt auf:

$$\lambda_n(t) := \sum_{i=1}^n \frac{1}{R_n(X_i)} \text{Kr} \left(\frac{t - X_i}{R_n(X_i)} \right) \frac{\delta_i}{\#\{j : L_j \leq X_i \leq X_j\}}. \quad (3.14)$$

Mit der Konvergenzrate $\mathcal{O}(\sqrt{\log(n) p_n / n})$ ist die wichtigste Annahme für die Konvergenz von (3.14) erfüllt. Zusätzlich wird angenommen, dass $\lambda_{\min} \leq$

$\lambda(\cdot) \leq \lambda_{max}$ Lipschitz-stetig auf $[A, B]$ ist und eine Lipschitz-Konstante L_λ existiert. Die Annahmen (K1)-(K4) für $\Lambda(\cdot)$ und $\lambda(\cdot)$ sind dann erfüllt und die Konvergenz von $\lambda_n(\cdot)$ kann für verschiedene Glätter $\tilde{\Psi}(\cdot)$ gezeigt werden. Es existiert dann eine Konstante D_0 , so dass

$$P \left\{ \limsup_{n \rightarrow \infty} \frac{\sup_{x \in [A, B]} |\lambda_n(x) - \lambda(x)|}{\sqrt{\log(n)/(np_n) + p_n}} = D_0 \right\} = 1.$$

Für den linearen Glätter (2.7) mit einer konstanten Bandbreite $R_n = p_n/\tilde{c}$, $\tilde{c} > 0$ gilt $D_0 \leq \max\{D_1, D_2\}$, wobei

$$D_1 = 2(\sqrt{2 \cdot (2M + \Lambda(B))} + 2M) \cdot \tilde{c}^{1/2} \text{Var}(\text{Kr}) \lambda_{max}^{1/2},$$

$$D_2 = L_\lambda \tilde{c}^{-1},$$

$M = \sup_{x \in [A, B]} [P(L_i \leq x \leq X_i | L_i \leq X_i)]^{-1}$ und die Kernfunktion $\text{Kr}(\cdot)$ die Voraussetzung (K6) erfüllt.

Für die variable Bandbreite wird hier $\tilde{\Psi}(\cdot) = \Lambda(\cdot)$ gewählt. Die Voraussetzungen für den Glätter sind dabei identisch zu diesen von $\Lambda(\cdot)$. Es gilt $D_0 \leq \max\{D_1 + D_2, D_3\}$, wobei

$$D_1 = 4(\sqrt{2 \cdot (2M + \Lambda(B))} + 2M) \lambda_{min}^{-2} \lambda_{max}^3 [\sup(\text{Kr}) + L_{Kr} \lambda_{min}^{-1} \lambda_{max}],$$

$$D_2 = 2(\sqrt{2 \cdot (2M + \Lambda(B))} + 2M) \text{Var}(\text{Kr}) \lambda_{min}^{-1/2} \lambda_{max}^{3/2},$$

$$D_3 = 2L_{Kr} L_\lambda \lambda_{min}^{-5} \lambda_{max}^4 + 2 \sup(\text{Kr}) L_\lambda \lambda_{min}^{-4} \lambda_{max}^3 + L_\lambda \lambda_{min}^{-1}$$

und die Kernfunktion $\text{Kr}(\cdot)$ die Voraussetzungen (K6) und (K7) erfüllt.

3.4 Markov-Prozess

3.4.1 Modell

Eine Erweiterung der Überlebensprozesse stellen Markov-Prozesse dar. Bei diesen Prozessen entstehen alle Objekte, genauso wie bei dem Überlebensprozess, nach einem bestimmten Ereignis und existieren dann unendlich lange.

Allerdings können diese Objekte sich in mehr als zwei Zuständen befinden, wobei Migrationen von einem Zustand in den anderen und das dauerhafte Verbleiben in einem der Zustände beobachtet werden.

Die Menge aller möglichen Zustände bildet den Zustandsraum \mathbb{S} des Markov-Prozesses. Der Einfachheit halber betrachten wir in dieser Arbeit nur Prozesse mit finitem Zustandsraum, wobei die Menge \mathbb{S} endlich und abzählbar ist und aus K Elementen besteht.

Definition 3.4.1. *Ein stochastischer Prozess $\{X_t\}_{t \in [0, \infty)}$ mit finitem Zustandsraum \mathbb{S} wird auf dem Wahrscheinlichkeitsraum $\{\Omega, \mathcal{A}, \mathbf{P}\}$ als $X : \Omega \times [0, \infty) \rightarrow \mathbb{S}$, $(\omega, t) \mapsto X_t(\omega)$ definiert. Dieser Prozess heißt Markov-Prozess, wenn die Markov-Eigenschaft*

$$P(X_{t_{n+1}} = h \mid X_{t_n} = i, \dots, X_{t_1} = i_1) = P(X_{t_{n+1}} = h \mid X_{t_n} = i) \quad (3.15)$$

für beliebige $0 \leq t_1 < t_2 < \dots < t_n < t_{n+1}$ aus $[0, \infty)$ und beliebige $i_1, \dots, i_{n-1}, i, j$ aus \mathbb{S} mit $P(X_{t_n} = i, \dots, X_{t_2} = i_2, X_{t_1} = i_1) > 0$ erfüllt wird. (vgl. Anderson (1991))

Für einen Markov-Prozess $\{X_t\}_{t \in [0, \infty)}$ mit finitem Zustandsraum \mathbb{S} bezeichnen wir für alle $t, s \in [0, \infty)$ und $i, h \in \mathbb{S}$ die Funktionen

$$p^{ih}(t, s) \stackrel{\text{def}}{=} P(X_s = h \mid X_t = i) \quad (3.16)$$

als Übergangswahrscheinlichkeiten,

$$p^{ih}(t) \stackrel{\text{def}}{=} p^{ih}(t, t + dt) = P(X_{t+dt} = h \mid X_t = i) \quad (3.17)$$

als momentane Übergangswahrscheinlichkeiten und

$$m_i(t) \stackrel{\text{def}}{=} P(X_t = i) \quad (3.18)$$

als Zustandswahrscheinlichkeiten.

Es wird angenommen, dass alle Übergangs- und Zustandswahrscheinlich-

keiten stetige Funktionen sind.

Lemma 3.4.1. *Bei jedem Markov-Prozess $\{X_t\}_{t \in [0, \infty)}$ können Wahrscheinlichkeiten*

$$P(X_{t_1} = i_1, \dots, X_{t_{n-1}} = i_{n-1}, X_{t_n} = i_n) \neq 0 \quad (3.19)$$

für beliebige $0 \leq t_1 < t_2 < \dots < t_n$ aus $[0, \infty)$ und beliebige i_1, i_2, \dots, i_n aus \mathbb{S} nur über Zustands- und Übergangswahrscheinlichkeiten als

$$P(X_{t_1} = i_1, \dots, X_{t_n} = i_n) = m_{i_1}(t_1) \prod_{j=0}^{n-2} p^{i_{n-j-1}, i_{n-j}}(t_{n-j-1}, t_{n-j}) \quad (3.20)$$

ausgedrückt werden.

Beweis. Bei $P(X_{t_1} = i_1, \dots, X_{t_{n-1}} = i_{n-1}, X_{t_n} = i_n) \neq 0$ ist der Beweis trivial und folgt direkt aus der Markov-Eigenschaft (3.15). □

Für die meisten Markov-Prozesse wird angenommen, dass die Ableitung

$$q_{ih}(t) \stackrel{\text{def}}{=} \lim_{\Delta t \rightarrow 0} \frac{P(X_{t+\Delta t} = h \mid X_t = i) - P(X_t = h \mid X_t = i)}{\Delta t} \quad (3.21)$$

für alle $t \geq 0$ und alle $i, h \in \mathbb{S}$ existiert. Die Familie von Funktionen $\{q_{ih}(t)\}_{i, h \in \mathbb{S}}$ bezeichnen wir als Migrations- oder Übergangsintensitäten. Es gilt also:

$$q_{ih}(t)dt = p_{ih}(t, t+dt) - p_{ih}(t, t). \quad (3.22)$$

Die Eigenschaften von Migrationsintensitäten sind offensichtlich und werden in Johansen (1978) aufgelistet:

1. $q_{ih}(t) \geq 0$, falls $i \neq h$
2. $q_{ih}(t) \leq 0$, falls $i = h$
3. $\sum_{h=1}^K q_{ih}(t) = 0$

Auf dem nächsten Schritt betrachten wir den Zusammenhang zwischen Migrationsintensitäten, Zustands- und Übergangswahrscheinlichkeiten. Es werden dabei zeitabhängige $K \times K$ -Matrizen von Migrationsintensitäten

$$(\mathbf{Q}(t))_{ih} := (q_{ih}(t))_{ih} \quad (3.23)$$

und Übergangswahrscheinlichkeiten

$$(\mathbf{P}(t, s))_{ih} := (p_{ih}(t, s))_{ih} \quad (3.24)$$

und ein zeitabhängiger K -Vektor der Zustandswahrscheinlichkeiten

$$(\mathbf{m}(t))_i := (P(X_t = i))_i \quad (3.25)$$

für alle $t \geq 0$ eingeführt.

Aus Johansen (1978) ist bekannt, dass die Matrix der Übergangswahrscheinlichkeiten als Produktintegral

$$\mathbf{P}(t, s) = \prod_{(t, s]} (\mathbf{I} + \mathbf{Q}(z) dz) \quad (3.26)$$

dargestellt werden kann. Wegen Lemma 3.4.1 können dann Wahrscheinlichkeiten $P(X_{t_1} = i_1, \dots, X_{t_{n-1}} = i_{n-1}, X_{t_n} = i_n) \neq 0$ für beliebige $0 \leq t_1 < t_2 < \dots < t_n$ aus $[0, \infty)$ und beliebige i_1, i_2, \dots, i_n aus \mathbb{S} nur durch Zustandswahrscheinlichkeiten und Migrationsintensitäten ausgedrückt werden.

Lemma 3.4.2. *Für jeden stetigen Markov-Prozess existiert folgende Beziehung zwischen Migrationsintensitäten und Zustandswahrscheinlichkeiten (siehe Strook (2000, S. 133)):*

$$\frac{d}{dt} \mathbf{m}(t) = \mathbf{Q}(t)' \mathbf{m}(t) \quad (3.27)$$

Beweis. Da Strook keinen Beweis für das Lemma angibt, wird die Behauptung

tung (3.27) weiter unten gezeigt.

Der Beweis folgt aus der Tatsache, dass

$$\begin{aligned} \sum_{i=1}^K P(X_{t+dt} = h \mid X_t = i) \cdot P(X_t = i) &= \\ \sum_{i=1}^K P(X_t = i, X_{t+dt} = h) &= P(X_{t+dt} = h) \end{aligned} \quad (3.28)$$

für alle $i, h \in \mathbb{S}$ und alle $t \geq 0$ gilt.

Somit kann die Gleichung (3.28) durch das Einsetzen von Migrationsintensitäten und Zustandswahrscheinlichkeiten als

$$m_h(t) + \sum_{i=1}^K q_{ih}(t)m_i(t)dt = m_h(t + dt)$$

umgeschrieben werden. Dabei gilt schließlich

$$\frac{m_h(t + dt) - m_h(t)}{dt} = \frac{d}{dt}m_h(t) = \sum_{i=1}^K q_{ih}(t)m_i(t) \quad (3.29)$$

Es ist leicht zu bemerken, dass der Ausdruck (3.27) direkt aus der Gleichung (3.29) folgt. \square

Wir nehmen an, dass der Ausdruck (3.20) auch für die Wahrscheinlichkeiten

$$P(X_{t_1} = i_1, \dots, X_{t_{n-1}} = i_{n-1}, X_{t_n} = i_n) = 0 \quad (3.30)$$

bei beliebigen $0 \leq t_1 < t_2 < \dots < t_n$ aus $[0, \infty)$ und beliebigen i_1, i_2, \dots, i_n aus \mathbb{S} gilt. Dann folgt es aus Lemma 3.4.2, dass nur Migrationsintensitäten $\mathbf{Q}(t)$ und Anfangswahrscheinlichkeiten $\mathbf{m}(0)$ zur Berechnung jeder Wahrscheinlichkeit (3.19) und (3.30) ausreichend sind. Gemäß Johansen (1978) bezeichnen wir solche Prozesse als Markov-Prozesse mit integrierten Migrationsintensitäten und Anfangswahrscheinlichkeiten.

Liegen nur die Zustandswahrscheinlichkeiten vor, so können Migrationsintensitäten nicht eindeutig aus dem Differenzialsystem (3.27) bestimmt wer-

den. Trotz eines Zusammenhangs zwischen $\mathbf{Q}(t)$ und $\mathbf{m}(t)$ erfüllt eine große Familie von Matrizen $\{\mathbf{Q}_1(t), \dots, \mathbf{Q}_m(t)\}$ die Gleichung (3.27). Somit ist die Schätzung von Migrationsintensitäten sogar bei den geschätzten Zustandswahrscheinlichkeiten ziemlich problematisch. Viele Autoren beschäftigen sich allerdings nur mit den homogenen Markov-Prozessen, bei denen $\forall t \geq 0 : \mathbf{Q}(t) = \text{const}$ angenommen wird. Wie in Weißbach *et al.* (2009) gezeigt wurde, ist diese Annahme bei den empirischen Studien nicht immer gerechtfertigt. Wir streben somit an, den konsistenten Kernschätzer für die Migrationsintensitäten bei den inhomogenen Markov-Prozessen herzuleiten. Als inhomogen bezeichnen wir einen Markov-Prozess, bei dem die Matrix der Migrationsintensitäten $\mathbf{Q}(t)$ nicht konstant sein kann.

3.4.2 Kumulative gewichtete Migrationsintensität

Da die Migrationsintensitäten $q_{ih}(t)$ bei $i \neq h$ nichtnegative Funktionen sind, sind auch die Funktionen

$$\tilde{Q}^{ih}(t) = \int_0^t \psi(s) q_{ih}(s) ds$$

nichtnegativ und monoton wachsend für alle $t \geq 0$ und beliebige nichtnegative Funktionen $\psi(t) < \infty$. Es ist somit sinnvoll den Schätzer (2.12) für solche Funktionen zu verwenden. So besteht dann die Möglichkeit aus diesem den Kernschätzer für $\psi(t)q_{ih}(t)$ bzw. $q_{ih}(t)$ zu bekommen.

Angenommen, es liegen n stochastisch unabhängige Markov-Prozesse

$$\{X_t^1 \in \mathbb{S}\}_{t \in [0, \infty)}, \dots, \{X_t^n \in \mathbb{S}\}_{t \in [0, \infty)}$$

mit integrierten Migrationsintensitäten $\mathbf{Q}(t)$ und Anfangswahrscheinlichkeiten $\mathbf{m}(0)$ vor. Wenn die Anzahl der beobachtbaren Prozesse n für alle $t \geq 0$

konstant bleibt, können empirische Prozesse $N_t^{ih;n} = \int_0^t dN_s^{ih;n}$ mit Sprüngen

$$dN_t^{ih;n} := \frac{\#\{j \in [1, n] \mid X_t^j = i, X_{t+dt}^j = h\}}{n} = \sum_{j=1}^n \frac{\mathbf{1}_{\{X_t^j=i, X_{t+dt}^j=h\}}}{n}$$

für alle $i, h \in \mathbb{S}$, $i \neq h$ wie in Abbildung 3.3 eingeführt werden.

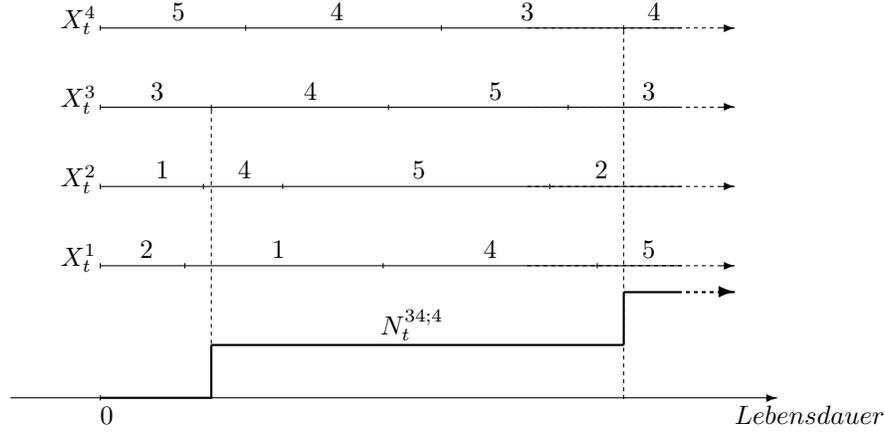


Abbildung 3.3: Empirische Funktion $N_t^{34;4}$ für vier Markov-Prozesse

Als kumulative gewichtete Migrationsintensität bezeichnen wir die Funktion

$$\tilde{Q}^{ih}(t) := \int_0^t q_{ih}(s) m_i(s) ds. \quad (3.31)$$

Bekanntlich ist $m_i(t)$ eine nichtnegative Funktion für alle $t \geq 0$. Wegen

$$E \left(\int_0^t dN_s^{ih;n} \right) = \int_0^t \frac{1}{n} \sum_{j=1}^n E(\mathbf{1}_{\{X_s^j=i, X_{s+ds}^j=h\}}) = \int_0^t q_{ih}(s) m_i(s) ds$$

wird dann ein erwartungstreuer Schätzer für $\tilde{Q}^{ih}(t)$ wie folgt definiert:

$$\widehat{\tilde{Q}}_n^{ih}(t) := \int_0^t dN_s^{ih;n} = N_t^{ih;n}.$$

3.4.3 Beobachtungsschranken

In der Realität sind Markov-Prozesse natürlich nicht auf der ganzen Zeitachse zu beobachten, sondern auf einem kurzen Zeitabschnitt. Darüber hinaus ist die Anzahl der beobachtbaren Prozesse nur stückweise konstant. Zur Beschreibung der Beobachtungsgrenzen werden stochastisch unabhängige Zufallsvariablen L und C mit Verteilungsfunktionen $F^L(t) := P(L \leq t)$ und $F^C(t) := P(C \leq t)$ eingeführt. Als untere und obere Beobachtungsschranke werden dann stochastisch abhängige Zufallsvariablen jeweils \tilde{L} und \tilde{C} mit der gemeinsamen Verteilungsfunktion

$$P(\tilde{L} \leq x, \tilde{C} \leq y) = P(L \leq x, C \leq y \mid L \leq C)$$

bezeichnet.

Der empirische Prozess $N_t^{ih;n}$ aus dem vorigen Abschnitt basiert im Wesentlichen auf der Annahme, dass die Anzahl der Beobachtungen auf der ganzen Zeitachse konstant ist. Bei den stochastischen Beobachtungsschranken wird diese Annahme nicht mehr gewährleistet. Allerdings kann ein Zeitabschnitt $[A, B] \subset [0, \infty)$ existieren, auf dem alle vorliegende Objekte bei beliebigen \tilde{L} und \tilde{C} beobachtbar sind. Dabei dürfen keine Objekte zwischen A und B in die Studie eintreten oder unbeobachtbar werden. Intuitiv kann dies durch die speziellen Bedingungen an die Verteilungsfunktionen $F^L(t)$ und $F^C(t)$ erreicht werden, indem $\lim_{t \nearrow A} F^L(t) = 1$ und $\lim_{t \searrow B} F^C(t) = 0$ vorausgesetzt werden. Für eine beliebige Anzahl von gleich definierten Markov-Prozessen und für ihre beliebigen Realisationen sind dann zwischen A und B alle Objekte in der Studie beobachtbar.

Sind die vorliegenden n Markov-Prozesse nur auf einem begrenzten Zeitintervall $[A, B] \subset [0, \infty)$ beobachtbar, so besteht nur die Möglichkeit die empirische Funktion

$$\widehat{\tilde{Q}}_{n,[A,B]}^{ih}(t) := \begin{cases} \widehat{\tilde{Q}}_n^{ih}(t) - \widehat{\tilde{Q}}_n^{ih}(A) = \int_A^t dN_s^{ih;n} & A \leq t \leq B \\ 0 & \text{sonst} \end{cases} \quad (3.32)$$

auf $[A, B]$ zu bestimmen. Von dem Interesse ist nun, unter welchen Bedingungen die Funktion $\widehat{\widetilde{Q}}_{n;[A,B]}^{ih}(t)$ lokal konsistent für

$$\widetilde{Q}_{[A,B]}^{ih}(t) := \begin{cases} \widetilde{Q}^{ih}(t) - \widetilde{Q}^{ih}(A) = \int_A^t q_{ih}(s)m_i(s)ds & A \leq t \leq B \\ 0 & \text{sonst} \end{cases} \quad (3.33)$$

auf dem Intervall $[A, B]$ ist.

Die Realisationen von vier Markov-Prozessen, die Beobachtungsgrenzen und ein möglicher Verlauf der empirischen Funktion $\widehat{\widetilde{Q}}_{n;[A,B]}^{ih}(t)$ sind in Abbildung 3.4 zu sehen.

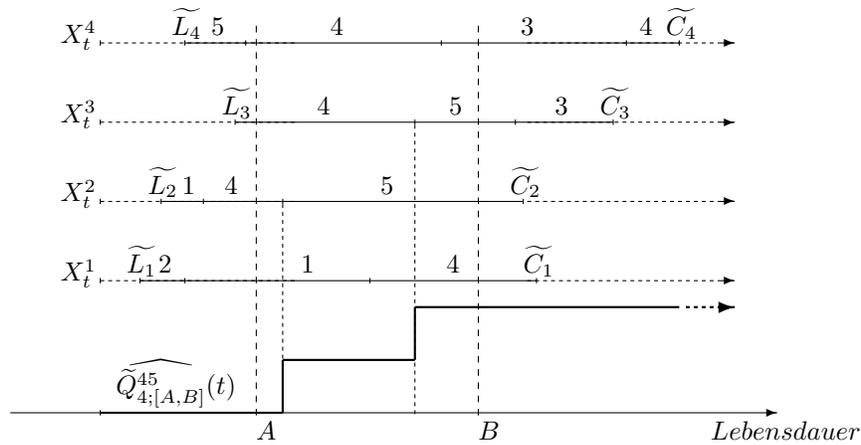


Abbildung 3.4: Empirische Funktion $\widehat{\widetilde{Q}}_{4;[A,B]}^{45}(t)$ für vier Markov-Prozesse

3.4.4 Annahmen

Bevor wir zum lokal konsistenten Schätzer für $\widetilde{Q}_{[A,B]}^{ih}(t)$ kommen, werden Annahmen für das Markov-Modell nochmals kurz zusammengefasst.

(M1) $\mathbb{S} := \{1, \dots, K\}$ für ein $K \in \mathbb{N}_{\geq 2}$ ist ein Zustandsraum.

(M2) $\{X_t^1 \in \mathbb{S}\}_{t \in [0, \infty)}, \dots, \{X_t^n \in \mathbb{S}\}_{t \in [0, \infty)}$ sind stochastisch unabhängige Markov-Prozesse mit integrierten Migrationsintensitäten $\mathbf{Q}(t)$ und Anfangswahrscheinlichkeiten $\mathbf{m}(0)$.

(M3) L und C sind stochastisch unabhängige Zufallsvariablen mit Verteilungsfunktionen $F^L(x)$ und $F^C(x)$.

(M4) $\{(\tilde{L}_j, \tilde{C}_j) \in [0, \infty) \times [0, \infty)\}_{j=1, \dots, n}$ sind Folgen unabhängiger und identisch verteilter Paare von Zufallsvariablen mit der Verteilungsfunktion

$$F^{\tilde{L}\tilde{C}}(x, y) := P(\tilde{L} \leq x, \tilde{C} \leq y) = P(L \leq x, C \leq y \mid L \leq C).$$

(M5) $\tilde{A} > 0$ ist eine Konstante, für die gilt $F^L(\tilde{A}) = 1$.

(M6) $\tilde{B} > \tilde{A}$ ist eine Konstante, für die gilt $F^C(\tilde{B}) = 0$.

(M7) $\exists [A, B] \subseteq [\tilde{A}, \tilde{B}] \forall t \in [A, B] \forall i, h \in \mathbb{S}, i \neq h, q_{ih}(t) \neq 0 :$

- $0 < \tilde{q}_{ih}^{\min} < q_{ih}(t)m_i(t) \leq \tilde{q}_{ih}^{\max} < \infty,$
- $q_{ih}(t)m_i(t)$ ist Lipschitz-stetig mit der Konstante $L_{ih},$
- $m_i^{\min} \leq m_i(t),$
- $\sum_{j=1}^n \mathbf{1}_{\{X_t^j=i\}} > 0,$
- $\int_A^B \mathbf{1}_{\{X_s^j=i, X_{s+ds}^j=h\}} \leq K^{ih}$ für $j = 1, \dots, n,$
d. h. die Anzahl der Übergänge von i nach h muss auf $[A, B]$ beschränkt sein.

(M8) J ist ein Intervall mit $J := [a, b] \subseteq [A, B].$

(M9) $\frac{0}{0} = 0$ wird vereinbart.

3.4.5 Kernschätzer für Migrationsintensitäten

Analog zur kumulativen gewichteten Migrationsintensität (3.33) wird die gewichtete Migrationsintensität als

$$\tilde{q}_{ih}(t) := q_{ih}(t)m_i(t) \quad (3.34)$$

definiert. Die wichtigste Bedingung zur Bestimmung von Konsistenzeigenschaften des Schätzers für (3.34) ist die lokale Konvergenz der empirischen kumulativen gewichteten Migrationsintensität (3.32).

Bei der lokalen Konvergenz handelt es sich dabei um die Konvergenz der empirischen Maße

$$\widehat{\tilde{Q}_{n;[A,B]}^{ih}}(J) := \int_a^b dN_s^{ih;n} = \widehat{\tilde{Q}_{n;[A,B]}^{ih}}(b) - \widehat{\tilde{Q}_{n;[A,B]}^{ih}}(a)$$

bezüglich der theoretischen

$$\tilde{Q}_{[A,B]}^{ih}(J) := \int_a^b q_{ih}(s)m_i(s)ds = \tilde{Q}_{[A,B]}^{ih}(b) - \tilde{Q}_{[A,B]}^{ih}(a)$$

für die Intervalle $J \subseteq [A, B]$.

Satz 3.4.1. *Seien die Annahmen (M1) bis (M9) erfüllt und die Folge (p_n) erfülle die Voraussetzung (K5).*

Für beliebige $i, h \in \mathbb{S}, i \neq h$ existiert dann eine Konstante

$$D \leq 2\sqrt{2 \cdot (2K^{ih} + \tilde{Q}_{[A,B]}^{ih}(B))},$$

so dass

$$P \left\{ \limsup_{n \rightarrow \infty} \frac{\sup_{J \subseteq [A,B], \widehat{\tilde{Q}_{[A,B]}^{ih}}(J) \leq p_n} \left| \tilde{Q}_{[A,B]}^{ih}(J) - \widehat{\tilde{Q}_{n;[A,B]}^{ih}}(J) \right|}{\sqrt{\log(n)p_n/n}} = D \right\} = 1.$$

Beweis. Um den Schätzer $\widehat{\tilde{Q}_{n;[A,B]}^{ih}}(\cdot)$ an die Form des Schätzers (2.12) anzu-

passen, machen wir für $j = 1, \dots, n$ und alle $t \in [A, B]$ folgende Setzungen:

- $S_j^1 := A$,
- Als S_j^k , $k = 2, \dots, (K^{ih} + 1)$ definieren wir den Zeitpunkt des $(k-1)$ -ten Übergangs ih oder -1 , falls dieser nicht vorkommt,
- $\Delta_j^t := \sum_{k=2}^{K^{ih}+1} \mathbf{1}_{\{A \leq S_j^k \leq t\}} = \int_A^t \mathbf{1}_{\{X_s^j=i, X_{s+ds}^j=h\}} \leq K^{ih} =: \Delta_{max}$,
- $\Psi(t) := \int_A^t P(X_s = i, X_{s+ds} = h) = \tilde{Q}_{[A,B]}^{ih}(t) \leq \tilde{Q}_{[A,B]}^{ih}(B) =: \Psi_{max}$,
- $G(\cdot) := 1$ und $G_n(\cdot) := 1$,
- $D_G := 0$.

Wegen Annahme (M7) ist $\Psi(\cdot)$ offensichtlich streng monoton wachsend. Für große n , alle $t \in [A, B]$ und ein $C > D_G$ gilt $|G_n(t) - G(t)| = 0 \leq C\sqrt{\log(n)/n}$. Bei festen $t \in [A, B]$ sind $\Delta_1^t, \dots, \Delta_n^t$ stochastisch unabhängig wegen der Unabhängigkeit der Markov-Prozesse.

Die Voraussetzung (L5) ist erfüllt wegen $G(t) = G_n(t)$. Wegen $G(t)$ ist auch die Voraussetzung (L4) erfüllt.

Nun muss noch die Voraussetzung (L6) für jedes feste $t \in [A, B]$ überprüft werden:

$$E \left(\frac{\mathbf{1}_{\{S_j^1 \leq t\}} \cdot \Delta_j^t}{G(S_j)} \right) = E(\Delta_j^t) = \int_A^t P(X_s = i, X_{s+ds} = h) = \Psi(t).$$

Die Voraussetzungen (L1) bis (L8) sind erfüllt und nach Korollar 2.2.1 existiert ein $D \leq 2\sqrt{2 \cdot (K^{ih} + \tilde{Q}_{[A,B]}^{ih}(B))}$, so dass

$$P \left\{ \limsup_{n \rightarrow \infty} \frac{\sup_{J \subseteq [A,B], \tilde{Q}_{[A,B]}^{ih}(J) \leq p_n} \left| \tilde{Q}_{[A,B]}^{ih}(J) - \widehat{\tilde{Q}_{n;[A,B]}^{ih}}(J) \right|}{\sqrt{\log(n)p_n/n}} = D \right\} = 1.$$

□

Der Kernschätzer für die kumulative Migrationsintensität kann auf der Basis der Funktion (3.32) aus der generalisierten Form (2.8) wie folgt hergeleitet werden:

$$\begin{aligned}
\widehat{\tilde{q}}_{ih;n}(t) &:= \int_A^B \frac{1}{R_n(s)} \text{Kr} \left(\frac{t-s}{R_n(s)} \right) d\widehat{\tilde{Q}}_{n;[A,B]}^{ih}(s) \\
&= \int_A^B \frac{1}{R_n(s)} \text{Kr} \left(\frac{t-s}{R_n(s)} \right) dN_s^{ih;n} \\
&= \frac{1}{n} \sum_{j=1}^n \int_A^B \frac{1}{R_n(s)} \text{Kr} \left(\frac{t-s}{R_n(s)} \right) \mathbf{1}_{\{X_s^j=i, X_{s+ds}^j=h\}}
\end{aligned} \tag{3.35}$$

Wegen Annahme (M7) sind die Annahmen (K1)-(K4) für $\tilde{Q}_{[A,B]}^{ih}(\cdot)$ und $\tilde{q}_{ih}(\cdot)$ erfüllt. Wir zeigen die Konvergenz von $\widehat{\tilde{q}}_{ih;n}(\cdot)$ wie üblich für verschiedene Glätter. Es existiert eine Konstante D_0 , so dass

$$P \left\{ \limsup_{n \rightarrow \infty} \frac{\sup_{t \in [A,B]} |\widehat{\tilde{q}}_{ih;n}(t) - \tilde{q}_{ih}(t)|}{\sqrt{\log(n)/(np_n) + p_n}} = D_0 \right\} = 1.$$

Für den linearen Glätter (2.7) mit einer konstanten Bandbreite $R_n = p_n/\tilde{c}$, $\tilde{c} > 0$ gilt $D_0 \leq \max\{D_1, D_2\}$, wobei

$$\begin{aligned}
D_1 &= 2\sqrt{2 \cdot (2K^{ih} + \tilde{Q}_{[A,B]}^{ih}(B)) [q_{ih}^{max}]^{1/2} \tilde{c}^{1/2} \text{Var}(\text{Kr})}, \\
D_2 &= L_{ih} \tilde{c}^{-1}
\end{aligned}$$

und die Kernfunktion $\text{Kr}(\cdot)$ die Voraussetzung (K6) erfüllt.

Falls die variable Bandbreite $\tilde{\Psi}(\cdot) = \tilde{Q}_{[A,B]}^{ih}(\cdot)$ gewählt wird, sind die Voraussetzungen für den Glätter identisch zu diesen von $\tilde{Q}_{[A,B]}^{ih}(\cdot)$. Es gilt $D_0 \leq \max\{D_1 + D_2, D_3\}$, wobei

$$\begin{aligned}
D_1 &= 4\sqrt{2 \cdot (2K^{ih} + \tilde{Q}_{[A,B]}^{ih}(B)) [q_{ih}^{min}]^{-2} [q_{ih}^{max}]^3 [\text{sup}(\text{Kr}) + L_{Kr} [q_{ih}^{min}]^{-1} [q_{ih}^{max}]]}, \\
D_2 &= 2\sqrt{2 \cdot (2K^{ih} + \tilde{Q}_{[A,B]}^{ih}(B)) \text{Var}(\text{Kr}) [q_{ih}^{min}]^{-1/2} [q_{ih}^{max}]^{3/2}}, \\
D_3 &= 2L_{Kr} L_{ih} [q_{ih}^{min}]^{-5} [q_{ih}^{max}]^4 + 2 \text{sup}(\text{Kr}) L_{ih} [q_{ih}^{min}]^{-4} [q_{ih}^{max}]^3 + L_{ih} [q_{ih}^{min}]^{-1}
\end{aligned}$$

und die Kernfunktion $\text{Kr}(\cdot)$ die Voraussetzungen (K6) und (K7) erfüllt.

Da die gewichtete Migrationsintensität (3.34) ein Produkt der Migrationsintensität und Zustandswahrscheinlichkeit ist, werden auch Schätzer für die Zustandswahrscheinlichkeiten $m_i(t)$ benötigt.

$$\widehat{m}_i(t) := \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{\{X_t^j=i\}} \quad (3.36)$$

Während der Schätzer (3.36) offensichtlich erwartungstreu bezüglich $m_i(t)$ ist, wird seine globale Konsistenz weiter gezeigt.

Lemma 3.4.3. *Die Annahmen (M1) bis (M9) seien erfüllt. Dann existiert eine Konstante $D \leq 2$, so dass*

$$P \left\{ \limsup_{n \rightarrow \infty} \frac{\sup_{t \in [A, B]} |m_i(t) - \widehat{m}_i(t)|}{\sqrt{\log(n)/n}} = D \right\} = 1$$

für alle $i \in \mathbb{S}$ gilt.

Beweis. Für feste $t \in [A, B]$ und $j = 1, \dots, n$ wird Folgendes gesetzt:

- $H_j := \{s : X_s^j = i\}$
- $\Gamma_j^t = \Gamma_{max} := 1$
- $G(t) := P(X_t^j = i) = m_i(t) \leq 1 =: G_{max}$

Die Voraussetzungen (G1) bis (G4) werden überprüft.

Wegen der Unabhängigkeit der vorliegenden Markov-Prozesse sind Bernoulli-Zufallsvariablen $\mathbf{1}_{\{t \in H_j\}} = \mathbf{1}_{\{X_t^j=i\}}$ bei festen t stochastisch unabhängig. Die Voraussetzung (G4) ergibt sich für jedes feste $t \in [A, B]$ wie folgt:

$$E(\mathbf{1}_{\{t \in H_j\}} \cdot \Gamma_j^t) = P(X_t^j = i) = m_i(t) = G(t).$$

Die Behauptung folgt direkt aus Korollar 2.3.1 □

Schließlich kann der konsistente Schätzer für die Migrationsintensität als das Quotient von den Schätzern für die gewichtete Migrationsintensität (3.35) und für die Zustandswahrscheinlichkeiten (3.36) wie folgt ausgedrückt werden:

$$\widehat{q}_{ih}(t) := \begin{cases} \widehat{\tilde{q}}_{ih;n}(t)/\widehat{m}_i(t) & \widehat{m}_i(t) > 0 \\ 0 & \widehat{m}_i(t) = 0 \end{cases} \quad (3.37)$$

Satz 3.4.2. *Seien die Annahmen (M1) bis (M9) erfüllt und die Folge (p_n) erfülle die Voraussetzung (K5).*

Für beliebige $i, h \in \mathbb{S}$, $i \neq h$ mit $q_{ih}(\cdot) \neq 0$ auf $[A, B]$ existiert dann eine Konstante

$$0 \leq D \leq \frac{2q_{ih}^{max} + 2D_0 + m_i^{min}D_0}{(m_i^{min})^2},$$

so dass

$$P \left\{ \limsup_{n \rightarrow \infty} \frac{\sup_{t \in [A, B]} |q_{ih}(t) - \widehat{q}_{ih}(t)|}{\varepsilon_n + \epsilon_n + \varepsilon_n \epsilon_n} = D \right\} = 1,$$

wobei $\varepsilon_n := \sqrt{\log(n)/(np_n)} + p_n$ und $\epsilon_n := \sqrt{\log(n)/n}$.

Beweis. Die Differenz $|q_{ih}(t) - \widehat{q}_{ih}(t)|$ kann weiter folgenderweise ausgeschrieben werden:

$$\begin{aligned} \sup_{t \in [A, B]} |q_{ih}(t) - \widehat{q}_{ih}(t)| &= \sup_{t \in [A, B]} \left| \frac{\tilde{q}_{ih}(t)}{m_i(t)} - \frac{\widehat{\tilde{q}}_{ih;n}(t)}{\widehat{m}_i(t)} \right| \leq \\ &\sup_{t \in [A, B]} \left| \frac{\tilde{q}_{ih}(t)}{m_i(t)} - \frac{\widehat{\tilde{q}}_{ih;n}(t)}{m_i(t)} \right| + \sup_{t \in [A, B]} \left| \frac{\widehat{\tilde{q}}_{ih;n}(t)}{m_i(t)} - \frac{\widehat{\tilde{q}}_{ih;n}(t)}{\widehat{m}_i(t)} \right| \end{aligned} \quad (3.38)$$

Nach Lemma A.1.1 existiert ein $\tilde{C}_1 > D_0$, so dass

$$\sup_{t \in [A, B]} \left| \tilde{q}_{ih}(t) - \widehat{\tilde{q}}_{ih;n}(t) \right| \leq \tilde{C}_1 \varepsilon_n \quad (3.39)$$

für große n f.s. gilt.

Für den ersten Summand aus (3.38) gilt dann folgende Ungleichung:

$$\sup_{t \in [A, B]} \left| \frac{\tilde{q}_{ih}(t)}{m_i(t)} - \frac{\widehat{\tilde{q}_{ih;n}}(t)}{m_i(t)} \right| \leq \frac{\sup_{t \in [A, B]} |\tilde{q}_{ih}(t) - \widehat{\tilde{q}_{ih;n}}(t)|}{m_i^{\min}} \leq C_1 \epsilon_n$$

für ein $C_1 > D_0/m_i^{\min}$.

Aus (3.39) können Schranken für $\widehat{\tilde{q}_{ih;n}}(t)$ wie folgt hergeleitet werden:

$$\widehat{\tilde{q}_{ih;n}}(t) \leq \tilde{q}_{ih}(t) + \tilde{C}_1 \epsilon_n \leq \tilde{q}_{ih}^{\max} + \tilde{C}_1 \epsilon_n$$

Aus Lemma A.1.1 und Satz 3.4.3 folgt ebenfalls, dass ein $\tilde{C}_2 > 2$ existiert, so dass

$$\sup_{t \in [A, B]} |m_i(t) - \widehat{m}_i(t)| \leq \tilde{C}_2 \epsilon_n \quad (3.40)$$

für große n f.s. gilt.

Die Schranken für $\widehat{m}_i(t)$ sind aus (3.40) ersichtlich:

$$m_i(t) - \tilde{C}_2 \epsilon_n \leq \widehat{m}_i(t) \leq m_i(t) + \tilde{C}_2 \epsilon_n.$$

Bei großen n kann auch die Ungleichung

$$\frac{1}{\widehat{m}_i(t)} \leq \frac{1}{m_i(t) - \tilde{C}_2 \epsilon_n} \leq \frac{1}{m_i^{\min} - \tilde{C}_2 \epsilon_n}$$

geschrieben werden, da $\epsilon_n \rightarrow 0$ und $m_i(t)$ auf $[A, B]$ großer Null ist.

Für den zweiten Summand aus (3.38) gilt dann

$$\begin{aligned} \sup_{t \in [A, B]} \left| \frac{\widehat{\tilde{q}_{ih;n}}(t)}{m_i(t)} - \frac{\widehat{\tilde{q}_{ih;n}}(t)}{\widehat{m}_i(t)} \right| &\leq (\tilde{q}_{ih}^{\max} + \tilde{C}_1 \epsilon_n) \cdot \sup_{t \in [A, B]} \left| \frac{1}{m_i(t)} - \frac{1}{\widehat{m}_i(t)} \right| \\ &\leq \frac{\tilde{q}_{ih}^{\max} + \tilde{C}_1 \epsilon_n}{m_i^{\min}} \cdot \sup_{t \in [A, B]} \left| \frac{\widehat{m}_i(t) - m_i(t)}{\widehat{m}_i(t)} \right| \\ &\leq \frac{\tilde{q}_{ih}^{\max} + \tilde{C}_1 \epsilon_n}{m_i^{\min}} \cdot \frac{\tilde{C}_2 \epsilon_n}{m_i^{\min} - \tilde{C}_2 \epsilon_n} \end{aligned}$$

für große n fast sicher.

Der Summand $\tilde{C}_2 \epsilon_n$ im Nenner kann bei großen n vernachlässigt werden.

Die obere Ungleichung wird dann wie folgt vereinfacht:

$$\sup_{t \in [A, B]} \left| \frac{\widehat{\tilde{q}}_{ih;n}(t)}{m_i(t)} - \frac{\widehat{\tilde{q}}_{ih;n}(t)}{\widehat{m}_i(t)} \right| \leq \frac{\tilde{C}_2(q_{ih}^{max} + \tilde{C}_1 \varepsilon_n) \epsilon_n}{(m_i^{min})^2} = C_2(q_{ih}^{max}/m_i^{min} + C_1 \varepsilon_n) \epsilon_n,$$

wobei $C_2 > 2/m_i^{min}$.

Für ein $C > \frac{2q_{ih}^{max} + 2D_0 + m_i^{min} D_0}{(m_i^{min})^2}$ gilt dann

$$\begin{aligned} \sup_{t \in [A, B]} |q_{ih}(t) - \widehat{q}_{ih}(t)| &= \sup_{t \in [A, B]} \left| \frac{\tilde{q}_{ih}(t)}{m_i(t)} - \frac{\widehat{\tilde{q}}_{ih;n}(t)}{m_i(t)} \right| + \sup_{t \in [A, B]} \left| \frac{\widehat{\tilde{q}}_{ih;n}(t)}{m_i(t)} - \frac{\widehat{\tilde{q}}_{ih;n}(t)}{\widehat{m}_i(t)} \right| \\ &\leq C(\varepsilon_n + \epsilon_n + \varepsilon_n \epsilon_n) \end{aligned}$$

bei großen n fast sicher und die Behauptung folgt aus Lemma A.1.1. \square

Kapitel 4

Simulationen stochastischer Prozesse

Neben den analytischen Methoden stellen Simulationen eine alternative Möglichkeit dar, das Konvergenzverhalten der Kernschätzer zu überprüfen. In dieser Arbeit widmen wir die Aufmerksamkeit den Kernschätzern für die Hazardrate bei linktrunkierten und rechtszensierten Daten aus Kapitel 3.3 und für die Migrationsintensitäten aus Kapitel 3.4. Im Unterschied zu den beiden genannten Schätzern ist das Konvergenzverhalten der empirischen Verteilungsfunktionen wohl bekannt und stellt für uns weniger Interesse dar. Es wird also angestrebt, die Simulationen der Überlebensprozesse bzw. Markov-Prozesse mit bekannten funktionalen Charakteristiken durchzuführen und dann die Kernschätzer dieser Funktionen zu berechnen.

Die Simulationen werden schrittweise durchgeführt. Beginnend mit 10 000 Beobachtungen wird die Beobachtungsanzahl auf jedem Schritt immer wieder um 20 000 Beobachtungen erhöht, bis die Anzahl von 90 000 Beobachtungen erreicht ist. Aus allen bereits erzeugten Pfaden werden funktionale Charakteristiken geschätzt und mit den theoretischen grafisch verglichen. Dabei werden konstante und variable Bandbreiten bei $k_{min} = 10$ und $k_{min} = 100$ (siehe Abschnitt 2.1.3) untersucht. Verschiedene Bandbreiten lassen den Einfluss

von k_{min} auf die Anpassungsgüte visualisieren.

Wie aus Kapitel 2.1 bekannt, hängt die Form des Kernschätzers vom gewählten Glätter, sowie von der Folge (p_n) mit $p_n > 0$, $p_n \rightarrow 0$ und $np_n/\log(n) \rightarrow \infty$ ab. Die Glätter und die Folge (p_n) werden jeweils gemäß den Abschnitten 2.1.2 und 2.1.3 gewählt.

Die Formel (2.11) liefert die optimale Folge (p_n) als:

$$p_n^{(\tilde{n})} = k_{min} \cdot \sup_{s \in [A, B]} (d\Psi_{\tilde{n}}(s)) (\log \tilde{n}/\tilde{n})^{-a} (\log n/n)^a$$

für alle $n \geq \tilde{n} \in \mathbb{N}$ und $0 < a < 1$. Für den Schätzer $\Psi_n(\cdot)$ in der Form (2.12) gilt $\sup_{s \in [A, B]} (d\Psi_n(s)) = \frac{\Delta_{max}}{n \cdot \inf_{S_i^1 \in [A, B], i=1, \dots, n} G_n(S_i^1)}$ und folglich

$$p_n^{(\tilde{n})} = \frac{k_{min} \cdot \Delta_{max}}{\inf_{S_i^1 \in [A, B], i=1, \dots, \tilde{n}} G_{\tilde{n}}(S_i^1)} \cdot \frac{1}{(\log \tilde{n})^a \tilde{n}^{1-a}} \cdot (\log n/n)^a. \quad (4.1)$$

Daher ist $\sup_{s \in [A, B]} (d\Psi_{\tilde{n}}(s)) \cdot (\log \tilde{n}/\tilde{n})^{-a} \rightarrow 0$ monoton fallend. Wir setzen weiter $a = 0.9$ und $\tilde{n} = 10\,000$, wobei $(\log \tilde{n})^{-a} \tilde{n}^{a-1} = 0.11$.

Da die Schätzung für konstante und variable Bandbreiten durchgeführt wird, muss eine beschränkte Kernfunktion mit $\sup(\text{Kr}) < \infty$ gewählt werden. Daher wird weiter die Epanechnikov-Kernfunktion $\text{Kr}(x) = \frac{3}{4}(1-x^2)\mathbf{1}_{\{|x| \leq 1\}}$ verwendet.

4.1 Überlebensprozess

Die Simulation in diesem Kapitel dient zu einer alternativen Untersuchung des Konvergenzverhaltens des Kernschätzers für die Hazardrate bei linkstrun-kierten und rechtszensierten Daten. Dabei ist ein Simulationsalgorithmus zur Erzeugung von linkstrun-kierten und rechtszensierten Beobachtungen vom Interesse.

Der Überlebensprozess aus Abschnitt 3.3.1 kann recht einfach auf der Basis der Annahmen (Z1) bis (Z7) simuliert werden. Die Beobachtungsgrenzen,

sowie die Lebensdauer sind dann stochastisch unabhängige Zufallsvariablen, deren Realisationen leicht erzeugt werden können. Danach müssen nur die unbeobachtbaren Kombinationen $C < L$, sowie $T < L$ eliminiert werden. Zu beachten ist auch, dass die Zufallsvariablen T und C positiv sind.

Der Kernschätzer für die Hazardrate wird durch einen grafischen Vergleich mit der theoretischen Hazardrate untersucht. Die Simulation muss dabei auf der exakten funktionalen Angabe der Hazardrate basieren. Damit die Erzeugung der Beobachtungen der Lebensdauer möglich ist, müssen die Verteilungsfunktion der Lebensdauer und ihre exakte Inverse aus der Hazardrate berechenbar sein.

Der Zusammenhang zwischen Hazardrate und Verteilung von T ist aus Abschnitt 3.3.2 bereits bekannt. Die Formel (3.3) kann wie folgt umgeschrieben werden:

$$F^T(t) = 1 - e^{-\int_0^t \lambda(s) ds}.$$

Damit die Funktion $F^T(t)$ und ihre Inverse eine exakte Darstellung haben, muss offensichtlich das exakte Integral $\int_0^t \lambda(s) ds$ existieren. Die Hazardrate $\lambda(t)$ wird dementsprechend gewählt. Als eine mögliche Wahl passt dabei

$$\lambda(t) = \begin{cases} \frac{1}{t+1} & t \geq 0 \\ 0 & \text{sonst.} \end{cases} \quad (4.2)$$

Die folgende exakte Verteilung und ihre exakte Inverse ergeben sich:

$$F^T(t) = \begin{cases} 1 - \frac{1}{t+1} & t \geq 0 \\ 0 & \text{sonst} \end{cases} \quad (4.3)$$

und

$$[F^T]^{-1}(y) = \frac{1}{1-y} - 1. \quad (4.4)$$

Für die Beobachtungsschranken L und C sind die wohl bekannten Verteilungen $L \sim \mathcal{N}(0, 1)$ und $C \sim \mathcal{LN}(0, 36)$ ganz akzeptabel. Es gelten dabei offensichtlich $F^L(0) > 0$ und $F^X(5) < 1$ wegen der Unbeschränktheit von $F^T(\cdot)$ und $F^C(\cdot)$, so dass die Konvergenz auf $[0, 5]$ gezeigt werden kann. Die Voraussetzungen (Z1) bis (Z7) sind damit erfüllt.

Zur Berechnung des Kernschätzers aus den simulierten Daten ist nur noch die Folge (p_n) zu spezifizieren. Für die Ausgangsstichprobe der Größe $m = 10\,000$ und für $a = 0.9$ wird die optimale Folge (4.1) als

$$p_n^{(10\,000)} = 1.1 \cdot k_{min} \cdot (\log n/n)^{0.9},$$

dargestellt, wobei $\inf_{X_i \in [0,5], i=1, \dots, 10\,000} G_{10\,000}(X_i) = 0.1$ in der aktuellen Simulation gilt und $\Delta_{max} = 1$ gemäß dem Beweis zum Satz 3.3.1 ist. Bei der variablen Bandbreite wird die kumulative Hazardrate (3.3) als Glätter gewählt. Weiter sind Kernschätzer für die theoretische Hazardrate (4.2) bei verschiedenen Bandbreiten und Datenumfängen dargestellt.

In Abbildungen 4.1 und 4.2 fällt sofort eine viel bessere Anpassung für die konstante Bandbreite auf, so dass die Schätzungen von der theoretischen Funktion nur schwer zu unterscheiden sind. Bei der variablen Bandbreite sind dagegen große Fluktuationen des Schätzers zu bemerken. Dies wird durch die Unterschätzung der variablen Bandbreite erklärt. Während die variable Bandbreite immer an die kleinste Anzahl der Sprünge k_{min} angepasst wird, nimmt die konstante Bandbreite den größten Wert vom ganzen Zeitintervall an und wird dadurch viel besser angepasst. Offensichtlich ist $k_{min} = 100$ viel zu wenig für vorliegenden Datenumfänge. Durch die Erhöhung der Mindestanzahl der Sprünge k_{min} für die variable Bandbreite wird das Problem der Unterschätzung gelöst. Beim hinreichend großen k_{min} liefert die variable Bandbreite eine gute Anpassung. Im Gegenteil dazu kann die Erhöhung dieses Parameters für die konstante Bandbreite zur Verzerrung des Schätzers führen.

Ansonsten weisen beide Abbildungen eine ganz gute Anpassung auf. Die

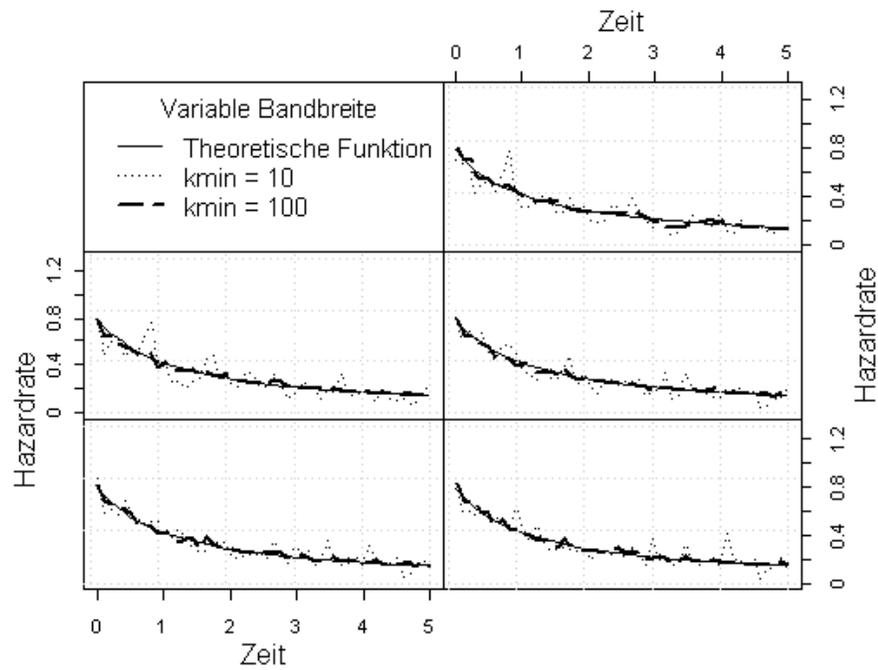


Abbildung 4.1: Kernschätzer für $\lambda(t)$ bei variabler Bandbreite mit $k_{min} = 10$ und $k_{min} = 100$ für verschiedene Datenumfänge.

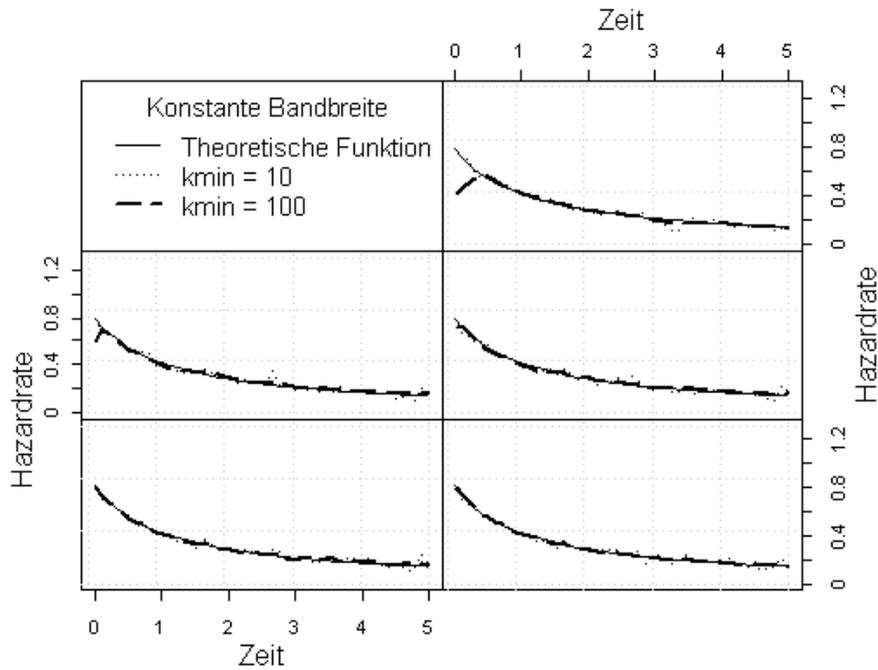


Abbildung 4.2: Kernschätzer für $\lambda(t)$ bei konstanter Bandbreite mit $k_{min} = 10$ und $k_{min} = 100$ für verschiedene Datenumfänge.

geschätzten Funktionswerte schwanken gleichmäßig um die theoretischen und die Ausschläge werden kleiner mit der Vergrößerung der Beobachtungsanzahl.

4.2 Markov-Prozess

Die Konvergenz des Kernschätzers für die Migrationsintensitäten steht im Hintergrund dieses Kapitels. Das Ziel ist dabei eine solche Simulation des inhomogenen Markov-Prozesses zu entwickeln, bei der auch die Übergangsintensitäten, sowie alle nötigen Wahrscheinlichkeitsverteilungen in jedem $t \in [0, \infty)$ exakt berechenbar sind. Wie wir weiter sehen, ist diese Aufgabe im Unterschied zu den Überlebensprozessen nicht trivial. Weder eine exakte Angabe des Markov-Prozesses ist bekannt, noch ein deutlicher Simulationsalgorithmus ist in der Literatur zu finden. Daher wird die Aufgabe in zwei geteilt und in zwei Abschnitten vorgestellt.

4.2.1 Testprozess aus exakten Migrationsintensitäten

In dieser Arbeit ist es besonders wichtig die Migrationsintensitäten möglichst exakt anzugeben, damit der direkte Vergleich von theoretischen Funktionen mit ihren Schätzern ermöglicht wird. Dabei ist die Angabe der Matrix von Migrationsintensitäten ziemlich problematisch, da komplizierte Zusammenhänge dieser Matrix zugrunde liegen. Recht einfach ist die direkte Angabe von Zustandswahrscheinlichkeiten. Allerdings ist die Herleitung von Migrationsintensitäten aus dem Differenzialsystem (3.27) generell nur numerisch möglich, was die Gefahr der Verzerrung der theoretischen Funktion erhöht. In diesem Abschnitt wird deswegen die Aufmerksamkeit der exakten Angabe von Matrizen $\mathbf{m}(0)$ und $\mathbf{Q}(t)$ gewidmet. Die Angabe eines Markov-Prozesses, bei der die Migrationsintensitäten und Zustandswahrscheinlichkeiten in jedem $t \geq 0$ exakt berechenbar sind, wird als Testprozess bezeichnet.

An dieser Stelle konzentrieren wir uns nur auf den Markov-Prozessen mit einem absorbierenden Zustand, dem sogenannten „Todeszustand“. Der

biologische Tod oder die Pleite können z. B. durch einen solchen Zustand beschrieben werden. Die Modelle mit einem absorbierenden Zustand sind auch in der Realität überwiegend vertreten, weil die meisten biologischen, politischen oder wirtschaftlichen Prozesse eine endliche Dauer haben.

Aus dieser Vorstellung folgt, dass die Wahrscheinlichkeit des „Todeszustandes“ mit der Zeit gegen Eins strebt, wobei die Wahrscheinlichkeiten aller anderen Zustände nach Null streben sollen. Wir nehmen an, dass der „Todeszustand“ der letzte Zustand K auf dem Zustandsraum \mathbb{S} ist. Dann kann diese Bedingung für die Zustandswahrscheinlichkeiten wie folgt ausgedrückt werden:

$$\mathbf{m}(t) \rightarrow (0, 0, \dots, 0, 0, 1)'. \quad (4.5)$$

Da die Auferstehung durch das Modell mit einem absorbierenden Zustand nicht vorgesehen ist, gibt es auch kein Zurück aus dem „Todeszustand“ und folglich gilt

$$q_{Kh}(t) = 0, \text{ für alle } h \in \mathbb{S} \quad (4.6)$$

Das weitere Ziel ist solche Funktionen $m_i(t)$ zu finden, für die Funktionen $q_{ih}(t)$, $i, h \in \mathbb{S}$ bei allen $t \geq 0$ exakt berechenbar sind. Die folgende Darstellung für $\mathbf{m}(t)$ wird angenommen:

$$\mathbf{m}(t) := \left(\frac{c_1}{\sum_{j=1}^{K-1} c_j + \theta(t)}, \dots, \frac{c_{K-1}}{\sum_{j=1}^{K-1} c_j + \theta(t)}, \frac{\theta(t)}{\sum_{j=1}^{K-1} c_j + \theta(t)} \right)',$$

wobei c_1, \dots, c_{K-1} Koeffizienten und $\theta(t)$ eine reellwertige Funktion sind.

Die Eigenschaften von $\theta(t)$ folgen aus der Tatsache, dass der Vektor $\mathbf{m}(t)$ die Wahrscheinlichkeiten enthält, wobei

$$0 < \frac{c_i}{\sum_{j=1}^{K-1} c_j + \theta(t)} < 1$$

und

$$0 < \frac{\theta(t)}{\sum_{j=1}^{K-1} c_j + \theta(t)} < 1$$

für alle $i \in \mathbb{S}$ und alle $t \in [0, \infty)$ gelten müssen. Somit ist die Funktion $\theta(t)$ entweder streng positiv oder streng negativ auf $[0, \infty)$, während die Koeffizienten c_1 bis c_{K-1} nicht gleich Null sind und dieselben Vorzeichen wie $\theta(t)$ haben.

Um weitere Eigenschaften von $\theta(t)$ deutlich zu machen, beziehen wir uns auf die angenommene Konvergenzeigenschaft (4.5). Damit $c_i / (\sum_{j=1}^{K-1} c_j + \theta(t)) \rightarrow 0$ und $\theta(t) / (\sum_{j=1}^{K-1} c_j + \theta(t)) \rightarrow 1$ gelten, muss offensichtlich $|\theta(t)| \rightarrow \infty$ gelten. Auch die Differenzierbarkeit von $\theta(t)$ auf $[0, \infty)$ wird vorausgesetzt, damit die erste Ableitung von $\mathbf{m}(t)$ existiert.

Bei bekannten Anfangswahrscheinlichkeiten $\mathbf{m}(0)$ lässt sich der Vektor $\mathbf{c} := (c_1, \dots, c_{K-1})'$ aus dem Gleichungssystem

$$\begin{cases} c_1(1 - m_1(0)) & -m_1(0) \sum_{i=2}^{K-1} c_i = \theta(0)m_1(0) \\ & \vdots \\ c_j(1 - m_j(0)) & -m_j(0) \sum_{i=1, i \neq j}^{K-1} c_i = \theta(0)m_j(0) \\ & \vdots \\ c_{K-1}(1 - m_{K-1}(0)) & -m_{K-1}(0) \sum_{i=1}^{K-2} c_i = \theta(0)m_{K-1}(0) \end{cases} \quad (4.7)$$

berechnen. Das Gleichungssystem (4.7) hat folgende Matrix-Form:

$$(\mathbf{I}_{K-1} - \mathbf{m}^* \cdot \mathbf{1}'_{K-1}) \cdot \mathbf{c} = \theta(0) \cdot \mathbf{m}^*$$

mit der Lösung

$$\mathbf{c} = \theta(0) \cdot (\mathbf{I}_{K-1} - \mathbf{m}^* \cdot \mathbf{1}'_{K-1})^{-1} \cdot \mathbf{m}^*, \quad (4.8)$$

wobei der Vektor \mathbf{m}^* als

$$\mathbf{m}^* := (m_1(0), \dots, m_{K-1}(0))' \quad (4.9)$$

definiert wird.

Der weitere Rechenprozess wird erleichtert, indem $\theta(t)$ durch $\theta^*(t) := \sum_{j=1}^{K-1} c_j + \theta(t)$ ersetzt wird. Die ersten Ableitungen beider Funktionen bleiben gleich, so dass $\dot{\theta}(t) = \dot{\theta}^*(t)$ gilt.

Der neue Vektor $\mathbf{m}(t)$ und seine erste Ableitung haben letztlich folgende Gestalt:

$$\mathbf{m}(t) = \left(\frac{c_1}{\theta^*(t)}, \dots, \frac{c_{K-1}}{\theta^*(t)}, 1 - \frac{\sum_{j=1}^{K-1} c_j}{\theta^*(t)} \right)' \quad (4.10)$$

und

$$\frac{d\mathbf{m}(t)}{dt} = \left(-\frac{c_1 \dot{\theta}^*(t)}{(\theta^*(t))^2}, \dots, -\frac{c_{K-1} \dot{\theta}^*(t)}{(\theta^*(t))^2}, \frac{\dot{\theta}^*(t) \sum_{j=1}^{K-1} c_j}{(\theta^*(t))^2} \right)'.$$

Aus der Formel (3.29) lassen sich die gewichteten Summen der Übergangintensitäten für alle $h = 1, \dots, K - 1$ ausdrücken:

$$\sum_{i=1}^{K-1} \frac{c_i}{c_h} q_{ih}(t) = -\frac{\dot{\theta}^*(t)}{\theta^*(t)}. \quad (4.11)$$

Bekanntlich, stellen die Diagonalelemente der Matrix $\mathbf{Q}(t)$ die negierten Summen der zugehörigen Zeilenelemente dar:

$$q_{ii}(t) = - \sum_{h=1, h \neq i}^K q_{ih}(t). \quad (4.12)$$

Daher besteht der lineare Zusammenhang zwischen mindestens einer Spalte der Matrix $\mathbf{Q}(t)$ und allen anderen. Die Spalte K kann somit entfernt werden. Da die Zeile K dieser Matrix wegen der Eigenschaft (4.6) nur Nullen

enthält, ist sie ebenfalls nicht vom Interesse.

Die neue $(K - 1) \times (K - 1)$ -Matrix $\mathbf{Q}^*(t)$ wird folgenderweise definiert:

$$(\mathbf{Q}^*(t))_{ih} := \frac{c_i}{c_h} q_{ih}(t) = q_{ih}^*(t).$$

Die Hauptdiagonalelemente von $\mathbf{Q}^*(t)$ sind negativ und die sonstigen Elemente sind positiv. Damit die Elemente der aus $\mathbf{Q}(t)$ ausgeschlossenen Spalte K positiv sind, müssen Summen der mit c_h/c_i gewichteten Zeilenelemente von $\mathbf{Q}^*(t)$ wegen (4.12) immer kleiner Null sein. Die Summen der Spaltenelemente müssen dabei gemäß der Eigenschaft (4.11) immer gleich $-\dot{\theta}^*(t)/\theta^*(t)$ sein.

Auch die kumulative gewichtete Migrationsintensität (3.31) muss bei beliebigen $t \in [0, \infty)$ exakt berechenbar sein. Wegen der Bedingung (4.5) und weil $q_{ih}(\cdot)$ und $m_i(\cdot)$ gemäß der Definition positive Funktionen sind, wird

$$0 < \int_0^\infty q_{ih}(s)m_i(s)ds < \infty \quad (4.13)$$

für alle $i, h \in \mathbb{S} \setminus K$, $i \neq h$ vorausgesetzt. Offensichtlich muss das Integral $\left| \int_0^\infty \frac{q_{ih}^*(s)}{\theta^*(s)} ds \right|$ existieren.

Für alle $i, h \in \mathbb{S} \setminus K$, $i \neq h$ werden weiter die gewichteten Migrationsintensitäten als $\tilde{q}_{ih}(t) := q_{ih}(t)m_i(t) = \frac{c_h q_{ih}^*(t)}{\theta^*(t)}$ definiert, wobei $0 < \int_0^\infty \tilde{q}_{ih}(s)ds < \infty$ gilt. Dann können auch Elemente von $\mathbf{Q}^*(t)$ über $\tilde{q}_{ih}(t)$ wie folgt ausgedrückt werden:

$$q_{ih}^*(t) = \frac{\tilde{q}_{ih}(t)\theta^*(t)}{c_h}. \quad (4.14)$$

Über $q_{ih}^*(t)$ können nun die Hauptdiagonalelemente von $\mathbf{Q}^*(t)$ folgenderweise berechnet werden:

$$q_{hh}^*(t) = -\frac{\dot{\theta}^*(t)}{\theta^*(t)} - \frac{\theta^*(t)}{c_h} \sum_{j=1, j \neq h}^{K-1} \tilde{q}_{jh}(t) \leq 0, \quad (4.15)$$

wobei

$$-c_h \frac{\dot{\theta}^*(t)}{(\theta^*(t))^2} \leq \sum_{i=1, i \neq h}^{K-1} \tilde{q}_{ih}(t) \quad (4.16)$$

gilt.

Wenn die Funktionen $\theta^*(t)$ und $\dot{\theta}^*(t)$ dieselben Vorzeichen haben, gilt die Ungleichung (4.16) immer. Weiter sind die gewichteten Summen der Zeilenelemente von dem Interesse. Nach der Substitution von (4.15) ergibt sich folgende Ungleichung für die j -te Zeile:

$$\sum_{h=1}^{K-1} \frac{c_h}{c_j} q_{jh}^*(t) = \frac{\theta^*(t)}{c_j} \sum_{h=1, h \neq j}^{K-1} \tilde{q}_{jh}(t) - \frac{\dot{\theta}^*(t)}{\theta^*(t)} - \frac{\theta^*(t)}{c_j} \sum_{i=1, i \neq j}^{K-1} \tilde{q}_{ij}(t) \leq 0,$$

wobei die Bedingung

$$-c_j \frac{\dot{\theta}^*(t)}{(\theta^*(t))^2} \leq \sum_{i=1, i \neq j}^{K-1} (\tilde{q}_{ij}(t) - \tilde{q}_{ji}(t))$$

für jedes $j = 1, \dots, K-1$ gelten muss. Bei den gleichen Vorzeichen von $\theta^*(t)$ und $\dot{\theta}^*(t)$ genügt es, dass $\tilde{q}_{ij}(t) = \tilde{q}_{ji}(t)$ für alle $i, j \in \mathbb{S} \setminus K, i \neq j$ gilt.

Die Übergangintensitäten $q_{iK}(t)$ sind dann aus den Formeln (4.14) und (4.15) folgenderweise zu bestimmen:

$$q_{iK}(t) = 0 - \sum_{h=1, h \neq i}^{K-1} q_{ih}(t) - q_{ii}(t) = \frac{\dot{\theta}^*(t)}{\theta^*(t)}. \quad (4.17)$$

Offensichtlich kann nun der erwünschte Testprozess nur über Funktionen $\theta(t)$ und $\tilde{q}_{ih}(t)$, $i, h \in \mathbb{S} \setminus K, i \neq h$ ermittelt werden.

Weiter wird die Reihenfolge zur Erstellung eines Testprozesses aus der bekannten Matrix der Übergangintensitäten, sowie die notwendigen Voraussetzungen an die Funktionen zusammengefasst.

1. Wähle eine Funktion $\theta(t)$, die folgende Voraussetzungen auf $[0, \infty)$ er-

füllt:

- (a) $\theta(t)$ ist entweder streng positiv oder streng negativ
- (b) $\lim_{t \rightarrow \infty} |\theta(t)| = \infty$
- (c) $\theta(t)$ ist differenzierbar
- (d) $\operatorname{sgn}(\theta(t)) = \operatorname{sgn}(\dot{\theta}(t))$

2. Bestimme die Anzahl der Zustände K und somit den Zustandsraum $\mathbb{S} = \{1, \dots, K\}$
3. Erstelle die Matrix $\mathbf{D}(t)$

$$(\mathbf{D}(t))_{ih} := \begin{cases} \tilde{q}_{ih}(t) & i, h \in \mathbb{S} \setminus \{K\}, i < h \\ \tilde{q}_{hi}(t) = \tilde{q}_{ih}(t) & i, h \in \mathbb{S} \setminus \{K\}, i > h \\ 0 & \text{sonst} \end{cases}$$

wobei Funktionen $\tilde{q}_{ih}(t)$ folgende Voraussetzungen auf $[0, \infty)$ erfüllen.

- (a) $\tilde{q}_{ih}(t) \geq 0$
 - (b) $0 < \int_0^\infty \tilde{q}_{ih}(s) ds < \infty$
4. Bestimme einen Vektor der Anfangswahrscheinlichkeiten $\mathbf{m}(0)$ und gemäß (4.9) den Hilfsvektor \mathbf{m}^* .
 5. Durch (4.8) berechne den Vektor der Koeffizienten $\mathbf{c} = (c_1, \dots, c_{K-1})'$ als

$$\mathbf{c} = \theta(0) \cdot (\mathbf{I}_{K-1} - \mathbf{m}^* \cdot \mathbf{1}'_{K-1})^{-1} \cdot \mathbf{m}^*$$

6. Berechne $\mathbf{Q}^*(t)$ als

$$\mathbf{Q}^*(t) = - \frac{\dot{\theta}(t)}{\sum_{i=1}^{K-1} c_i + \theta(t)} \cdot \mathbf{I}_{K-1} + \left(\sum_{i=1}^{K-1} c_i + \theta(t) \right) (\mathbf{D}(t) - \text{diag}(\mathbf{D}(t) \cdot \mathbf{1}_{K-1})) \cdot (\text{diag}(\mathbf{c}))^{-1}$$

7. Berechne die letzte Spalte von $\mathbf{Q}(t)$ als

$$\mathbf{q}_{.K} = \mathbf{0}_{K-1} - [\text{diag}(\mathbf{c})]^{-1} \cdot \mathbf{Q}^*(t) \cdot \text{diag}(\mathbf{c}) \cdot \mathbf{1}_{K-1}$$

8. Konstruiere die Übergangsmatrix $\mathbf{Q}(t)$ als

$$\mathbf{Q}(t) = \left(\begin{array}{c|c} [\text{diag}(\mathbf{c})]^{-1} \cdot \mathbf{Q}^*(t) \cdot \text{diag}(\mathbf{c}) & \mathbf{q}_{.K} \\ \hline \mathbf{0}'_{K-1} & 0 \end{array} \right)$$

Letztlich stellt sich natürlich die Frage, ob die Bedingungen für die Funktionen $\theta(t)$ und $\tilde{q}_{ih}(t)$ erfüllbar sind. Um alle Zweifel auszuschließen werden ein paar Beispiele angeführt die alle Annahmen erfüllen.

Für $\theta(t)$ kann sowohl eine exponentielle Funktion e^{a+bt} als auch eine polynomiale Funktion $at^2 + bt + c$ mit Parametern $a, b, c \in \mathbb{R}_+$ gewählt werden. Beide Funktionen sind positiv, differenzierbar und streben nach unendlich auf $[0, \infty)$. Offensichtlich sind die ersten Ableitungen beider Funktionen, nämlich be^{a+bt} und $2at + b$, genauso wie Ausgangsfunktionen positiv auf $[0, \infty)$.

Es passen gut Funktionen der Art $\tilde{q}_{ih}(t) = (t + a_{ih})^{-2b_{ih}}$ mit Parametern $a_{ih} \in \mathbb{R}_+$ und $b_{ih} \in \mathbb{N}_+$. Diese Funktionen besitzen auf $[0, \infty)$ endliche positive

Integrale

$$\begin{aligned} \int_0^\infty \tilde{q}_{ih}(t) dt &= \int_0^\infty (t + a_{ih})^{-2b_{ih}} dt = -\frac{(t + a_{ih})^{-2b_{ih}+1}}{2b_{ih} - 1} \Big|_0^\infty \\ &= \frac{a_{ih}^{-2b_{ih}+1}}{2b_{ih} - 1} < \infty \end{aligned} \quad (4.18)$$

Auch bei Funktionen $\tilde{q}_{ih}(t) = \dot{\theta}^*(t)/(\theta^*(t))^2$ sind alle Voraussetzungen an die gewichteten Migrationsintensitäten erfüllt. Folgendes endliches Integral ergibt sich:

$$\int_0^\infty \tilde{q}_{ih}(t) dt = \int_0^\infty \frac{\dot{\theta}^*(t)}{(\theta^*(t))^2} dt = -\frac{1}{\theta^*(t)} \Big|_0^\infty = \frac{1}{\theta^*(0)} < \infty \quad (4.19)$$

Die Matrix der Migrationsintensitäten wird dann nur durch die Parameter der Funktion $\theta(t)$ definiert, wobei die Komplexität des Testprozesses reduziert wird.

4.2.2 Simulationsalgorithmus

Wie aus Abschnitt 3.4.1 bekannt, kann ein inhomogener Markov-Prozess über die Anfangswahrscheinlichkeiten $\mathbf{m}(0)$ und Migrationsintensitäten $\mathbf{Q}(t)$ definiert werden. Leider geben diese zwei Matrizen keine Auskunft darüber, wie unabhängige Realisationen des gegebenen Markov-Prozesses erzeugt werden können. Offensichtlich müssen die Funktionen aus der Basisdefinition so transformiert werden, dass die schrittweise Erzeugung von unabhängigen Realisationen des Markov-Prozesses ermöglicht wird. Daher wird weiter angestrebt, einen Algorithmus zur Simulation von Markov-Prozessen mit integrierten Anfangswahrscheinlichkeiten und Migrationsintensitäten zu entwickeln.

Für eine Realisation des Markov-Prozesses müssen nur die Reihenfolge der Zustände und die Dauer des Verbleibens in jedem Zustand erzeugt werden. Der erste Zustand wird aus den Anfangswahrscheinlichkeiten erhalten, die

in der Basisdefinition festgelegt sind. Die Verweildauer im aktuellen Zustand kann aus einer Verteilung erzeugt werden, die auf das Verbleiben im aktuellen Zustand in t bedingt ist. Die Entscheidung darüber, welcher Zustand als nächster kommt, wird anhand der bedingten Wahrscheinlichkeiten getroffen. Diese sind auf den Übergang vom aktuellen Zustand in den anderen zwischen t und $t + dt$ bedingt. Die drei oben genannten Typen von Verteilungen sind ausreichend zum Simulieren eines Markov-Prozesses. Die Herleitung dieser Verteilungen aus den Zustandswahrscheinlichkeiten und Migrationsintensitäten wird weiter in diesem Abschnitt erläutert.

Die Verweildauer im Zustand i wird durch die Zufallsvariable U_t^i bezeichnet. Dabei wird angenommen, dass der Prozess sich im aktuellen Zustand i im Zeitpunkt t befindet. Die bedingte Verteilung von U_t^i kann durch die Funktion $G_t^i(u) := P(U_t^i \geq u \mid X_t = i) = P(X_s = i, t \leq s \leq t + u \mid X_t = i)$ aus Johansen (1978) ausgedrückt werden.

Lemma 4.2.1. *Sei $\mathbb{S} := \{1, \dots, K\}$ für ein $K \in \mathbb{N}_{\geq 2}$ ein Zustandsraum und stochastisch unabhängige Markov-Prozesse $\{X_t^1 \in \mathbb{S}\}_{t \in [0, \infty)}, \dots, \{X_t^n \in \mathbb{S}\}_{t \in [0, \infty)}$ mit integrierten Migrationsintensitäten $\mathbf{Q}(t)$ und Anfangswahrscheinlichkeiten $\mathbf{m}(0)$ liegen vor.*

Dann gilt

$$G_t^i(u) = P(U_t^i \geq u \mid X_t = i) = \exp \int_t^{t+u} q_{ii}(s) ds \quad (4.20)$$

für beliebige $i \in \mathbb{S}$.

Beweis. Tatsächlich kann bei jedem beliebigen Markov-Prozess die Wahrscheinlichkeit $P(X_s = i, t \leq s \leq t + u \mid X_t = i)$ durch ein Produkt wie folgt

ausgeschrieben werden:

$$\begin{aligned}
P(X_s = i, t \leq s \leq t + u \mid X_t = i) &= \frac{P(X_s = i, t \leq s \leq t + u)}{P(X_t = i)} \\
&= \frac{P(X_t = i) \prod_{s=t+}^{t+u} P(X_s = i \mid X_{s-ds} = i)}{P(X_t = i)} \quad (4.21) \\
&= \exp \int_t^{t+u} \log P(X_s = i \mid X_{s-ds} = i) = \exp \int_t^{t+u} \log(q_{ii}(s)ds + 1)
\end{aligned}$$

Es wird nun $\lim_{x \rightarrow 0} \log(x+1)/x = 0$ verwendet. Wegen $q_{ii}(s)ds \rightarrow 0$ kann $\log(q_{ii}(s)ds + 1)$ in (4.21) durch $q_{ii}(s)ds$ ersetzt werden.

Schließlich ergibt sich

$$\begin{aligned}
&P(X_s = i, t \leq s \leq t + u \mid X_t = i) \\
&= \exp \int_t^{t+u} \log(q_{ii}(s)ds + 1) = \exp \int_t^{t+u} q_{ii}(s)ds
\end{aligned}$$

Damit ist das Lemma bewiesen. \square

Die Funktion (4.20) kann folgenderweise ausgeschrieben werden:

$$\begin{aligned}
G_t^i(u) &= P(U_t^i \geq u \mid X_t = i) \\
&= P(U_t^i \geq u, U_t^i = \infty \mid X_t = i) + P(U_t^i \geq u, U_t^i \neq \infty \mid X_t = i) \\
&= P(U_t^i = \infty \mid X_t = i) + P(U_t^i \geq u, U_t^i \neq \infty \mid X_t = i) \\
&= G_t^i(\infty) + P(U_t^i \geq u, U_t^i \neq \infty \mid X_t = i).
\end{aligned}$$

Bei der Erzeugung der Verweildauer muss beachtet werden, dass sich einige Objekte unendlich lange in einem (absorbierenden) Zustand i befinden können. Ein Zufallszahlengenerator kann dagegen nur reelle Werte der Verweildauer ergeben. Daher muss anhand der Wahrscheinlichkeit $P(U_t^i = \infty \mid X_t = i)$ zuerst entschieden werden, ob das Objekt weiter unendlich lange im aktuellen Zustand bleibt. Ist es nicht der Fall, so kann eine Realisation von

$U_t^i < \infty$ aus der Verteilungsfunktion

$$\begin{aligned} F_t^i(u) &:= P(U_t^i \leq u \mid X_t = i, U_t^i \neq \infty) = 1 - P(U_t^i \geq u \mid X_t = i, U_t^i \neq \infty) \\ &= 1 - \frac{G_t^i(u) - G_t^i(\infty)}{1 - G_t^i(\infty)} = \frac{1 - G_t^i(u)}{1 - G_t^i(\infty)} \end{aligned}$$

erzeugt werden. Die Existenz der Wahrscheinlichkeit $P(U_t^i = \infty \mid X_t = i) > 0$ folgt direkt aus der Integrierbarkeit von $q_{ii}(t)$. Gilt $\int_t^\infty q_{ii}(s)ds < \infty$, so auch $P(U_t^i = \infty \mid X_t = i) > 0$.

Zur Erzeugung der Realisationen von U_t^i sind die exakte Inverse von $F_t^i(u)$ bzw. das exakte Integral $\int q_{ii}(s)ds$ gefragt, was sich wiederum auf das Problem der Wahl von Funktionen $q_{ii}(t)$ bezieht. Wird die gewichtete Migrationsintensität $\tilde{q}_{ih}(t) := \dot{\theta}^*(t)/(\theta^*(t))^2$ für alle $i, h \in \mathbb{S} \setminus K$, $i \neq h$ gewählt, so gelten auch $q_{ih}(t) = [c_i]^{-1} \dot{\theta}(t)/\theta^*(t)$, $q_{iK}(t) = \dot{\theta}(t)/\theta^*(t)$ und folglich $q_{ii}(t) = -[(K-2)/c_i + 1] \dot{\theta}(t)/\theta^*(t)$ für alle $i \in \mathbb{S}$, $h \in \mathbb{S} \setminus K$, $i \neq h$. Folgendes Integral ergibt sich:

$$\begin{aligned} \int_t^{t+u} q_{ii}(s)ds &= \log |\theta^*(t+u)|^{(2-K-c_i)/c_i} - \log |\theta^*(t)|^{(2-K-c_i)/c_i} \\ &= \log \left| \frac{\theta(t+u) + \sum_{j=1}^{K-1} c_j}{\theta(t) + \sum_{j=1}^{K-1} c_j} \right|^{(2-K-c_i)/c_i} \end{aligned}$$

für alle $i \in \mathbb{S}$. Für die Funktion $G_t^i(u)$ gilt dann

$$G_t^i(u) = \left[\frac{\theta(t+u) + \sum_{j=1}^{K-1} c_j}{\theta(t) + \sum_{j=1}^{K-1} c_j} \right]^{(2-K-c_i)/c_i}$$

Wird zusätzlich $\theta(t) := e^{at+b}$ für $a, b \in \mathbb{R}_+$ angenommen, so kann die Funktion $G_t^i(u)$ weiter wie folgt ausgeschrieben werden:

$$G_t^i(u) = \left[\frac{e^{-at-b} \sum_{j=1}^{K-1} c_j + e^{au}}{e^{-at-b} \sum_{j=1}^{K-1} c_j + 1} \right]^{(2-K-c_i)/c_i}$$

Unter Berücksichtigung der Tatsache, dass $G_t^i(\infty) = 0$ gilt, ergibt sich die Verteilungsfunktion für U_t^i als $F_t^i(t) = 1 - G_t^i(t)$. Die Inverse $[F_t^i]^{-1}(y)$ von $F_t^i(t)$ ist dann folgenderweise herzuleiten:

$$[F_t^i]^{-1}(y) = \frac{\log [(e^{-at-b} \sum_{j=1}^{K-1} c_j + 1)(1 - y)^{c_i/(2-K-c_i)} - e^{-at-b} \sum_{j=1}^{K-1} c_j]}{a}$$

Nun betrachten wir die Wahrscheinlichkeitsverteilungen, anhand deren die Entscheidungen über den nächsten Zustand getroffen werden. Wechselt der Markov-Prozess zwischen t und $t + dt$ seinen Zustand sicher von i zum anderen, so ist der nächste Zustand nicht mehr i , sondern ein $h \in \mathbb{S}$, $h \neq i$. Hier muss erwähnt werden, dass das Verbleiben in demselben Zustand als der Übergang in den gleichen Zustand betrachtet wird. Aus diesem Grund passen die üblichen momentanen Übergangswahrscheinlichkeiten zur Beschreibung des Wechsels in den anderen Zustand nicht. Die Wahrscheinlichkeit des Übergangs in den anderen Zustand h wird also für ein $h \neq i$ wie folgt ausgedrückt:

$$\begin{aligned} S_{ih}(t) &:= P(X_t = i, X_{t+dt} = h \mid X_t = i, X_{t+dt} \neq i) \\ &= \frac{P(X_t = i, X_{t+dt} = h)}{P(X_t = i, X_{t+dt} \neq i)} = \frac{P(X_{t+dt} = h \mid X_t = i)}{P(X_{t+dt} \neq i \mid X_t = i)} \\ &= \frac{P(X_{t+dt} = h \mid X_t = i)}{\sum_{j=1, j \neq i}^K P(X_{t+dt} = j \mid X_t = i)} = \frac{q_{ih}(t)}{\sum_{j=1, j \neq i}^K q_{ij}(t)} \end{aligned}$$

Bei $\tilde{q}_{ih}(t) := \dot{\theta}^*(t)/(\theta^*(t))^2$ sind die Wahrscheinlichkeiten $S_{ih}(t)$ konstant für alle $t \geq 0$ und haben folgende Gestalt:

$$S_{ih}(t) = \begin{cases} \frac{1}{K-2-c_i} & h \neq K, h \neq i \\ \frac{c_i}{K-2-c_i} & h = K, i \neq K \\ 0 & \text{sonst} \end{cases} \quad (4.22)$$

Zum Schluss wird der Algorithmus zur Simulation des Markov-Prozesses zusammengefasst:

1. Definiere $\mathbf{m}(0)$ und $\mathbf{Q}(t)$.

2. Wähle die Beobachtungsdauer B_{max} .
3. Aus $\mathbf{m}(0)$ erzeuge den Anfangszustand i ; $t := 0$.
4. Durch $P(U_t^i = \infty \mid X_t = i)$ entscheide, ob der Zustand i unendlich lange dauert. Wenn ja, wähle ein neues Objekt (insgesamt n Objekte) und gehe zu 3.
5. Durch $P(U_t^i \leq u \mid X_t = i, U_t^i \neq \infty)$ erzeuge die Verweildauer u_t^i im Zustand i . $t := t + u_t^i$. Gilt $t \geq B_{max}$, so wähle ein neues Objekt (insgesamt n Objekte) und gehe zu 3.
6. Aus der Wahrscheinlichkeitsverteilung $\{S_{i1}(t), \dots, S_{iK}(t)\}$ erzeuge den nächsten Zustand h ; $i := h$, gehe zu 4.

Ein Beispiel der Simulation betrachten wir im nächsten Abschnitt.

4.2.3 Simulation eines Markov-Prozesses

In diesem Abschnitt wird angestrebt, die Pfade eines inhomogenen Markov-Prozesses mit integrierten Anfangswahrscheinlichkeiten und Migrationsintensitäten zu erzeugen. Der Zustandsraum wird als $\mathbb{S} = \{1, \dots, 7\}$ definiert, wobei der letzte 7. Zustand absorbierend ist. Auf der Basis dieser Simulation werden die zugrunde liegenden Migrationsintensitäten geschätzt. Für die Beobachtungsschranken L und C werden jeweils $F^L(0) = 1$ und $F^C(5) = 0$ angenommen, so dass die Schätzung auf einem festen Zeitabschnitt $[0, 5]$ erfolgen kann.

Die Simulation wird nach dem Algorithmus aus Abschnitt 4.2.2 durchgeführt. Zur Angabe des Markov-Prozesses benutzen wir das Verfahren aus Abschnitt 4.2.1. Es werden dabei die Anfangswahrscheinlichkeiten $\mathbf{m}(0)$ und die parametrische Funktion $\theta(\cdot)$ wie folgt angegeben:

$$\mathbf{m}(0) := (0.35, 0.05, 0.2, 0.05, 0.15, 0.19, 0.01)'$$

und

$$\theta(t) := \exp(1.1t + 0.005).$$

Daraus ergeben sich die Koeffizienten c_1, \dots, c_6 als

$$\mathbf{c} = (35.18, 5.03, 20.10, 5.03, 15.08, 19.10)'$$

Damit die Konvergenz durch die Simulation nachweisbar ist, müssen die Voraussetzungen aus Abschnitt 3.4.4 für die theoretischen Migrationsintensitäten erfüllt sein. Da der Zustandsraum \mathbb{S} angegeben ist und die Simulation auf einem festen Zeitabschnitt $[0, 5]$ erfolgt, sind die Annahmen (M1) bis (M6) erfüllt.

Bei der Überprüfung der Voraussetzung (M7) ganz problematisch ist die Bestimmung der maximalen Anzahl der Übergänge auf $[0, 5]$, da $\int_0^5 \mathbf{1}_{\{X_s^j=i, X_{s+ds}^j=h\}}$ eine infinite Summe von abhängigen und nicht identisch verteilten Bernoulli-Zufallsvariablen darstellt. Nur in den absorbierenden 7. Zustand kann höchstens ein einziger Übergang auf $[0, 5]$ passieren. Daher gilt $\int_0^5 \mathbf{1}_{\{X_s^j=i, X_{s+ds}^j=7\}} \leq 1 =: K^{i7}$ für $i = 1, \dots, 6$. Aus diesem Grund können wir nur das Konvergenzverhalten für Kernschätzer von Migrationsintensitäten $q_{17}(\cdot) = \dots = q_{67}(\cdot)$ untersuchen. Dabei gilt $\sum_{j=1}^n \mathbf{1}_{\{X_t^j=i\}} > 0$, $i = 1, \dots, 6$ auf $[0, 5]$ für die vorliegende Simulation.

Gemäß Kapitel 3.4 wird der angegebene Markov-Prozess durch die Funktionen

$$\tilde{q}_{i7}(t) = q_{i7}(t)m_i(t) = c_i \cdot \frac{1.1 \cdot \exp(1.1t + 0.005)}{(\exp(1.1t + 0.005) + 99.52)^2} = c_i \cdot \frac{1.1 \cdot \theta(t)}{\theta^*(t)^2}$$

und

$$m_i(t) = \frac{c_i}{(\exp(1.1t + 0.005) + 99.52)} = \frac{c_i}{\theta^*(t)}$$

für $i = 1, \dots, 6$ mit $q_{i7}(\cdot) \neq 0$ charakterisiert, wobei

$$d\tilde{q}_{i7}(t)/dt = c_i \cdot 1.1\dot{\theta}(t)\theta^*(t) \cdot [\theta^*(t) - 2\theta(t)]/\theta^*(t)^4.$$

Die Voraussetzung (M7) ist erfüllt, weil folgende Schranken für diese Funktionen auf $[0, 5]$ existieren:

- $0 < \tilde{q}_{i7}^{min} := 0.000109 \cdot c_i \leq \tilde{q}_{i7}(t) \leq 0.002763 \cdot c_i =: \tilde{q}_{i7}^{max} < \infty$
- $|d\tilde{q}_{i7}(t)/dt| < 0.00117 \cdot c_i =: L_{i7} < \infty$
- $0 < m_i^{min} := 0.101835 \cdot c_i \leq m_i(t)$

Um den hohen Rechenaufwand zu vermeiden, untersuchen wir weiter nur den Schätzer $\hat{q}_{17}(\cdot)$. Zuerst muss allerdings die Folge (p_n) spezifiziert werden. Die Folge (4.1) vom Anfang des Kapitels kann bei $m = 10\,000$ und $a = 0.9$ als

$$p_n^{(10\,000)} = 0.11 \cdot k_{min} \cdot (\log n/n)^{0.9} \quad (4.23)$$

dargestellt werden, da $\Delta_{max} = K^{i7} = 1, \forall i = 1, \dots, 6$ und $G_n(\cdot) = 1$ für die Markov-Prozesse gemäß dem Beweis zum Satz 3.4.1 gelten.

Weiter sind die geschätzten Migrationsintensitäten für die Datenumfänge von 10 000 bis 90 000 Beobachtungen bei $k_{min} = 10$ und $k_{min} = 100$ dargestellt. Bei der variablen Bandbreite wird dabei die kumulative gewichtete Migrationsintensität $\tilde{Q}_{[0,5]}^{17}(\cdot)$ als Glätter gewählt.

Ähnlich zum Kernschätzer für die Hazardrate ist in Abbildungen 4.3 und 4.4 zu ersehen, dass die Abweichungen der Schätzer mit der wachsenden Beobachtungsanzahl immer geringer werden. Allerdings fallen in diesem Fall noch starke Ausschläge auf den rechten Seiten aller Grafiken auf. Dieser Effekt ist als Randverhalten der Kernschätzer bekannt und ist z. B. in (Wand und Jones, 1995, S. 46) beschrieben. Zur Bekämpfung oder zumindest Verminderung dieses Effektes ist die optimale Folge (p_n) gefragt. Da eine optimale Gestalt dieser Folge heutzutage noch nicht bekannt ist, sollte man die Bereiche des Kernschätzers sehr dicht an die Beobachtungsgrenzen am liebsten nicht verwenden.

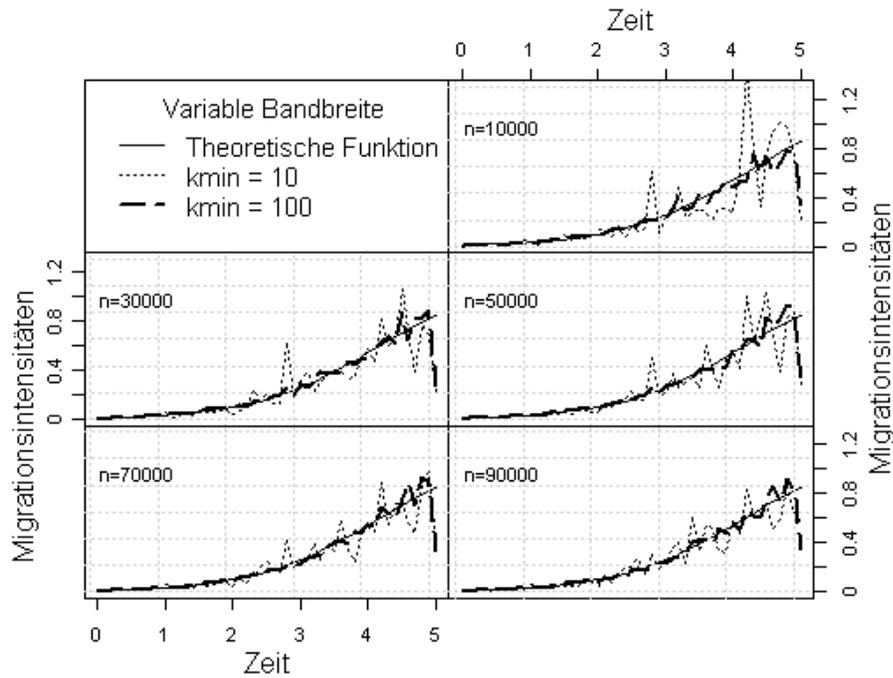


Abbildung 4.3: Kernschätzer für $q_{17}(t)$ bei variabler Bandbreite mit $k_{min} = 10$ und $k_{min} = 100$ für verschiedene Datenumfänge.

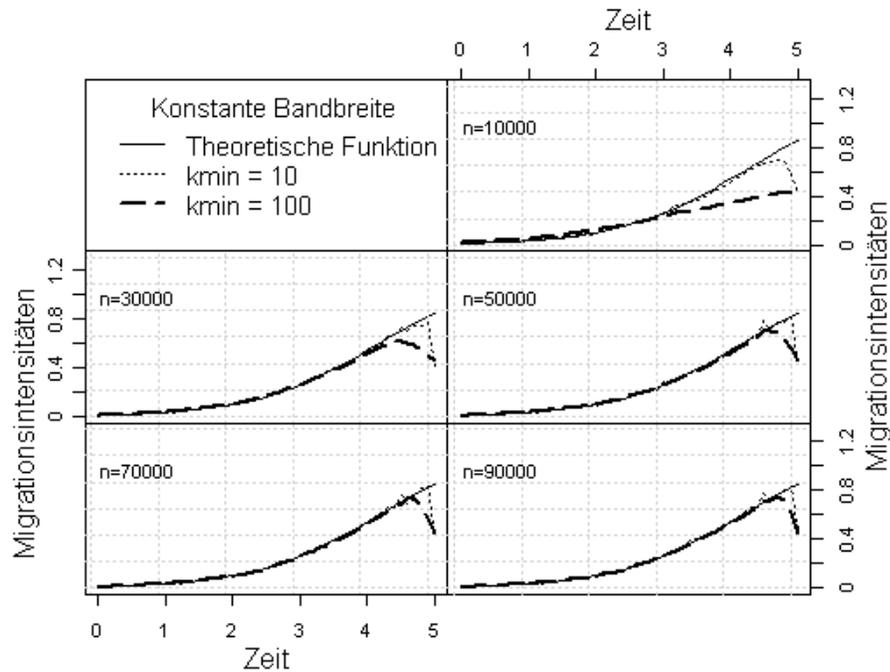


Abbildung 4.4: Kernschätzer für $q_{17}(t)$ bei konstanter Bandbreite mit $k_{min} = 10$ und $k_{min} = 100$ für verschiedene Datenumfänge.

Kapitel 5

Anwendung: Risikobewertung in Kreditportfolios

Auf dem Finanzmarkt gehört u. a. die Risikobewertung zum Anwendungsbereich der stochastischen Prozesse. In diesem Kapitel zeigen wir, wie der entwickelte Kernschätzer für Migrationsintensitäten bei der Risikobewertung von Kreditportfolios eingesetzt werden kann.

Angenommen, die Bonität jedes Emittenten wird durch ein Ratingsystem $\{1, \dots, K\}$ skaliert, wobei K die Insolvenz bedeutet. Im Rahmen eines Kreditportfolios wird zusätzlich angenommen, dass die Emittenten voneinander unabhängig handeln und die Migrationen zwischen Kreditratings demselben Markov-Prozess $\{X_t\}_{t=0, \dots, \infty}$ folgen. Unter diesen Annahmen wird weiter ein Beitrag zur Abschätzung eines risikodeckenden Kreditrisikoaufschlags (Credit Spread), sowie einer Risikoprämie bei der digitalen Konkursversicherung (Digital Default Put) geleistet. Anschließend werden empirische Daten aus einem Teilportfolio von *WestLB* untersucht und entsprechend bewertet.

5.1 Kreditrisikoaufschlag (Credit Spread)

In diesem Abschnitt betrachten wir eine Erweiterung des Rating-basierten Ansatzes zur Berechnung von Credit Spreads. Die Grundidee dieser Methode stammt aus Jarrow und Turnbull (1995) und besteht darin, den Insolvenzprozess exogen zu modellieren. Dabei wird den Ratingsverläufen im Rahmen eines Kreditportfolios ein stochastischer Prozess unterstellt. Nach einer verallgemeinerten Modellierung des Insolvenzprozesses sehen wir, wie der Credit Spread unter der Markov-Annahme berechnet werden kann. Im Unterschied zu Jarrow *et al.* (1997) wird in dieser Arbeit angenommen, dass Markov-Prozesse inhomogen sind.

Für den Auszahlungszeitpunkt T kann das Risiko von einem Emittenten durch den Kreditrisikoaufschlag (engl. Credit Spread) $r_s^p, 0 \leq s \leq T$ ausgedrückt werden. Beim risikofreien Zinssatz $r_s^f, 0 \leq s \leq T$ wird der risikobehaftete als $r_s^f + r_s^p, 0 \leq s \leq T$ berechnet. In der Tat wird der tatsächliche Credit Spread nicht durch das Portfolio, sondern durch den Finanzmarkt beeinflusst. Für jeden neuen Emittenten streben wir also an, die kleinste risikodeckende Schranke von r_s^p bzgl. des Portfolios abzuschätzen. Daraus kann geschlossen werden, ob der tatsächliche Kreditrisikoaufschlag die Risiken des neuen Emittenten abdeckt.

Wir betrachten eine Anleihe mit einer einmaligen Zahlung B_T , die fest oder zufällig sein kann. Im Zeitpunkt t beträgt der abdiskontierte Preis dieser Anleihe $B(t, T) = B_T e^{-\int_t^T r_s^f}$ (Amman, 2001, S. 33). Falls für einen Emittenten ein Ausfallrisiko besteht, so wird in T der Betrag $I_{(T, \infty)} B_T$ ausgezahlt. Für den Zeitpunkt des Ausfalls τ zeigt dabei die Indikatorfunktion $I_{(s, t)} := \mathbf{1}_{\{s \leq \tau \leq t\}}$ die Insolvenz zwischen s und t an. Unter Annahme der Unabhängigkeit des Ausfalls von der Spot Rate $r_s^f, 0 \leq s \leq T$ und der Zahlung

B_T hat die Anleihe in $t < T$ den folgenden Wert:

$$\begin{aligned}
\bar{B}_0(t, T) &= E \left(B_T e^{-\int_t^T r_s^f ds} I_{(T, \infty)} \mid \mathcal{F}_t \right) \\
&= E \left(B_T e^{-\int_t^T r_s^f ds} (1 - I_{(0, t)})(1 - I_{(t, T)}) \mid \mathcal{F}_t \right) \\
&= (1 - I_{(0, t)}) E \left(B_T e^{-\int_t^T r_s^f ds} (1 - I_{(t, T)}) \mid \mathcal{F}_t \right) \\
&= (1 - I_{(0, t)}) E \left(B_T e^{-\int_t^T r_s^f ds} E(1 - I_{(t, T)} \mid \mathcal{F}_t, B_T, r_s^f, 0 \leq s \leq T) \mid \mathcal{F}_t \right) \\
&= (1 - I_{(0, t)}) E \left(B_T e^{-\int_t^T r_s^f ds} E(1 - I_{(t, T)} \mid \mathcal{F}_t) \mid \mathcal{F}_t \right).
\end{aligned}$$

Die innere Erwartung kann folgenderweise geschrieben werden:

$$\begin{aligned}
E(1 - I_{(t, T)} \mid \mathcal{F}_t) &= 1 - \int_t^T f_{\tau, \mathcal{F}_t}(s) ds \\
&= \exp \left(\log \left(1 - \int_t^T f_{\tau, \mathcal{F}_t}(s) ds \right) \right) \\
&= \exp \left(- \int_t^T \frac{f_{\tau, \mathcal{F}_t}(s)}{1 - \int_s^T f_{\tau, \mathcal{F}_t}(z) dz} ds \right),
\end{aligned} \tag{5.1}$$

wobei $f_{\tau, \mathcal{F}_t}(s)$ die bedingte Dichtefunktion von τ ist. Somit beträgt der Wert der Anleihe in t :

$$\bar{B}_0(t, T) = (1 - I_{(0, t)}) E \left(B_T e^{-\int_t^T \left(r_s^f + \frac{f_{\tau, \mathcal{F}_t}(s)}{1 - \int_s^T f_{\tau, \mathcal{F}_t}(z) dz} \right) ds} \mid \mathcal{F}_t \right).$$

Daher hat die bis t ausgefallene Anleihe den Wert 0 in t , ansonsten kann sie weiter als eine risikofreie mit dem erhöhten Zinssatz behandelt werden. Dies impliziert die gefragte kleinste risikodeckende Schranke bzgl. des aktuellen Zeitpunktes t als

$$r_{s,t}^p = \begin{cases} \frac{f_{\tau, \mathcal{F}_s}(s)}{1 - \int_s^T f_{\tau, \mathcal{F}_s}(z) dz} & \text{für } s \leq t, \\ \frac{f_{\tau, \mathcal{F}_t}(s)}{1 - \int_s^T f_{\tau, \mathcal{F}_t}(z) dz} & \text{für } s > t. \end{cases} \tag{5.2}$$

Die Formel (5.2) beschreibt einen Credit Spread, der Risiken bzgl. der aktuellen Situation auf dem Finanzmarkt (und im Portfolio) abdecken sollte. Wegen eines stochastischen Charakters der risikofreien Spot Rate r_s^f muss dieser Credit Spread an die ständig wechselnde Situation angepasst werden und wird daher bzgl. des aktuellen Zeitpunktes t berechnet. Dabei steht der Teil $s \leq t$ für das bereits realisierte Risiko und der Teil $s > t$ für das noch kommende bzgl. der Situation in t . Im Unterschied zum minimalen risikodeckenden Credit Spread $r_{s,t}^p$ zeigt r_s^p einen Credit Spread, der tatsächlich vorkommt also stochastisch ist.

Unter der Markov-Annahme für die Ratingmigrationen muss nun die Funktion $I_{(s,t)}$ spezifiziert werden. Wegen eines einmaligen Vorkommens des Ausfalls übertritt die gesamte Anzahl der Übergänge in die Insolvenz auf $[s, t]$ den Wert 1 nicht. Dabei ist diese Anzahl nur dann gleich 1, wenn auf $[s, t]$ der Ausfall passiert ist. Daher kann $I_{(s,t)}$ wie folgt geschrieben werden:

$$I_{(s,t)} := \sum_{j=1}^{K-1} \int_s^t \mathbf{1}_{\{X_z=j, X_{z+dz}=K\}}. \quad (5.3)$$

Die Markov-Eigenschaft impliziert u. a. die folgende Erwartung:

$$\begin{aligned} E(I_{(t,T)} | \mathcal{F}_t) &= E \left(\sum_{j=1}^{K-1} \int_t^T \mathbf{1}_{\{X_s=j, X_{s+ds}=K\}} | \mathcal{F}_t \right) \\ &= \sum_{j=1}^{K-1} \int_t^T P(X_s = j, X_{s+ds} = K | \mathcal{F}_t) = \sum_{j=1}^{K-1} \int_t^T P(X_s = j, X_{s+ds} = K | X_t) \\ &= \sum_{j=1}^{K-1} \sum_{i=1}^{K-1} \mathbf{1}_{\{X_t=i\}} \int_t^T P(X_s = j, X_{s+ds} = K | X_t = i) \\ &= \sum_{j=1}^{K-1} \sum_{i=1}^{K-1} \mathbf{1}_{\{X_t=i\}} \int_t^T \frac{P(X_s = j, X_{s+ds} = K, X_t = i)}{P(X_t = i)} \\ &= \sum_{j=1}^{K-1} \sum_{i=1}^{K-1} \mathbf{1}_{\{X_t=i\}} \int_t^T \frac{P(X_{s+ds} = K | X_s = j, X_t = i) P(X_s = j, X_t = i)}{P(X_t = i)} \end{aligned}$$

$$\begin{aligned}
&= \sum_{j=1}^{K-1} \sum_{i=1}^{K-1} \mathbf{1}_{\{X_t=i\}} \int_t^T P(X_{s+ds} = K \mid X_s = j) P(X_s = j \mid X_t = i) \\
&= \sum_{j=1}^{K-1} \sum_{i=1}^{K-1} \mathbf{1}_{\{X_t=i\}} \int_t^T q_{jK}(s) p_{ij}(t, s) ds,
\end{aligned}$$

wobei $p_{ij}(t, s)$ die Übergangswahrscheinlichkeiten und $q_{jK}(s)$ Migrationsintensitäten sind. Aus (5.1) folgt die bedingte Dichte von τ als

$$f_{\tau, \mathcal{F}_t}(s) = \sum_{j=1}^{K-1} \sum_{i=1}^{K-1} \mathbf{1}_{\{X_t=i\}} p_{ij}(t, s) q_{jK}(s). \quad (5.4)$$

Unter der Markov-Eigenschaft kann die kleinste risikodeckende Schranke für den Credit Spread folgenderweise präzisiert werden:

$$r_{s,t}^p = \begin{cases} \frac{\sum_{i=1}^{K-1} \mathbf{1}_{\{X_s=i\}} q_{iK}(s)}{1 - \sum_{j=1}^{K-1} \sum_{i=1}^{K-1} \mathbf{1}_{\{X_s=i\}} \int_s^T p_{ij}(s, z) q_{jK}(z) dz} & \text{für } s \leq t, \\ \frac{\sum_{j=1}^{K-1} \sum_{i=1}^{K-1} \mathbf{1}_{\{X_t=i\}} p_{ij}(t, s) q_{jK}(s)}{1 - \sum_{j=1}^{K-1} \sum_{i=1}^{K-1} \mathbf{1}_{\{X_t=i\}} \int_s^T p_{ij}(t, z) q_{jK}(z) dz} & \text{für } s > t. \end{cases} \quad (5.5)$$

Angenommen, im Portfolio befinden sich bereits historische Daten über n Emittenten. Die Schätzung der Migrationsintensitäten $q_{iK}(t)$, $i = 1, \dots, K-1$ stellt offensichtlich kein Problem dar. Aus Abschnitt 3.4.5 resultiert folgender Schätzer:

$$\hat{q}_{iK}(t) := \frac{\frac{1}{R_n \cdot n} \sum_{j=1}^n \int_0^{T_{max}} \text{K}\Gamma\left(\frac{t-s}{R_n}\right) \mathbf{1}_{\{X_s^j=i, X_{s+ds}^j=K\}}}{\frac{1}{n} \sum_{j=1}^n \mathbf{1}_{\{X_t^j=i\}}},$$

wobei T_{max} das größte Fälligkeitsdatum im Portfolio ist und für die rechte Beobachtungsschranke steht. Das geht natürlich nur dann, wenn keine Zensurierung von Daten vorliegt. Auf die Erfüllung der Voraussetzungen (M1) bis (M9) für solche Intensitäten wurde bereits in Abschnitt 4.2.3 eingegangen. Zur Schätzung von Übergangswahrscheinlichkeiten $p_{ij}(t, s)$, $i, j =$

$1, \dots, K - 1$, $t < s$ kann der Johansen-Aalen-Schätzer (Johansen, 1978) eingesetzt werden. Liegt ein Datenmangel vor, so kann auch ein folgender Schätzer für die Übergangswahrscheinlichkeiten verwendet werden:

$$\hat{p}_{ij}(t, s) = \frac{\sum_{j=1}^n \mathbf{1}_{\{X_t^j=i, X_s^j=j\}}}{\sum_{j=1}^n \mathbf{1}_{\{X_t^j=i\}}}, \quad (5.6)$$

wobei $p_{ij}(t, s) = P(X_s = j \mid X_t = i) = P(X_s = j, X_t = i) / P(X_t = i)$ gilt.

Wir bezeichnen weiter $f_{ij}(t, s) = \int_s^T p_{ij}(t, z) q_{jK}(z) dz$, $i, j = 1, \dots, K - 1$, $t \leq s$ als kumulative bedingte momentane Übergangswahrscheinlichkeit. Wegen $f_{ij}(t, s) = \int_s^T P(X_z = j, X_{z+dz} = K \mid X_t = i)$ bekommen wir den folgenden Schätzer:

$$\hat{f}_{ij}(t, s) = \frac{\sum_{k=1}^n \int_s^T \mathbf{1}_{\{X_z^k=j, X_{z+dz}^k=K, X_t^k=i\}}}{\sum_{k=1}^n \mathbf{1}_{\{X_t^k=i\}}}.$$

Der Vergleich des geschätzten kleinsten risikodeckenden Kreditrisikoaufschlags mit dem tatsächlichen gibt die Auskunft über die Zeitintervalle, in denen das Ausfallrisiko durch den erhöhten Zinssatz nicht abgedeckt wird. Solche Zeitintervalle stellen eine besondere Gefahr dar. Die obere Auslegung kann zum Schluss grafisch veranschaulicht werden.

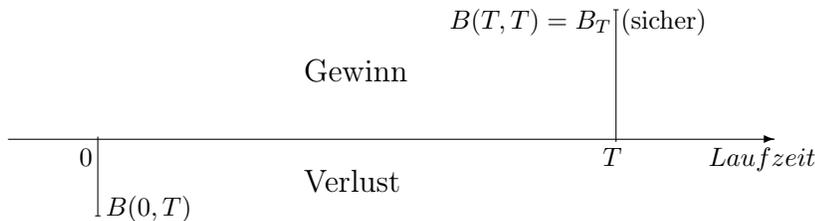


Abbildung 5.1: Lebensdauer einer risikofreien Anleihe mit dem Nominal B_T und dem Fälligkeitsdatum in T .

In Abbildungen 5.1 und 5.2 ist deutlich zu sehen, dass der Ankaufspreis für die risikobehaftete Anleihe niedriger als für die risikofreie ist. Dafür ist die Auszahlung in T für die risikobehaftete Anleihe nicht gesichert. Zur Sicherung

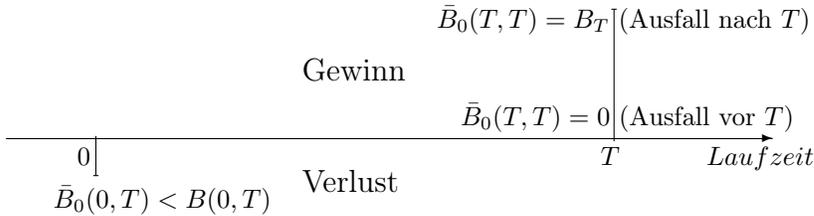


Abbildung 5.2: Lebensdauer einer risikobehafteten Anleihe mit dem Nominal B_T und dem Fälligkeitsdatum in T .

des Gewinns werden verschiedene Hedgingsstrategien verwendet. Eine davon betrachten wir im nächsten Abschnitt.

5.2 Digitale Konkursversicherung (Digital Default Put)

Die Existenz von Zeitintervallen mit dem erhöhten Ausfallrisiko bringt uns zur Notwendigkeit die Auszahlung in T abzusichern. Eine der Möglichkeiten ist die digitale Konkursversicherung (engl. Digital Default Put oder kurz DDP). Bei einem DDP wird das Nominal der abgesicherten Anleihe im Zeitpunkt der Insolvenz an den Investor ausbezahlt und sonst nichts. Um die Verluste vom Versicherer abzudecken, wird eine Risikoprämie für die erfolgreiche Durchführung des Geschäfts vom Investor verlangt. Unser Ziel ist in diesem Abschnitt diese Verluste und somit auch die kleinste risikodeckende Prämie zu berechnen.

Wir betrachten weiter die digitale Konkursversicherung der Anleihe aus Abschnitt 5.1. Angenommen, die Anleihe muss auf dem Zeitabschnitt $[t, T]$, $t \geq 0$ von einer Bank versichert werden. Im vereinfachten Fall findet die Auszahlung in T statt. Der von der Bank auszubehaltende Betrag $B_T I_{(0, T)}$ hat

dann in t den folgenden Wert

$$\begin{aligned} D(t, T) &= E(B_T I_{(0, T)} e^{-\int_t^T r_z^f dz} \mid \mathcal{F}_t) \\ &= E(B_T (1 - I_{(T, \infty)}) e^{-\int_t^T r_z^f dz} \mid \mathcal{F}_t) \\ &= B(t, T) - \bar{B}_0(t, T). \end{aligned}$$

Daher können die Verluste der Bank durch die Replikation von zwei Anleihen abgesichert werden, wobei die Risikoprämie behilflich wird.

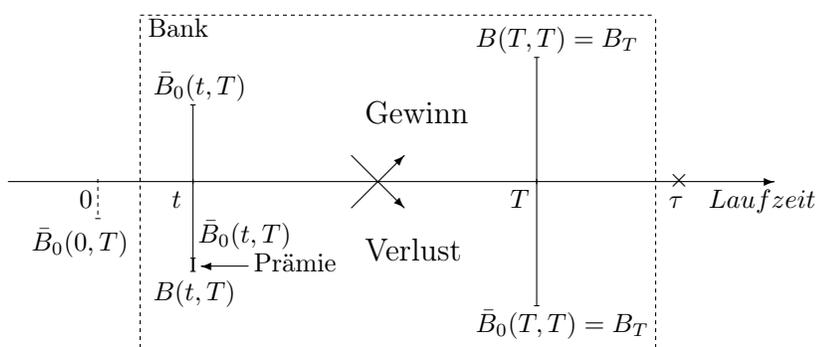


Abbildung 5.3: Hedgingstrategie für eine risikobehaftete Anleihe mit Nominal B_T und Fälligkeit in T : Ausfall nach T .

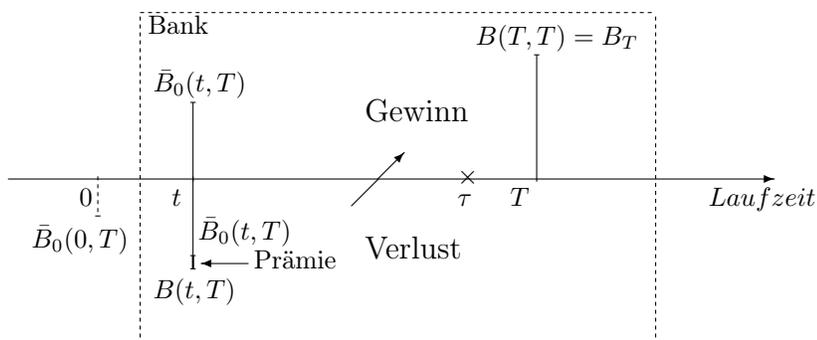


Abbildung 5.4: Hedgingstrategie für eine risikobehaftete Anleihe mit Nominal B_T und Fälligkeit in T : Ausfall in $[t, T]$, Auszahlung in T .

Zunächst leiht die Bank in t die Anleihe vom Investor und verkauft sie.

Für den Erlös zusammen mit der Risikoprämie wird gleichzeitig eine risikofreie Anleihe mit dem Nominal B_T und der Fälligkeit in T gekauft (siehe Abbildungen 5.3 und 5.4). Dadurch werden die Verluste zwischen t und T abgesichert. Bei der Insolvenz wird das Nominal der risikofreien Anleihe an den Investor in T ausgezahlt. Falls die Insolvenz nicht eintritt, wird die geliehene Anleihe in T wiedergekauft und an den Investor zurückgegeben. Da die Konkursversicherung nach dem Ausfall sinnlos ist, bekommen wir die endgültige Risikoprämie als $P(t, T) := (1 - I_{(0,t)}) \cdot D(t, T)$.

Allerdings findet die Auszahlung bei einem DDP nicht am Fälligkeitstag, sondern im Zeitpunkt des Ausfalls statt. Die Berechnung der Verluste wird in diesem Fall komplizierter. Der Betrag $B_T I_{(0,T)}$ muss nun nicht mehr in T , sondern in τ ausgezahlt werden. In t hat die Auszahlung den Barwert $D_\tau(t, T) = E(B_T I_{(0,T)} e^{-\int_t^\tau r_z^f dz} \mid \mathcal{F}_t)$ und die Risikoprämie $P_\tau(t, T) = (1 - I_{(0,t)}) \cdot D_\tau(t, T)$ wird verlangt.

Da $P_\tau(t, T)$ durch einen Erwartungswert ausgedrückt wird, brauchen wir vor allem die bedingte Dichte von τ zur Berechnung der Risikoprämie. Allerdings wurde die bedingte Dichte $f_{\tau, \mathcal{F}_t}(s)$ im vorigen Abschnitt als (5.4) bereits berechnet. Unter der Annahme, dass der Ausfall sowohl von der Spot Rate als auch vom Nominal der Anleihe unabhängig ist, schreiben wir die kleinste Risikoprämie weiter wie folgt aus:

$$\begin{aligned}
P_\tau(t, T) &= (1 - I_{(0,t)}) \cdot D_\tau(t, T) \\
&= E(B_T \mathbf{1}_{\{t < \tau \leq T\}} e^{-\int_t^\tau r_z^f dz} \mid \mathcal{F}_t) \\
&= E(B_T E(\mathbf{1}_{\{t < \tau \leq T\}} e^{-\int_t^\tau r_z^f dz} \mid \mathcal{F}_t, B_T, r_s^f, 0 \leq s \leq T) \mid \mathcal{F}_t) \\
&= E(B_T E(\mathbf{1}_{\{t < \tau \leq T\}} e^{-\int_t^\tau r_z^f dz} \mid \mathcal{F}_t) \mid \mathcal{F}_t) \tag{5.7} \\
&= E(B_T \int_t^T e^{-\int_t^s r_z^f dz} f_{\tau, \mathcal{F}_t}(s) ds \mid \mathcal{F}_t) \\
&= \mathbf{1}_{\{t < \tau\}} \int_t^T E(B_T e^{-\int_t^s r_z^f dz} f_{\tau, \mathcal{F}_t}(s) \mid \mathcal{F}_t) ds
\end{aligned}$$

Durch das Einsetzen von $f_{\tau, \mathcal{F}_t}(\cdot)$ aus der Formel (5.4) bekommen wir:

$$P_{\tau}(t, T) = \mathbf{1}_{\{t < \tau\}} \sum_{j=1}^{K-1} \sum_{i=1}^{K-1} \mathbf{1}_{\{X_t=i\}} \int_t^T p_{ij}(t, s) q_{jK}(s) E(B_T e^{-\int_t^s r_z^f dz} | \mathcal{F}_t) ds.$$

Dementsprechend kann die Risikoprämie für einen neuen Emittenten mit den deterministischen Spot Rate $r_s^f, 0 \leq s \leq T$ und Nominal B_T folgenderweise geschätzt werden:

$$\hat{P}_{\tau}(t, T) = \mathbf{1}_{\{t < \tau\}} B_T \sum_{j=1}^{K-1} \sum_{i=1}^{K-1} \mathbf{1}_{\{X_t=i\}} \frac{\sum_{k=1}^n \int_t^T \mathbf{1}_{\{X_s^k=j, X_{s+ds}^k=K, X_t^k=i\}} e^{-\int_t^s r_z^f dz}}{\sum_{k=1}^n \mathbf{1}_{\{X_t^k=i\}}},$$

wobei historische Daten durch X_t^k und die aktuelle Anleihe durch X_t bezeichnet werden.

5.3 Beispiel einer Risikobewertung

Um die praktische Anwendung der Methoden aus Abschnitten 5.1 und 5.2 zu präsentieren, führen wir ein Beispiel der Risikoanalyse durch. Im Laufe der Analyse wird angestrebt, die Charakteristiken eines Kreditportfolios anhand der historischen Daten unter der Markov-Annahme zu bestimmen. Dies ermöglicht dann die Entscheidung über den Eintritt eines neuen Emittenten ins Portfolio, sowie über die Kreditversicherung und die Risikoprämie.

Zur Analyse wird ein Datensatz benötigt, in dem der zeitliche Ablauf von Ratingmigrationen für jeden Emittenten aus dem Portfolio gespeichert wird. Hier muss beachtet werden, dass solche Datensätze von den finanziellen Institutionen streng geheim gehalten werden. Bedingt zugänglich sind nur Teilportfolios mit dem begrenzten Datenumfang.

In dieser Arbeit steht ein Teilportfolio von *WestLB* mit der Ratingsgeschichte von 358 Emittenten für die Periode vom Januar 1997 bis Dezember 2003 zur Analyse. Die Ratings sind von 1 bis 10 skaliert. Die Ratingsklassen 1 und 10 sind dabei absorbierend und bedeuten jeweils volle Auszahlung

und die Insolvenz. Zur Vereinfachung des Modells nehmen wir an, dass die Ratingmigrationen einem Markov-Prozess folgen. Dann kann anhand der geschätzten Migrationsintensitäten auch die Risikobewertung für das Portfolio durchgeführt werden. Zu beachten ist dabei, dass die Schätzung von Migrationsintensitäten wesentlich vereinfacht wird, falls sie konstant sind. Für das vorliegende Portfolio ist es allerdings nicht der Fall. Das inhomogene Charakter von Ratingmigrationen wurde für eine erweiterte Version dieses Datensatzes in Weißbach *et al.* (2009) gezeigt. Daher bleibt uns nur die einzige Option, den Kernschätzer für die Migrationsintensitäten zu verwenden.

Natürlich sind 358 Beobachtungen im Vergleich zur Simulation aus Abschnitt 4.2.3 äußerst wenig. Darüber hinaus zeigt eine tiefere Analyse, dass nur 17 Fälle der Insolvenz in diesem Datensatz registriert werden und von diesen liegt manchmal sogar nur eine Beobachtung für die einzelnen Übergänge vor. Da es allerdings die einzigen vorliegenden Daten sind, wird die Analyse auf ihrer Basis durchgeführt.

Aus dieser Perspektive wird klar, dass die Schätzung nur für eine kleinere Ratingsanzahl möglich ist. Daher werden die Ratingsklassen so aggregiert, dass die Ratings 2 bis 5 in die neue Klasse 2' und 6 bis 9 in 3' fallen. Den absorbierenden Ratings 1 und 10 werden jeweils die Klassen 1' und 4' zugeordnet. So entsteht ein neuer Prozess mit nur 4 Ratingsklassen, den wir weiter untersuchen. Wir beachten dabei, dass die Ergebnisse der bevorstehenden Analyse wegen der mangelnden Daten nicht repräsentativ sind und in dieser Arbeit nur zur Veranschaulichung des Analyseverfahrens dienen.

Als Basiseigenschaften des vorliegenden Portfolios gelten Migrationsintensitäten für alle Übergänge, wovon nur die Übergänge in die Insolvenz ($2'-4'$ und $3'-4'$) zur Berechnung von Risikoprämien gebraucht werden. Bevor diese Intensitäten geschätzt werden, müssen Beobachtungsschranken festgelegt werden. Wir nehmen an, dass alle Emittenten auf dem festen Zeitabschnitt $[0, 7]$ (in Jahren) beobachtet werden. Diese Annahme ist zulässig, da alle Ratingsverläufe mit dem absorbierenden Zustand enden und der maxima-

le Zeitpunkt des Übergangs in einen solchen Zustand etwa 6.99 Jahre nach der Kreditausgabe ist. Da die Anzahl der Übergänge in den 4. Zustand beschränkt ist, können nun die Migrationsintensitäten $q_{2'4'}(\cdot)$ und $q_{3'4'}(\cdot)$ unter den Annahmen (M1) bis (M9) durch die Formel (3.37) geschätzt werden. Dabei wird die minimale Anzahl der Sprünge in der Bandbreite auf 4 beschränkt. Die geschätzten Migrationsintensitäten sind in Abbildung 5.5 präsentiert.

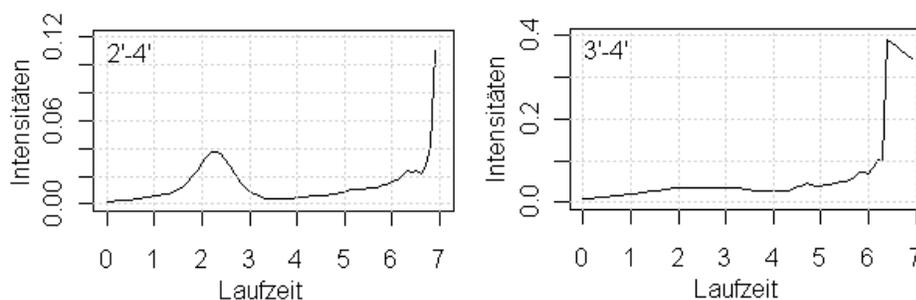


Abbildung 5.5: Kernschätzer für $q_{2'4'}(\cdot)$ (links) und $q_{3'4'}(\cdot)$ (rechts) bei der variablen Bandbreite mit $k_{min} = 4$.

Für beide Schätzer in Abbildung 5.5 ist ein großes Ausschlag im rechten Teil auffällig, was sowohl durch das Randverhalten der Kernschätzer als auch durch eine große Varianz wegen der Datenmangel verursacht wird. Am liebsten sollten in diesem Fall nur die Anleihen mit der Laufzeit bis 6 Jahre betrachtet werden.

Als Nächstes wird auf einem Beispiel gezeigt, wie eine Entscheidung über den Eintritt eines neuen Emittenten ins Portfolio getroffen werden kann. Angenommen, bei einem Emittenten im 2. Rating wird die Anleihe mit dem Startwert von 1 000 US-Dollars und der Laufzeit von 5 Jahren gekauft. Die Informationen über die Risikofreien US-Bonds-Zinssätze, sowie über die Credit Spreads für das Ende 2003 sind z. B. in Pape und Schlecker (2008) zu finden. Wegen des Informationsmangels nehmen wir hier an, dass ähnliche Kreditrisikoaufschläge auch für unsere interne Ratingsklassen gelten. Dieser Annahme zufolge entspricht der Credit Spread (r_t^p) für die Ratingsklasse 2'

dem für BBB-. Die risikofreie Spot Rate (r_t^f) übernehmen wir ebenfalls aus diesem Artikel. Beide abgeschätzten Kurven sind in Abbildung 5.6 gezeigt.

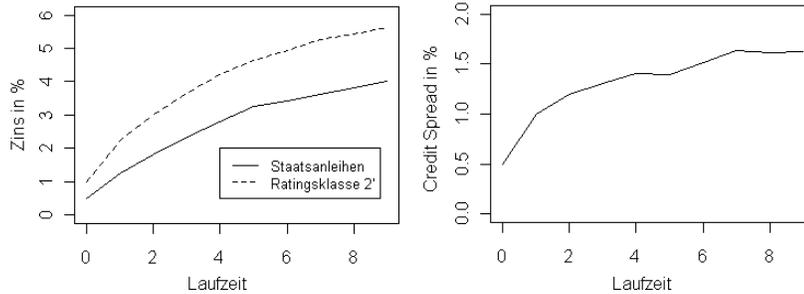


Abbildung 5.6: Spot Rates für risikofreie und risikobehaftete (Ratingsklasse 2') Anleihen (links) und Credit Spread für die Ratingsklasse 2' (rechts).

Gemäß der Spot Rate für den 2. Rating liegt das Nominal der gekauften Anleihe bei $B_T = 1000e^{\int_0^5 (r_s^f + r_s^p) ds} = 1435.8$ US-Dollars. Wird die Inflationsrate außer Acht gelassen, so gewinnt der Kreditur etwa 435.8 US-Dollars am Ende der Laufzeit. Den minimalen risikodeckenden Credit Spread für die gekaufte Anleihe berechnen wir nach der Formel (5.5) und vergleichen ihn mit dem tatsächlichen in Abbildung 5.7. Während die ersten anderthalb Jahre ohne großen Risiken verlaufen, wird das Risiko auf dem Zeitabschnitt von 1.6 bis 2.6 Jahre nicht bedeckt. Daher muss die Anleihe auf diesem Zeitabschnitt eventuell durch ein DDP gesichert werden.

Passiert allerdings der Übergang in einen schlechteren Rating in $t = 1.2$, so dehnt sich der Unsicherheitsbereich aus. Nach dem Übergang in den Rating 3' liegt der Bereich mit dem ungedeckten Risiko von 1.2 bis 3.1, was in Abbildung 5.8 zu sehen ist. In diesem Fall muss ein DDP bereits im Zeitpunkt $t = 1.2$ auf die Anleihe angewendet werden.

Die minimale Risikoprämie für DDP berechnen wir nach der Methode aus Abschnitt 5.2. Dafür bestimmen wir zuerst die Emittenten, die im Zeitpunkt 1.2 im Rating 3' waren und dann auf dem Zeitabschnitt von 1.2 bis 5 in den Konkurs gegangen sind. Von 57 Schuldner, die im 3. Rating in 1.2

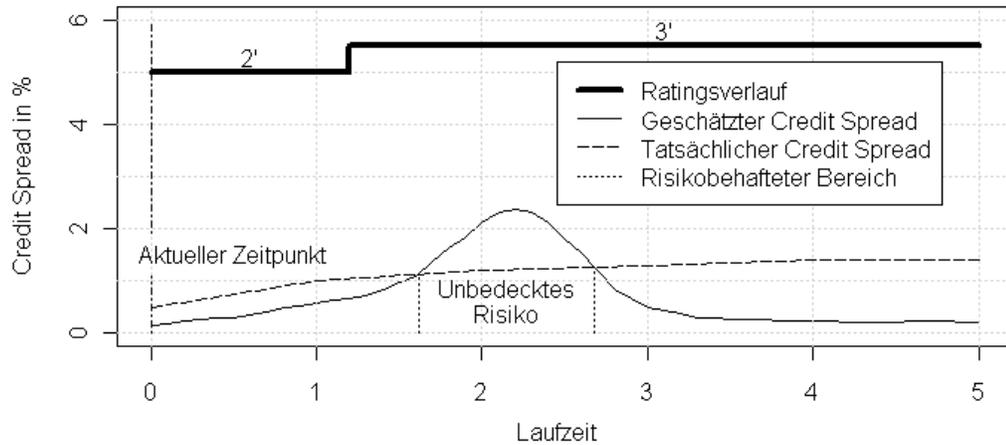


Abbildung 5.7: Vergleich des risikodeckenden Credit Spreads mit dem tatsächlichen bzgl. des aktuellen Zeitpunktes in 0.

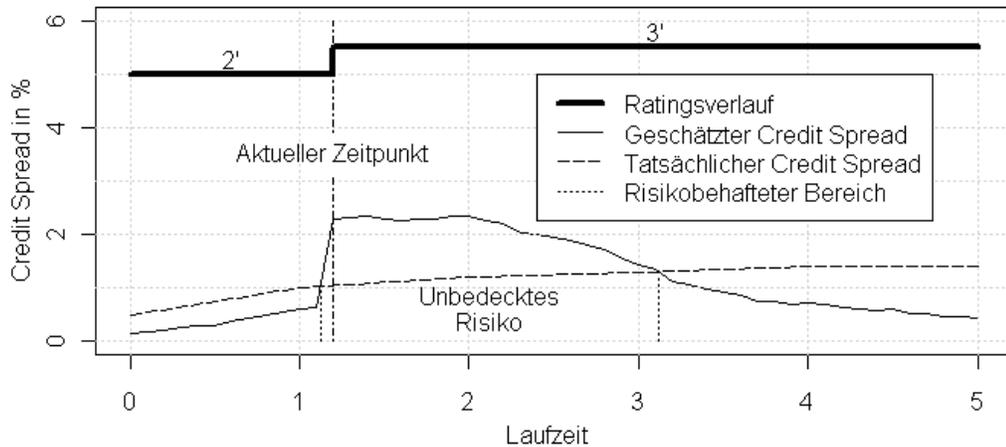


Abbildung 5.8: Vergleich des risikodeckenden Credit Spreads mit dem tatsächlichen bzgl. des aktuellen Zeitpunktes in 1.2.

waren, liegen im aktuellen Datensatz nur zwei Übergänge von 3' in 4', nämlich in Zeitpunkten 2.04 und 4.55 vor. Die Risikoprämie wird dann wie folgt geschätzt:

$$\hat{P}_\tau(1.2, 5) = 1435.8 \cdot \frac{e^{-\int_{1.2}^{2.04} r_z dz} + e^{-\int_{1.2}^{4.55} r_z dz}}{57} \approx 36.57,$$

d. h. der Kreditor muss mindestens 36.57 US-Dollars für die Versicherung der Anleihe ausgeben. Dies vermindert den Gesamtgewinn, der in diesem Fall höchstens 399.23 US-Dollars ausmacht.

Kapitel 6

Diskussion

Die Rating-basierten Modelle zur Risikobewertung in Kreditportfolios gewinnen heutzutage immer mehr an Popularität. Zu den Vorteilen dieses Ansatzes gehört eine einfache Modellierung der Insolvenzprozesse durch die stochastischen Prozesse, wobei die beobachtbaren Kreditratings ins Modell herangezogen werden.

Ausgehend von der Wichtigkeit der stochastischen Prozesse in der Risikoanalyse beschäftigte sich die vorliegende Arbeit mit der Entwicklung einer möglichst allgemeinen Methode zur Schätzung von funktionalen Charakteristiken dieser Prozesse. Als ein passender Ausgangspunkt wurde der generalisierte Kernschätzer mit der variablen Bandbreite aus Weißbach (2006) gewählt, der unter der Einhaltung von gewissen Modellannahmen mit der globalen Rate von $\mathcal{O}(\sqrt{\log(n)/(np_n)} + p_n)$ gegen die zu schätzende Charakteristik konvergiert. Eine der wichtigsten Modellannahmen ist dabei die lokale Konvergenzrate des geschätzten Integrals dieser Charakteristik. Daher wurde weiter der Fokus auf die Konstruktion eines generalisierten Schätzers mit der lokalen Konvergenzrate von $\mathcal{O}(\sqrt{\log(n)p_n/n})$ gelegt. Durch eine weitere Modellannahme wird die Konvergenzgeschwindigkeit der positiven Nullfolge (p_n) als $\mathcal{O}(\log(n)/n)$ festgelegt.

Eine mögliche Konstruktion (2.12) eines solchen Schätzers wurde als ei-

ne Verallgemeinerung der empirischen Verteilungsfunktion und kumulativen Hazardrate für rechtszeinsierte Daten vorgestellt. Bekanntlich konvergieren diese empirischen Funktionen mit der lokalen Geschwindigkeit von $\mathcal{O}(\sqrt{\log(n)p_n/n})$ gegen die theoretischen. Die erwünschte lokale Konvergenzrate dieses Schätzers wird dadurch erreicht, dass eine empirische Funktion mit der globalen Konvergenzrate von $\mathcal{O}(\sqrt{\log(n)/n})$ in die Konstruktion herangezogen wird. Eine allgemeine Konstruktion (2.18) einer solchen empirischen Funktion wird aus der empirischen Verteilungsfunktion hergeleitet, da diese eine passende globale Konvergenzrate von $\mathcal{O}(\sqrt{\log(n)/n})$ hat.

Eine wichtige Komponente des generalisierten Kernschätzers ist die variable Bandbreite, die eine bessere Anpassungsgüte gewährleistet. Wie auch im Artikel von Weißbach (2006) benutzen wir hier die Nächste-Nachbar-Bandbreite. Bei diesem Ansatz wird zuerst eine glättende Funktion festgelegt und dann wird die Bandbreite für einen festen Zeitpunkt so gewählt, dass der gesamte Sprung des Glätters innerhalb der Bandbreite durch (p_n) beschränkt wird.

Offenbar hängt auch die Konvergenzrate des Kernschätzers von der Gestaltung der Folge (p_n) . Um die Optimalitätskriterien für (p_n) herzuleiten, wurde in der vorliegenden Arbeit (p_n) als ein Produkt einer Konstante und einer Nullfolge dargestellt. Der Einfachheit halber wurde die Nullfolge dann als $([\log n/n]^a)$ mit $0 < a < 1$ (hier $a = 0.9$) gewählt, da ihre Geschwindigkeit leicht zu steuern ist. Die Herleitung einer optimalen Konstante ist etwas komplizierter und hängt sehr stark vom gewählten Glätter ab. Einerseits muss die Konstante möglichst klein sein, um die Konvergenzrate des Kernschätzers zu minimieren, andererseits erhöhen viel zu kleine Werte die Varianz des Kernschätzers. In dieser Arbeit wurde vorgeschlagen diese Konstante so zu wählen, dass die Anzahl der Sprünge der geschätzten integrierten Charakteristik konstant innerhalb der Bandbreite ist. Allerdings bleibt die Frage der Optimalität von (p_n) immer noch offen, da die angegebene Abschätzung offenbar nicht optimal ist.

Die resultierende Konstruktion des Kernschätzers ermöglicht einen schnellen Aufbau von Funktionalschätzern mit der bekannten Konvergenzrate für eine breite Klasse von stochastischen Prozessen. Die Anwendung des generalisierten Kernschätzers wurde auf den Beispielen der Überlebens- und Markov-Prozesse gezeigt, wobei neue konsistente Schätzer jeweils für die Hazardrate bei linkstrunkierten und rechtszensierten Daten und für die Migrationsintensitäten geliefert wurden.

Im Falle der linkstrunkierten und rechtszensierten Daten wurde gezeigt, dass der Schätzer für die kumulierte Hazardrate aus Cao *et al.* (2005) ein Spezialfall des generalisierten Schätzers (2.12) darstellt. Demzufolge besitzt er die gewünschte lokale Konvergenzrate und der generalisierter Kernschätzer kann zur Schätzung der Hazardrate eingesetzt werden.

Im Unterschied zu den Hazardraten ist noch kein nichtparametrischer Schätzer für die Migrationsintensitäten von inhomogenen und zeitstetigen Markov-Prozessen in der Literatur beschrieben. Daher wurde in der vorliegenden Arbeit angestrebt, diese Lücke zu schließen und einen Kernschätzer für die Migrationsintensitäten herzuleiten. Leider zeigte die vorläufige Untersuchung, dass der Nelson-Aalen-Schätzer für die kumulierten Migrationsintensitäten aus Andersen *et al.* (1993) keinen Spezialfall des generalisierten Schätzers (2.12) darstellt. Um eine passende Konstruktion des Schätzers aufzuspüren, wurden die so genannten gewichteten Migrationsintensitäten eingeführt. Diese stellen zeitabhängige Produkte von Intensitäten und Zustandswahrscheinlichkeiten dar und ihre Integrale können als ein Spezialfall von (2.12) auch unter der Berücksichtigung von unvollständigen Daten geschätzt werden. Ein Schätzer für die Migrationsintensität ergibt sich dann, indem der Kernschätzer für die gewichtete Intensität durch die geschätzte Zustandswahrscheinlichkeit geteilt wird.

Es bleibt allerdings noch ungeklärt, wie die Schätzbarkeit der theoretischen Migrationsintensitäten zu überprüfen ist. Dafür muss die Anzahl der Übergänge auf dem beobachtbaren Zeitabschnitt auf die Beschränktheit ge-

testet werden, wobei die Verteilung einer infiniten Summe von stochastisch abhängigen und nicht identisch verteilten seltenen Bernoulli-Ereignissen zu bestimmen ist. Diese Frage ist in der aktuellen Arbeit offen geblieben, da eine solche Verteilung in der Literatur nicht zu finden ist. Im empirischen Falle kann eine beschränkte Anzahl der Übergänge einfach angenommen werden, wenn nichts anderes dagegen spricht. Darüber hinaus beachten wir, dass die Migrationen in die absorbierenden Zustände nur einmal vorkommen können.

Um die analytischen Beweise der Konvergenz von Kernschätzern besser nachzuvollziehen, wurde eine Reihe von Simulationsstudien durchgeführt. Die Simulationsalgorithmen wurden dabei so entwickelt, dass alle notwendigen Verteilungen exakt dargestellt werden und auf den gegebenen zu schätzenden Charakteristiken basieren. Dadurch wurde ermöglicht, numerische Fehler beim Simulieren zu vermeiden. Der grafische Vergleich von theoretischen und geschätzten Charakteristiken hat die Konsistenz der entwickelten Schätzer bestätigt. Allerdings benötigt man einen relativ großen Datenumfang für eine „gute“ Schätzung, was zu den Nachteilen der Kernschätzer gehört. Auch ziemlich starke Ausschläge an den Rändern des beobachtbaren Intervalls erschweren die Schätzung. Vermutlich lassen sich diese Probleme beheben, indem die oben erwähnte Nullfolge optimal abgeschätzt wird.

Zum Schluss wurde eine Anwendung der Kernschätzer in der Risikoanalyse präsentiert. Es wurde gezeigt, wie der Insolvenzprozess in einem Kreditportfolio durch die Markov-Prozesse modelliert werden kann. Im Unterschied zu der üblichen Anwendung von Überlebensprozessen ermöglichen die Markov-Prozesse im vollen Umfang die Informationen aus Kreditratings zu benutzen. Bei der Berechnung des minimalen risikodeckenden Credit Spreads für einen neuen Emittenten werden sowohl sein Kreditrating als auch die geschätzten Migrationsintensitäten für das gesamte Portfolio eingesetzt.

Zusätzlich wurde im letzten Teil der Arbeit eine Erweiterung der Konkursversicherung von Anleihen (DDP) unter der Markov-Annahme erläutert. Bei der Berechnung der Risikoprämie für DDP wurde ersichtlich, dass ein

generalisierter Ansatz zur konsistenten und stetigen Schätzung auch für die integrierten Charakteristiken der stochastischen Prozesse benötigt wird. Anhand des existierenden Ansatzes zur Kernschätzung kann auch diese Aufgabe zukünftig gelöst werden.

Die Weiterentwicklung des Kernschätzers betrifft vor allem die Herleitung einer optimalen Methode zur Berechnung der Nächste-Nachbar-Bandbreite. Auch die Verteilung der Summe von stochastisch abhängigen Bernoulli-Variablen zur sicheren Schätzung von Migrationsintensitäten muss genauer untersucht werden. Als ein weiterer Schritt können die Anpassung des generalisierten Kernschätzers an die anderen stochastischen Prozesse, sowie die Anwendung dieser Methode in den anderen Einsatzbereichen wie z. B. Demographie erfolgen.

Literaturverzeichnis

- Amman, M. (2001): *Credit risk valuation*, Springer.
- Andersen, P. K., Borgan, Ø., Gill, R. D. und Keiding, N. (1993): *Statistical Models Based on Counting Processes*, Springer.
- Anderson, F., W. (1991): *Continuous-Time Markov Chains: an applications-oriented approach*, Springer.
- Black, F. und Scholes, M. (1973): The pricing of options and corporate liabilities, *Journal of Political Economy* **81**, 637–654.
- Cao, R., Janssen, P. und Veraverbeke, N. (2005): Relative hazard rate estimation for right censored and left truncated data, *Sociedad de Estadística e Investigación Operativa Test* **14**, 257–280.
- Devroye, L. und Penrod, C. (1984): The consistency of automatic kernel density estimates, *Annals of Statistics* **12**, 1231–1249.
- Duffie, D. und Singleton, K. (1999): Modeling term structures of defaultable bonds, *Review of Financial Studies* **12**, 197–226.
- Epanechnikov, V. (1969): Nonparametric estimate of a multivariate probability density, *Theory of Probability and Its Applications* **14**, 153–158.
- Failing, K. (1983): *Die Verallgemeinerung und Konsistenz von nichtparametrischen Dichte- und Hazard-Raten-Schätzern für den Fall zensierter Daten*, Doktorarbeit, Universität Gießen.

- Földes, A., Rejtő, L. und Winter, B. (1981): Strong consistency properties of nonparametric estimators for randomly censored data. part ii: Estimation of density and failure rate, *Periodica Mathematica Hungarica* **12**, 15–29.
- Gasser, T., Müller, H.-G. und Mammitzsch, V. (1985): Kernels for nonparametric curve estimation, *Journal of the Royal Statistical Society B* **47**, 238–352.
- Hougaard, P. (2000): *Analysis of Multivariate Survival Data*, Springer.
- Jarrow, R. A., Lando, D. und Turnbull, S. M. (1997): A Markov model for term structure of credit risk spread, *Review of Financial Studies* **10**, 481–522.
- Jarrow, R. A. und Turnbull, S. M. (1995): Pricing derivatives on financial securities subject to credit risk, *Journal of Finance* **50**, 53–85.
- Johansen, S. (1978): The product limit estimator as maximum likelihood estimator, *Scandinavian Journal of Statistics* **5**, 195–199.
- Madan, D. und Unal, H. (2000): A two-factor hazard rate model for pricing risky debt and the term structure of credit spreads, *Journal of Financial and Quantitative Analysis* **35**, 43–65.
- Merton, R. (1974): On the pricing of corporate debt: The risk structure of interest rates, *Journal of Finance* **29**, 449–470.
- Moore, O. und Yackel, J. (1977): Consistency properties of nearest neighbor density function estimators, *Annals of Mathematical Statistics* **5**, 143–154.
- Nadaraya, E. (1965): On nonparametric estimates of density functions and regression curves, *Theory of Probability and its Applications* **10**, 186–190.
- Pape, U. und Schlecker, M. (2008): Berechnung des credit spreads, *Finanz Betrieb* **10**, 658–665.

- Parzen, E. (1962): On the estimation of a probability density function and the mode, *Annals of Mathematical Statistics* **33**, 1065–1076.
- Ramlau-Hansen, H. (1983): Smoothing counting process intensities by means of kernel functions, *Annals of Statistics* **11**, 453–466.
- Rao, B. L. S. P. (1983): *Nonparametric functional estimation*, Academic Press, Orlando, Florida.
- Rosenblatt, M. (1956): Remarks on some nonparametric estimates of a density function, *Annals of Mathematical Statistics* **27** (3), 832–837.
- Schäfer, H. (1985): A note on data-adaptive kernel estimation of the hazard and density function in the random censorship situation, *Annals of Statistics* **13**, 818–820.
- Schäfer, H. (1986a): *Die Konvergenz des variablen Kernschätzers und die Geschwindigkeit der lokalen Konvergenz empirischer Maße bei zufällig zensurierten Daten*, Doktorarbeit, Mathematisch-Naturwissenschaftliche Fakultät, Universität Düsseldorf.
- Schäfer, H. (1986b): Local convergence of empirical measures in the random censorship situation with application to density and rate estimators, *Annals of Statistics* **14**, 1240–1245.
- Strook, D. W. (2000): *An Introduction to Markov Processes*, Springer.
- Stute, W. (1982): The oscillation behaviour of empirical processes, *Annals of Probability* **10**, 86–107.
- Tanner, M. und Wong, W. (1983): The estimation of the hazard function from randomly censored data by kernel method, *Annals of Statistics* **11**, 989–993.
- Wand, M. und Jones, M. (1995): *Kernel Smoothing*, Chapman & Hall, London.

- Weißbach, R. (2006): A general kernel functional estimator with general bandwidth - strong consistency and applications, *Nonparametric Statistics* **18** (1), 1–12.
- Weißbach, R. und Dette, H. (2007): Kolmogorov-Smirnov-type testing for the partial homogeneity of Markov processes – with application to credit risk, *Stochastic Models in Business and Industry* **23**, 223–234.
- Weißbach, R., Tschiersch, P. und Lawrenz, C. (2009): Testing time-homogeneity of rating transitions after origination of debt, *Empirical Economics* **36**, 575–596.
- Yandell, B. (1983): Non-parametric inference for rates and densities with censored serial data, *Annals of Statistics* **11**, 1119–1135.

Anhang A

Theoretische Grundlagen

A.1 Limes superior Formulierung der Konvergenz

Lemma A.1.1. $(Y_i)_{i \in \mathbb{N}}$ sei eine Folge unabhängiger identisch verteilten Zufallsvektoren $Y_i : \Omega \rightarrow \mathbb{R}^d$.

$S_n : (\mathbb{R}^d)^n \rightarrow \mathbb{R}^+$ sei eine Folge von Borel-meßbaren symmetrischen Abbildungen, d. h. $S_n(y_1, \dots, y_n) = S_n(y_{\sigma(1)}, \dots, y_{\sigma(n)})$ für alle $y_1, \dots, y_n \in \mathbb{R}^d$ und alle Permutationen $\sigma : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$. Sei (a_n) eine Nullfolge positiver reeller Zahlen.

Dann sind folgende Konvergenzaussagen äquivalent:

(i) Für alle $\alpha > 1$ ist

$$P(\omega : \exists N \in \mathbb{N} \forall n > N : S_n(Y_1(\omega), \dots, Y_n(\omega)) \leq \alpha a_n) = 1$$

(ii) Es existiert eine Konstante $D \leq 1$ mit

$$P(\omega : \limsup_{n \rightarrow \infty} S_n(Y_1(\omega), \dots, Y_n(\omega))/a_n = D) = 1$$

Beweis. Siehe Beweis zum Lemma 1.1.5 in Schäfer (1986a). \square

A.2 Verteilung des absoluten Mittelwertes von u.i.v. Zufallsvariablen

Satz A.2.1. T_1, \dots, T_n seien beschränkte unabhängige Zufallsvariablen mit Erwartungswert 0 und gleicher Varianz σ^2 , also

$$E(T_i) = 0$$

$$|T_i| \leq b$$

$$\sigma^2 := \text{Var}(T_i)$$

für $i = 1, \dots, n$.

Sei

$$\bar{T} := \frac{1}{n} \sum_{i=1}^n T_i$$

das arithmetische Mittel.

Dann gilt für jedes $\varepsilon > 0$

$$P(|\bar{T}| \geq \varepsilon) \leq 2 \exp\left(-\frac{n\varepsilon^2}{2\sigma^2 + 2b\varepsilon/3}\right)$$

Beweis. Siehe Beweis zum Satz 1.2.1 in Schäfer (1986a). \square