

Lebensdauermodelle mit flexibler Hazardrate

Dissertation

an der Fakultät für Mathematik, Informatik und Statistik

der Ludwig–Maximilians–Universität München

zur Erlangung des Grades Doctor rerum naturalium (Dr. rer. nat.)

vorgelegt von

Stefan Pilz

eingereicht am 15. November 2005

Erstgutachter: Prof. Dr. Fahrmeir
Zweitgutachter: PD. Dr. Heumann
Externer Gutachter: Prof. Dr. Kauermann
Tag des Rigorosums: 26.4.2006

Vorwort

Diese Arbeit entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Institut für Statistik der Ludwig-Maximilians-Universität München. Ich möchte mich bei allen bedanken, die in vielerlei Hinsicht zum Gelingen der Arbeit beigetragen haben.

Mein besonderer Dank gilt Prof. Dr. Ludwig Fahrmeir für die Betreuung der Arbeit und dass er stets die Zeit fand für konstruktive Besprechungen in einer angenehmen und ruhigen Atmosphäre. Weiter möchte ich mich bei PD. Dr. Christian Heumann bedanken, der das Koreferat übernahm und für sein Interesse an der Arbeit sowie zahlreiche fruchtbare Gespräche.

Weiter möchte ich mich bei Prof. Dr. Leo Knüsel bedanken, der mir stets die Freiräume für die Forschung ließ sowie für die gute Zusammenarbeit bei der gemeinsamen Betreuung des studentischen Computernetzwerkes, wobei ich auch die Mitarbeit von Dipl.- Stat. Volker Martin sehr zu schätzen wußte.

Zudem möchte ich mich bedanken bei Prof. Dr. Thomas Augustin und Prof. Dr. Helmut Küchenhoff, die mich in Zeiten knapper Budgets unterstützt haben, deren vertrauensvolle Zusammenarbeit mir angenehm im Gedächtnis bleiben wird.

Bei Dr. Jens Oehlschlägel und Dr. Korbinian Strimmer möchte ich mich ganz herzlich bedanken für fachliche Diskussionen und Tips zu dem Softwarepaket *R*, letzterem auch für \LaTeX -Tips. Nicht zuletzt möchte ich mich bei Dipl.- Stat. Sebastian Maier und Dr. Thomas Nittner für das Korrekturlesen und die moralische Unterstützung bedanken.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Problemstellung und Aufbau	1
1.2	Daten und Analyse	3
2	Theoretische Grundlagen	10
2.1	Stochastische Prozesse	10
2.2	Elemente der diskreten stochastischen Analysis	13
2.2.1	Diskrete Martingale	14
2.2.2	Stochastische Integration	15
2.2.3	Doob–Meyer–Zerlegung	17
2.3	Elemente der stetigen stochastischen Prozesse	18
2.3.1	Stetige Martingale und stochastische Integration	21
2.3.2	Doob–Meyer–Zerlegung	22
2.3.3	Punktprozesse und Kompensatoren	24
2.3.4	Der Likelihoodprozeß	25
3	Modellierung und Statistische Inferenz bei Lebensdauermodellen	29
3.1	Zensierung von Beobachtungen	30
3.2	Multiplikatives Intensitätsmodell	31
3.3	Multiplikatives Hazardmodell	32
3.4	Modelle mit zeitabhängigen Parametern	34
3.5	Schätzfunktionen	34
3.5.1	Konzept der Schätzfunktionen	35
3.5.2	Schätzfunktionen und Zählprozesse	37
3.5.3	Schätzfunktion und Lebensdauermodelle	40
3.6	Bayesianische Inferenz für Lebensdauermodelle	40
3.6.1	Anmerkungen zu Bayesianischen Verfahren bei hochdimensionalen Modellen	41
3.6.2	Simulationsmethoden	42
3.6.3	MCMC Verfahren	43
3.6.4	Hybride Monte Carlo Verfahren	46

4	Lebensdauermodelle mit flexibler Hazardrate	51
4.1	Modell einer allgemeinen Hazardrate	51
4.2	Splinefunktionen	52
4.2.1	Definition und Basisfunktionen	53
4.2.2	Penalisierte Splines	54
4.3	Gemischte Modelle und Zufallseffekte	55
4.3.1	Definition der Gemischten Modelle	55
4.3.2	Gemischte Modelle mit penalisierten Splines	59
4.3.3	Gemischte Modelle für die Lebensdaueranalyse	62
4.4	Bemerkungen zur Wahl des Splinegrades	63
4.5	Modellierung und Schätzung der Parameter	64
4.5.1	Modellierung der Intensitätsrate	64
4.5.2	Schätzung der Intensitätsrate	66
4.6	Schätzung im bayesianischen Kontext	91
4.6.1	Darstellung der a posteriori Verteilung	91
4.6.2	Modifikation des Leapfrog Algorithmus	94
5	Implementierung der Methoden und Anwendung	96
5.1	Einführende Beispiele	96
5.1.1	Ein naives Beispiel	96
5.1.2	Beispiel zur komplexeren Schätzung	98
5.2	Bayesianische Verfahren	99
5.2.1	Simulationstudien	99
5.2.2	Ergebnisse zu Datensatz PBC	126
5.3	Verfahren mit Schätzfunktionen	135
5.3.1	Modell mit zeitunabhängigen Kovariableneffekten	135
5.3.2	Modell mit zeitabhängigen Kovariableneffekten	136
5.3.3	Lokale Modellierung und numerische Kondition	137
6	Zusammenfassung und Ausblick	139
A	Ergebnisse zum Cox Modell	141
B	Ergebnisse zur Numerischen Integration	144
C	Ergebnisse der Simulationsstudien	152
D	Ergebnisse der Analyse des Datensatzes	155

D.1	Modell mit flexibler Baseline-Hazardrate	155
D.2	Modell mit flexibler Baseline-Hazardrate, sowie glatten Kovariablenef- fekten für $\log(\text{bili})$, $\log(\text{protime})$	159
D.2.1	Modellierung mit Polynomen vom Grad 3	159
D.2.2	Modellierung mit Splines vom Grad 2	164
D.3	Modell mit flexibler Baseline-Hazardrate, sowie glatten Kovariablenef- fekten für $\log(\text{bili})$, edema , $\log(\text{protime})$	167
D.3.1	Modellierung mit Polynomen vom Grad 3	167
D.3.2	Modellierung mit Splines vom Grad 2	173
D.4	Ergebnisse der Schätzfunktionen	176
	Literatur	177

Abbildungsverzeichnis

1.1	glatter Effekt der Variable <i>age</i> mit partiellen Schoenfeldresiduen	6
1.2	glatter Effekt der Variable <i>log(albumin)</i> mit partiellen Schoenfeldresiduen	7
1.3	glatter Effekt der Variable <i>log(bili)</i> mit partiellen Schoenfeldresiduen	7
1.4	glatter Effekt der Variable <i>edema</i> mit partiellen Schoenfeldresiduen	8
1.5	glatter Effekt der Variable <i>log(protime)</i> mit partiellen Schoenfeldresiduen	8
1.6	Martingalresiduen mit geglättetem (<i>lowess</i>) funktionalen Verlauf	9
4.1	Beträge der Chebychev Koeffizienten für erste Stammfunktion von <i>f</i> , Darstellung auf logarithmischer Skala	86
4.2	Beträge der Chebychev Koeffizienten für iterierte Stammfunktionen von <i>f</i> , Darstellung auf logarithmischer Skala	87
4.3	Beträge der Chebychev Koeffizienten für erste Stammfunktion von <i>g</i> , Darstellung auf logarithmischer Skala	88
4.4	Beträge der Chebychev Koeffizienten für iterierte Stammfunktionen von <i>g</i> , Darstellung auf logarithmischer Skala	89
4.5	Beträge der Chebychev Koeffizienten für erste Stammfunktion von <i>h</i> , Darstellung auf logarithmischer Skala	90
4.6	Beträge der Chebychev Koeffizienten für iterierte Stammfunktionen von <i>h</i> , Darstellung auf logarithmischer Skala	90
5.1	‘quartilswise’ Schätzung der Hazardrate	97
5.2	Hazardrate, theoretisch und zwei Schätzungen	99
5.3	Polynom Grad 1, $n = 200$	104
5.4	Polynom Grad 1, Samplingpfade der Hazardrate, $n = 200$	104
5.5	Polynom Grad 2, $n=200$	105
5.6	Spline Grad 1 mit 9 Gitterpunkten, $n = 200, \mu = -10$	105
5.7	Polynom Grad 1, $n = 500$	105
5.8	Polynom Grad 1, $n = 1000$	106
5.9	Spline Grad 1 mit 9 Gitterpunkten, $n = 500$	106

5.10	Spline Grad 1 mit 9 Gitterpunkten, $n = 1000$	106
5.11	Polynom Grad 1, $n = 200$	107
5.12	Polynom Grad 2, $n = 200$	107
5.13	Polynom Grad 3, $n = 200$	108
5.14	Spline Grad 1, 9 Gitterpunkte, $n = 200$	108
5.15	Spline Grad 1, 9 Gitterpunkte, $n = 1000$	109
5.16	Polynom Grad 1, $n = 200$	109
5.17	Polynom Grad 2	110
5.18	Spline Grad 1, 9 Gitterpunkte	110
5.19	Polynom Grad 1, $n = 200$	111
5.20	Polynom Grad 2	111
5.21	Polynom Grad 3	111
5.22	Spline Grad 1, 9 Gitterpunkte, $n = 200, \mu = -6$	112
5.23	Spline Grad 1, 9 Gitterpunkte, $n = 200, \mu = +5$	112
5.24	Spline Grad 1, $\mu_0 = -10$	113
5.25	Spline Grad 1, $\mu_0 = +5$	114
5.26	Boxplot für Datensatz mit $n = 200$	116
5.27	Polynom vom Grad 1	116
5.28	Polynom Grad 1	117
5.29	Polynom Grad 1	118
5.30	Polynom vom Grad 1	118
5.31	Spline vom Grad 1	119
5.32	Polynom vom Grad 1	119
5.33	Spline vom Grad 1, 15 Gitterpunkte, ohne Zensierung	122
5.34	Spline vom Grad 1, 15 Gitterpunkte, ohne Zensierung	122
5.35	Spline vom Grad 1, 15 Gitterpunkte	123
5.36	Spline vom Grad 1, 15 Gitterpunkte, ohne Zensierung	123
5.37	Spline vom Grad 1, 7 Gitterpunkte, ohne Zensierung	123
5.38	Spline vom Grad 1, 7 Gitterpunkte, ohne Zensierung	124
5.39	Spline vom Grad 1, 7 Gitterpunkte	124
5.40	Baseline-Hazardrate, Polynom Grad 3	128
5.41	zeitvariierender Effekt für $\log(\text{bili})$, Polynom Grad 3	129
5.42	zeitvariierender Effekt für edema , Polynom Grad 3	129
5.43	zeitvariierender Effekt für $\log(\text{protime})$, Polynom Grad 3	130
5.44	Baseline-Hazardrate, Spline Grad 2, 4 Gitterpunkte	131
5.45	zeitvariierender Effekt für $\log(\text{bili})$, Spline Grad 2, 4 Gitterpunkte	132
5.46	zeitvariierender Effekt für edema , Spline Grad 2, 4 Gitterpunkte	132

5.47	zeitvariierender Effekt für $\log(\text{protime})$, Spline Grad 2, 4 Gitterpunkte	133
A.1	durchschnittliche Überlebenszeiten in Abhängigkeit von Kovariablen	143
D.1	Samplingpfade und geschätzte Dichte der Parameter der Hazardrate	157
D.2	Samplingpfade und geschätzte Dichte der Parameter der Hazardrate, modifiziertes Verfahren	158
D.3	Baseline-Hazardrate, Polynom, Grad 3	160
D.4	zeitvariierender Effekt für $\log(\text{protime})$, Polynom, Grad 3	160
D.5	zeitvariierender Effekt für $\log(\text{bili})$, Polynom, Grad 3	161
D.6	Samplingpfade und geschätzte Dichte der Parameter der Hazardrate, Polynom, Grad 3, modifiziertes Verfahren	161
D.7	Samplingpfade und geschätzte Dichte der Parameter für $\log(\text{bili})$, Polynom Grad 3, modifiziertes Verfahren	162
D.8	Samplingpfade und geschätzte Dichte der Parameter für $\log(\text{protime})$, Polynom Grad 3, modifiziertes Verfahren	162
D.9	Samplingpfade und geschätzte Dichte der Parameter für age , edema und $\log(\text{albumin})$, Polynom Grad 3, modifiziertes Verfahren	163
D.10	Baseline-Hazardrate, Spline, Grad 2	165
D.11	zeitvariierender Effekt für $\log(\text{bili})$, Spline, Grad 2	165
D.12	zeitvariierender Effekt für $\log(\text{protime})$, Spline, Grad 2	166
D.13	Samplingpfade und geschätzte Dichte der Parameter der Hazardrate, Polynom, Grad 3, modifiziertes Verfahren	168
D.14	Samplingpfade und geschätzte Dichte der Parameter für $\log(\text{bili})$, Polynom, Grad 3, modifiziertes Verfahren	169
D.15	Samplingpfade und geschätzte Dichte der Parameter für edema , Polynom, Grad 3, modifiziertes Verfahren	170
D.16	Samplingpfade und geschätzte Dichte der Parameter für $\log(\text{protime})$, Polynom, Grad 3, modifiziertes Verfahren	171
D.17	Samplingpfade und geschätzte Dichte der Parameter für age und $\log(\text{albumin})$, Polynom, Grad 3, modifiziertes Verfahren	172
D.18	Samplingpfade und geschätzte Dichte der Parameter der Baseline-Hazardrate, Spline Grad 2, 4 Gitterpunkte	174
D.19	Samplingpfade und geschätzte Dichte der Parameter der Baseline-Hazardrate, Spline Grad 2, 4 Gitterpunkte	174
D.20	Samplingpfade und geschätzte Dichte der Parameter für $\log(\text{bili})$, Spline Grad 2, 4 Gitterpunkte	175

D.21 Samplingpfade und geschätzte Dichte der Parameter für $\log(\text{bili})$, Spline Grad 2, 4 Gitterpunkte	175
---	-----

Tabellenverzeichnis

5.1	Parameterschätzungen für ein Polynom ersten Grades	117
A.1	Tabellen zum Datensatz PBC N=312	142
B.1	Fehler bei der Numerischen Integration eines exponentiellen Splines vom Grad 1	145
B.2	Fehler bei der Numerischen Integration eines exponentiellen Splines vom Grad 1	146
B.3	Fehler bei der Numerischen Integration eines exponentiellen Splines vom Grad 1	147
B.4	Fehler bei der Numerischen Integration eines Polynoms	148
B.5	Fehler bei der Numerischen Integration eines Polynoms	149
B.6	Fehler bei der Numerischen Integration eines Polynoms	150
B.7	Numerische Integration eines exponentiellen Splines	151
C.1	Werte für ein Cox Weibull Modell mit Spline Grad 1	153
D.1	Parameterschätzungen mit Polynom vom Grad 1, modifiziertes Verfahren	155
D.2	Parameterschätzungen mit Polynom vom Grad 1, modifiziertes Verfahren, Kovariablen zentriert	156
D.3	Parameterschätzungen mit Polynom vom Grad 1	156
D.4	Parameterschätzungen mit Polynom vom Grad 1, Kovariablen zentriert .	157
D.5	Parameterschätzungen (Modell: Polynom, Grad 3, modifiziertes Verfahren)	159
D.6	Parameterschätzungen Modell: Spline, Grad 2, 4 Gitterpunktemodifiziertes Verfahren)	164
D.7	Parameterschätzungen (Modell: Polynom, Grad 3, modifiziertes Verfahren)	167
D.8	Parameterschätzungen (Modell: Spline Grad 2, 4 Gitterpunkte)	173
D.9	Parameterschätzungen	176

1 Einleitung

Der erste Abschnitt beschreibt die Ziele und das Vorgehen in dieser Arbeit. Im Anschluß wird eine Studie vorgestellt. Zum Zweck der Motivation wird eine Datenanalyse mit herkömmlichen Methoden durchgeführt, die Anlaß zu einer allgemeineren Modellierung, die in dieser Arbeit behandelt wird, gibt.

1.1 Problemstellung und Aufbau

Diese Arbeit behandelt Lebensdauermodelle in einem allgemeinen Rahmen. Aufgezeichnet wird bei den Lebensdauermodellen der Zeitpunkt des Eintretens von Ereignissen, wobei nicht ausgeschlossen wird, dass keine Beobachtung erfolgt. In letzterem Fall spricht man von Zensierung. Die Beobachtung der Zeitpunkte der Ereignisse erfordert, dass zumindest approximativ zeitstetige Modelle verwendet werden. Derartige Daten sind z.B. typisch im medizinischen Bereich, aber auch im Finanzbereich. Allgemein dürfen die Daten das Ergebnis eines beliebigen (markierten) Punktprozesses sein. Neben den Ereigniszeitpunkten werden noch Kovariablen beobachtet, deren Einfluß auf die Ereigniszeitpunkte modelliert wird. Eine derartige Situation liegt z.B. dem Cox-Modell zugrunde. Voraussetzung hierfür sind allerdings (zeitlich) feste Effekte der Kovariablen. Man kann sich nun überlegen, ob der Einfluß einem zeitlichen Wandel unterliegt. So wird sich etwa der Effekt einer medizinischen Therapie über die Zeit ändern, oder es können Nebenwirkungen auftreten. Auch ohne detaillierte Erläuterung ist unmittelbar einsichtig, dass zeitvariierende Effekte wesentlich komplizierter sind als feste Effekte. Es ist also stets abzuwägen, ob ein einfaches Modell die Realität hinreichend gut modelliert, oder ob ein komplexeres Modell angemessener ist. Substanzwissenschaftliche Aussagen sollten hierbei berücksichtigt werden. Hinweise über die Güte der Modellierung liefert z.B. die Residuenanalyse (vgl. Therneau, Grambsch und Fleming (1990) bzw. Grambsch und Therneau (1994)).

Da es in der Statistik verschiedene Strömungen gibt, existieren auch verschiedene Modellierungsansätze, unter anderem ein frequentistischer Ansatz bzw. ein bayesianischer Ansatz. Die den beiden Ansätzen zugrundeliegenden Modelle sollen dargestellt und ihre

Ergebnisse verglichen werden.

Die Modellierung im zeitstetigen Fall erfordert ein mathematisch relativ technisches Instrumentarium mit entsprechend anspruchsvoller Literatur (vgl. Andersen, Borgan, Gill und Keiding (1993)), die vermutlich viele Nichtmathematiker abschreckt. Im zweiten Kapitel werden daher einige Grundlagen über stochastische Prozesse zusammengetragen. Es folgen Aussagen über Martingale, Punkt- und Zählprozesse, die Doob–Meyer–Zerlegung sowie über den Likelihoodprozeß als Verallgemeinerung des Likelihoodquotienten und den Intensitätsprozeß, den man als Verallgemeinerung der Hazardrate einer Lebensdauervertelung betrachten kann. Dies ist notwendig für die weitere Darstellung und stellt zudem den Bezug zur Literatur her.

Das dritte Kapitel beschreibt Lebensdauermodelle sowie die Behandlung von zensierten Beobachtungen. Das Konzept der Schätzfunktionen wird als Basis für den frequentistischen Ansatz kurz eingeführt. Es folgen Methoden aus dem bayesianischen Bereich, insbesondere Markov Chain Monte Carlo–Verfahren.

Das vierte Kapitel befaßt sich mit den behandelten Modellen. Für die zeitvariierenden Effekte werden Splines verwendet. Dazu benötigt man normalerweise einen Glattheitsparameter, der durch aufwendige Verfahren (z.B. Kreuzvalidierung oder AIC–Kriterium) bestimmt wird. Im Rahmen dieser Arbeit wird darauf verzichtet – in der Hoffnung günstigere Verfahren zu verwenden. Anstelle der genannten Verfahren folgt eine Einbettung in gemischte Modelle, die daher kurz ebenfalls im vierten Kapitel eingeführt werden. Es folgt eine ausführliche Darstellung der zu modellierenden Intensitätsrate sowie von Schätzverfahren im frequentistischen Sinne, die auf den vorherigen Kapiteln aufbauen. Die dazu notwendigen numerischen Verfahren werden kurz dargestellt – im einzelnen ein Verfahren zur Nullstellenbestimmung sowie ein Verfahren zur numerischen Integration. Es folgt eine ausführliche Erörterung der bayesianischen Modellierung mit Markov Chain Monte Carlo–Verfahren. Danach folgt eine Anwendung auf reale Daten. Es soll aufgezeigt werden, wie auf grafische Weise die Zeitabhängigkeit der Effekte – als Abweichung vom Cox–Modell – aufgezeigt werden kann.

Die beiden Ansätze sollen verglichen werden. Im Rahmen dieser Arbeit werden zwar verschiedene Modelle analysiert, es geht dabei aber nicht explizit um Modellwahl (vgl. hierzu Bunea (2004)). Eine (fundamentale) Kritik dazu findet man in Leeb und Pötscher (2005).

1.2 Daten und Analyse

Es handelt sich um eine ('doubleblinded') randomisierte Studie aus dem medizinischen Bereich, die von der Mayo-Klinik 1986 publiziert wurde (vgl. Fleming und Harrington (1991, S. 2) und Dickinson, Grambsch, Fleming, Fisher und Langworthy (1989)). Die Daten wurden im Zeitraum von Januar 1974 bis Mai 1986 erhoben. Die Personen leiden an einer primären Leberzirrhose (kurz: PBC). Die Behandlung erfolgte in der einen Gruppe mit einem Medikament (D-Penicillame), in der anderen Gruppe mit einem Placebo. Es liegen Daten von 312 Personen vor. Neben der Überlebenszeit existieren zwanzig Kovariablen aus den Bereichen

- Demographie,
- klinische Untersuchung,
- biochemische Befunde,
- Histologie.

Zu beachten ist, dass bei einigen Patienten Werte zu bestimmten Kovariablen fehlen, z.B für den Cholesterinspiegel. In der Literatur (Fleming und Harrington, 1991, S. 156 ff) wird zur Analyse ein Cox-Modell verwendet (vgl. Abschnitt 3.3). Die Berechnung erfolgt mit dem Statistikpaket R, Version 1.9. Eine gut lesbare Einführung findet sich z.B in Dalgaard (2002) bzw. in Fox (2002). Zur Modellierung wurden die 5 Kovariablen verwendet (`age`, `albumin`, `bili`, `edema`, `protime`), zu denen bei allen 312 Personen die Werte der Kovariablen vorliegen. Der Effekt der medikamentösen Behandlung erwies sich als nicht signifikant, daher wird er nicht weiter aufgeführt. Die Kovariablen betreffend biochemischer Einflußgrößen (`albumin`, `bili`, `protime`) wurden mit dem natürlichen Logarithmus transformiert. Die folgende Tabelle enthält die Ergebnisse des geschätzten Modells. Der Wert von 1.0338 besagt, dass bei sonst gleichen Kovariablen und einer Erhöhung der Variable `age` um ein Jahr, die Hazardrate für das Ereignis, an der Erkrankung zu sterben, um den Faktor 1.0338 (oder um 3.38%) steigt. Die nächste Spalte enthält den Standardfehler der Parameterschätzung, hinter z verbirgt sich der Quotient aus Parameterschätzung und Standardfehler, der asymptotisch unter der Nullhypothese ($\beta = 0$) normalverteilt ist. Beträgsmäßig große Werte sprechen dafür, dass der Parameter von Null verschieden ist. Die nächste Spalte enthält den p -Wert unter der Normalverteilungsannahme. Der Likelihoodquotienten- (bzw. der Wald- bzw. der Score-)Test haben als globale Nullhypothese, dass alle Parameter gleich Null sind, die gleichermaßen abgelehnt wird.

	coef	exp(coef)	se(coef)	z	p
age	0.0332	1.0338	0.00866	3.84	1.2e-04
log(albumin)	-3.0599	0.0469	0.72404	-4.23	2.4e-05
log(bili)	0.8801	2.4110	0.09874	8.91	0.0e+00
edema	0.7859	2.1943	0.29897	2.63	8.6e-03
log(protime)	3.0140	20.3690	1.02395	2.94	3.2e-03

	exp(coef)	exp(-coef)	lower .95	upper .95
age	1.0338	0.9673	1.0164	1.051
log(albumin)	0.0469	21.3260	0.0113	0.194
log(bili)	2.4110	0.4148	1.9868	2.926
edema	2.1943	0.4557	1.2213	3.943
log(protime)	20.3690	0.0491	2.7377	151.549

Rsquare= 0.472 (max possible= 0.983)

Likelihood ratio test= 199 on 5 df, p=0

Wald test = 198 on 5 df, p=0

Score (logrank) test = 270 on 5 df, p=0

Die oben zitierten Autoren argumentieren, dass das Modell mit den biologischen Aspekten der Krankheit im Einklang steht. Die negative Schätzung des Parameters für $\log(\text{albumin})$ spiegelt die Tatsache wider, dass die Leber bei fortschreitender Krankheit den Eiweißstoff Albumin nicht mehr herstellen kann. Der positive Effekt von $\log(\text{protime})$ steht dazu nicht im Widerspruch. Gemessen wird der Zeitraum bis zur Blutgerinnung, der sich erhöht, da die Leber das Protein Prothrombin nicht mehr in ausreichender Menge produziert. Der Effekt für bili ist positiv, da Bilirubin, ein Abbauprodukt des Blutfarbstoffes Hämoglobin, aufgrund der Lebererkrankung nur mehr verlangsamt abgebaut werden kann und damit der Bilirubinspiegel steigt. Die Parameterschätzung für edema entspricht dem Anschwellen der Gewebe im fortgeschrittenen Stadium der Krankheit.

Das Cox-Modell setzt voraus, dass der Effekt der Parameter zeitlich konstant ist. Eine relativ einfache Methode zur Überprüfung der Proportional-Hazard-Annahme dürfte wohl die Stratifizierung sein, d.h. man bildet Schichten bezüglich der Ausprägungen einer Kovariable und berechnet pro Schicht eine eigene Baseline-Hazardrate. Die

Baseline-Hazardrate ist eine Funktion der Zeit. Unterscheiden sich die Raten, spricht das für einen Kovariableneffekt, der zeitlich nicht konstant ist. Dieses Vorgehen ist insbesondere sinnvoll, falls die Kovariable nur wenige Ausprägungen hat. Testet man die Proportional-Hazard-Annahme erhält man folgende Tabelle.

	rho	chisq	p
age	-0.0369	0.1494	0.6991
log(albumin)	-0.0149	0.0288	0.8652
log(bili)	0.1564	2.8505	0.0913
edema	-0.1424	2.4558	0.1171
log(protime)	-0.1910	3.6932	0.0546
GLOBAL	NA	9.5632	0.0886

Die p -Werte für die Variablen $\log(\text{bili})$, edema , $\log(\text{protime})$ legen nahe, dass die Parametereffekte nicht unbedingt konstant über die Zeit sind. Der approximative Test basiert auf den skalierten Schoenfeldresiduen (Grambsch und Therneau, 1994). Sei $\beta_j(t) = \beta + \theta_j g_j(t)$, dann ist die Proportional-Hazard-Annahme erfüllt (entspricht Nullhypothese), falls $\theta_j = 0$ gilt. Die Grafiken (Abb. (1.1) bis (1.5)) zeigen die Schoenfeldresiduen (vgl. Abschnitt (3.3)), jeweils mit Splineschätzungen der Funktionen $g_j(t)$ und Konfidenzbändern, entsprechend dem zweifachen Standardfehler. Systematische Abweichungen von einer horizontalen Linie indizieren eine Abweichung von der Annahme, dass die Parameter zeitlich konstant sind. Eine Vermutung für die zeitliche Abhängigkeit der Variable $\log(\text{protime})$ findet sich in Therneau et al. (1990). Es könnte sein, dass die Variable durch die medikamentöse Behandlung beeinflusst wird, ob-
schon die Behandlung selbst keine Wirkung auf die Überlebenszeit hat. Der zeitliche Effekt von protime ist übrigens vor der Logarithmustransformation wesentlich stärker ausgeprägt als danach. Zu der Residuenanalyse sei noch angemerkt, dass sie sich deutlich schwieriger gestaltet, als bei Regressionsmodellen, da es eine Vielzahl von Residuen gibt. Neben den Schoenfeldresiduen gibt es unter anderem noch Scorerresiduen bzw. Martingalresiduen. Weiter ist zu unterscheiden, ob man anhand der Residuen den funktionalen Einfluß der Kovariablen bzw. die Annahme des zeitlich konstanten Quotienten zweier Hazardraten überprüfen möchte. Eine Übersicht findet sich in Klein und Moeschberger (2003) oder in Grambsch und Therneau (2000). Ein Nachteil der Martingalresiduen ist, dass sie schief verteilt sind; sie werden lediglich zur Überprüfung des funktionalen Zusammenhangs verwendet (siehe Grafik (1.6)). Die logarithmische Transformation scheint gerechtfertigt. Der geschätzte Kovariableneffekt ist nach der Transfor-

mation nahezu linear. Die Scoreresiduen folgen nach einer geeigneten Standardisierung asymptotisch einer Brownschen Brücke. Sie unterscheidet sich dadurch vom Brownschen Prozeß, dass zum (zeitlichen) Endpunkt die Pfade den Wert Null annehmen. Es ist zu beachten, dass der Score- bzw. Schoenfeldresiduenprozeß mehrdimensional ist. In der oben genannten Literatur werden verschiedene Vorschläge zur (multivariaten) Standardisierung gemacht, die allerdings zu unterschiedlichen Ergebnissen führen. Daher wird in Grambsch und Therneau (2000) die Residuenanalyse des standardisierten Scoreprozesses als eher schwierig zu interpretieren eingestuft, eine Einschätzung die der Autor für – im Sinne einer Asymptotik verhältnismäßig – kleine Datensätze teilt. Für kritische Anmerkungen zu den Schoenfeldresiduen siehe etwa Winnett und Sasieni (2001). Eine neuere Veröffentlichung zu den Scoreresiduen findet man in Kvaløy und Reiersølmoen Neef (2004).

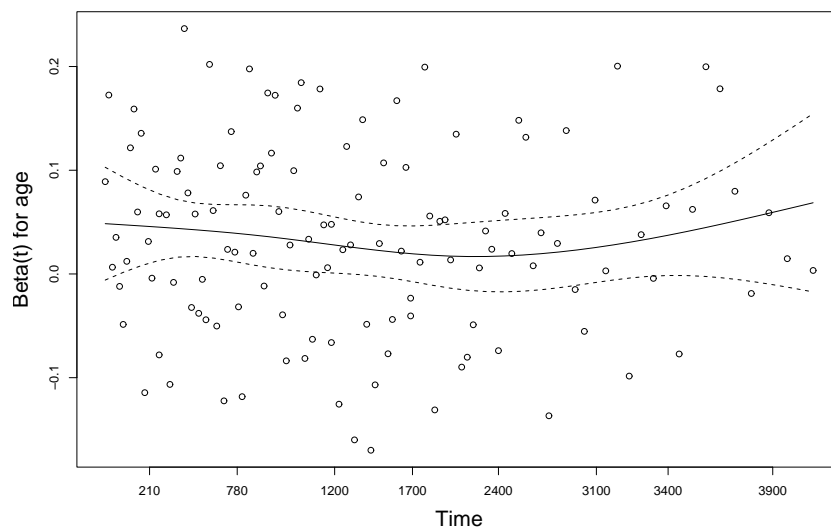


Abbildung 1.1: glatter Effekt der Variable age mit partiellen Schoenfeldresiduen

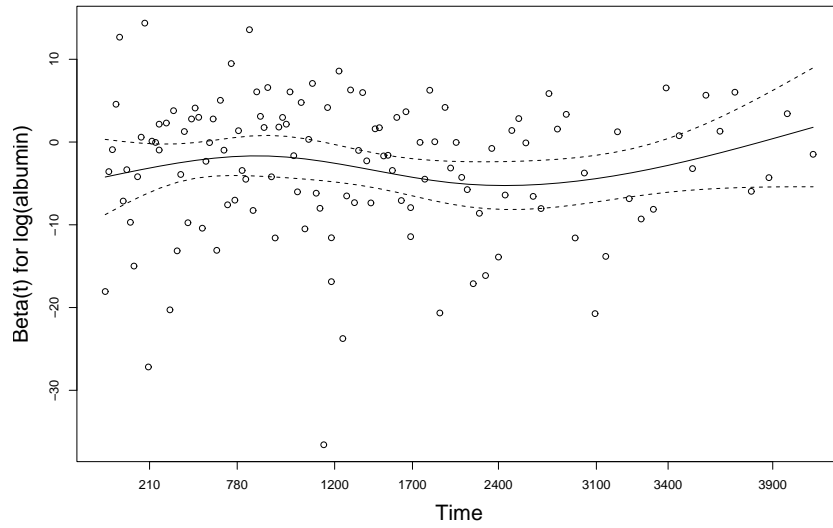


Abbildung 1.2: glatter Effekt der Variable log(albumin) mit partiellen Schoenfeldresiduen

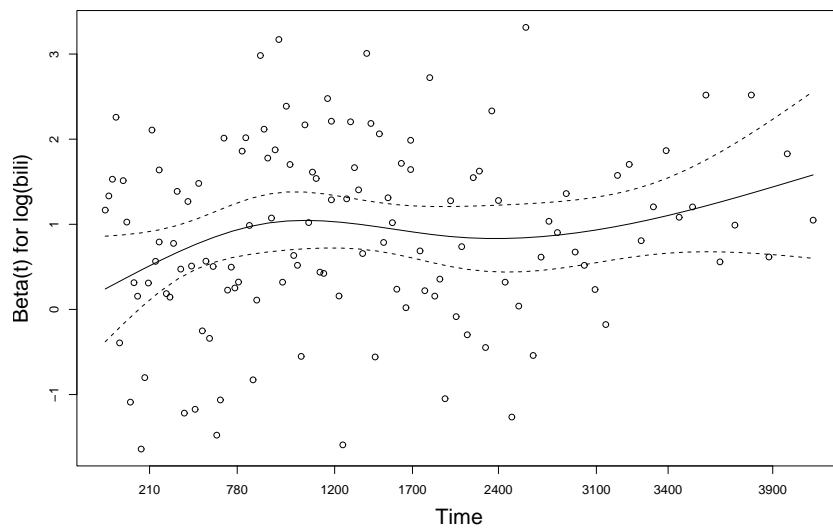


Abbildung 1.3: glatter Effekt der Variable log(bili) mit partiellen Schoenfeldresiduen

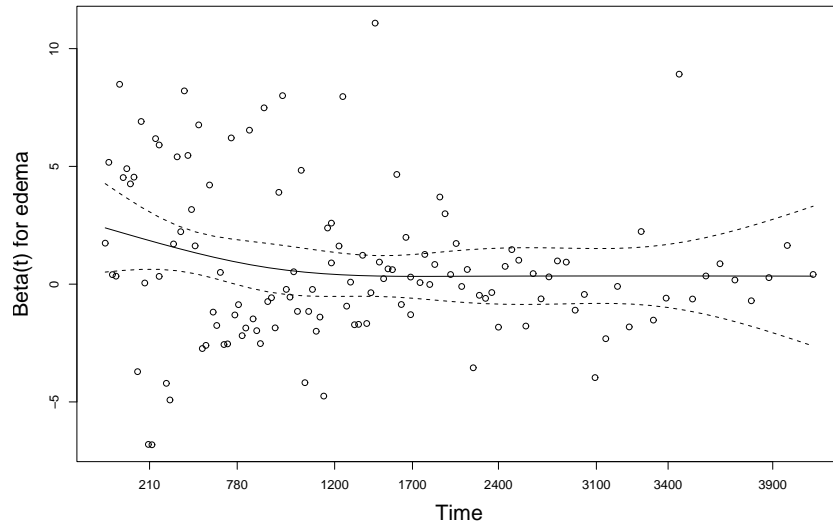


Abbildung 1.4: glatter Effekt der Variable edema mit partiellen Schoenfeldresiduen

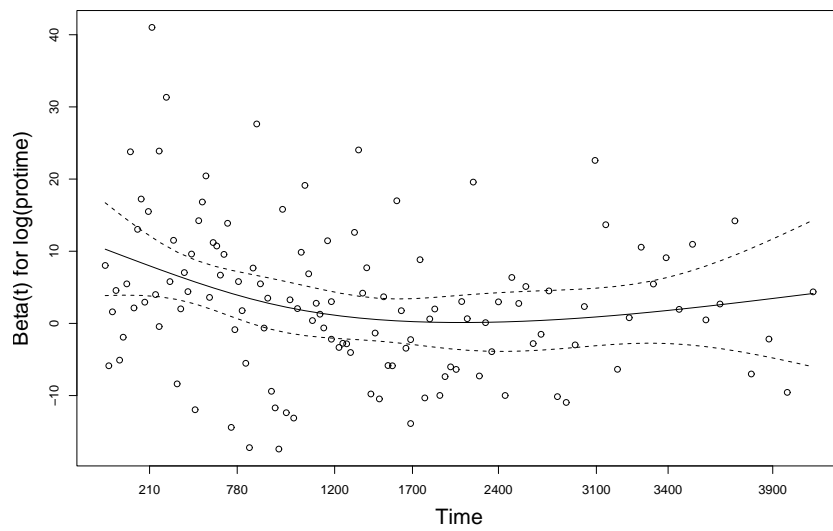


Abbildung 1.5: glatter Effekt der Variable log(protime) mit partiellen Schoenfeldresiduen

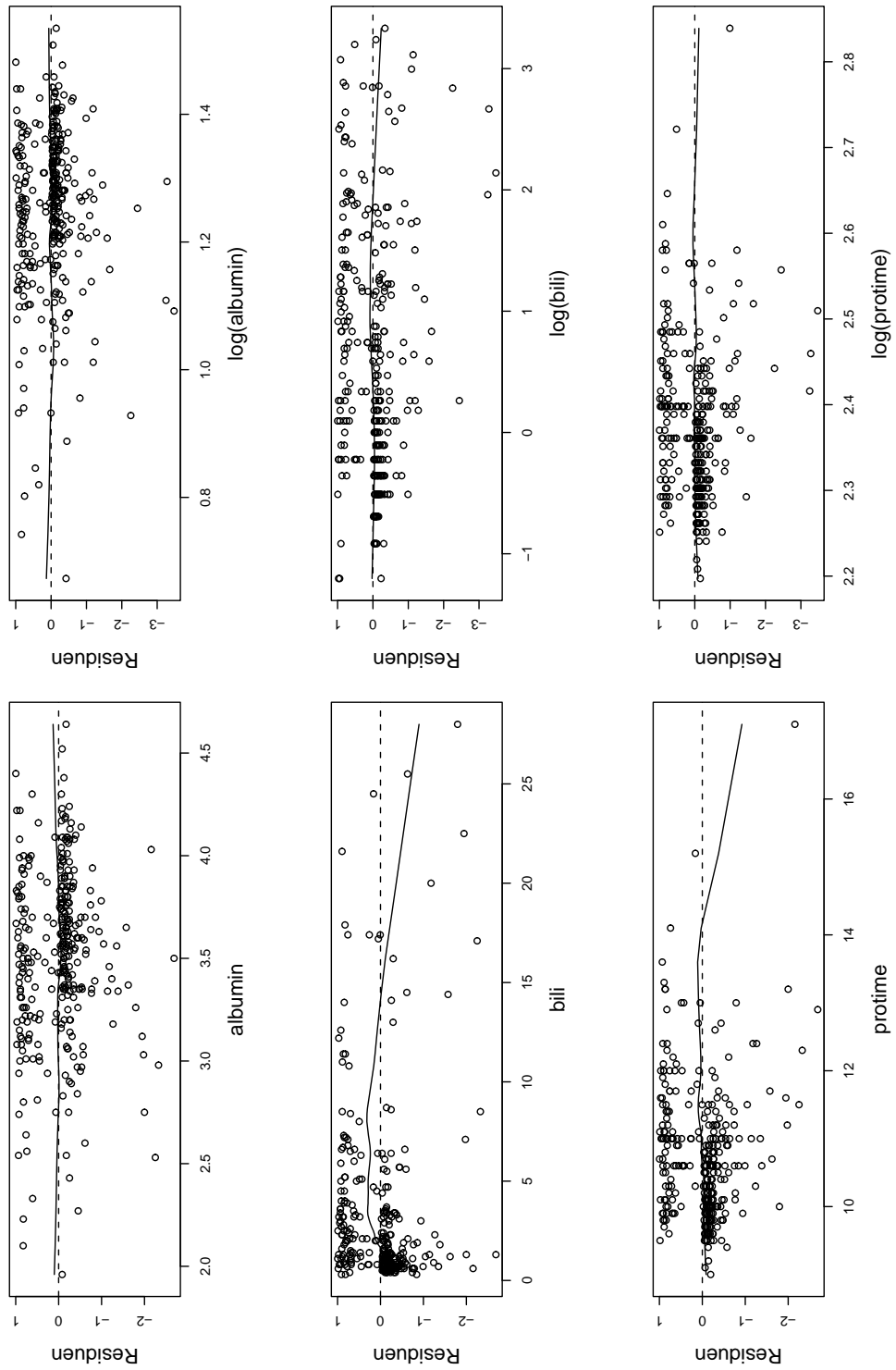


Abbildung 1.6: Martingalresiduen mit geglättetem (lowess) funktionalen Verlauf

2 Theoretische Grundlagen

2.1 Stochastische Prozesse

Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ und eine Familie von Zufallsgrößen $X = (X_t, t \in I)$ mit Werten in \mathbb{R}^m , $m \in \mathbb{N}$, so dass

$$X_t^{-1}(A) \in \mathcal{F}, \text{ für alle } A \in \mathcal{B}(\mathbb{R}^m), t \in I \quad (2.1)$$

gilt, wobei mit $\mathcal{B}(\mathbb{R}^m)$ die Borelsche σ -Algebra auf \mathbb{R}^m bezeichnet wird. Zur Vereinfachung der Notation wird im folgenden angenommen, dass X Werte in $(\mathbb{R}, \mathcal{B})$ annimmt. Das Kriterium (2.1) besagt, dass die Zufallsgrößen für festes t in $\omega \in \Omega$ meßbar sind. Wählt man I als Teilmenge von \mathbb{R}_+ , so läßt sich t als Zeitindex interpretieren. Bei festem ω erhält man durch

$$X : t \rightarrow X_t, t \in I \quad (2.2)$$

einen Pfad oder eine Trajektorie, die den zeitlichen Verlauf des stochastischen Prozesses modelliert. Allgemeiner lassen sich auch zeitlich räumliche Strukturen modellieren, z.B. durch die Wahl von $I = \mathbb{R}_+ \times \mathbb{R}^3$. Im Fall $I \subset \mathbb{R}_+$ läßt sich die Verteilung \mathbb{P}_X des stochastischen Prozesses $X = (X_t, t \in I)$ auf dem Bildraum $(\mathbb{R}^I, \mathcal{B}(\mathbb{R}^I))$ definieren durch

$$\mathbb{P}_X := \mathbb{P}(\omega : X(\omega) \in B), \quad B \in \mathcal{B}(\mathbb{R}^I) \quad (2.3)$$

Mit den Zeitpunkten $t_1 < t_2 < \dots < t_n, t_i \in I$, erhält man die endlich dimensionale Verteilung des Prozesses X durch

$$\mathbb{P}_{t_1, \dots, t_n}(B) = \mathbb{P}(\omega : (X_{t_1}, \dots, X_{t_n}) \in B), \quad B \in \mathcal{B}(\mathbb{R}^I) \quad (2.4)$$

bzw. mit $B = (-\infty, x_1] \times \dots \times (-\infty, x_n]$ erhält man die endlich dimensionale Verteilungsfunktion

$$F_{t_1, \dots, t_n}(x_1, \dots, x_n) = \mathbb{P}(\omega : (X_{t_1}, \dots, X_{t_n}) \in B) \quad , B \in \mathcal{B}(\mathbb{R}^I). \quad (2.5)$$

Sei $B = J_{t_1} \times \dots \times J_{t_n}$, wobei $J_i, i \in \mathbb{N}$ Intervalle in \mathbb{R} darstellen, dann umfaßt B alle Funktionen, für die zu den Zeitpunkten t_1, \dots, t_n gilt, dass $X_{t_i} \in J_{t_i}, i \in I$. Stochastische

Prozesse mit unabhängigen und stationären Zuwächsen bilden relativ einfache Beispiele.

Beispiel

- Brownscher Prozeß (ohne Drift)

$$\begin{aligned} I &= [0, \infty), \quad t_0, \dots, t_n \in I, \\ 0 &\leq t_0 \leq \dots \leq t_n, \quad \sigma^2 > 0, \quad X_0 = 0 \\ X_{t_k} - X_{t_{k-1}} &\sim N(0; (t_k - t_{k-1}) \cdot \sigma^2) \end{aligned}$$

- Poisson-Prozeß (homogen)

$$\begin{aligned} I &= [0, \infty), \quad t_0, \dots, t_n \in I, \\ 0 &\leq t_0 \leq \dots \leq t_n, \quad \lambda > 0, \quad X_0 = 0 \\ X_{t_k} - X_{t_{k-1}} &\sim Po((t_k - t_{k-1}) \cdot \lambda) \quad . \end{aligned}$$

Bemerkenswert ist, dass sich alle Prozesse mit unabhängigen und stationären Zuwächsen aus diesen beiden Prozessen zusammensetzen (Shiryayev und Prokhorov, 2000). Der Poisson-Prozeß ist der einfachste Punktprozeß. Allgemeiner läßt sich ein Punktprozeß auch als eine Folge von Zufallsgrößen $(\tau_n), n \in \mathbb{N}$, mit Werten in \mathbb{R}_+ auffassen, so dass

- (1) $\tau_1 > 0$, fast sicher
- (2) $\tau_{n+1} > \tau_n$, $\{\tau_n < \infty\}$ fast sicher
- (3) $\tau_{n+1} = \tau_n$, $\{\tau_n = \infty\}$ fast sicher

gilt. Die Bedingung (2) besagt, dass mehrere Ereignisse mit positiver Wahrscheinlichkeit im Endlichen nicht gleichzeitig auftreten. Innerhalb dieser Verallgemeinerung (gegenüber dem Poissonprozeß) lassen sich bereits Erneuerungsprozesse betrachten. Da alle Ereignisse vom gleichen Typ sind, spricht man auch von univariaten Punktprozessen. Demnach lassen sich Punktprozesse mit verschiedenen Ereignistypen durch eine Doppelfolge $(\tau_n, \xi_n), n \in \mathbb{N}$ charakterisieren, wobei wie bisher τ_n die Ereigniszeitpunkte und ξ_n den Ereignistyp modellieren. Nimmt ξ_n endlich viele Werte an, so spricht man von multivariaten Punktprozessen. Falls ξ_n überabzählbar viele Werte annehmen kann, spricht man von markierten Punktprozessen. Multivariate Punktprozesse werden z.B. in der Bedienungs- und Warteschlangentheorie zur Modellierung der eintreffenden und abgehenden Kundenströme verwendet. Darüber hinaus läßt sich hiermit auch der Verlauf von Krankheitsstadien modellieren, falls man als Ereignistyp den Übergang zwischen zwei Stadien bezeichnet. Markierte Punktprozesse spielen z.B. in der Versicherungsmathematik eine Rolle, da bei eintretenden Schadensfällen typischerweise der Zeitpunkt

und die Schadenshöhe zufällig sind. Eine gewisse Rolle spielen sie auch in der Extremwerttheorie (vgl. Coles (2001), Stichwort: peaks over thresholds). Man denke auch an ein Beispiel aus der stochastischen Geometrie: Sei $I = [0, 1] \times [0, 1]$ eine Fläche, etwa eine Käsescheibe (z.B. Emmentaler), dann ist die Anzahl und Größe der Löcher im Käse erfahrungsgemäß zufällig. Unter der Annahme, dass es sich bei den Löchern um disjunkte Kreise handelt, entspricht der Ereignistyp dem Radius. Gelegentlich bezeichnet man den Bildraum von ξ_n auch als Markenraum und die Realisierungen auch als Marken $\xi_n(\omega)$. In enger Verwandtschaft zu den Punktprozessen stehen die Zählprozesse $N_t(\omega) \equiv N_t$, mit

$$N_t = \sum_{n \geq 1} \mathbb{1}_{\{\tau_n \leq t\}}, \quad t \geq 0, \quad t \in I \subset \mathbb{R}_+ \quad (2.6)$$

bzw. im multivariaten Fall durch

$$N_t = \sum_{j \geq 1} N_{j,t}, \quad 1 \leq j \leq k \quad (2.7)$$

mit

$$N_{j,t} = \sum_{n \geq 1} \mathbb{1}_{\{\tau_n \leq t\}} \cdot \mathbb{1}_{\{\xi_n = j\}}, \quad t \geq 0, \quad t \in I \subset \mathbb{R}_+. \quad (2.8)$$

Der Prozeß $N_{j,t}$ zählt alle Ereignisse vom Typ j bis t , wogegen N_t alle Ereignisse bis t zählt. Analog definiert man den Zählprozeß für markierte Punktprozesse

$$N_t(B) = \sum_{n \geq 1} \mathbb{1}_{\{\tau_n \leq t\}} \cdot \mathbb{1}_{\{\xi_n \in B\}}, \quad t \geq 0, \quad t \in I \subset \mathbb{R}_+, \quad B \in \mathcal{B}(\mathbb{R}). \quad (2.9)$$

Die Pfade der Zählprozesse als Funktionen in t sind isoton, rechtsstetig mit linksseitigen Limiten. Im univariaten Fall gilt offensichtlich

$$N_{\tau_n} = n, \quad \mathbb{P} - \text{fast sicher.} \quad (2.10)$$

Gilt $N_t = \infty$ für endliches $t \in I$, so nennt man den Prozeß explodierend. Die Pfadeneigenschaften sind wichtig, sie strukturieren den Bildraum, der sich als ein Funktionenraum auffassen läßt. Das Verhalten der Funktionen ist im allgemeinen von überabzählbar vielen Punkten abhängig. Es ist also fraglich, ob z.B. die Menge $\{X, \sup X_t < \infty\}$ meßbar ist und im allgemeinen ist dies nicht der Fall. Jede Menge $A \in \mathcal{B}^I$ besitzt nämlich die Gestalt $A = \{X : (X_{t_1}, \dots, X_{t_n}, \dots) \in B\}$, $B \in \mathcal{B}^I$ mit abzählbar vielen Punkten t_1, \dots in I , d.h. die Meßbarkeit hängt nur von abzählbar vielen Koordinaten ab. Der Raum der Funktionen von I nach \mathbb{R} mit linksseitigen Limiten und rechtsseitig stetigen Pfaden wird im folgenden mit $D(I)$ bezeichnet. Seine Bedeutung erlangt er dadurch, dass

er einerseits hinreichend allgemein ist, so enthält er die Pfade der Zählprozesse, andererseits werden pathologische Pfade mit beliebiger Oszillation auf beschränkten Intervallen ausgeschlossen. Als pathologisches Beispiel nehme man etwa die Funktion $\sin(\frac{1}{x})$, $x \in I = [0, T)$, $T \in \mathbb{R}_+$, die nicht in $D(I)$ enthalten ist, da der rechtsseitige Grenzwert für $x = 0$ nicht existiert, wohingegen die Funktion $x \cdot \sin(\frac{1}{x})$, $x \in I = [0, T)$, $T \in \mathbb{R}_+$ (mit unendlicher Variation, vgl. Dothan (1990, S. 177)) in $D(I)$ enthalten ist. Da die endlich dimensionalen Verteilungen einen stochastischen Prozeß nicht eindeutig charakterisieren, beschränkt man sich jeweils auf Versionen der stochastischen Prozesse, deren Pfade in $D(I)$ liegen.

2.2 Elemente der diskreten stochastischen Analysis

Im folgenden werden einige Elemente der stochastischen Analysis zusammengetragen, unter anderem der Begriff der Filtration, des Martingals, der Vorhersagbarkeit, des stochastischen Integrals, der vorhersagbaren quadratischen Variation sowie der Doob–Meyer–Zerlegung. Dies geschieht für zeitdiskrete Strukturen, in diesem Fall lassen sich die entsprechenden Aussagen noch relativ einfach formulieren und die wesentlichen Strukturen erkennen. Der spätere zeitstetige Fall benötigt hingegen weitere Regularitätsbedingungen, die zum Teil sehr technischer Natur sind und auf deren Herleitung nur knapp eingegangen wird. Als Referenz und Motivation sollen daher zumindest ansatzweise die im diskreten Fall formulierten Aussagen dienen. In vielen Fällen läßt sich die Indexmenge von stochastischen Prozessen als Zeit interpretieren. Die zu jedem Zeitpunkt n bekannte Information, d.h. die Gegenwart und die Vergangenheit, lassen sich durch Folgen von σ -Algebren \mathcal{F}_n formalisieren, die im Zeitverlauf immer reichhaltiger werden. Passend dazu betrachtet man Folgen von Zufallsgrößen, die bezüglich dieser σ -Algebren meßbar sind. Eine wichtige Klasse bilden die Martingale. Diese Klasse ist allgemein genug und es existiert eine breit ausgebaute Theorie z.B. für Grenzwertsätze. Wichtig sind nur die Integrierbarkeitsvoraussetzungen. Die Verteilungsannahmen oder die Unabhängigkeit der Zufallsgrößen sind nur für die jeweiligen Modelle von Interesse. Daneben gibt es noch vorhersagbare Prozesse, die bereits bezüglich der Vergangenheit, also \mathcal{F}_{n-1} -meßbar sind. Für diese Prozesse wird das stochastische Integral eingeführt. Es folgen einige Aussagen, die für die Integrierbarkeit von Interesse sind. Von besonderer Bedeutung ist die Doob–Meyer–Zerlegung, die im zeitstetigen Fall auf Zählprozesse angewendet wird.

2.2.1 Diskrete Martingale

Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$. Eine Filtration $\mathcal{F}_n, n \in \mathbb{N}$ oder aufsteigende Folge von σ -Algebren werde durch die Eigenschaften

$$\mathcal{F}_n \subset \mathcal{F}, n \in \mathbb{N}, \mathcal{F}_0 = \{\Omega, \emptyset\} \quad (2.11)$$

$$\mathcal{F}_n \subset \mathcal{F}_p, n \leq p \in \mathbb{N}. \quad (2.12)$$

charakterisiert. Eine Folge X_0, X_1, \dots von Zufallsgrößen auf $(\Omega, \mathcal{F}, \mathbb{P})$ mit \mathcal{F}_n -meßbarem X_n und $\mathbb{E}(|X_n|) < \infty$ heißt Martingal, falls für alle $n \geq 0$

$$\mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n \text{ fast sicher} \quad (2.13)$$

bzw. Submartingal, falls

$$\mathbb{E}(X_{n+1}|\mathcal{F}_n) \geq X_n \text{ fast sicher.} \quad (2.14)$$

Eine stochastische Folge X_n , die bzgl. \mathcal{F}_n -meßbar ist, heißt Supermartingal, falls $-X_n$ bzgl. \mathcal{F}_n ein Submartingal ist. Sei die Folge X_n ein Martingal, dann definiert man die Zuwächse $Y_n = \Delta X_n = X_n - X_{n-1}, n \geq 1, (Y_0 = X_0)$. Es gilt $\mathbb{E}(|Y_n|) < \infty$ sowie $\mathbb{E}(Y_{n+1}|\mathcal{F}_n) = 0, \mathcal{F}_n = \sigma(Y_0, \dots, Y_n)$. Sei $y^n = (y_0, \dots, y_n)$ und f eine \mathcal{F}_n -meßbare Funktion, so dass $y_{n+1} \cdot f(y^n)$ integrierbar ist, dann gilt

$$\begin{aligned} \mathbb{E}[Y_{n+1} \cdot f(y^n)] &= \mathbb{E}[\mathbb{E}(Y_{n+1} \cdot f(y^n)|\mathcal{F}_n)] \\ &= \mathbb{E}[f(y^n) \cdot \mathbb{E}(Y_{n+1}|\mathcal{F}_n)] \\ &= 0, \end{aligned} \quad (2.15)$$

das heißt, die Zuwächse sind orthogonal für eine beliebige \mathcal{F}_n -meßbare Funktion f . Im Vergleich zur Unabhängigkeit ist dies eine Abschwächung, da

$$\mathbb{E}(g(Y_{n+1}) \cdot f(y^n)) = \mathbb{E}(f(y^n)) \cdot \mathbb{E}(g(Y_{n+1})) = 0,$$

gilt, mit g meßbar, $\mathbb{E}(g(Y_{n+1})) = 0$ und $g \cdot f$ integrierbar (Doob, 1953). Die Orthogonalität der Zuwächse des Martingals wird ausgenutzt bei der stochastischen Integration und zentralen Grenzwertsätzen (für Martingale).

Als Beispiel folgt der Likelihoodquotient, der gewissermaßen die Vorstufe des Likelihoodprozesses ist. Dieses Beispiel motiviert auch den Zusammenhang zwischen Like-

lihoodquotientenprozessen und Martingalen.

Beispiel

Gegeben sei der Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ mit einer Filtration $(\mathcal{F}_n), n \in \mathbb{N}$ und einem weiteren Wahrscheinlichkeitsmaß Q . Ohne vorauszusetzen, dass Q bezüglich \mathbb{P} absolut stetig ist, definiert man die Restriktionen $Q_n = Q|\mathcal{F}_n$ und $\mathbb{P}_n = \mathbb{P}|\mathcal{F}_n$ bezüglich der aufsteigenden Folge von σ -Algebren, so dass Q_n absolut stetig bzgl. \mathbb{P}_n ist. Sei

$$X_n = \frac{dQ_n}{d\mathbb{P}_n},$$

dann gilt wegen $\mathbb{E}_{\mathbb{P}}(X_{n+1}|\mathcal{F}_n) = X_n$ (Kopp, 1984, S. 41), dass (X_n) ein Martingal ist .

Neben der Charakterisierung durch die Orthogonalität der Zuwächse läßt sich das Martingal auch durch ein faires Spiel motivieren. Sei Y_0, Y_1, \dots eine Folge von unabhängig identisch verteilten Zufallsgrößen mit $\mathbb{E}(Y_0) = \mu < \infty$ und $\mathcal{F}_n = \sigma_n(Y_0, \dots, Y_n)$ die kleinste σ -Algebra bezüglich derer Y_0, \dots, Y_n meßbar sind, dann gilt mit $X_n = \sum_{0 \leq i \leq n} Y_i$

$$\mathbb{E}_{\mathbb{P}}(X_{n+1}|\mathcal{F}_n) = \mathbb{E}_{\mathbb{P}}(X_n + Y_{n+1}|\mathcal{F}_n) = X_n + \mu \quad , \tag{2.16}$$

letzteres wegen der \mathcal{F}_n -Meßbarkeit von X_n und der Unabhängigkeit der Y_n . Für $\mu = 0$ ist X_n ein Martingal, d.h. der zu erwartende Gewinn ist Null (bei bekannter Vergangenheit).

2.2.2 Stochastische Integration

Als Erweiterung zu den vorherigen Ausführungen kann man noch zulassen, dass der Einsatz in jedem Spiel variabel ist. Zu jedem Zeitpunkt n wird der Spieleinsatz V_n für das n -te Spiel von den ersten $n - 1$ Zeitpunkten abhängen. Allgemein bezeichnet man eine Folge $(V_n)_{n \in \mathbb{N}}$ von \mathcal{F}_n -meßbaren Zufallsgrößen, die sogar \mathcal{F}_{n-1} -meßbar sind, als vorhersagbar. Weiter definiert man den Prozeß $Z := V * X$ mit

$$Z_n = V_0 X_0 + V_1(X_1 - X_0) + \dots + V_n(X_n - X_{n-1}), \tag{2.17}$$

bzw.

$$Z_n = V_0 X_0 + \oint_1^n V dX \tag{2.18}$$

als stochastisches Integral. Aufgrund der Vorhersagbarkeit von V_n gilt, dass Z_n bzgl. \mathcal{F}_n ein Martingal darstellt, falls Z_n integrierbar ist, d.h. das Spiel bleibt fair. Eine interessante Spielstrategie, die die Integrierbarkeitsvoraussetzung verletzt, besteht im Verdoppeln des Einsatzes nach jedem verlorenen Spiel (Sirjaev, 1988, S. 469). Man beendet das Spiel nach dem ersten Sieg. Wählt man $V_i = 2^{i-1}$, falls man die ersten $i - 1$ Spiele verloren hat, so berechnet sich der Gesamtverlust nach n Spielen zu

$$\sum_{1 \leq i \leq n} 2^{i-1} = 2^n - 1 \quad V_1 = 1, X_0 = 0.$$

Im Falle eines Sieges im n -ten Spiel gilt

$$Z_n = Z_{n-1} + V_n = -(2^n - 1) + 2^n = 1.$$

Die Martingaleigenschaft geht also verloren. Dies liegt daran, dass V_n unbeschränkt wächst und damit die Integrationsvoraussetzung verletzt. Die Spielstrategie, nämlich das Spiel zu einem zufälligen, nur von der Vergangenheit abhängenden Zeitpunkt abbrechen, motiviert folgende Definition einer Stoppzeit.

Definition Sei τ eine Zufallsgröße auf $(\Omega, \mathcal{F}, \mathbb{P})$, mit $\tau : \Omega \rightarrow \mathbb{N} \cup \{\infty\}$ und $(\mathcal{F}_n)_{n \in \mathbb{N}}$ eine aufsteigende Filtration mit $\mathcal{F}_n \subset \mathcal{F}$. Gilt

$$\{\tau \leq n\} \in \mathcal{F}_n, \quad \text{für alle } n \in \mathbb{N}, \quad (2.19)$$

so nennt man τ eine Stoppzeit.

Die Größe τ hängt also nicht von der Zukunft ab. Mittels Stoppzeiten lassen sich Mengensysteme definieren, die den Informationsstand zu einem zufälligen, durch die Stoppzeit τ bestimmten Zeitpunkt, widerspiegeln,

$$\mathcal{F}_\tau = \{A \in \mathcal{F} : A \cap \{\tau = n\} \in \mathcal{F}_n, n \in \mathbb{N}\} \quad (2.20)$$

Es läßt sich zeigen, dass es sich hierbei um eine σ -Algebra handelt. Für \mathcal{F}_n -meßbare Zufallsgrößen X_n auf $(\Omega, \mathcal{F}, \mathbb{P})$ mit Werten in \mathbb{R} definiert man

$$X_\tau = \sum_{n \geq 0} X_n \mathbb{1}_{\{\tau \geq n\}} \quad . \quad (2.21)$$

Die Zufallsgröße X_τ ist \mathcal{F} -meßbar, sie ist sogar \mathcal{F}_τ -meßbar, denn für $B \in \mathcal{B}(\mathbb{R})$ gilt mit $A = \{X_\tau \in B\}$ die Zerlegung $A = \bigcup_{n \geq 0} \{X_n \in B\} \cap \{\tau \geq n\}$. Im obigen Beispiel wurde

schon darauf hingewiesen, dass bei zufälligen Stoppzeiten die Martingaleigenschaft verloren gehen kann. Hinreichend ist z.B. die gleichgradige Integrierbarkeit (Sirjaev, 1988). Eine weitere Methode zur Sicherstellung der Integrierbarkeit ist das Verwenden von lokalisierenden Stoppzeiten. Dazu betrachtet man eine Folge von Stoppzeiten $(\tau_k)_{k \in \mathbb{N}}$ mit $\tau_k \leq \tau_{k+1}$ fast sicher und $\tau_k \rightarrow \infty$ für $k \rightarrow \infty$. Bildet man dann die gestoppten Folgen $X_{\tau_k \wedge n}$ mit

$$X_{\tau_k \wedge n} = \sum_{0 \leq m \leq n-1} X_m \mathbb{1}_{\{\tau_k = m\}} + X_n \mathbb{1}_{\{\tau_k \geq n\}} \quad , \quad (2.22)$$

so heißt X_n lokales Martingal (Submartingal), falls eine Folge von Stoppzeiten $(\tau_k)_{k \in \mathbb{N}}$ existiert, so dass $X_{\tau_k \wedge n}$ ein Martingal (Submartingal) bildet.

2.2.3 Doob–Meyer–Zerlegung

Für Submartingale existiert eine bemerkenswerte, fast sicher eindeutige Zerlegung, die für Zählprozesse von Bedeutung ist, dort allerdings in einer zeitstetigen Version. Sei (X_n, \mathcal{F}_n) ein Submartingal, dann existiert ein Martingal (M_n, \mathcal{F}_n) und eine vorhersagbare isotone Folge A_n , d.h. A_n ist \mathcal{F}_{n-1} -meßbar mit $A_{n+1} \geq A_n$, so dass

$$X_n = M_n + A_n \quad . \quad (2.23)$$

Die Folge A_n wird auch als Kompensator bezeichnet. Da M_n im Gegensatz zu A_n nicht vorhersagbar ist, bezeichnet man es auch als Innovationsmartingal, sozusagen als Komplement des vorhersagbaren Anteils des Submartingals. Definiert man als duale vorhersagbare Projektion (Dothan, 1990, S. 123) $(\pi^* X)$ eines integrierbaren Prozesses X mit $\Delta X_n = X_n - X_{n-1}$, $n \geq 1$

$$(\pi^* X)_n = \begin{cases} X_0 & , \quad n = 0 \\ X_0 + \sum_{s=1}^n \mathbb{E}(\Delta X_s | \mathcal{F}_{s-1}) & , \quad \text{sonst} \end{cases} \quad (2.24)$$

so gilt

$$A_n = (\pi^* X)_n$$

und

$$M_n = (\pi^* X)_n - X_n \quad . \quad (2.25)$$

Die Isotonie von A_n läßt sich etwa wie folgt zeigen

$$\begin{aligned} A_n &= (\pi^* X)_n \\ &= (\pi^* X)_{n-1} + \mathbb{E}(\Delta X_n | \mathcal{F}_{n-1}) \\ &= A_{n-1} + \mathbb{E}(\Delta X_n | \mathcal{F}_{n-1}) \\ &\geq A_{n-1} \quad . \end{aligned}$$

Für den späteren zeitstetigen Fall wird noch die vorhersagbare quadratische Variation (alternativer Begriff: Projektion) eingeführt. Sei X ein quadratintegrierbarer Prozeß, mit $\Delta X_n = X_n - X_{n-1}$, $n \geq 1$, dann bezeichnet man $\langle X \rangle$ als quadratische vorhersagbare Projektion (Dothan, 1990, S. 95)

$$\langle X \rangle_n = \begin{cases} \mathbb{E}(X_0^2) & , \quad n = 0 \\ \langle X \rangle_0 + \sum_{s=1}^n \mathbb{E}(\Delta X_s)^2 | \mathcal{F}_{s-1}) & , \quad \text{sonst} \end{cases} \quad (2.26)$$

Beide Prozesse beschreiben die bedingten ersten und zweiten Momente eines Prozesses und werden sich bei den Zählprozessen als sehr charakteristische Kenngrößen erweisen mit Nutzen für die statistische Inferenz.

2.3 Elemente der stetigen stochastischen Prozesse

Bisher wurde $X = (X_t, t \in I)$ als Familie von Zufallsgrößen auf $(\Omega, \mathcal{F}, \mathbb{P})$ aufgefasst. Alternativ kann man X auch als eine Abbildung auf dem Produktraum $I \times \Omega$ mit Werten in \mathbb{R} betrachten und fordern, dass X meßbar bezüglich $\mathcal{B}(I) \times \mathcal{F}$ ist. Die Nützlichkeit einer aufsteigenden Filtration $(\mathcal{F}_t, t \in I)$ mit $\mathcal{F}_t \subset \mathcal{F}$ wurde bereits für diskrete Martingale dargelegt (siehe Sirjaev (1988)). Dementsprechend heißt ein Prozeß adaptiert, falls die Zufallsgrößen X_t \mathcal{F}_t -meßbar sind. Ergänzend wird zusätzlich noch gefordert, dass

- \mathcal{F}_0 alle Nullmengen bzgl. \mathbb{P} enthält,
- \mathcal{F}_t rechtsstetig ist, d.h. $\mathcal{F}_t = \mathcal{F}_{t^+} := \bigcap_{t < s} \mathcal{F}_s$, $t, s \in I$.

Durch die erste Forderung entledigt man sich eventueller Meßbarkeitsprobleme bei Modifikationen von stochastischen Prozessen. Die zweite Forderung ist ebenso technischer Natur. Sie wird dadurch motiviert, dass man Bedingungen sucht, die die Existenz einer Modifikation mit rechts- bzw. linksstetigen Pfaden garantieren. Ist dies der Fall und $I \subset \mathbb{R}$, kann man sich bei den Beweisen auf die abzählbare Teilmenge der rationalen

Zahlen in I zurückziehen. Die Rechtsstetigkeit der Filtration spielt dabei eine besondere Rolle. Zusätzlich benötigt man Integrierbarkeitsbedingungen, z.B. die gleichgradige Integrierbarkeit, die von den im Rahmen dieser Arbeit betrachteten Prozessen für endliches t erfüllt werden. Setzt man $\mathcal{F}_t^X = \sigma(X_s, s \leq t)$, so ist X_t \mathcal{F}_t^X -meßbar. Diese spezielle Filtration wird auch als eigene Historie bezeichnet, da sie als Informationsgehalt die Prozeßvergangenheit beinhaltet. Unter allgemeinen Regularitätsbedingungen (Ω muß reichhaltig genug sein) kann auch die eigene Historie so gewählt werden, dass sie die oberen Bedingungen erfüllt (Lipster und Shiryaev, 1978, S. 244). Weitere Meßbarkeitskonzepte werden benötigt.

Definition Ein Prozeß heißt progressiv meßbar bzgl. $(\mathcal{F}_t, t \in I)$, falls für jedes t gilt, dass

$$X : (s, \omega) \rightarrow X(s, \omega), 0 \leq s \leq t \quad , \quad (2.27)$$

meßbar bzgl. $\mathcal{B}([0, t]) \times \mathcal{F}_t$ ist.

Zwar ist jeder progressiv meßbare Prozeß adaptiert, die Umkehrung ist aber falsch (Kopp, 1984, S. 83). Hinreichend für die progressive Meßbarkeit eines adaptierten Prozesses ist, dass die Pfade rechts- oder linksstetig sind. Der Beweis beruht auf einer Approximation von $X(s, \omega)$ durch rechts- oder linksstetige Treppenfunktionen mit Sprungstellen in der Menge der rationalen Zahlen. Die Nützlichkeit der progressiven Meßbarkeit liegt darin, dass ein gestoppter Prozeß $X^\tau = (X_t^\tau, t \in I)$ wieder progressiv meßbar ist, falls X progressiv meßbar ist mit

$$X_t^\tau := \begin{cases} X_{\tau \wedge t}, & \tau < \infty \\ X_\tau \mathbb{1}_{\{\tau < \infty\}}, & \tau = \infty \end{cases} \quad ,$$

wobei

$$\tau : \Omega \rightarrow [0, \infty]$$

eine Stoppzeit bzgl. \mathcal{F}_t ist, d.h.

$$\{\tau \leq t\} \in \mathcal{F}_t, t \in I$$

und

$$\mathcal{F}_\tau = \{A \in \mathcal{F} : A \cap \{\tau \leq t\} \in \mathcal{F}_t, t \in I\}$$

die σ -Algebra darstellt, die alle Informationen bis zum zufälligen Zeitpunkt τ enthält. Des weiteren benötigt man die progressive Meßbarkeit bei Aussagen vom Typ Fubini, z.B. $\mathbb{E}(\int_0^t V_s dB_s(\omega))$, bei der es auf die Meßbarkeit in beiden Argumenten (ω, t) ankommt. Dabei ist der innere Ausdruck, das stochastische Integral (siehe unten Formel (2.32)) eine Zufallsgröße, deren Erwartungswert berechnet wird. Sei τ eine Stoppzeit und \mathcal{F}_τ die korrespondierende σ -Algebra. Dann definiert man

$$\mathcal{F}_{\tau-} := \sigma(\mathcal{F}_0, A \cap \{\tau > t\}, t \geq 0, A \in \mathcal{F}_t)$$

als σ -Algebra der strikten Vergangenheit der Stoppzeit τ . Offensichtlich gilt

$$\mathcal{F}_{\tau-} \subset \mathcal{F}_\tau$$

und aufgrund der Rechtsstetigkeit der Filtration ist τ auch $\mathcal{F}_{\tau-}$ -meßbar. Man kann im allgemeinen nicht erwarten, dass X_τ $\mathcal{F}_{\tau-}$ -meßbar ist, falls X progressiv meßbar ist. Ist die Stoppzeit endlich, gilt dies allerdings, falls X meßbar bzgl. der vorhersagbaren σ -Algebra $\mathcal{P}(\mathcal{F}_t)$ (Last und Brandt, 1995, S. 56)

$$\mathcal{P}(\mathcal{F}_t) := \sigma((s, t], s < t, A \in \mathcal{F}_s)$$

ist (man beachte, dass $A \in \mathcal{F}_s$, nicht $A \in \mathcal{F}_t$). Pragmatisch ausgedrückt ist ein Prozeß X vorhersagbar, falls er adaptiert und linksstetig ist. Die σ -Algebra $\mathcal{F}_{\tau-}$ enthält alle Informationen bis kurz vor dem Zeitpunkt τ . Angewendet auf multivariate Zählprozesse heißt dies, dass die σ -Algebra \mathcal{F}_t bereits die Information eines Ereigniseintritts beinhaltet, nicht aber den Ereignistyp (Karr, 1991, S. 56). Definiert man die vorhersagbare quadratische Variation $\langle X \rangle$ (Dothan, 1990, S. 169) für zeitstetige \mathcal{F}_t -adaptierte quadratintegrierbare stochastische Prozesse X mittels gleichmäßiger Limiten bzgl. t , d.h. bildet man den zeitstetigen Grenzübergang, so erhält man ein wichtiges Beispiel für vorhersagbare Prozesse. Die Nützlichkeit der Meßbarkeit bzgl. der σ -Algebra $\mathcal{F}_{\tau-}$ und $\mathcal{P}(\mathcal{F}_t)$ wird deutlich, wenn man die Doob–Meyer–Zerlegung eines Submartingals S_n in ein Martingal M_n und einen vorhersagbaren isotonen Prozeß A_n als diskretes Analogon behandelt. Dort wurde die \mathcal{F}_{n-1} -Meßbarkeit des Kompensators A_n gefordert. Betrachtet man \mathcal{F}_{n-1} als letzte Vergangenheit im diskreten Fall, so überträgt sich diese Interpretation im zeitstetigen Fall auf die σ -Algebra $\mathcal{F}_{\tau-}$. Die Meßbarkeit eines Kompensators bzgl. $\mathcal{F}_{\tau-}$ erfordert, dass der Kompensator vorhersagbar ist, d.h. er muß meßbar bzgl. $\mathcal{P}(\mathcal{F}_t)$ sein. Mit diesem feinsinnigen Konzept läßt sich nun auch eine Zerlegung eines zeitstetigen Submartingals in ein Martingal und einen Kompensator beweisen. Da Zählprozesse Submartingale darstellen, liefert die Doob–Meyer–Zerlegung, insbesondere der Kompensa-

tor, interessante Einblicke in die Struktur von Zähl- und Punktprozessen. Weiter ist die Zerlegung von Nutzen für die statistische Inferenz, etwa für Schätzfunktionen. Ein kurzes Kapitel über zeitstetige Martingale soll die Doob–Meyer–Zerlegung vorbereiten. Zusätzlich wird der stochastische zeitstetige Fall noch kurz erörtert, einerseits weil bei Zählprozessen der Kompensator als Integrator auftaucht, andererseits wird für Schätzfunktionen die stochastische Integration benötigt.

2.3.1 Stetige Martingale und stochastische Integration

Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum mit einer aufsteigenden und rechtsstetigen Filtration $(\mathcal{F}_t), t \in I, \mathcal{F}_t \subset \mathcal{F}$. Eine Familie $M_t, t \in I$ von Zufallsgrößen heißt Martingal (Submartingal), falls gilt

- M_t ist \mathcal{F}_t -meßbar, $t \in I$,
- $\mathbb{E}(|M_t|) < \infty, t \in I$,
- $\mathbb{E}(M_t | \mathcal{F}_s) = M_s$, fast sicher, $s < t$, (Martingal),
- $\mathbb{E}(M_t | \mathcal{F}_s) \geq M_s$, fast sicher, $s < t$ (Submartingal) .

Durch Vorzeichenwechsel erhält man aus einem Submartingal ein Supermartingal. Unter gewissen Regularitätsbedingungen läßt sich für Super- und Submartingale – für Martingale ohne weitere Bedingungen – zeigen, dass jeweils Modifikationen mit rechtsstetigen Pfaden existieren. Deshalb lassen sich viele Aussagen von diskreten Martingalen auf den zeitstetigen Fall übertragen. Als eine Anwendung erhält man für Martingale mit rechtsstetigen Pfaden, dass die linksseitigen Grenzwerte fast sicher existieren, d.h. die Pfade sind in $D(I)$ und damit hinreichend ‘gutartig’. Für die Doob–Meyer–Zerlegung wird noch die Definition eines τ_∞ -lokalen Martingals benötigt. Sei $M = (M_t, \mathcal{F}_t)$ ein adaptierter Prozeß und τ_∞ eine Stoppzeit bzgl. \mathcal{F}_t . M heißt τ_∞ -lokales Martingal, falls eine Folge $(\tau_n)_{n \in \mathbb{N}}$ von Stoppzeiten existiert mit (Lipster und Shiryaev, 1977)

$$\mathbb{P}(\tau_n < \tau_{n+1} < \tau_\infty) = 1, n \in \mathbb{N}, \text{ bzw.} \quad (2.28)$$

$$\mathbb{P}(\lim_{n \rightarrow \infty} \tau_n = \tau_\infty) = 1, n \in \mathbb{N} \quad (2.29)$$

und $M_{t \wedge \tau_n}, t < \tau_\infty$ ist ein gleichgradig integrierbares Martingal. Der Zweck der lokalisierenden Folge von Stoppzeiten besteht darin, dass die gleichgradige Integrierbarkeit und die Martingaleigenschaft des gestoppten Prozesses für $t < \tau_\infty$ hergestellt werden.

Die stochastische Integration wurde im zeitdiskreten Fall schon kurz angesprochen (sie-

he Formel (2.18)). Definiert man $Z_t := (V * X)_t = \int_0^t V_s dX_s$ so hat man zu beachten, dass dies nur sinnvoll ist, wenn

$$\int_0^t |V_s| d|X_s| < \infty \quad (2.30)$$

gilt (Karr, 1991, S. 91). Man erhält pfadweise Lebesgue–Stieltjes–Integrale. In diesem Sinn lassen sich alle folgenden Integrale betrachten, die einen Zählprozeß oder einen sonstigen isotonen Prozeß als Integrator haben. Die Voraussetzung der Endlichkeit ist notwendig, da die Eindeutigkeit des Integrals verloren geht, wenn X_s unendliche Variation hat, siehe etwa das Ito–Integral mit dem Wiener Prozeß B_s (Kopp, 1984, S. 11). Dort betrachtet man Prozesse V_s für die

$$\mathbb{E}\left(\int_0^t V_s^2 ds\right) < \infty \quad (2.31)$$

bzw.

$$\mathbb{E}\left(\int_0^t V_s^2 ds\right) = \mathbb{E}\left(\int_0^t V_s dB_s\right)^2 \quad (2.32)$$

gilt. Beachtet man, dass für die vorhersagbare quadratische Variation $\langle B \rangle_s = s$ gilt, so ist es naheliegend, stochastische Integrale $\int_0^t V_s dX_s$ (Andersen et al., 1993, S. 71) zuzulassen mit

$$\mathbb{E}\left(\int_0^t V_s^2 d\langle X \rangle_s\right) = \mathbb{E}\left(\int_0^t V_s dX_s\right) < \infty \quad (2.33)$$

Es gilt nun in Analogie zum zeitdiskreten Fall, dass $(V * X)_t$ ein Martingal ist, falls V_s vorhersagbar und X_s ein Martingal ist (Gill, 1984, S. 444).

2.3.2 Doob–Meyer–Zerlegung

Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum mit einer aufsteigenden rechtsstetigen Familie von σ –Algebren $\mathcal{F}_t \subset \mathcal{F}$. Ein rechtsstetiger Prozeß $A_t, t \geq 0$ heißt ansteigend, falls er adaptiert ist und $A_0 = 0$ sowie $A_s \leq A_t, s \leq t$ fast sicher gilt. Der Prozeß A heißt natürlicher ansteigender Prozeß, falls für jedes beschränkte positive Martingal $Y = (Y_t, \mathcal{F}_t)$ mit

rechtsstetigen Pfaden und linksseitigen Limiten

$$\mathbb{E} \left(\int_0^\infty y_s dA_s \right) = \mathbb{E}(y_\infty A_\infty) \quad (2.34)$$

gilt. Sei $N = (N_t, \mathcal{F}_t), t \geq 0$ ein Punktprozeß. Dann existieren ein τ_∞ -lokales Martingal M_t und ein natürlicher isotoner Prozeß A_t , so dass folgende bis auf stochastische Äquivalenz eindeutige Zerlegung für $t < \tau_\infty$

$$N_t = M_t + A_t$$

existiert. Der Prozeß A_t wird analog zum diskreten Fall als Kompensator bezeichnet. Diese technische Voraussetzung (2.34) garantiert die Vorhersagbarkeit des Kompensators. Die Doob–Meyer–Zerlegung soll anhand des Poisson–Prozesses verdeutlicht werden. Sei $X = (X_t, \mathcal{F}_t)$ ein Poisson–Prozeß mit Parameter λ . Demnach ist X_t Poisson–verteilt mit Erwartungswert λt . Wegen der Unabhängigkeit der Zuwächse gilt

$$\begin{aligned} \mathbb{E}(X_t - \lambda t | \mathcal{F}_s) &= \mathbb{E}(X_s - \lambda s | \mathcal{F}_s) + \mathbb{E}(X_t - X_s - \lambda(t - s) | \mathcal{F}_s) \\ &= X_s - \lambda s \quad . \end{aligned} \quad (2.35)$$

Daraus folgt, dass

$$A_t = \lambda t$$

gilt. In diesem Fall ist A_t eine deterministische Funktion. Allgemein gilt, dass Punktprozesse mit deterministischem Kompensator unabhängige Zuwächse haben (Lipster und Shiryaev, 1978, S. 279). Die Doob–Meyer–Zerlegung hängt in entscheidender Weise von der Filtration \mathcal{F}_s ab. Wählt man $\mathcal{F}_s \equiv \mathcal{F}$ für alle s , so gilt (Karr, 1991, S. 62)

$$A_t = N_t \quad .$$

Demgemäß ist es naheliegend, als Filtration die eigene Historie $\mathcal{F}_t^N = \sigma(N_s, s \leq t)$ zu verwenden.

2.3.3 Punktprozesse und Kompensatoren

Die Stoppzeiten lassen sich unterscheiden in vorhersagbare und nicht vorhersagbare Stoppzeiten. Eine Stoppzeit T heißt vorhersagbar bzgl. \mathcal{F}_t , falls der Einpunktprozeß

$$N_t = \mathbb{1}_{\{T \leq t\}}$$

vorhersagbar ist (Lipster und Shiryaev, 1978, S. 239). Im Gegensatz dazu heißt eine Stoppzeit T vollständig unerreichbar, falls für jede vorhersagbare Stoppzeit S gilt, dass

$$\mathbb{P}(S = T < \infty) = 0 \quad . \quad (2.36)$$

So sind z.B. alle endlichen Sprungzeiten des Poisson–Prozesses vollständig unerreichbar (Lipster und Shiryaev, 1978, S. 240). Es läßt sich zeigen, dass der Kompensator eines Punktprozesses genau dann stetig ist (fast sicher), wenn die Sprungzeitpunkte vollständig unerreichbar sind. Dies steht auch im Einklang mit obigem Beispiel. Wählt man nämlich $\mathcal{F}_t \equiv \mathcal{F}$, so sind alle Sprungzeitpunkte vorhersagbar und der Kompensator ist unstetig. Wählt man die minimale Filtration $\mathcal{F}_t = \mathcal{F}_t^N$, so läßt sich weiter zeigen, dass der Kompensator A_t fast sicher (absolut) stetig ist, falls die bedingten Verteilungen der T_n , gegeben die vorherigen Sprungzeitpunkte, (absolut) stetig sind (Lipster und Shiryaev, 1978, S. 243). Im Falle des Poisson–Prozesses gilt

$$A_t = \lambda t = \int_0^t \lambda_s ds \quad .$$

Punktprozesse $N = (N_t, \mathcal{F}_t)$, deren Kompensator $A = (A_t, \mathcal{F}_t)$ der Gestalt

$$A_t = \int_0^t \lambda_s db_s$$

ist, nennt man auch Prozesse vom Typ Poisson, wobei λ_s ein nichtnegativer vorhersagbarer Prozeß und b_s eine deterministische nichtnegative isotope und rechtsstetige Funktion ist. Der Prozeß λ_s wird auch als stochastische Intensität bezeichnet. Nach einem Satz von Aalen (1978) gilt (Fleming und Harrington, 1991, S. 131)

$$\lambda_{t+} = \lim_{h \searrow 0} \frac{\mathbb{P}(N_{t+h} - N_t \geq 1 | \mathcal{F}_t)}{h} \quad , \quad (2.37)$$

wobei mit λ_{t+} der rechtsseitige Limes bezeichnet wird. Dies mag den Begriff der Intensität verdeutlichen. Er ist ein Analogon zur Hazardrate bei Lebensdauerverteilungen. Für

Zählprozesse mit absolut stetigem Kompensator gilt folgende Beziehung (M_t ist quadratintegrierbar, (Karr, 1991, S. 64))

$$\langle M \rangle_t = A_t \quad ,$$

wobei es sich bei M_t um das Martingal aus der Doob–Meyer–Zerlegung handelt, dessen Bedeutung durch ein Zitat wiedergegeben wird (Karr, 1991, S. 61):

[...] *this martingale [...] is schizophrenic. In some context it is a noise process for which estimation is impossible, while in others it contains all relevant information [...].*

Bemerkung Alternativ kann ein Kompensator A_t bzgl. (N_t, \mathcal{F}_t^N) auch durch folgende Eigenschaften definiert werden (Karr, 1991, S. 60)

- A_t \mathcal{F}_t^N -vorhersagbar und
- für jeden nichtnegativen \mathcal{F}_t^N -vorhersagbaren Prozeß f gilt $\mathbb{E}(\int_0^\infty f dN) = \mathbb{E}(\int_0^\infty f dA)$.

Wählt man $f = 1$, so gilt für die jeweiligen Zuwächse

$$\mathbb{E}(\mathbb{E}(N_t - N_s | \mathcal{F}_{s-}^N)) = \mathbb{E}(A_t - A_s) \quad ,$$

wofür gelegentlich die heuristische Schreibweise

$$dA_s = \mathbb{E}(dN_s | \mathcal{F}_{s-}^N)$$

bzw.

$$\lambda_s ds = \mathbb{E}(dN_s | \mathcal{F}_{s-}^N) \quad ,$$

verwendet wird, falls eine stochastische Intensität λ_t existiert. Wegen

$$\mathbb{E}([dN_t - \mathbb{E}(dN_t | \mathcal{F}_{t-})]^2 | \mathcal{F}_{t-}) = \mathbb{E}([dM_t]^2 | \mathcal{F}_{t-}) = d\langle M \rangle_t = dA_t$$

gilt, dass sich Zählprozesse mit absolut stetigem Kompensator aufgrund der Gleichheit des Erwartungswertes und der Varianz lokal und bedingt wie Poisson–Prozesse verhalten.

2.3.4 Der Likelihoodprozeß

Im diskreten Fall wurde der Likelihoodprozeß als ein Beispiel für ein Martingal genannt. Der Likelihoodprozeß läßt sich auch im zeitstetigen Fall darstellen. Mit Hilfe der Cha-

Charakterisierung von Punktprozessen durch den Kompensator kann man notwendige und hinreichende Bedingungen angeben, die die Absolutstetigkeit von zwei Punktprozessen garantieren. Der Likelihoodprozeß erweist sich somit als Radon–Nikodym–Dichte. Die Darstellung beschränkt sich hier auf Prozesse mit stetigem Kompensator, da man sich meist auf den (homogenen) Poisson–Prozeß mit Parameter $\lambda = 1$ bezieht.

Satz

Sei $N = (N_t, \mathcal{F}_t)$ ein Punktprozess mit der Verteilung π^N sowie mit dem stetigen Kompensator $A^N = (A_t^N, \mathcal{F}_t)$ und $\pi^{\tilde{N}}$ ein weiteres Wahrscheinlichkeitsmaß. Die Restriktionen von $\pi^N, \pi^{\tilde{N}}$ auf \mathcal{F}_t werden mit $\pi_t^N, \pi_t^{\tilde{N}}$ bezeichnet. Für die absolute Stetigkeit von $\pi^{\tilde{N}}$ bzgl. π^N ($\pi^{\tilde{N}}$ fast sicher) ist notwendig und hinreichend, dass folgende Bedingungen erfüllt sind (Lipster und Shiryaev, 1978, S. 315)

$$A_t^{\tilde{N}} = \int_0^t \lambda_s dA_s^N, \quad t < T_\infty, \quad \int_0^{T_\infty} (1 - \sqrt{\lambda_s})^2 dA_s^N < \infty \quad (2.38)$$

wobei λ_s ein nichtnegativer vorhersagbarer Prozeß ist. In diesem Fall gilt

$$\begin{aligned} L_t &= \frac{d\pi_t^{\tilde{N}}}{d\pi_t^N} \\ &= \exp \left\{ \int_0^t \ln \frac{dA_s^{\tilde{N}}}{dA_s^N} dN_s - [A_t^{\tilde{N}} - A_t^N] \right\} \end{aligned} \quad (2.39)$$

Dieser Satz ermöglicht es, alle Prozesse anzugeben, die absolut stetig sind bzgl. des homogenen Poisson–Prozesses mit $\lambda = 1$.

Korollar

(Lipster und Shiryaev, 1978, S. 317)

Es gelten die oben genannten Bezeichnungen. Weiter sei $\tilde{N} = (\tilde{N}_t, \mathcal{F}_t)$ ein Punktprozeß mit absolut stetigem Wahrscheinlichkeitsmaß $\pi^{\tilde{N}}$ bzgl. π^N . Dann hat der Kompensator von \tilde{N} folgende Gestalt

$$A_t^{\tilde{N}} = \int_0^t \lambda_s ds, \quad t < \infty, \quad ,$$

wobei λ_s ein nichtnegativer vorhersagbarer Prozeß ist mit

$$\int_0^\infty (1 - \sqrt{\lambda_s})^2 ds < \infty$$

und

$$L_t = \frac{d\pi_t^{\tilde{N}}}{d\pi_t^N} = \exp \left\{ \int_0^t \ln \lambda_s dN_s + \int_0^t (1 - \lambda_s) ds \right\} .$$

Man beachte, dass im Poissonfall ($\lambda = 1$) $A_t \equiv t$ und $T_\infty = \infty$ gilt.

Bemerkung Der rechtsseitige Ausdruck läßt sich umformen zu

$$L_t = \prod_{\{T_n \leq t\}} \lambda_{T_n} \cdot \exp \left\{ \int_0^t (1 - \lambda_s) ds \right\} ,$$

da nur die Sprungzeitpunkte T_n einen Beitrag zur Integration bezüglich dN_s liefern. Zur Motivation (Karr, 1991, S. 72) betrachte man den Punktprozeß als überabzählbares Bernoulliexperiment auf $[0, T]$ mit unabhängigen Erfolgswahrscheinlichkeiten λ_t und Beobachtungen T_1, \dots, T_n , deren Wahrscheinlichkeit

$$\prod_{T_i \leq t} \lambda_{T_i} = \exp \left\{ \int_0^t \ln \lambda_s dN_s \right\}$$

beträgt und den Nichterfolgswahrscheinlichkeiten $1 - \lambda_s$, $T_{j-1} < s < T_j$, für das Intervall (T_{j-1}, T_j) , deren überabzählbares Produkt sich als Produktintegral (Andersen et al., 1993)

$$\prod_{T_{j-1}}^{T_j} (1 - \lambda_s ds) = \exp \left\{ - \int_{T_{j-1}}^{T_j} \lambda_s ds \right\}$$

darstellen läßt, wobei vorausgesetzt wird, dass λ_s stetig in s ist. Die Gesamtwahrscheinlichkeit beträgt

$$\exp \left\{ \int_0^t \ln \lambda_s dN_s - \int_0^t \lambda_s ds \right\} .$$

Ein analoges Vorgehen für den Poisson-Prozeß mit $\lambda_s \equiv 1$ liefert den Faktor

$$\exp t = \exp \int_0^t 1 ds .$$

Unter der genannten Bedingung erhält man alle Punktprozesse, die bezüglich des homogenen Poissonprozesses eine Dichte haben, d.h. dominiert sind. Daher betrachtet man meistens statt des Likelihoodquotientenprozesses nur den Zähler

$$\exp \left\{ \int_0^t \ln \lambda_s dN_s - \int_0^t \lambda_s ds \right\} = \prod_{\{T_n \leq t\}} \lambda_{T_n} \cdot \exp \left\{ - \int_0^t \lambda_s ds \right\} .$$

Ist der Intensitätsprozeß λ_s in der statistischen Anwendung von einem Parametervektor $\theta \in \Theta$ abhängig, so schreibt man den Likelihoodprozeß $L_t = L_t(\theta)$ als

$$L_t(\theta) = \prod_{\{T_n \leq t\}} \lambda_{T_n}(\theta) \cdot \exp \left\{ - \int_0^t \lambda_s(\theta) ds \right\} \quad . \quad (2.40)$$

Nach dem Logarithmieren und einigen Umformungen erhält man auch die in der Literatur gebräuchliche Schreibweise

$$l_t(\theta) = \ln L_t(\theta) = \int_0^t \ln(\lambda_s(\theta)) dN_s - \int_0^t \lambda_s(\theta) ds \quad .$$

Falls man Integration und Differentiation vertauschen kann, erhält man den Score-Prozeß:

$$\frac{\partial}{\partial \theta} l_t(\theta) = \int_0^t \frac{\partial}{\partial \theta} \ln(\lambda_s(\theta)) dN_s - \int_0^t \frac{\partial}{\partial \theta} \lambda_s(\theta) ds \quad . \quad (2.41)$$

Der Likelihoodprozeß sowie der Score-Prozeß besitzen die Martingaleigenschaft (Andersen et al. (1993, S. 100) bzw. Fleming und Harrington (1991, S. 46, 65,77)).

3 Modellierung und Statistische Inferenz bei Lebensdauermodellen

Lebensdauermodelle zeichnen sich dadurch aus, dass der Zeitpunkt T des Eintretens eines Ereignisses aufgezeichnet wird. Oft beobachtet man dazu Kovariablen X . Sind die Kovariablen zeitabhängig, so spricht man auch von einem Kovariablenprozeß X_t . Können verschiedene Ereignisse eintreten, so handelt es sich um multivariate Lebensdauermodelle, die mit h indiziert werden, also T_h für das Eintreten des Ereignisses vom Typ h . Stillschweigend wird davon ausgegangen, dass nicht mehrere Ereignisse gleichzeitig auftreten können. Oft beobachtet man die Ereignisse für eine Population und betrachtet pro Individuum i den Kovariablenprozeß X_{it} sowie das Auftreten eines Ereignisses T_{ih} . Abgesehen von dem engen Fall der parametrischen Lebensdauermodelle, bei denen die Wartezeit für ein Ereignis z.B. weibullverteilt ist, sind semiparametrische Modelle von besonderer Bedeutung. Grundsätzlich kann man die Überlebenswahrscheinlichkeiten oder die Hazardrate modellieren. Bei der Modellierung der Überlebenswahrscheinlichkeiten ist die Parameteridentifikation schwieriger, siehe Klein und Moeschberger (2003). Daher erfolgt die Modellbildung anhand der Hazardrate. Die Hazardrate steht im engen Zusammenhang zum Intensitätsprozeß, wie sich beim multiplikativen Intensitätsmodell zeigt. Ein mittlerweile klassisches Modell geht auf Cox (1972) zurück, bei dem die Intensität in einen zeitabhängigen und zeitunabhängigen Teil aufgespalten wird. Es erlaubt allerdings nur die Schätzung von festen Effekten β – der zeitabhängige Anteil wird nur nichtparametrisch mitgeschätzt. Aus theoretischer wie praktischer Sicht ist allerdings eine Schätzung von zeitabhängigen Effekten β_t von Interesse. Da typischerweise Kovariablen bzw. Kovariablenprozesse beobachtet werden, erfolgt eine Modellierung der Intensität ähnlich wie bei Regressionsmodellen. Daher wird vorgeschlagen, entweder im allgemeinen Rahmen mit Schätzfunktionen, die den Likelihoodfall beinhalten, zu arbeiten oder aber einen bayesianischen Ansatz zu wählen. Die Schätzfunktionen fügen sich ‘nahtlos’ in die Theorie der Zählprozesse ein, wohingegen ein bayesianischer Ansatz vom Likelihoodprozeß ausgeht. Bei beiden wird der Intensitätsprozeß modelliert. Für die Umsetzung des bayesianischen Ansatzes werden sich Markov Chain Monte Carlo–Methoden (MCMC) als hilfreich erweisen. Ein besonderes Problem bei Lebensdauermodellen, vor allem in der praktischen Anwendung, ist die Zensierung, d.h. man

beobachtet das Ereignis nicht. Bei Studien werden meistens nur für einen bestimmten Zeitraum Daten erhoben; es ist demnach möglich, dass bis zum Ende der Studie für ein Individuum kein Ereignis beobachtet werden kann. In der Praxis treten verschiedene Formen der Zensierung auf.

3.1 Zensierung von Beobachtungen

Für jedes Individuum i liegen im Fall vollständiger Beobachtungen die Daten (T_{ih}, X_{it}) , $i = 1, \dots, n$ vor. Zur Berücksichtigung der Zensierung führt man eine Indikatorvariable C_i ein mit

$$C_i = \begin{cases} 1 & , \text{keine Zensierung} \\ 0 & , \text{Ereignis zensiert für Individuum } i \end{cases}$$

und beschreibt die Daten durch (T_{ih}, X_{it}, C_i) , $i = 1, \dots, n, h = 1, \dots, H, t = 1, \dots, T$. Allgemeiner könnte die Indikatorvariable C_i noch von h abhängen. Die Zensierungsarten sollen anhand einer Studie besprochen werden, wobei vorausgesetzt wird, dass Beobachtungen nur während der genau festgelegten Studiendauer gemacht werden. Man spricht von Rechtszensierung, falls für ein Individuum bis zum Ende der Studie keine Beobachtung gemacht wird. Linkszensierung tritt dann ein, wenn das Ereignis beobachtet wird, allerdings die wahre Wartezeit nicht ermittelt werden kann, da die Beobachtung erst mit Beginn der Studie einsetzt. Man denke etwa an den Ausbruch einer Krankheit. Intervallzensierung liegt vor, wenn man den genauen Zeitpunkt des Ereignisses nicht ermitteln kann, da die Daten nicht stetig, sondern nur z.B. periodisch erhoben werden, d.h. man ermittelt nur den nächsten auf den Ereigniseintritt folgenden Beobachtungstermin. In diesem Fall verwendet man zeitdiskrete Lebensdauermodelle (Blossfeld, Hamerle und Mayer, 1989). Im folgenden wird nur der Fall der Rechtszensierung betrachtet. Dieser Fall läßt sich dahingehend verallgemeinern, dass die Zensierung an einem zufälligen zukünftigen Zeitpunkt U_i erfolgt. Entsprechend verallgemeinert man C_i zu $C_{it} := \mathbb{1}_{\{t \leq U_i\}}$. Dies ist z.B. der Fall, wenn man die Studie nach dem r -ten Eintreten des Ereignisses h abbricht. Die Indikatorvariablen C_{it} sind linksstetige Stoppzeiten, sie müssen aber nicht notwendigerweise adaptiert bezüglich der Filtration \mathcal{F}_t (des unzensierten Falles) sein. Vergrößert man die Filtration zu $\mathcal{G}_t = \sigma(\mathcal{F}_t, C_{is}, s \leq t, i = 1, \dots, n)$, so sind die Variablen C_{it} adaptiert bezüglich \mathcal{G}_t und aufgrund der Linksstetigkeit sind sie sogar \mathcal{G}_t vorhersagbar. Die Zensierung hängt nur von der Vergangenheit ab. Man spricht nun von unabhängiger Rechtszensierung, wenn der Kompensator A bezüglich \mathcal{F}_t und \mathcal{G}_t identisch ist (obwohl der Wert des Kompensators für t im Zensierungsfall unterschiedlich

ist). Bezeichne N_{it}, A_{it}, M_{it} für ein Individuum i die jeweiligen Größen im Nichtzensierungsfall und $N_{it}^c, A_{it}^c, M_{it}^c$ im Zensierungsfall mit Stoppzeit C_{it} . Mit

$$N_{it}^c = \int_0^t C_{is} dN_{is}$$

gilt dann die Doob–Meyer–Zerlegung im Zensierungsfall

$$N_{it}^c = \int_0^t C_{is} dA_{is} + \int_0^t C_{is} dM_{is} = A_{it}^c + M_{it}^c,$$

wobei die Vorhersagbarkeit und die Beschränktheit von C_{it} wesentlich sind. Ein wichtiger Spezialfall der unabhängigen Rechtszensierung ist die nicht informative Zensierung. In diesem Fall läßt sich der Likelihoodprozeß L_t in einen Teil aufspalten, der nur von θ abhängt und einen Teil, der den Zensierungsmechanismus beschreibt.

Eine Darstellung für die Zensierung, die sich eher am Intensitätsprozeß orientiert, findet man in Fleming und Harrington (1991). In der Literatur findet man auch folgende Darstellung mit $t_i = T_i \wedge U_i$ und $\delta_i = \mathbb{1}_{\{T_i \leq U_i\}}$. Damit lautet der Beitrag zur Likelihood je Individuum i ($dN_{it}^c = \delta_i dN_{it}$ und $t_i \leq t$)

$$\exp \left\{ \int_0^{t_i} \ln \lambda_s dN_{is}^c - \int_0^{t_i} \lambda_s ds \right\} = \lambda_{t_i}^{\delta_i} \exp \left\{ - \int_0^{t_i} \lambda_s ds \right\}. \quad (3.1)$$

3.2 Multiplikatives Intensitätsmodell

Beim multiplikativen Intensitätsmodell, das auf Aalen (1978) zurückgeht, wird der (\mathcal{F}_t -vorhersagbare) Intensitätsprozeß λ_t bzw. λ_{it} modelliert mit Index i für das jeweilige Individuum. Es erfolgt eine multiplikative Aufspaltung in

$$\lambda_{it} = Y_{it} \cdot h_{it} \quad i = 1, \dots, n, \quad (3.2)$$

wobei die Funktionen h und Y nichtnegativ, linksstetig mit rechtsseitigen Grenzwerten, Y \mathcal{F}_t -vorhersagbar sowie h integrierbar sind. Betrachtet man den Einpunktprozeß, d.h., beobachtet man nur eine Lebensdauer T mit Verteilungsfunktion F und Dichte f , also $N_t = \mathbb{1}_{\{T \leq t\}}$, so erhält man

$$\lambda_t = \mathbb{1}_{\{t \leq T\}} \cdot \frac{f(t)}{1 - F(t)} \quad (3.3)$$

mit $Y_t = \mathbb{1}_{\{t \leq T\}} = 1 - N_{t-}$ und h_t als Hazardrate $h_t = \frac{f(t)}{1-F(t)}$. Im Fall von unabhängiger Rechtszensierung mit C_t als Zensierungsvariable gilt für den zensierten Intensitätsprozeß λ_s^c (Andersen et al., 1993, S. 140)

$$\lambda_{it}^c = Y_{it}^c \cdot h_{it} \quad (3.4)$$

mit

$$Y_{it}^c = Y_{it} \cdot C_t. \quad (3.5)$$

Üblicherweise hängt λ_t noch von einem Parameter $\theta \in \Theta$ ab, die zensierte Variante schreibt sich also

$$\lambda_s^c(\theta) = Y_{it}^c \cdot h_{it}(\theta). \quad (3.6)$$

Entscheidend ist die Funktion $h_{it}(\theta)$. Zerlegt man h multiplikativ in einen zeitabhängigen (für alle Individuen gemeinsamen Teil) und einen Teil, der nur von θ abhängt, so erhält man das Cox-Modell.

3.3 Multiplikatives Hazardmodell

Wählt man für $h_{it}(\theta)$, $\theta = (\gamma, \beta)$ eine multiplikative Form mit Kovariablen x_{it} der Form

$$h_{it}(\theta) = h_t^0(\gamma) \cdot \exp(\beta^t x_{it}), \quad (3.7)$$

so spricht man vom Cox-Modell. In der Literatur wird es auch als multiplikatives Hazardmodell bezeichnet. Die Funktion $h_t^0(\gamma)$ wird auch als *Baseline-Hazard* bezeichnet. Das primäre Interesse gilt der Schätzung der Regressionsparameter β mittels der partiellen Loglikelihoodfunktion $pl_t(\beta)$ (Andersen et al., 1993, S. 482)

$$pl_t(\beta) = \sum_{i=1}^n \left[\int_0^t \beta^t x_{is} dN_{is}^c - \int_0^t \log \left(\sum_{i=1}^n Y_{is} \exp(\beta^t x_{is}) \right) dN_{is}^c \right] \quad (3.8)$$

bzw. den partiellen Ableitungen

$$\frac{\partial}{\partial \beta} pl_t(\beta) = \sum_{i=1}^n \left[\int_0^t x_{is} - \frac{\sum_{i=1}^n Y_{is} x_{is} \exp(\beta^t x_{is})}{\sum_{i=1}^n Y_{is} \exp(\beta^t x_{is})} dN_{is}^c \right]. \quad (3.9)$$

Der Gradient von $pl_t(\boldsymbol{\beta})$ läßt sich (vgl. Gill (1984)) umformen zu

$$\frac{\partial}{\partial \boldsymbol{\beta}} pl_t(\boldsymbol{\beta}) = \sum_{i=1}^n \left[\int_0^t \mathbf{x}_{is} - \frac{\sum_{i=1}^n Y_{is} \mathbf{x}_{is} \exp(\boldsymbol{\beta}' \mathbf{x}_{is})}{\sum_{i=1}^n Y_{is} \exp(\boldsymbol{\beta}' \mathbf{x}_{is})} dM_{is}^c \right] \quad (3.10)$$

und die Vorhersagbarkeit des Intensitätsprozesses liefert die Martingaleigenschaft von $\frac{\partial}{\partial \boldsymbol{\beta}} pl_t(\boldsymbol{\beta})$. Asymptotische Aussagen finden sich in Gill (1984), (bzw. ausführlicher in Andersen et al. (1993), Kapitel 7). Die im allgemeinen unbekannt Funktion $h_t^0(\boldsymbol{\gamma})$ wird nur mitgeschätzt (Breslow–Aalen–Schätzer). Dies geschieht nichtparametrisch. Sie ist ein (unendlich-dimensionaler) Störparameter. Eine alternative Darstellung der partiellen Loglikelihoodfunktion ohne Bezug zu den Zählprozessen lautet

$$pl(\boldsymbol{\beta}) = \sum_{i=1}^n C_i \left[\boldsymbol{\beta}' \mathbf{x}_i - \ln \left\{ \sum_{j \in \mathcal{R}_i} \exp(\boldsymbol{\beta}' \mathbf{x}_j) \right\} \right] \quad (3.11)$$

mit Risikomenge $\mathcal{R}_i = \{j : T_j \geq T_i\}$, die auch die zensierten Fälle enthält. Dem korrespondierenden Gradienten

$$\frac{\partial}{\partial \boldsymbol{\beta}} pl(\boldsymbol{\beta}) = \sum_{i=1}^n C_i (\mathbf{x}_i - \bar{\mathbf{x}}_i) \quad (3.12)$$

mit

$$\bar{\mathbf{x}}_i = \frac{\sum_{j \in \mathcal{R}_i} \mathbf{x}_j \exp(\boldsymbol{\beta}' \mathbf{x}_j)}{\sum_{j \in \mathcal{R}_i} \exp(\boldsymbol{\beta}' \mathbf{x}_j)} \quad (3.13)$$

entnimmt man, das $\hat{\boldsymbol{\beta}}$ so zu wählen ist, dass die Summe der Differenzen aus Kovariablen und gewichteten durchschnittlichen Kovariableneffekten Null ergeben muß. Von den Ereigniszeitpunkten – dem eigentlichen Response – wird nur die Reihenfolge benötigt. Aus diesem Blickwinkel ergeben sich in natürlicher Weise die vektorwertigen Schoenfeldresiduen (Schoenfeld, 1982) pro Individuum ($\bar{\mathbf{x}}_i = \bar{\mathbf{x}}_i(\hat{\boldsymbol{\beta}})$, $\hat{\mathbf{r}}_i = \hat{\mathbf{r}}_i(\hat{\boldsymbol{\beta}})$)

$$\hat{\mathbf{r}}_i = \mathbf{x}_i - \bar{\mathbf{x}}_i \quad . \quad (3.14)$$

Betrachtet man den Quotienten für zwei Individuen A und B,

$$\frac{h_{At}}{h_{Bt}} = \exp(\boldsymbol{\beta}' \mathbf{x}_{At} - \boldsymbol{\beta}' \mathbf{x}_{Bt}) \quad (3.15)$$

$$= \exp(\boldsymbol{\beta}' (\mathbf{x}_{At} - \mathbf{x}_{Bt})), \quad (3.16)$$

so hängt das relative Risiko nur von den Parametern und den Kovariablen ab. Ein Vorteil der multiplikativen Form ist gerade diese eingängige Interpretation, insbesondere im Fall von zeitunabhängigen Kovariablen. Praktische Hinweise zu dem (komplexeren) Umgang im Fall von zeitabhängigen Kovariablen findet man in Vermunt (1996, S. 142). In einigen Situationen ist das Modell evtl. ungeeignet, da etwa für zeitunabhängige Kovariablen das relative Risiko konstant über die Zeit ist. Es besteht daher ein Interesse, Modelle mit einer größeren Klasse von Funktionen als $\exp(\beta' x_{it})$ zu benutzen.

3.4 Modelle mit zeitabhängigen Parametern

Es existieren verschiedene Ansätze in der Literatur, etwa Karr (1991),

$$h_{it}(\theta) = h_t^0(\gamma) \cdot \exp(\eta_{it}) \quad (3.17)$$

mit

$$\eta_{it} = \beta_t \cdot z_i, \quad (3.18)$$

wobei zur Vereinfachung der Notation β eindimensional ist. Die Kovariablen sind zeitunabhängig. Die Schätzung (Karr, 1991) erfolgt mit der partiellen Likelihoodfunktion, allerdings ergänzt um einen Penalisierungsterm (vgl. auch Gray (1992)). Für aktuellere Literatur siehe etwa Martinussen, Scheike und Skovgaard (2002), dort werden kumulierte Regressionsfunktionen mit Hilfe von Kerndichteschätzern berechnet. In Tian, Zucker und Wei (2005) werden ebenfalls Kerndichteschätzer für zeitvariierende Effekte im Cox-Modell propagiert. In Kauermann (2005) verzichtet man auf eine Einbettung der zeitvariierenden Effekte in ein Cox-Modell. Die auftretenden Integrale in der Likelihood werden mit der Sehnen-Trapezregel numerisch berechnet, was approximativ zu einer Poisson Regression führt (vgl. auch Cai, Hyndman und Wand (2002)). Weitere Literaturhinweise findet man auch in Kapitel (4) bzw. in Abschnitt (4.1).

3.5 Schätzfunktionen

Der Abschnitt beinhaltet eine kurze Einführung in Schätzfunktionen und die Gewinnung von Schätzern, wobei der Schwerpunkt auf dem Gebiet der stochastischen Prozesse liegt. Dies erscheint sinnvoll, da bei Lebensdauermodellen mit Zufallseffekten keine (vollen)

Likelihoodschätzungen möglich sind. Die Schätzer werden über ihre asymptotische Normalität motiviert. Es folgt ein Abschnitt zur Anwendung in Hinblick auf Zählprozesse, anschließend eine Erörterung der Schätzungen in Lebensdauermodellen. Die Darstellung orientiert sich an Heyde (1997).

3.5.1 Konzept der Schätzfunktionen

Es wird kurz die Gewinnung von Schätzern $\hat{\theta} \in \mathbb{R}^p$ für $\theta \in \Theta \subset \mathbb{R}^p$ erörtert. Dies geschieht im Hinblick auf die Anwendung im Lebensdauerbereich in einem etwas allgemeineren Rahmen als im Likelihoodfall mittels Schätzfunktionen $G_T \in \mathcal{G}, G_T \in \mathbb{R}^p$, wobei $p \in \mathbb{N}$ der Dimension des Schätzers bzw. des unbekannt Parametervektors entspricht, obwohl die Scorefunktion nicht als Schätzfunktion ausgeschlossen wird. Es kann jedoch sinnvoll sein, einfachere bzw. robustere Schätzfunktionen zu betrachten, wobei man allerdings einen Effizienzverlust für die Schätzer in Kauf nehmen muß. Um verschiedene Schätzer $\hat{\theta}$ zu vergleichen, die man als Lösung der Gleichung $G_T(\theta) = 0$ erhält, beschränkt man sich auf eine Teilmenge, die Klasse der erwartungstreuen Schätzfunktionen, die asymptotisch normalverteilt sind. Mit \mathcal{G} bezeichnet man die Klasse der quadratintegrierbaren Schätzfunktionen $G_T \equiv G_T(\mathbf{X}, \theta) (= G_T(\theta))$ mit Parameter θ und Beobachtungen $\{X_t(\omega), t \leq T\}$. Es gelte $\mathbb{E}_\theta(G_T(\theta)) = 0$ sowie die Vertauschbarkeit von Integration und Differentiation, d.h.

$$\frac{\partial}{\partial \theta} \mathbb{E}_\theta(G_T(\theta)) = \mathbb{E}_\theta \left(\frac{\partial}{\partial \theta} G_T(\theta) \right) = \mathbb{E}_\theta(\dot{G}_T(\theta)) \quad ,$$

wobei beide Matrizen für alle θ den vollen Rang haben. Bezeichnet θ_0 den unbekannt wahren Parameter, so gilt approximativ

$$0 = G_T(\hat{\theta}) \approx G_T(\theta_0) + (\hat{\theta} - \theta_0) \mathbb{E}_{\theta_0}(\dot{G}_T(\theta_0)) \quad .$$

Auflösen nach $\hat{\theta}$ liefert

$$\hat{\theta} - \theta_0 \approx - [\mathbb{E}_{\theta_0}(\dot{G}_T(\theta_0))]^{-1} G_T(\theta_0)$$

und nach Bildung des Erwartungswertes gilt unter der o. g. Bedingung

$$\mathbb{E}_{\theta_0}(\hat{\theta}) \approx \theta_0 \quad .$$

Zusätzlich fordert man noch den vollen Rang p von $\mathbb{E}_\theta(G_T G_T^t)$. Nun kann versucht werden, eine untere Schranke für die Varianz der Schätzfunktion zu finden, um so Effizienz zu definieren. Alternativ könnte ein asymptotischer Effizienzbegriff entwickelt werden. Die beiden Effizienzbegriffe müssen sich nicht ausschließen. Im allgemeinen wird man sich allerdings mit einer asymptotischen Verteilung des Schätzers zufrieden geben müssen. Deswegen wird die asymptotische Effizienz im weiteren verwendet. Beschränkt man sich auf Schätzfunktionen G_T , die zusätzlich noch die Martingaleigenschaft besitzen, so kann man die Martingaltheorie für Grenzwertsätze benutzen. Im Hinblick auf die zeitstetigen Zählprozesse geschieht dies gleich für zeitstetige Martingale. Es sei also G_T ein p -dimensionales Martingal bezüglich \mathcal{F}_T und $\langle G \rangle_T$ die dazu korrespondierende $(p \times p)$ -wertige vorhersagbare quadratische Variation. Sei M diejenige Teilmenge von \mathcal{G} , für die der multivariate zentrale Grenzwertsatz

$$(\langle G \rangle_T)^{-\frac{1}{2}} G_T \xrightarrow{d} NV(\mathbf{0}, \mathbf{I}_p), \quad T \rightarrow \infty, \quad (3.19)$$

gilt, sowie mit $\hat{\theta}$ als einer Lösung von $G_T(\theta) = 0$ mittels Taylor-Approximation

$$(\langle G(\theta) \rangle_T)^{-\frac{1}{2}} \frac{\partial}{\partial \theta} G(\theta)_T (\hat{\theta} - \theta) \xrightarrow{d} NV(\mathbf{0}, \mathbf{I}_p), \quad T \rightarrow \infty. \quad (3.20)$$

In der Literatur, (Heyde, 1997) wird vorgeschlagen, anstelle von $\frac{\partial}{\partial \theta} G(\theta)_T = \dot{G}(\theta)_T$ den \mathcal{F}_t -vorhersagbaren Prozeß

$$\bar{G}_t(\theta) = \int_0^t \mathbb{E}(d\dot{G}_s(\theta) | \mathcal{F}_{s-})$$

zu verwenden. Damit schreibt sich der zentrale Grenzwertsatz als

$$(\langle G(\theta) \rangle_T)^{-\frac{1}{2}} \bar{G}_T(\theta) (\hat{\theta} - \theta) \xrightarrow{d} NV(\mathbf{0}, \mathbf{I}_p), \quad T \rightarrow \infty.$$

Es folgt unmittelbar

$$(\hat{\theta} - \theta)' \bar{G}_T^t(\theta) \langle G(\theta) \rangle_T^{-1} \bar{G}_T(\theta) (\hat{\theta} - \theta) \xrightarrow{d} \chi_p^2.$$

Der Ausdruck

$$I_G(\theta) = \bar{G}_T^t(\theta) \langle G(\theta) \rangle_T^{-1} \bar{G}_T(\theta) \quad (3.21)$$

wird als Martingal-Information bezeichnet und läßt sich als Verallgemeinerung der Fisher-Information betrachten. Man erhält für maximales $I_G(\theta)$ minimale asymptoti-

sche Konfidenzbereiche. Derartige Schätzfunktionen bezeichnet man als asymptotisch effizient, wobei der Vergleich von Matrizen über die Definitheit der Differenz erfolgt, d.h. $G_T^*(\theta)$ wird als asymptotisch effizient bezeichnet, wenn für $\theta \in \Theta$

$$I_{G^*}(\theta) - I_G(\theta) \geq 0 \quad \text{fast sicher} \quad (3.22)$$

für alle Schätzfunktionen $G_T(\theta)$, $T > 0$ gilt, die die obigen Aussagen erfüllen. Es ist nicht ohne weiteres ersichtlich, wie eine optimale Schätzfunktion gefunden werden kann. Bildet man zu einem Prozeß X_t die Doob–Meyer–Zerlegung $X_t = A_t + M_t$ mit $A_0 = M_0 = 0$ (Existenz vorausgesetzt) bezüglich \mathcal{F}_t und beschränkt sich auf die Klasse der Schätzfunktionen mit

$$M = \left\{ G_t(\theta) : \int_0^t \alpha_s(\theta) dM_s(\theta), \quad \alpha_s \text{ vorhersagbar} \right\} \quad (3.23)$$

d.h. auf Martingale mit integraler Darstellung, so existiert die Hutton–Nelson–Lösung (Hutton und Nelson, 1986), mit $d\bar{M}_s = \mathbb{E}(dM_s | \mathcal{F}_{s-})$,

$$G_t^*(\theta) = \int_0^t (d\bar{M}_s)^t (d\langle M \rangle_s)^- dM_s, \quad (3.24)$$

wobei $(\cdot)^-$ eine verallgemeinerte Inverse bezeichnet. Falls man die Doob–Meyer–Zerlegung für univariate Prozesse durchführt, sind die Prozesse dM_s und $d\langle M \rangle_s$ skalar. Der Prozeß $d\bar{M}_s^t$ hat die Dimension des Parametervektors θ , ebenso wie die optimale Schätzfunktion G_T^* .

3.5.2 Schätzfunktionen und Zählprozesse

Die hier betrachteten Zählprozesse gestatten noch einige Umformungen. Zur Vereinfachung der Notation werden in diesem Abschnitt nur univariate Zählprozesse betrachtet. Da sich für Zählprozesse der Kompensator A_t auch schreiben läßt als

$$A_t = \langle M \rangle_t, \quad ,$$

bzw. mit dem vorhersagbaren Intensitätsprozeß $\lambda_t = \lambda_t(\theta)$

$$dA_t = d\langle M \rangle_t = \lambda_t dt$$

und (unter Beachtung, dass nachdifferenziert werden muß)

$$d\bar{M}_t = \mathbb{E}(d\dot{M}_t | \mathcal{F}_{t-}) = -d\dot{A}_t = -\dot{\lambda}_t(\theta)dt = -\frac{\partial}{\partial\theta}\lambda_t(\theta)dt \quad ,$$

vereinfacht sich die Schätzfunktion $G_t^*(\theta)$ zu (das Minuszeichen ist irrelevant)

$$G_t^*(\theta) = \int_0^t \frac{\dot{\lambda}_s(\theta)}{\lambda_s(\theta)} dM_s \quad . \quad (3.25)$$

Wegen $dM_s = dN_s - dA_s = dN_s - \lambda_s ds$ läßt sich dies umformulieren zu

$$G_t^*(\theta) = \int_0^t \frac{\dot{\lambda}_s(\theta)}{\lambda_s(\theta)} dN_s - \int_0^t \frac{\dot{\lambda}_s(\theta)}{\lambda_s(\theta)} \lambda_s(\theta) ds \quad . \quad (3.26)$$

Den Schätzer $\hat{\theta}$ erhält man über die Gleichung $G_t^*(\theta) = 0$.

$$\int_0^t \frac{\dot{\lambda}_s(\theta)}{\lambda_s(\theta)} dN_s = \int_0^t \frac{\dot{\lambda}_s(\theta)}{\lambda_s(\theta)} \lambda_s(\theta) ds \quad . \quad (3.27)$$

Dies läßt sich weiter vereinfachen zu

$$\sum_{\{s \leq t: N_s \neq N_{s-}\}} \frac{\dot{\lambda}_s(\theta)}{\lambda_s(\theta)} = \int_0^t \lambda_s(\theta) ds \quad , \quad (3.28)$$

d.h. die Integration entspricht einer Summation über die Sprungstellen ($N_s \neq N_{s-}$) des Zählprozesses. Zur Berechnung der asymptotischen Varianz des Schätzers muß die Martingal-Information $I_G(\theta)$ berechnet werden. Mit

$$b_s(\theta) := \frac{\dot{\lambda}_s(\theta)}{\lambda_s(\theta)}$$

erhält man nach Vertauschung von Integration und Differentiation unter Zuhilfenahme von $\dot{M}_s(\theta) = \frac{\partial}{\partial\theta} M_s(\theta)$

$$\frac{\partial}{\partial\theta} G_t(\theta) = \dot{G}_t(\theta) = \int_0^t \dot{b}_s(\theta) dM_s(\theta) + \int_0^t b_s(\theta) d\dot{M}_s(\theta) \quad , \quad (3.29)$$

sowie

$$d\dot{G}_s(\theta) = \dot{b}_s(\theta) dM_s(\theta) + b_s(\theta) d\dot{M}_s(\theta) \quad . \quad (3.30)$$

Wegen $\mathbb{E}(\dot{b}_s(\theta)dM_s(\theta)|\mathcal{F}_{s-}) = 0$ folgt nach Übergang zu den vorhersagbaren Zuwächsen

$$\mathbb{E}(d\bar{G}_s(\theta)|\mathcal{F}_{s-}) = \mathbb{E}(b_s(\theta)d\dot{M}_s(\theta)|\mathcal{F}_{s-}) = -b_s(\theta)\lambda_s(\theta)ds \quad .$$

Somit erhält man

$$\bar{G}_t(\theta) = - \int_0^t b_s(\theta)\lambda_s(\theta)ds \quad .$$

Zur Berechnung von $\langle G \rangle_t$ verwendet man (Andersen et al., 1993, S. 71) die integrale Struktur der Schätzfunktion. Dies führt zu (vgl. (2.33))

$$\langle G(\theta) \rangle_t = \oint_0^t \langle b(\theta) \rangle_s \quad d\langle M(\theta) \rangle_s \quad (3.31)$$

und mit $d\langle M(\theta) \rangle_s = \lambda_s(\theta)ds$ sowie der Vorhersagbarkeit von $b_s(\theta)$ zu

$$\langle G(\theta) \rangle_t = \int_0^t (b_s(\theta))^2 \lambda_s(\theta)ds = -\bar{G}_t(\theta) \quad . \quad (3.32)$$

Die Martingal-Information vereinfacht sich zu

$$I_G(\theta) = \int_0^t b_s(\theta)\lambda_s(\theta)ds \quad . \quad (3.33)$$

Beispiel

- Homogener Poisson-Prozeß mit Parameter $\lambda_s = \theta, \theta > 0$ und $N_t = n$; dies führt mit $\lambda_s(\theta) = 1$ zu

$$\frac{n}{\theta} = \int_0^t 1ds \quad \Rightarrow \quad \hat{\theta} = \frac{n}{t} \quad .$$

Für die Martingal-Information $I_G(\theta)$ erhält man

$$I_G = \int_0^t \frac{1}{\theta} ds = \frac{t}{\theta} \quad \Rightarrow \quad \text{Var}(\hat{\theta}) = \frac{\theta}{t} \quad .$$

- Cox-Modell:

Falls $a_s(\theta)$ in (3.23) nicht von der Baseline Hazard abhängt, erhält man als Schätzfunktionen (siehe Chang und Hsiung (1990)) die gleiche Lösung wie beim Cox-Modell.

Bemerkungen

1. In Fall des homogenen Poisson–Prozesses erhält man als Lösung für die Punktschätzung die Maximum–Likelihood–Schätzung (Snyder, 1975, S. 76) bzw. für seine Varianz (Snyder, 1975, S. 85). Im allgemeinen erhält man allerdings keine in θ lineare Gleichung, etwa im Fall $\lambda_s = \exp(x_s\theta)$ mit x_s als vorhersagbarem Kovariablenprozeß und Parameter θ . Ebenso wenig läßt sich das Integral auf der rechten Seite im allgemeinen analytisch lösen, falls der Kovariablenprozeß zeitabhängig ist. Dies gelingt etwa für $x_s = s$. Man hat also in der Praxis zwei Probleme: Die Berechnung des Integrals sowie die Lösung der nichtlinearen Gleichung, letzteres dann allerdings mehrdimensional.
2. In der Praxis wird man für die Varianz einen Varianzschätzer verwenden, im obigen Fall etwa, indem θ durch $\hat{\theta}$ ersetzt wird.
3. Aufgrund der asymptotischen Normalverteilung lassen sich asymptotische Konfidenzbereiche angeben. Die Güte der Approximation ist abhängig davon, in welchem Umfang bereits im Endlichen die Asymptotik greift. In der Literatur Heyde (1997) wird deshalb alternativ vorgeschlagen, Konfidenzbereiche unmittelbar anhand der Schätzfunktionen G zu konstruieren. Man vermeidet dadurch den Fehler, den man bei der Taylorapproximation begeht, d.h. beim Übergang von $G_T(\theta)$ nach $\dot{G}_T(\theta)(\hat{\theta} - \theta)$. Die Konfidenzbereiche lassen sich dann über die χ^2 –Verteilung konstruieren:

$$\{\theta : G_T^t(\theta) \langle G(\theta) \rangle_T^{-1} G_T(\theta) \leq \chi_{p,\alpha}^2\}$$

3.5.3 Schätzfunktion und Lebensdauermodelle

Eine genaue Betrachtung erfolgt nach der Beschreibung der jeweiligen Lebensdauermodelle. Die bisherigen Ausführungen haben bereits aufgezeigt, dass die Inferenz teilweise identisch bzw. parallel zum Likelihoodfall verläuft, was auch im folgenden zu erwarten ist. Der Vorteil der Schätzfunktionen ist ihre Flexibilität, z.B. kann man sie kombinieren (Heyde, 1997) oder sie für gemischte Modelle anwenden (Heyde, 1997).

3.6 Bayesianische Inferenz für Lebensdauermodelle

Unter der Annahme, dass die Modellparameter Zufallsgrößen sind, lassen sich bayesianische Verfahren anwenden. Man spricht von der a priori Verteilung der zufälligen Parameter, wobei die a priori Verteilung entsprechend dem Vorwissen gewählt wird.

Die Modellierung folgt über die a posteriori Verteilung bzw. die a posteriori Dichte, die proportional dem Produkt aus der Likelihoodfunktion und der a priori Dichte ist. Zur Gleichheit anstelle der Proportionalität bedarf es noch einer Normierungskonstante, die im allgemeinen von den Beobachtungen abhängt und je nach Dimension der Parameter schwer zu berechnen ist. Einfache Modelle erhält man bei konjugierten Verteilungsfamilien, da die a priori und die a posteriori Dichte zur gleichen Verteilungsfamilie gehören. Bei diesem Ansatz entfällt die Berechnung der Normierungskonstante, die a priori Information kann zusammen mit der Likelihood in geschlossener Form verarbeitet werden. Dieser Zugang ist im Rahmen der Lebensdauermodelle zu eng, zum einen da die Einschränkung auf konjugierte Verteilungsfamilien sehr restriktiv ist, zum anderen da nicht klar ist, wie die zeitstetige Modellierung mit Kovariablen erfolgen soll. Es ist nahelegend die Struktur auszunutzen, die von der Zählprozeßtheorie geliefert wird. Damit stehen der Likelihood- und der Intensitätsprozeß zur Verfügung. Wurden bisher die unbekanntenen Größen $\theta \in \Theta$ des Intensitätsprozesses bzw. der Hazardrate als unbekannteste feste Parameter aufgefaßt, so betrachtet man sie nun als Zufallsgrößen mit a priori Verteilung ν auf Θ . Die Dichte $f_\pi(\theta)$ der a posteriori Verteilung π lautet (Aven, 1986)

$$f_\pi(\theta) = \frac{L_t(\theta) \cdot \nu(\theta)}{\int_{\Theta} L_t(\theta) d\nu(\theta)} \quad ,$$

wobei die Martingaleigenschaft von L_t wesentlich ist. Analog zur frequentistischen Inferenz lassen sich multiplikative Intensitätsmodelle oder Cox-Modelle (Hjort, 1986) formulieren. Es existieren asymptotische Aussagen (Rao, 1999) betreffend der Konsistenz des Schätzers sowie bezüglich der asymptotischen Normalität, man vergleiche dies mit dem Maximum-Likelihood Fall (Borgan, 1984).

3.6.1 Anmerkungen zu Bayesianischen Verfahren bei hochdimensionalen Modellen

Bei der Bayesianischen Inferenz spielt die a posteriori Verteilung eine wesentliche Rolle. Von Interesse sind dabei z.B. der a posteriori Modus, der Erwartungswert, die Varianz oder die Quartile. Sie geben auf Basis der a priori Verteilung und der Daten in Form der Likelihood Auskunft über die Verteilung der Parameter. Ein Problem dabei ist die Normierungskonstante, die man bei der Integration der Likelihood über die zufälligen Parameter mit der a priori Verteilung erhält, da der Wert meist nicht auf analytischem Weg gewonnen werden kann. Ein einfaches Verfahren, das zu approximativen Lösungen der Integration führt, ist die Laplace Approximation (MacKay, 2003). Sie wird kurz

bei den gemischten Modellen skizziert (s. u.). Alternativ können Gitterverfahren zur Bestimmung des Integrals verwendet werden. Neben den Vollgitterverfahren gibt es noch Dünngitterverfahren (Gerstner und Griebel, 1998), für die gelegentlich auch der Ausdruck Smolyak–Quadratur verwendet wird. Sie wurden eingeführt, da bei Vollgitterverfahren der Aufwand exponentiell mit der Anzahl der Dimensionen wächst. Dünngitterverfahren wurden bisher in der Statistik kaum benutzt (Ausnahme: (Garcke, Griebel und Thess, 2001)).

Üblicherweise werden Simulationsverfahren bei der bayesianischen Inferenz verwendet, da das Dimensionsproblem bei Monte Carlo Verfahren eher von untergeordneter Bedeutung ist. Entscheidend ist die Verwendung effizienter Simulationsverfahren. Außerdem sei als Vorgriff schon gesagt, daß bei den verwendeten Verfahren die Normierungskonstante keine Rolle spielt, da man Quotienten von a posteriori Verteilungen betrachtet.

3.6.2 Simulationsmethoden

Die Realisierung beliebig verteilter Zufallsgrößen erfolgt am Computer. Das Vorgehen erfolgt in zwei Schritten: Im ersten Schritt werden unabhängige auf $(0,1)$ gleichverteilte Zufallsgrößen realisiert. Das Interesse an den gleichverteilten Zufallsgrößen wird durch folgende Inversionsmethode motiviert: Sei $\xi \sim GV(0, 1)$ und F eine stetige und streng monotone Verteilungsfunktion, $\eta := F^{-1}(\xi)$; dann folgt wegen

$$\mathbb{P}(\eta \leq t) = \mathbb{P}(F^{-1}(\xi) \leq t) = \mathbb{P}(\xi \leq F(t)) = F(t) \quad .$$

Man kann also im zweiten Schritt mittels der Transformation F^{-1} Zufallsgrößen erzeugen, die gemäß F verteilt sind. Bei diesem Vorgehen wird idealisierend vorausgesetzt, dass unabhängige, auf $(0,1)$ gleichverteilte Zufallsgrößen, realisiert werden können. Dies ist am Computer nicht möglich, da dieser deterministisch arbeitet; man spricht daher auch von Pseudozufallszahlen. Eine Vielzahl von Zufallszahlengeneratoren beruht auf der linearen Kongruenzmethode. Die Zufallszahlen werden dabei rekursiv berechnet, die benutzte Abbildung hat die Eigenschaft, ergodisch zu sein – zumindest für reelle Zahlen (Sirjaev, 1988, S. 396). Da am Rechner nur mit endlicher Zahlengenauigkeit gearbeitet werden kann, geht diese Eigenschaft zwar verloren, die Zufallszahlengeneratoren gelten allerdings als gut untersucht. Eine Methode zum Ziehen von Zufallsgrößen, die ohne explizite Angabe von F bzw. F^{-1} funktioniert, deren Berechnung oft nur numerisch möglich ist, ist die sogenannte *Rejection Methode*. Bei hochdimensionalen Verteilungen wie sie für die Bayes'sche Modellierung benötigt werden, versagen allerdings diese klassischen Verfahren. Als Ausweg verwendet man Markov Chain Monte

Carlo-Verfahren.

3.6.3 MCMC Verfahren

Wenn kein exaktes Verfahren zum Ziehen von Zufallsgrößen ξ mit Verteilungsfunktion F und Verteilung ν existiert, kann versucht werden, Zufallsgrößen $(\xi_n)_{n \in \mathbb{N}}$, jeweils mit Verteilungsfunktion F_n , $n \in \mathbb{N}$, bzw. Verteilung ν_n , $n \in \mathbb{N}$ zu erzeugen, so dass $F_n(x)$ die gewünschte Verteilungsfunktion $F(x)$ approximiert. Beschränkt man sich dabei auf Folgen $(\xi_n)_{n \in \mathbb{N}}$, die die Markov Eigenschaft erfüllen, d.h.

$$\mathbb{P}(\xi_{n+1} \in B \mid \xi_1 = x_1, \dots, \xi_n = x_n) = \mathbb{P}(\xi_{n+1} \in B \mid \xi_n = x_n) \quad , \quad (3.34)$$

so existiert eine gut entwickelte Theorie, die im folgenden kurz dargestellt wird.

Markovketten

Markovketten $(\xi_n)_{n \in \mathbb{N}}$ mit Werten in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ heißen homogen oder stationär, falls für die Übergangswahrscheinlichkeiten oder Markovkerne

$$\mathbb{P}(\xi_1 \in B \mid \xi_0 = x) = \mathbb{P}(\xi_{n+1} \in B \mid \xi_n = x), \quad B \in \mathcal{B}, \quad n \in \mathbb{N} \quad (3.35)$$

gilt. Stillschweigend sei vorausgesetzt, dass die Übergangswahrscheinlichkeiten jeweils eine Version einer regulären bedingten Verteilung $P(A \mid x)$ sind, deren Existenz aus den topologischen Eigenschaften des Borel'schen Raumes $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ folgt (Sirjaev, 1988, S. 237). Die Übergangswahrscheinlichkeiten bestimmen zusammen mit einer Startverteilung $\mathbb{P}(A) = \mathbb{P}(\xi_0 = A)$ die endlich dimensionalen Verteilungen der Markovkette (Breiman, 1992, S. 130). Die Übergangswahrscheinlichkeit $P(\xi_1 \in B \mid \xi_0 = x)$ kann als diejenige Wahrscheinlichkeit interpretiert werden, in einem Schritt vom Startwert x_0 in die Menge B zu gelangen. Allgemeiner kann die Menge B in n Schritten (zum ersten mal) erreicht werden und es gilt

$$P^n(\xi_n \in B \mid \xi_0 = x) \quad ,$$

wobei die Übergangswahrscheinlichkeit eine Version der regulären bedingten Verteilung $P^n(A \mid x)$ ist. Es existiere nun ein Wahrscheinlichkeitsmaß $\bar{\pi}$, so dass die Gleichung

$$\bar{\pi}(A) = \int P(A \mid x) \bar{\pi}(dx), \quad A \in \mathcal{B} \quad (\bar{\pi} = \mathbb{P}\bar{\pi}) \quad (3.36)$$

erfüllt ist. Maße $\bar{\pi}$ mit dieser Eigenschaft nennt man invariant für $P(A | x)$. Für sie gilt auch (mittels vollständiger Induktion)

$$\bar{\pi}(A) = \int P^n(A | x) \bar{\pi}(dx), \quad A \in \mathcal{B} \quad (\bar{\pi} = P^n \bar{\pi}) \quad . \quad (3.37)$$

Invariante Maße erscheinen als geeignete ‘Kandidaten’ für die Grenzverteilung, wobei noch geklärt werden muß, welche Voraussetzungen $P(A | x)$ bzw. $P^n(A | x)$ zusätzlich erfüllen müssen, so dass für beliebige Startverteilungen ν

$$P^n \nu \rightarrow \bar{\pi} \quad \text{für } n \rightarrow \infty \quad \bar{\pi} \text{ fast sicher}$$

gilt. In der Literatur findet man eine Vielzahl theoretischer Aussagen dazu, siehe z. B. (Tierney, 1994). Es erscheint allerdings zweckmäßiger, sich auf die praktischen Probleme zu konzentrieren. Abgesehen davon sind in vielen Fällen die weiteren Voraussetzungen (Irreduzibilität und Aperiodizität) als erfüllt anzusehen. Sie (Irreduzibilität und Aperiodizität) sollen ein gutes Mischungsverhalten sicherstellen, d.h. dass man jede meßbare Menge, die keine Nullmenge ist, mit positiver Wahrscheinlichkeit in endlich vielen Schritten erreicht. Zu überprüfen ist die Invarianz, worauf allerdings meistens verzichtet wird; statt dessen prüft man die (leichter zu beweisende) Zeitumkehrbedingung (Tierney, 1994), die hinreichend für die Invarianz ist. Von großer praktischer Bedeutung wären natürlich Aussagen über die Konvergenzgeschwindigkeit. Für Markovketten, deren Zustandsraum eine kompakte Menge ist, gibt es derartige Aussagen (Winkler, 1995, S. 87). Sie lassen sich jedoch nicht ohne weiteres auf nichtkompakte Zustandsräume verallgemeinern. Ein wichtiges Hilfsmittel in diesem Fall sind die Lyapunov–Funktionen (Tong, 1990) bzw. (Lasota und Mackey, 1985).

Metropolis–Hastings Algorithmus

Unter der Annahme, dass die invariante Verteilung zu jeder Startverteilung gegen die Grenzverteilung konvergiert, sucht man eine Übergangswahrscheinlichkeit, so dass die dazu invariante Verteilung π die gesuchte Verteilung ist, von der man Zufallsgrößen simulieren möchte. Bei der Wahl des Übergangskerns kann das Metropolis–Hastings Verfahren verwendet werden. Die zu simulierende Verteilung sei π , mit positiver Dichte $f(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^d$. Bei bayesianischen Verfahren handelt es sich oft um die a posteriori Verteilung der zufälligen Parameter. Q sei ein Markovkern mit positiver Dichte $q(\mathbf{x} | \mathbf{x}')$, $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^d$. Der Metropolis–Hastings Algorithmus besteht nun aus den folgenden zwei Schritten:

1. Vorschlagsschritt: Realisiere ein \mathbf{x}' gemäß Q bei gegebenen \mathbf{x}
2. Akzeptanzschritt: $a(\mathbf{x}', \mathbf{x}) = 1 \wedge \frac{f(\mathbf{x}') \cdot q(\mathbf{x} | \mathbf{x}')}{f(\mathbf{x}) \cdot q(\mathbf{x}' | \mathbf{x})}$

Mit Wahrscheinlichkeit $a(\mathbf{x}', \mathbf{x})$ wird \mathbf{x}' als neuer Wert akzeptiert, anderenfalls behält man \mathbf{x} bei. Einen Beweis für die Invarianz des Markovkerns findet man in Tierney (1994). Es gibt eine Vielzahl von Varianten des Metropolis–Hastings Algorithmus. Anstelle dem Vorgehen, alle Koordinaten in einem Schritt zu ändern, kann dies auch komponentenweise vorgenommen werden. Zu beachten ist hierbei, dass die Irreduzibilität verloren gehen kann, vgl. Besag, Green, Higdon und Mengersen (1995). Entscheidend für ein komponentenweises Vorgehen ist die Akzeptanzrate, die im allgemeinen sehr gering ist, falls alle Koordinaten in einem Schritt neu berechnet werden. Andererseits wird die Konvergenzgeschwindigkeit negativ beeinflusst, falls man jeweils nur eine Koordinate in einem Schritt neu berechnet, da man –bildlich gesprochen– an den alten Werten ‘kleben’ bleibt. Für die Wahl von Q existieren ebenfalls verschiedene Vorschläge, wobei der Varianz der Vorschlagsdichte eine besondere Bedeutung zukommt. In der Literatur findet man z.B.

- (a) independence sampler (Tierney, 1994):
 $q(\mathbf{x}' | \mathbf{x}) = q(\mathbf{x}')$,
- (b) random walk (Metropolis, Rosenbluth, Rosenbluth, Teller und Teller, 1953):
 $\mathbf{x}' = \mathbf{x} + \epsilon, \quad \epsilon \sim F$,
- (c) Langevin Methode (Liu, 2001): ähnlich dem random walk, zusätzlich wird der logarithmierte Gradient der Dichte der gewünschten Verteilung mitverwendet,
- (d) Gibbs sampler (Geman und Geman, 1984): Neue Werte werden komponentenweise gezogen mit den exakten bedingten Verteilungen.

Die Qualität des Metropolis–Hastings Verfahrens wird unter anderem anhand der Samplingpfade beurteilt. Sie liefern Hinweise auf die Konvergenz bzw. Konvergenzgeschwindigkeit. Eine koordinatenweise hohe Autokorrelation des Parametervektors bzw. eine hohe Korrelation zwischen den Koordinaten indiziert ein schlechtes Konvergenzverhalten. Man versucht dem vorzubeugen, indem man blockweise neue Werte vorschlägt (Knorr-Held, 1999) und die Beobachtungen ‘ausdünt’, d.h. man verwendet nur jeweils die m -ten Werte. Als Alternative kann man Hybride Monte Carlo Verfahren verwenden. Neben diesen praktisch orientierten Hinweisen existieren auch theoretische Aussagen. So spielt für das Mischungsverhalten des Metropolis–Hastings Verfahrens der (betragsmäßig) zweitgrößte Eigenwert des Übergangskerns eine Rolle (Winkler, 1995). Im Fall

von endlichen Zustandsräumen muß dazu nur der entsprechende Eigenwert der Übergangsmatrix betrachtet werden. Im allgemeinen bedarf es dazu numerischer Techniken, siehe etwa Mason und Handscomb (2003) oder Fischer (2003).

3.6.4 Hybride Monte Carlo Verfahren

Der Begriff Hybride Monte Carlo (HMC) Verfahren wird nicht ganz einheitlich verwendet; eine Einführung findet man etwa in Liu (2001). Sie werden manchmal auch als Hamilton Monte Carlo Verfahren bezeichnet. Die Hybriden Monte Carlo Verfahren haben ihren Ursprung in der Molekulardynamik. Man modelliert den Zustand eines geschlossenen Systems, indem man das Energieniveau einer Vielzahl von Teilchen im Zeitverlauf betrachtet. Jede Konfiguration der Teilchen entspricht genau einem Energieniveau. Die Gesamtenergie bleibt idealisierend konstant für alle Teilchen über die Zeit betrachtet. Man versucht durch die Beobachtungen im zeitlichen Verlauf Rückschlüsse auf das System zu ziehen. Hier tauchen bereits erste Parallelen zur Wahrscheinlichkeitstheorie auf, nämlich zu ergodischen Prozessen, da die (räumliche) Mittelung über alle Teilchenkonfigurationen ersetzt wird durch eine zeitliche Mittelung. Die Modellierung des zeitlichen Verlaufes erfolgt mittels Differentialgleichungen unter Beachtung des Hamilton'schen Prinzips (Forster, 1993, S. 93), (Courant und Hilbert, 1993, S. 210), sich von einem Zustand mit minimaler Energie zum nächsten Zustand zu bewegen. In der praktischen Umsetzung werden die Differentialgleichungen durch Differenzgleichungen approximiert. Es wurde vorgeschlagen (Duane, Kennedy, Pendelton und Roweth, 1987), mittels dieser Differenzgleichungen neue Werte zu generieren und in einem Akzeptanzschritt über ihre Annahme oder Ablehnung zu entscheiden (s. u.). Hybride Monte Carlo Verfahren zeichnen sich im Gegensatz zu anderen Markov Chain Monte Carlo Verfahren also dadurch aus, dass man auf eine Vorschlagsdichte verzichtet und neue Werte aufgrund von physikalischen Prinzipien ermittelt. Den Akzeptanzschritt könnte man auch als Fehlerkorrektur der zeitlich diskretisierten Molekulardynamik betrachten. Die HMC Methoden haben in der Statistik bisher wenig Anklang gefunden, abgesehen von Neal (1993), Gustafson, MacNab und Wen (2004), Ishwaran (2000) und Heumann (2003). Ordnet man jedem Teilchen i , $i = 1, \dots, d$, $d \in \mathbb{N}$, seine (der Einfachheit halber eindimensionalen) Raumkoordinaten x_i und seinen Impuls p_i zu, so kann man das Energieniveau $H(\mathbf{x}, \mathbf{p})$ des Systems zu einem festen Zeitpunkt in den Vektoren $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$ und $\mathbf{p} = (p_1, \dots, p_d) \in \mathbb{R}^d$ beschreiben durch

$$H(\mathbf{x}, \mathbf{p}) = U(\mathbf{x}) + k(\mathbf{p}) \quad ,$$

wobei mit $U(\mathbf{x})$ die potentielle Energie und mit $k(\mathbf{p})$ die kinetische Energie bezeichnet wird. Führt man noch die Zeit t ein, so erhält man die Raumkoordinaten und den Impuls als Funktionen in der Zeit. Man schreibt deshalb auch $\mathbf{x}(t)$ bzw. $\mathbf{p}(t)$. Da die Gesamtenergie konstant ist, gleichen sich die Veränderungen der kinetischen und der potentiellen Energie über die Zeit für alle Teilchen aus. Unter Verwendung des Hamilton'schen Prinzips lassen sich die Veränderungen der Zustände über die Zeit beschreiben durch die Hamilton'schen Gleichungen

$$\frac{\partial}{\partial t}\mathbf{x}(t) = \frac{\partial}{\partial \mathbf{p}}H(\mathbf{x}, \mathbf{p}) \quad (3.38)$$

und

$$\frac{\partial}{\partial t}\mathbf{p}(t) = -\frac{\partial}{\partial \mathbf{x}}H(\mathbf{x}, \mathbf{p}) \quad . \quad (3.39)$$

Wählt man für die kinetische Energie $k(\mathbf{p})$

$$k(\mathbf{p}) = \sum_{i=1}^d \frac{p_i^2}{2},$$

so vereinfachen sich die Ableitungen. Der sogenannte Leapfrog-Algorithmus basiert auf dem Diskretisierungsschema ($\delta > 0$, $i = 1, \dots, d$)

$$x_i(t + \delta t) - x_i(t) = \delta t \cdot \left(p_i \left(t + \frac{\delta t}{2} \right) \right) \quad (3.40)$$

beziehungsweise

$$p_i \left(t + \frac{\delta t}{2} \right) - p_i \left(t - \frac{\delta t}{2} \right) = \delta t \cdot \frac{\partial}{\partial x} H|_t \quad . \quad (3.41)$$

Die Approximationen haben noch weitere Eigenschaften (Liu, 2001), die dazu führen, dass die Zeitumkehrbedingung erfüllt wird. Neue Werte \mathbf{x}' und \mathbf{p}' werden mittels des Leapfrog-Algorithmus generiert, um dann bei dem Akzeptanzschritt

$$a(\mathbf{x}', \mathbf{x}) = 1 \wedge \exp(-H(\mathbf{x}', \mathbf{p}') + H(\mathbf{x}, \mathbf{p})) \quad (3.42)$$

zu entscheiden, ob sie angenommen werden. Für bayesianische Simulationen ersetzt man beim HMC die Teilchen durch den Parametervektor, die a posteriori Verteilung

$\pi(\mathbf{x})$ schreibt man als $\exp(-U(\mathbf{x}))$. Zusätzlich erhält man noch einen Vektor unabhängiger standardnormalverteilter Zufallsgrößen mit Dichte $\exp(-k(\mathbf{p}))$, die allerdings nur für den Leapfrog Algorithmus und den Akzeptanzschritt benötigt werden. Betrachtet man in diesem Kontext nochmal den Leapfrog-Algorithmus, so erkennt man, dass durch $\frac{\partial}{\partial \mathbf{x}} H$ Informationen des Gradienten der logarithmierten a posteriori Dichte verarbeitet werden. Durch den Vektor der Impulsvariablen schafft man sich über die Gesamtenergie des Systems einen zusätzlichen Spielraum für die Parameter, d.h. sie dürfen sich freier voneinander bewegen. Dazu läßt man den Leapfrog-Algorithmus über mehrere Zeitschritte laufen. Da die Impulsvariablen auch im Akzeptanzschritt ihr Gewicht erhalten, wird verhindert, dass nur solche Parameterwerte gezogen werden, die sehr nahe am (unbekannten) Maximum (einer Lösung der Differentialgleichung (Forster, 1993, S. 154)) der a posteriori Dichte liegen, was die Verwendung des Gradienten der logarithmierten a posteriori Dichte vermuten lassen würde. Die Idee dieses Algorithmus ist es also, mittels Hilfsvariablen ein höheres Energieniveau zu erreichen, so dass die vorgeschlagenen und akzeptierten Werte den Bereich der a posteriori Verteilung gut abdecken und sich dabei unabhängig voneinander möglichst frei bewegen können, womit man die bei den MCMC Verfahren sonst übliche hohe Korrelation bzw. Autokorrelation vermeiden möchte. Dafür nimmt man den höheren Aufwand in Kauf. So muß man in jedem Leapfrog-Schritt den Gradienten neu berechnen. Ausgehend von einem Parametervektor $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$ sieht ein Schritt des HMC Algorithmus wie folgt aus:

1. Realisiere einen Impulsvektor $\mathbf{p} = (p_1, \dots, p_d) \in \mathbb{R}^d$ mit i.i.d. standardnormalverteilten Zufallsgrößen als Komponenten;
2. Ausgehend von (\mathbf{x}, \mathbf{p}) : Generiere neue Werte $(\mathbf{x}', \mathbf{p}')$ mit Schrittweite δ und einer Schrittzahl L ;
3. Akzeptiere den neuen Wert \mathbf{x}' mit Wahrscheinlichkeit

$$a(\mathbf{x}', \mathbf{x}) = 1 \wedge \exp(-H(\mathbf{x}', \mathbf{p}') + H(\mathbf{x}, \mathbf{p})). \quad (3.43)$$

Der Vektor \mathbf{p}' wird außer im Akzeptanzschritt nicht weiter benötigt. Es folgen einige kritische Anmerkungen sowie Hinweise.:

- Durch die Schrittweite δ sowie die Anzahl der Schritte L erhält man zwei Einstellgrößen des Leapfrog-Algorithmus. Im allgemeinen ist der Approximierungsfehler von Differenzgleichungen um so größer, je größer das Gitter, in diesem Fall δ , ist. Große Werte sind demnach ungünstig, kleine Werte hingegen entspre-

chend ‘kostspielig’ (mit größeren Werten von L), da in jedem Schritt der Gradient neu berechnet werden muß. Die Qualität des Leapfrog Algorithmus hängt auch vom Gradienten der logarithmierten a posteriori Dichte ab. Da man mit einem für alle Parameter konstanten zeitlichen Gitter δ arbeitet, kann es zu unerwünschten Verzerrungen kommen, wenn die Werte des Gradienten für einzelne Koordinaten stark voneinander abweichen.

- Wählt man für $L = 1$, so erhält man ein MCMC Verfahren, bei dem die vorgeschlagenen Werte nach dem Langevin Verfahren bestimmt werden.
- Bei dem hier vorgestellten HMC Verfahren wurde simultan für alle Parameter ein neuer Wert vorgeschlagen – dies läßt sich sicherlich jedoch auch blockweise gestalten, z.B. für Hyperparameter.
- Es wurde bisher angenommen, dass die Impulsvariablen am Anfang alle unabhängig identisch standardnormalverteilt sind. Als Erweiterung kann man auch Gewichte m_i zulassen, d.h. man erhält unterschiedliche Massen (Varianzen) für die Impulsvariablen. Die kinetische Energie schreibt sich dann als

$$k(\mathbf{p}) = \mathbf{p}' M^{-1} \mathbf{p} \quad , \quad (3.44)$$

wobei M eine positiv definite, symmetrische Matrix ist. Eine Rolle könnte dies spielen, falls der Gradient komponentenweise stark unterschiedliche Werte annimmt. Eventuell kann man durch eine geschickte Wahl der Varianzen die Skalierung des Gradienten ausgleichen. In Choo (2000) wird vorgeschlagen, die Hessematrix der zweiten Ableitungen von $U(\mathbf{x})$ für M zu verwenden oder vereinfachend nur die Hauptdiagonalelemente m_i . Das Differenzenschema (3.40) schreibt sich nun anstelle von δ mit $\delta_i = \delta \frac{1}{m_i}$. Dies soll zu einer besseren Skalierung führen. Der Nachteil dabei ist, dass man pro Leapfrog Durchlauf zumindest approximativ die Hesse Matrix berechnen muß. Vorschläge für M finden sich in Abschnitt (4.6).

- In der Literatur Liu (2001) wird noch vorgeschlagen, anstelle von $U(\mathbf{x})$ mit $U'(\mathbf{x}) = U(\mathbf{x})/c$, $c > 1$ im Leapfrog–Algorithmus zu arbeiten, falls die Dichte viele Barrieren hat. Es sei dahingestellt, inwiefern man mit einem Parameter derartige Barrieren überwinden kann (die man im allgemeinen nicht kennt) und wie man den Parameter dabei zu wählen hat. Die Idee rührt vermutlich vom Simulated Annealing.
- In MacKay (2003, S. 388) findet man einen Vergleich (anhand der Samplingpfade) zum random walk Metropolis Algorithmus. Es handelt sich um ein künstli-

ches Beispiel, einer zweidimensionalen Normalverteilung, wobei die Komponenten sehr hoch miteinander korreliert sind.

- Ein Verfahren, das auf die Verwendung von Ableitungen verzichtet, geht auf Skilling zurück (MacKay, 2003), leider ist es nicht für hochdimensionale Probleme geeignet.

4 Lebensdauermodelle mit flexibler Hazardrate

Dieses Kapitel befaßt sich mit der Modellierung des Intensitätsprozesses. Wenn von der Hazardrate gesprochen wird, ist der zu modellierende Teil des Intensitätsprozesses gemeint. Zu Beginn wird ein allgemeines Modell aufgestellt, es folgen technische Details über Splines und penalisierte Splines sowie ein kurzer Abschnitt über gemischte Modelle. Im Anschluß wird die Verwendung penalisierter Splines bei gemischten Modellen erörtert. Eine Anwendung dazu findet sich in Kneib und Fahrmeir (2004). Bei diesem Ansatz benötigt man nicht die sonst üblichen rechenintensiven Verfahren der Kreuzvalidierung zur Berechnung der Glättungsparameter der penalisierten Splines.

4.1 Modell einer allgemeinen Hazardrate

Im Hinblick auf eine stochastische Modellierung in einem für die Praxis hinreichend allgemeinen Rahmen werden flexible Ansätze der Modellierung des Intensitätsprozesses

$$\lambda_{it} = Y_{it} \cdot h_{it} \quad i = 1, \dots, n \quad (4.1)$$

untersucht. Die Modellierung erfolgt zunächst ohne Berücksichtigung der Zensierung. Betrachtet man Verallgemeinerungen, die dem Cox-Modell angelehnt sind, ließe sich $h_{it} = h_{it}(\theta)$ aufspalten in

$$h_{it}(\theta) = h_t(\gamma) \exp(\eta_{it}) \quad , \quad (4.2)$$

wobei die positive Funktion $h_t(\gamma)$ als (unendlich- dimensionaler) Störparameter aufgefaßt wird. Alternativ schreibt man

$$h_{it} = \exp(\eta_{it} + \log(h_{it})) \quad . \quad (4.3)$$

Die Wahl der Exponentialfunktion für h_{it} ist vorteilhaft, da diese strikt positiv ist. Die eigentliche Modellierung geschieht mittels Prädiktoren η_{it} , wobei der Index i sich auf das i -te Individuum bezieht und der Index t auf den Zeitpunkt t . Der Prädiktor für das

Individuum i lautet

$$\eta_{it}(\theta) = \mathbf{x}_i^t \boldsymbol{\beta} + \sum_{j=1}^p f_j(x_{2ji}) + \sum_{j=1}^q g_j(t)x_{3ji} \quad . \quad (4.4)$$

Der Vektor θ enthält alle Parameter. Im Fall von f_j spricht man von nichtlinearen, bei g_j von zeitvariierenden Effekten, die mittels ‘glatter’ Funktionen modelliert werden. In der Literatur (z.B. Hastie und Tibshirani (1993)) werden häufig Splines für die Modellierung der glatten Funktionen verwendet. Eine Anwendung im geoadditiven Survivalbereich mit bayesianischen Methoden findet sich in Hennerfeind, Brezger und Fahrmeir (2005). Splines ergeben sich als optimale Schätzung für Funktionen in Sobolevräumen bezüglich einer Halbnorm (siehe unten Formel (4.7)). Die Splinefunktionen gestatten eine Darstellung der geschätzten Funktionen \hat{f}_j und \hat{g}_j mit jeweils endlich vielen Parametern α_{lj} bzw. deren Schätzungen $\hat{\alpha}_{lj}$ und Basisfunktionen B_{lj} , d.h.

$$\hat{f}_j(x) = \sum_{l=1}^{K(f_j)} \hat{\alpha}_{lf_j} B_{lj}(x) \quad (4.5)$$

beziehungsweise

$$\hat{g}_j(t) = \sum_{l=1}^{K(g_j)} \hat{\alpha}_{lg_j} B_{lj}(t) \quad . \quad (4.6)$$

Bei den Termen $K(f_j)$ bzw. $K(g_j)$ handelt es sich lediglich um kompliziert aussehende Indizes, die die Obergrenze von l anzeigen. Die zur Darstellung der glatten Funktionen zu schätzenden Parameter werden dabei zum Teil als zufällig betrachtet. Das Modell beinhaltet also feste und zufällige Parameter (alternativ: Effekte). Solche Modelle werden auch als gemischte Modelle bezeichnet. Im folgenden werden einige Aussagen über Splines zusammengestellt, allerdings nur soweit, wie es ihre Anwendung für die Modellierung der Hazardrate erfordert. Sinngemäß gilt dies auch für die gemischten Modelle.

4.2 Splinefunktionen

Wird als Ausgangspunkt das Interpolationsproblem von Daten gewählt, so kann man die Splines als Funktionen betrachten, die den Interpolationspolynomen in dem Sinne überlegen sind, dass ihr Grad nicht von der Anzahl der Daten abhängt. Zu den Werten

$\mathbf{x} = (x_0, \dots, x_n)$ liegen Beobachtungen $\mathbf{y} = (y_0, \dots, y_n)$ vor. Partitioniert man den Datenbereich $\Delta = \{x_0 < x_1 < \dots < x_n\}$ in disjunkte Intervalle, wobei die Randpunkte der Intervalle als Knotenpunkte bezeichnet werden, so werden die Splines als stückweise Polynome definiert, so dass sie an den inneren Knotenpunkten gewisse Differenzierbarkeitseigenschaften aufweisen. Unter der Annahme, dass sich der Einfluß der Kovariablen auf die abhängigen Variablen funktional durch eine glatte Funktion darstellen läßt, versucht man, diesen mittels glatter Funktionen zu modellieren. Beschränkt man sich dabei auf Funktionen aus dem Sobolevraum $K_2^m[x_0, x_n]$, d.h. auf die Menge aller Funktionen g auf $[x_0, x_n]$, für die die $(m - 1)$ -te Ableitung $g^{(m-1)}$ absolutstetig ist und deren m -te Ableitung stückweise stetig und quadratintegrierbar ist, so läßt sich zeigen, dass das Kriterium

$$\psi(g) = \frac{1}{n} \sum_{i=0}^n (y_i - g(x_{ij}))^2 + \lambda \int_{x_0}^{x_n} (|g^{(m)}(x)|)^2 dx \quad , \lambda > 0 \quad , \quad (4.7)$$

sein eindeutiges Minimum über alle Funktionen des Sobolevraumes annimmt, falls g eine Splinefunktion ist. Zu dem Kriterium (4.7) sei noch erwähnt, dass der erste Term die Anpassung an die Daten steuert, der zweite Term die Glattheit der Funktion. Wählt man $m = 2$, so wird durch $|g^{(2)}|$ die Krümmung an der Stelle x approximiert und $\int_{x_0}^{x_n} |g^{(2)}|^2 dx$ kann als Maß der Gesamtkrümmung interpretiert werden, deren Einfluß durch λ gesteuert wird. Man bezeichnet daher λ auch als Glattheitsparameter. Im Gegensatz zu Interpolationspolynomen läßt sich für interpolierende Splinefunktionen unter gewissen Bedingungen zeigen, dass sie gegen die zu interpolierende Funktion konvergieren, falls das Gitter Δ hinreichend fein gewählt wird (Stoer, 1999, S. 109). Zur Darstellung einer Splinefunktion können unterschiedliche Systeme von Basisfunktionen verwendet werden, z.B. B-Splines (Stoer, 1999) oder einseitige Splines (Hämmerlin und Hoffmann, 1994) (englische Bezeichnung: Truncated Power Splines). Im Hinblick auf die numerische Stabilität ist die Verwendung von B-Splines als Basisfunktionen vorteilhaft (Deuffhard und Hohmann, 1991). Demgegenüber wird sich zeigen, dass die einseitigen Splines aus modelltheoretischer Sicht einfacher zu schätzen und interpretieren sind. Daher werden beide Klassen von Basisfunktionen dargestellt.

4.2.1 Definition und Basisfunktionen

Eine Funktion f heißt Splinefunktion der Ordnung $r + 1$ und vom Grad r zur Partition Δ , falls sie auf jedem Teilintervall von Δ mit einem Polynom vom Grad $\leq r$ übereinstimmen und an den inneren Knotenpunkten mindestens $(r - 1)$ -mal differenzierbar ist.

Es lässt sich zeigen, dass $S_{\Delta,r}$, d.h. die Menge aller Splinefunktionen vom Grad r zur Partition Δ der Feinheit n ein Vektorraum der Dimension $n + r$ ist. Die Elemente f des Vektorraums lassen sich darstellen durch

$$f(x) = \sum_{i=-r}^{n-1} \alpha_i B_{i,r+1}(x)$$

mit $n + r$ linear unabhängigen Basisfunktionen $B_{i,r+1}(x)$, die B -Splines genannt werden. Sie lassen sich rekursiv nach folgendem Schema berechnen:

$$t_{-r} := \dots := t_0 := x_0 < t_1 := x_1 < \dots < t_n := x_n \quad (4.8)$$

und $B_{i,1}(x) = \mathbb{1}_{\{t_i \leq x \leq t_{i+1}\}}$

$$B_{i,r}(x) = \frac{x - t_i}{t_{i+r-1} - t_i} B_{i,r-1}(x) + \frac{t_r - x}{t_{i+r} - t_{i+1}} B_{i+1,r-1}(x) \quad . \quad (4.9)$$

Der Träger der B -Splines $B_{i,r}(x)$ ist das Intervall (t_i, t_{i+r}) . Dies und die Tatsache, dass bei der Rekursion nur positive Zahlen addiert werden, verleiht den B -Splines eine hohe numerische Stabilität (vgl. Stoer (1999, S. 117)).

Die alternative Darstellung erfolgt mit einseitigen Splines (Hämmerlin und Hoffmann, 1994). Mit $(z)_+ = \max(0, z)$ schreibt sich dabei die Splinefunktion f als

$$f(x) = \sum_{l=0}^r \beta_l x^l + \sum_{k=1}^n \alpha_k (x - x_k)_+^r \quad (4.10)$$

mit den Monomen $1, \dots, x^r$ und den abgebrochenen Potenzen $(x - x_k)_+^r$, wobei es sich bei x_k um die Gitterpunkte von Δ handelt.

4.2.2 Penalisierte Splines

Das Kriterium (4.7) bezieht sich auf den Kleinste-Quadrate Schätzer. Es lässt sich erweitern, indem man die Quadratsumme durch die Loglikelihoodfunktion $l(\alpha, y)$ ersetzt. Beim Kriterium (4.7) wird davon ausgegangen, dass jede Beobachtung auf einem Knotenpunkt liegt. Diese Annahme wird im folgenden aufgegeben. In der Literatur (Eilers und Marx, 1996) wurde eine Vereinfachung des Penalisierungsterms vorgeschlagen, die

auf den Differenzen der Koeffizienten der B -Splines beruht,

$$\psi(\alpha) = l(\alpha, y) + \lambda \sum_{i=k+1}^n (\Delta^k \alpha_i)^2 \quad , \quad (4.11)$$

mit $\Delta^k \alpha_i = \Delta^{k-1}(\Delta \alpha_i)$, $\Delta \alpha_i = \alpha_i - \alpha_{i-1}$ als Differenzenoperator der Ordnung k . Dabei geht die Differenz benachbarter Basiskoeffizienten in den Strafterm ein. Man beachte, dass $\int B_{i,r}(x) dx = 1$ gilt. Die Standardwahl für k bei kubischen Splines ist $k = 2$. In diesem Fall läßt sich der vereinfachte Penalisierungsterm auch dadurch motivieren, dass man die zweiten Ableitungen diskret approximiert durch die zweiten Differenzen, wobei man sich auf die Koeffizienten der B -Splines beschränkt. Für die Wahl der Knoten gibt es keine genauen theoretischen Vorgaben, sie richten sich eher nach Gesichtspunkten des praktischen Modellierens. So möchte man etwa eine zu große Parameteranzahl, verursacht durch eine zu hohe Knotenanzahl, vermeiden. Ist die Knotenanzahl hingegen zu klein, wird man die unbekannte glatte Funktion nur ungenügend approximieren. Im Gegensatz zu Regressionssplines scheint die Anzahl der Knoten nicht so wichtig zu sein, da der Strafterm für eine gewisse Stabilität sorgt (Gray, 1992, S. 943). Wählt man die Knotenpunkte datenabhängig, so kommen empirische Quantile in Frage, man vergleiche etwa mit (Cai et al., 2002). Die Schätzung von λ erfolgt z.B. mit der Kreuzvalidierung (Hastie und Tibshirani, 1990) oder mit dem Kriterium von Akaike (Eilers und Marx, 1996). Eine weitere und weniger rechenintensive Alternative zur Bestimmung des Glättungsparameters λ beruht auf einem bayesianischen Ansatz mit gemischten Modellen.

4.3 Gemischte Modelle und Zufallseffekte

4.3.1 Definition der Gemischten Modelle

Man spricht von gemischten Modellen, wenn es sich bei den zu schätzenden Parametern um Zufallsgrößen handelt. Sie haben sich als sehr vielseitig erwiesen (vgl. etwa McCulloch und Searle (2001)). Ursprünglich wurden sie entwickelt für unbalancierte Versuchspläne (Laird und Ware, 1982), angewendet werden sie auch bei Longitudinaldaten. Sie eignen sich insbesondere für die Modellierung von geclusterten Daten zur Berücksichtigung von Abhängigkeiten, aber auch zur Berücksichtigung einer erhöhten Variation der Daten. Ein weiterer Aspekt ist die sparsame Parametrisierung, weil die Zufallseffekte durch ihre Verteilung beschrieben werden. In der Regel gibt es bei gemischten

Modellen feste und zufällige Parameter. Für $i = 1, \dots, n$ Beobachtungen bezeichne Y_i die Responsevariablen und $\mathbf{x}_i, \mathbf{z}_i$ die Kovariablenvektoren. Der korrespondierende Parametervektor $\boldsymbol{\beta}$ bezieht sich auf die festen Effekte und \mathbf{u} auf die zufälligen Effekte. Für generalisierte gemischte Modelle erhält man unter der Annahme, dass die bedingten Verteilungen $f_{Y_i|u_i}$ von Y_i gegeben u_i unabhängig sind und zu einer Exponentialfamilie gehören, folgenden linearen Prädiktor

$$\eta_i = \mathbf{x}_i^t \boldsymbol{\beta} + \mathbf{z}_i^t \mathbf{u} \quad . \quad (4.12)$$

Die bedingte Dichte für \mathbf{Y} lautet

$$f_{\mathbf{Y}|\mathbf{u}}(\mathbf{y}|\mathbf{u}, \mathbf{X}, \mathbf{Z}, \boldsymbol{\beta}) = \prod_{i=1}^n f_{Y_i|u_i}(y_i|\mathbf{u}, \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\beta}) \quad . \quad (4.13)$$

Sei $f_{\mathbf{u}}(\mathbf{u}|D)$ die Dichte von \mathbf{u} mit Parameter D . Da die Realisierungen der zufälligen Parameter nicht beobachtet werden können, kann man, außer bei linearen gemischten Modellen, nicht mit der vollständigen Likelihoodfunktion arbeiten. Integriert man über die zufälligen Effekte, so erhält man die marginale Likelihoodfunktion

$$L(\boldsymbol{\beta}, D|\mathbf{Y}) = \int \prod_{i=1}^n f_{Y_i|u_i}(y_i|\mathbf{u}, \mathbf{x}_i, \mathbf{z}_i, \boldsymbol{\beta}) f_{\mathbf{u}}(\mathbf{u}|D) d\mathbf{u} \quad . \quad (4.14)$$

Dies gelingt meist nicht analytisch; einen Ausweg bietet die sogenannte Laplace–Approximation (Tierney und Kadane, 1986), siehe auch (Liu und Pierce, 1994) für einen Zusammenhang zur Gauß Hermite Integration. Unter der Annahme, dass die Zufallseffekte einer multivariaten Gaußverteilung genügen, d.h.

$$f_{\mathbf{u}}(\mathbf{u}|D) \propto |D|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} \mathbf{u}^t D^{-1} \mathbf{u} \right] \quad (4.15)$$

und mit der Abkürzung

$$\exp(-\boldsymbol{\kappa}(\mathbf{u}, \boldsymbol{\beta}, D)) = f_{\mathbf{Y}|\mathbf{u}}(\mathbf{y}|\mathbf{u}, \mathbf{X}, \mathbf{Z}, \boldsymbol{\beta}) \cdot \exp \left[-\frac{1}{2} \mathbf{u}^t D^{-1} \mathbf{u} \right]$$

folgt, dass sich (4.14) schreiben läßt als

$$L(\boldsymbol{\beta}, D|\mathbf{Y}) \propto |D|^{-\frac{1}{2}} \int \exp(-\boldsymbol{\kappa}(\mathbf{u}, \boldsymbol{\beta}, D)) d\mathbf{u} \quad .$$

Man erhält dann (vgl. Breslow und Clayton (1993)) als Approximation für L bzw. $\ln L$

$$\ln L(\boldsymbol{\beta}, D|Y) \approx -\frac{1}{2} \ln(D) - \frac{1}{2} \ln |\boldsymbol{\kappa}''(\hat{\mathbf{u}}, \boldsymbol{\beta}, D)| - \boldsymbol{\kappa}'(\hat{\mathbf{u}}, \boldsymbol{\beta}, D), \quad (4.16)$$

wobei $\hat{\mathbf{u}}$ die Funktion $\boldsymbol{\kappa}(\hat{\mathbf{u}}, \boldsymbol{\beta}, D)$ maximiert. Mit $\boldsymbol{\kappa}'(\hat{\mathbf{u}}, \boldsymbol{\beta}, D)$ bzw. $\boldsymbol{\kappa}''(\hat{\mathbf{u}}, \boldsymbol{\beta}, D)$ werden die ersten beiden partiellen Ableitungen von $\boldsymbol{\kappa}$ bezüglich \mathbf{u} bezeichnet. Unter Vernachlässigung des Terms $\frac{1}{2} \ln |\boldsymbol{\kappa}''(\hat{\mathbf{u}}, \boldsymbol{\beta}, D)|$ (vgl. Breslow und Clayton (1993)) erhält man für die fixen bzw. zufälligen Effekte Schätzgleichungen (mit $\mathbb{E}(Y_i) = \mu_i$)

$$\sum_{i=1}^n \mathbf{x}_i^t (y_i - \mu_i), \quad \text{für } \boldsymbol{\beta} \quad , \quad (4.17)$$

$$\sum_{i=1}^n \mathbf{z}_i^t (y_i - \mu_i) - D^{-1} \mathbf{u}, \quad \text{für } \mathbf{u} \quad . \quad (4.18)$$

Aus (4.18) ist ersichtlich, dass die Varianz der a priori Verteilung die Funktion eines Strafterms übernimmt.

Anschließend müssen noch die Varianzkomponenten geschätzt werden (vgl. Breslow und Clayton (1993) oder Cai et al. (2002))

$$\ln L(D|Y) \approx -\frac{1}{2} \ln(D) - \frac{1}{2} \ln |\boldsymbol{\kappa}''(\hat{\mathbf{u}}, \hat{\boldsymbol{\beta}}, D)| - \boldsymbol{\kappa}'(\hat{\mathbf{u}}, \hat{\boldsymbol{\beta}}, D) \quad . \quad (4.19)$$

Die Funktion $\ln L(D|Y)$ ist keine vollständige Likelihoodfunktion, sie wird in der zitierten Literatur unterschiedlich eingeordnet. Wird über die Normalverteilungsapproximation argumentiert (Breslow und Clayton, 1993), so wird auf die Analogie zur REML Schätzfunktion im linearen Modell hingewiesen (siehe Patterson und Thompson (1971), Harville (1977)). Alternativ wird sie auch als marginale Likelihood bezeichnet (Cai et al., 2002). Die beiden genannten Ansätze sind Spezialfälle der Modified Profile Likelihood (siehe Barndorff-Nielsen (1983) bzw. Pawitan (2001)). Die Matrix $\boldsymbol{\kappa}''(\hat{\mathbf{u}}, \hat{\boldsymbol{\beta}}, D)$ des Strafterms $\frac{1}{2} \ln |\boldsymbol{\kappa}''(\hat{\mathbf{u}}, \hat{\boldsymbol{\beta}}, D)|$ läßt sich entsprechend als beobachtete Fisher Information auffassen. Sie soll dem Umstand Rechnung tragen, dass anstelle der ‘wahren’ Werte von $(\mathbf{u}, \boldsymbol{\beta})$ nur mit Schätzungen gearbeitet wird und führt zu einer Bias Korrektur der Profile Likelihood.

Meistens wird angenommen, dass D eine spezielle Struktur hat, im einfachsten Fall Unabhängigkeit, d.h. $D = I$, oder etwa compound symmetry bzw. eine autoregressive Struktur. Sei $\sigma_{u_i}^2$ die Varianz von u_i und $V^{\frac{1}{2}} = \text{Diag}(\sigma_{u_1}, \dots, \sigma_{u_n})$, dann gelte (siehe

Fahrmeir und Tutz (1994))

$$D = V^{\frac{1}{2}}R(\boldsymbol{\gamma})V^{\frac{1}{2}}, \quad (4.20)$$

wobei der Parameter $\boldsymbol{\gamma}$ die Struktur der Korrelationsmatrix R beschreibt. Alternativ sind auch Schätzfunktionen möglich, die auf den Pearsonschen Residuen beruhen (Jørgensen und Knudsen, 2004). Basierend auf den geschätzten Pearson Residuen $\hat{\mathbf{r}} = \hat{V}^{\frac{1}{2}}(\mathbf{y} - \hat{\boldsymbol{\mu}})$ lassen sich quadratische Schätzfunktionen (Jørgensen und Knudsen, 2004) entwickeln für Komponenten γ_h von $\boldsymbol{\gamma}$

$$g(h) = \mathbf{r}^t A_h^{-1} \mathbf{r} - \text{spur}(A_h^{-1} R), \quad (4.21)$$

wobei A_h nur von $\boldsymbol{\gamma}$ abhängt und so gewählt werden muß, dass $\mathbb{E}(\frac{\partial}{\partial \gamma_h} g(\gamma_h)) \neq 0$ ist. Dies gewährleistet die Regularität der Schätzfunktion. Wählt man

$$A_h^{-1} = -\frac{\partial}{\partial \gamma_h} R^{-1} = R^{-1} \frac{\partial}{\partial \gamma_h} R R^{-1}, \quad (4.22)$$

so ist die Schätzfunktion (4.21) äquivalent zu Maximum Likelihood Schätzfunktion für das lineare gemischte Modell mit Normalverteilung (McCulloch, Casella und Searle, 1992). Parallel zur Bias Korrektur bei der Modified Profile Likelihood wird empfohlen, die Schätzfunktion zumindest approximativ um den Bias zu korrigieren (Jørgensen und Knudsen, 2004)¹:

$$\tilde{g}(h) = g(h) + \tilde{p}_h, \quad (4.23)$$

mit $(\boldsymbol{\theta} = (\boldsymbol{\beta}, \mathbf{u}), M = (M_{ij})$ mit $M_{ij} = \frac{\partial}{\partial \theta_j} \mu_i$)

$$\tilde{p}_h = \text{spur}\left([M^t \hat{V}^{-\frac{1}{2}} A_h^{-1} \hat{V}^{\frac{1}{2}} M][M^t D M]^{-1}\right). \quad (4.24)$$

Der wesentliche Gedanke hierbei ist, den Bias der Varianzschätzung zu korrigieren, der durch Vernachlässigung der Abhängigkeitsstruktur des Response entsteht sowie der Tatsache, dass für die Mittelwertstruktur nur eine Schätzung vorliegt (s. a. die weiteren Erörterungen in Abschnitt (4.3.3) bzw. im entsprechenden Unterabschnitt von (4.5.2)). Die Schätzungen für $\boldsymbol{\beta}$, \mathbf{u} und $\boldsymbol{\gamma}$ werden dann wechselseitig iterativ durchgeführt. Möglicherweise ist es aus numerischer Sicht vorteilhaft, die Kovariablen zu standardisieren. Bei der Approximation des Integrals (4.19) wird ein Fehler in Kauf genommen, der

¹Stichwort: BAPE = 'Bias adjusted Pearson estimator'

um so größer ist, je mehr die Verteilung der abhängigen Variable von der Normalverteilung abweicht und je kleiner die Varianz der zufälligen Effekte ist. Bei einer geringen Dimension der zufälligen Effekte sind auch andere Integrationsverfahren, z.B. Gauss-Hermite Prozeduren (McCulloch und Searle, 2001, S. 269) möglich. Ein Ansatz, der in der Statistik bisher kaum verbreitet ist, beruht auf dem Smolyak-Algorithmus (Gerstner und Griebel, 1998), der verspricht, auch in höher dimensionalen Bereichen einsetzbar zu sein. Alternativ kann man bayesianische MCMC Methoden verwenden oder den EM Algorithmus, meist in einer Monte Carlo Variante (Terminus: MCEM), da man den Erwartungswert bezüglich der marginalen Verteilungen nicht analytisch berechnen kann (Mebane und Wand, 1999). Des weiteren gibt es noch simulative Varianten der Maximum Likelihood und Newton Raphson Verfahren (Booth, Hobert und Jank, 2001). Einen kurzen Überblick findet man in McCulloch und Searle (2001, S. 269 ff).

4.3.2 Gemischte Modelle mit penalisierten Splines

Werden die Penalisierungsterme $\mathbf{u}'D(\theta)^{-1}\mathbf{u}$ und $\lambda \sum_{i=k+1}^n (\Delta^k \alpha_i)^2$ betrachtet, so erkennt man, dass der Penalisierungsterm bei gemischten Modellen keinen Glättungsparameter beinhaltet, zumindest nicht explizit, sondern dass der Einfluß der glatten Komponenten über die Varianz der Zufallseffekte gesteuert wird. Dies ermöglicht eine datengesteuerte Schätzung ohne Verwendung von Methoden der Kreuzvalidierung. Daher werden die Koeffizienten $\boldsymbol{\alpha}$ im folgenden als zufällige Effekte betrachtet. Die Struktur der Kovarianzmatrix $D(\theta)$ steuert, in welchem Maß benachbarte Werte α_h zueinander ähnlich sind. Bisher wurden an die Zufallseffekte keine Nebenbedingungen gestellt. Im Fall von Nebenbedingungen oder Rangverlust von $D(\theta)$ kann es von Vorteil sein, anstelle der Kovarianzmatrix mit der Präzisionsmatrix $Q(\theta)$ zu arbeiten (vgl. Breslow und Clayton (1993)). Der Penalisierungsterm lautet dann $\frac{1}{\sigma_\alpha^2} \boldsymbol{\alpha}' Q(\theta) \boldsymbol{\alpha}$ (mit $Var(\alpha_k) = \sigma_\alpha^2$). Betrachtet man für $k = 1, 2$ die Summen $\sum_{i=k+1}^n (\Delta^k \alpha_i)^2$, die sich in Matrizenform auch als $\frac{1}{\sigma_\alpha^2} \boldsymbol{\alpha}' Q_1(\theta) \boldsymbol{\alpha}$ und $\frac{1}{\sigma_\alpha^2} \boldsymbol{\alpha}' Q_2(\theta) \boldsymbol{\alpha}$ für $k = 1, 2$ mit

$$Q_1(\theta) = \begin{pmatrix} 1 & -1 & 0 & \cdots & & & 0 \\ -1 & 2 & -1 & 0 & \cdots & & \\ 0 & -1 & 2 & -1 & 0 & \cdots & \\ \vdots & \ddots & \ddots & \ddots & \ddots & & \vdots \\ & \cdots & 0 & -1 & 2 & -1 & 0 \\ & & \cdots & 0 & -1 & 2 & -1 \\ 0 & & & \cdots & 0 & -1 & 1 \end{pmatrix} \quad (4.25)$$

beziehungsweise

$$Q_2(\theta) = \begin{pmatrix} 1 & -2 & 1 & 0 & \dots & & & & & 0 \\ -2 & 5 & -4 & 1 & \dots & & & & & \\ 1 & -4 & 6 & -4 & 1 & 0 & \dots & & & \\ 0 & 1 & -4 & 6 & -4 & 1 & 0 & \dots & & \\ \vdots & & \ddots & \ddots & \ddots & \ddots & & & & \vdots \\ & \dots & 0 & 1 & -4 & 6 & -4 & 1 & 0 & \\ & & \dots & 0 & 1 & -4 & 6 & -4 & 1 & \\ & & & \dots & 0 & 1 & -4 & 5 & -2 & \\ 0 & & & & \dots & 0 & 1 & -2 & 1 & \end{pmatrix} \quad (4.26)$$

schreiben lassen, so erkennt man, dass die Matrizen einen Rangverlust um einen bzw. um zwei Ränge haben. Dies ist nicht weiter verwunderlich, da die quadratische Form $\frac{\boldsymbol{\alpha}' Q_1(\theta) \boldsymbol{\alpha}}{\sigma_\alpha^2}$ der ersten Differenzen für Vektoren $\boldsymbol{\alpha}$ mit Komponenten $\alpha_h = c$, $c \in \mathbb{R}$, Null ergibt, im Fall der zweiten Differenzen sogar für ‘lineare’ Vektoren $\boldsymbol{\alpha}$ mit Komponenten $\alpha_h = c + d \cdot h$, $c, d \in \mathbb{R}$, ergibt. Dieser Effekt beruht darauf, dass der Penalisierungsterm $\int_{x_0}^{x_n} |g^{(k)}|^2 dx$ für lineare Funktionen (wegen der zweiten Ableitungen) den Wert Null annimmt – ähnlich einer Halbnorm. Er beruht nicht auf der Approximation von Marx/Eilers. Es können deshalb Probleme bei der Parameteridentifikation auftreten. Ähnliche Effekte treten auch bei additiven Modellen auf, falls man mehrere glatte Komponenten schätzt oder falls zusätzlich noch lineare Effekte geschätzt werden müssen. Diese Probleme bezeichnet man auch als *concurvity* (Buja, Hastie und Tibshirani, 1989), einer Verallgemeinerung der Kollinearität im linearen Modell, bei der der Mittelwert eindeutig bestimmt ist, die Parameter jedoch nicht eindeutig schätzbar sind, da die Matrix der Kovariablen nicht den vollen Rang hat. Die Autoren Buja et al. (1989) schlagen in diesem Fall vor, für diejenigen Kovariablen, für die man glatte Funktionen schätzt, auch lineare Effekte (parametrisch) zu schätzen. Anstelle von

$$f_j(x) = \sum_{l=1}^{K(f_j)} \alpha_{lf_j} B_{lj}(x) \quad \text{bzw.} \quad g_j(t) = \sum_{l=1}^{K(g_j)} \alpha_{lg_j} B_{lj}(t)$$

werden die ‘zentrierten’ Varianten

$$\beta_{f_j,0}^t \mathbf{x}_{ji} + f_j(x) \quad \text{bzw.} \quad \beta_{g_j,0}^t \mathbf{t}_{ji} + g_j(t) \quad (4.27)$$

verwendet, wobei man die Parameter $\beta_{f_j,0}$ und $\beta_{g_j,0}$ als fest betrachtet und beachtet, dass der Intercept in dem Vektor \mathbf{x}_{ji} (analog: $\mathbf{t}_{ji} = (1, t_i)$) enthalten ist. In dieser Situation

lassen sich die glatten Komponenten als Abweichung vom linearen Prädiktor auffassen. Auf diese Weise gelangt man zu eindeutigen Schätzungen für den linearen Teil und den glatten Teil des Prädiktors. Negativ ist allerdings eine gewisse Überparametrisierung, da man jetzt zusätzlich noch die Parameter für das lineare Submodell hat. Betrachtet man die Darstellung anhand einseitiger Splines, so erkennt man, dass die Basisfunktionen in natürlicher Weise eine Aufspaltung zulassen wie sie hier gebraucht wird:

$$f(x) = \sum_{l=0}^r \beta_l x^l + \sum_{k=1}^n \alpha_k (x - x_k)_+^r \quad . \quad (4.28)$$

Im Gegensatz zu den B-Splines braucht man bei den einseitigen Splines keine zusätzlichen Parameter. Die Monome bzw. die dazu korrespondierenden Effekte lassen sich parametrisch schätzen, der Einfluß der abgebrochenen Potenzen wird über zufällige Parameter modelliert. Als Nebenprodukt erhält man einfache Submodelle ohne glatte Komponenten (vgl. Gray (1992, S. 947)). Im Fall der B-Splines hat die Varianzmatrix $D(\theta)$ bzw. die Dispersionmatrix $Q(\theta)$ eine Bandstruktur. Dies ist zweckmäßig, da sich die Träger benachbarter Basisfunktionen überlappen und damit die dazu korrespondierenden (benachbarten) Parameter voneinander abhängig sind. Bei einseitigen Splines ist die Situation anders, die zu den Monomen korrespondierenden Parameter werden als fest vorausgesetzt, die abgebrochenen Potenzen werden abhängig vom jeweiligen x_k vollständig überdeckt (mit Ausnahme vom Fall $k = 1$). In der Literatur (Wand, 2003) wird für diese Situation vorgeschlagen, die zufälligen Effekte α_k als unabhängig normalverteilt anzusehen, d.h. es gilt

$$\alpha_k \sim NV(0, \sigma_\alpha^2) \quad . \quad (4.29)$$

Der Varianzparameter σ_α^2 der a priori Verteilung steuert das Verhalten, kleine Werte von σ_α^2 schränken den Spielraum der Parameter der glatten Komponente ein. Weitere Hinweise findet man in der Literatur, z.B. Lin und Zhang (1999) bzw. Verbyla, Cullis, Kenward und Welham (1999). In Kneib und Fahrmeir (2004) und insbesondere in Eilers und Marx (2004) finden sich Hinweise zu den verschiedenen Systemen von Basisfunktionen, u. a. auch eine Transformation der B-Splines mittels Projektionsmatrizen, die eine Aufspaltung in fixe Komponenten (Polynome) und zufällige Komponenten gestattet. Die letztgenannten Autoren favorisieren die Verwendung von B-Splines und die Schätzung der Smoothing Parameter über Kreuzvalidierung gegenüber Splines im Kontext von gemischten Modellen.

4.3.3 Gemischte Modelle für die Lebensdaueranalyse

Als Beispiel betrachte man ein Cox–Modell mit zufälligen Effekten \mathbf{u} , (Ripatti und Palmgren, 2000) für $i = 1, \dots, n$ Individuen mit Kovariablen \mathbf{x}_i und \mathbf{z}_i und festen Effekten $\boldsymbol{\beta}$. Solche Modelle werden auch als frailty Modelle bezeichnet (Hougaard, 2000). Die Hazardrate lautet dann

$$\lambda_{it}(\mathbf{u}) = h_t \cdot \exp(\mathbf{x}_i^t \boldsymbol{\beta} + \mathbf{z}_i^t \mathbf{u}) \quad ,$$

wobei davon ausgegangen wird, dass \mathbf{u} entweder gamma– oder normalverteilt ist. Im Fall, dass \mathbf{u} normalverteilt ist mit Erwartungswert 0 und Kovarianzmatrix $D(\theta)$, schreibt sich die vollständige Cox-Likelihood (Zensierungsindikator δ_i für Individuum i) als

$$L_i(\mathbf{u}, \boldsymbol{\beta}, h_t, D(\theta)) = \prod_{i=1}^n \lambda_i(\mathbf{u}, \boldsymbol{\beta}, h_t, D(\theta))^{\delta_i} \cdot \exp \left\{ - \int_0^{T_i} \lambda_s(\mathbf{u}, \boldsymbol{\beta}, h_t, D(\theta)) ds \right\} p(\mathbf{u}, D(\theta)) \quad . \quad (4.30)$$

Wird über die zufälligen Effekte \mathbf{u} mit der Dichte $p(\mathbf{u}, D(\theta))$ integriert, so erhält man die marginale Likelihoodfunktion

$$L_i(\boldsymbol{\beta}, h_t, D(\theta)) = \int \prod_{i=1}^n \lambda_i(\mathbf{u}, \boldsymbol{\beta}, h_t, D(\theta))^{\delta_i} \cdot \exp \left\{ - \int_0^{T_i} \lambda_s(\mathbf{u}, \boldsymbol{\beta}, h_t, D(\theta)) ds \right\} p(\mathbf{u}, D(\theta)) d\mathbf{u} \quad . \quad (4.31)$$

Nach Anwendung der Laplace–Approximation, bei der die Integration ersetzt wird durch eine Einpunkt–Approximation am Modalwert der Dichte von \mathbf{u} sowie unter Vernachlässigung der ersten beiden Terme in der Gleichung (4.19) – zumindest für die Schätzung von \mathbf{u} und $\boldsymbol{\beta}$ – in Ripatti und Palmgren (2000)² erhält man die partielle Likelihoodfunktion nach Cox, erweitert um einen Penalisierungsterm, d.h.

$$L_i(\mathbf{u}, \boldsymbol{\beta}, D(\theta)) \approx C_i(\mathbf{u}, \boldsymbol{\beta}) - \frac{1}{2} \mathbf{u}^t D(\theta)^{-1} \mathbf{u} \quad . \quad (4.32)$$

²Vorsicht, die Gleichungen (5) und (6) der Veröffentlichung sind falsch.

Dies führt zu Schätzfunktionen³ für $\boldsymbol{\beta}$ und \mathbf{u}

$$\sum_{i=1}^n \delta_i \left(\mathbf{x}_i^t - \frac{\sum_{j \in R(t_i)} \mathbf{x}_j^t \exp(\mathbf{x}_j^t \boldsymbol{\beta} + \mathbf{z}_j^t \mathbf{u})}{\sum_{j \in R(t_i)} \exp(\mathbf{x}_j^t \boldsymbol{\beta} + \mathbf{z}_j^t \mathbf{u})} \right) \quad \text{für } \boldsymbol{\beta} \quad , \quad (4.33)$$

$$\sum_{i=1}^n \delta_i \left(\mathbf{z}_i^t - \frac{\sum_{j \in R(t_i)} \mathbf{z}_j^t \exp(\mathbf{x}_j^t \boldsymbol{\beta} + \mathbf{z}_j^t \mathbf{u})}{\sum_{j \in R(t_i)} \exp(\mathbf{x}_j^t \boldsymbol{\beta} + \mathbf{z}_j^t \mathbf{u})} \right) - D^{-1} \mathbf{u} \quad \text{für } \mathbf{u} \quad . \quad (4.34)$$

Die Menge $R(t_i)$ enthält diejenigen Individuen j , für die zur Zeit t_i noch keine Beobachtung erfolgt ist. Dabei wird vorausgesetzt, dass die Variablen \mathbf{x}_i bzw. \mathbf{z}_i komponentenweise für mindestens zwei Individuen verschieden sind, anderenfalls sind die dazu korrespondierenden Parameter nicht schätzbar, da die Schätzgleichung automatisch den Wert Null annimmt. Dies ist ersichtlich, da in diesem Fall die Komponenten j von \mathbf{x}_i bzw. \mathbf{z}_i aus der inneren Summe gezogen werden können, Zähler und Nenner identisch sind und die innere Klammer den Wert Null annimmt.

Zusätzlich müssen noch die Varianzparameter geschätzt werden. Dazu wird für $\boldsymbol{\kappa}''$ (vgl. Gleichung (4.19)) eine Approximation basierend auf der Partial Likelihoodfunktion verwendet (siehe Ripatti und Palmgren (2000)).

4.4 Bemerkungen zur Wahl des Splinegrades

Die Glattheit der geschätzten Kurven hängt von dem Grad der gewählten Splines ab. Kubische Splines liefern demnach optisch sehr glatte Funktionen, Splines vom Grad Eins lediglich einen Polygonzug. Demgegenüber ist der numerische Aufwand bei kubischen Splines unter Umständen wesentlich höher. Je nach Wahl des Modells können die Schätzfunktionen Integrale enthalten, deren Wert approximiert werden muß, falls der Grad größer als Eins ist. Unabhängig davon wird in der Literatur (Gray, 1992) darauf hingewiesen, dass beim Cox-Modell für große Lebensdauern numerische Instabilitäten auftreten können. Der Grund hierfür ist, dass die Risikomenge entsprechend klein ist.

³So würde auch die korrekte Form der fehlerhaften Gleichungen des Artikels lauten.

4.5 Modellierung und Schätzung der Parameter

Bisher wurden hauptsächlich Aspekte der Modellierung aufgezeigt. Diese werden jetzt angewendet auf eine allgemeine Form einer Intensitätsrate. Dazu werden noch kurz die theoretischen Vorgaben aus Kapitel 2 angegeben. Unmittelbar im Anschluß erfolgen Hinweise über die Parameterschätzung. Dabei wird auf die Ergebnisse aus Kapitel 3 zurückgegriffen.

4.5.1 Modellierung der Intensitätsrate

Es wurden einige Eigenschaften des Intensitätsprozesses λ_s zusammengetragen, die gewährleisten, dass der modellierte Punktprozeß eine Radon–Nikodym Dichte bezüglich des homogenen Poissonprozesses mit Parameter $\lambda = 1$ besitzt. Es wird daher vorausgesetzt, das λ_s strikt positiv, quadratintegrierbar und stetig ist. Letztere Eigenschaft zieht die Vorhersagbarkeit nach sich. Zusammengenommen stellen diese Voraussetzungen sicherlich keine gravierende Einschränkung dar. Der einfachste Intensitätsprozeß ist sicherlich ein homogener Poissonprozeß mit $\lambda_s = const$, $const > 0$. Eine Verallgemeinerung auf den inhomogenen Poissonprozeß führt zu $\lambda_s = h(s)$, $h(s) > 0$, bzw. zu $\lambda_s = \exp(h'(s))$ mit $h'(s) = \ln(h(s))$. Unter Hinzunahme von Kovariablen gelangt man zu komplizierten Modellen

$$\lambda_{is}(\boldsymbol{\beta}, a, \boldsymbol{\alpha}_1, \mathbf{u}^{\alpha_1}, \boldsymbol{\alpha}_2, \mathbf{u}^{\alpha_2}, \boldsymbol{\alpha}_3, \mathbf{u}^{\alpha_3}) = \exp\left(\eta_i(\boldsymbol{\beta}) + \eta_{is}(a) + \eta_s(\boldsymbol{\alpha}_1, \mathbf{u}^{\alpha_1}) + \eta_{is}(\boldsymbol{\alpha}_2, \mathbf{u}^{\alpha_2}) + \eta_i(\boldsymbol{\alpha}_3, \mathbf{u}^{\alpha_3})\right)$$

mit festen Effekten $\boldsymbol{\beta}$

$$\eta_i(\boldsymbol{\beta}) = \mathbf{x}'_i \boldsymbol{\beta} \quad (4.35)$$

bzw. festen Effekten a zu zeitvariierenden Kovariablen

$$\eta_{is}(a) = \mathbf{x}'_{is} a \quad . \quad (4.36)$$

Der griechische Buchstabe η entspricht dem Prädiktor, eine Notation, die den Regressionsmodellen entnommen ist. Die Funktion $h'(s)$ wird mit Splines modelliert

$$h'_s(\boldsymbol{\alpha}_1, \mathbf{u}^{\alpha_1}) = \eta_s(\boldsymbol{\alpha}_1, \mathbf{u}^{\alpha_1}) = \sum_{l=0}^g \alpha_{1l} s^l + \sum_{k=0}^K u_k^{\alpha_1} (s - \kappa_k)_+^g, \quad ,$$

mit festen Parametern $\alpha_{(\cdot)}$ und zufälligen Parametern $u_{(\cdot)}^{\alpha_1}$, von denen angenommen wird, dass sie unabhängig normalverteilt sind mit

$$u_k^{\alpha_1} \sim NV(0, \sigma_{\alpha_1}^2) \quad , k = 1, \dots, K \quad .$$

Im folgenden wird dieser glatte Effekt als Typ 1 bezeichnet. Für die zeitvariierenden Effekte der j -ten Kovariable und Individuum i (Typ 2) gilt

$$\eta_{is}(\boldsymbol{\alpha}_2, \mathbf{u}^{\alpha_2}) = x_{ij} \left(\sum_{l=0}^g \alpha_{2l} s^l + \sum_{k=0}^K u_k^{\alpha_2} (s - \kappa_k)_+^g \right)$$

mit festen Parametern $\alpha_{(\cdot)}$ und zufälligen Effekten $u_{(\cdot)}^{\alpha_2}$, für die gilt

$$u_k^{\alpha_2} \sim NV(0, \sigma_{\alpha_2}^2) \quad , k = 1, \dots, K \quad .$$

Im Fall glatter Effekte (Typ 3) gilt

$$\eta_i(\boldsymbol{\alpha}_3, \mathbf{u}^{\alpha_3}) = \sum_{l=0}^g \alpha_{3l} x_{ij}^l + \sum_{k=0}^K u_k^{\alpha_3} (x_{ij} - \kappa_k)_+^g$$

mit festen Parametern $\alpha_{(\cdot)}$ und zufälligen Parametern $u_{(\cdot)}^{\alpha_3}$ mit

$$u_k^{\alpha_3} \sim NV(0, \sigma_{\alpha_3}^2) \quad , k = 1, \dots, K \quad .$$

Die drei Typen unterscheiden sich dahingehend, dass im Fall von Typ 1 und 2 ein glatter Effekt im zeitlichen Bereich, einmal ohne – sonst mit Kovariablen modelliert wird (zeitvariierende Effekte), bzw. bei Typ 3 ein glatter Effekt in der jeweiligen Kovariable modelliert wird. Zur Vereinfachung der Darstellung wird im folgenden stets nur ein Ef-

fekt vom Typ 2 bzw. 3 erwähnt, obwohl nicht ausgeschlossen ist, dass mehrere Effekte modelliert werden können. Im Hinblick auf den Zeitpunkt des Ereigniseintritts sowie auf die Kovariablenstruktur ist jeweils zu beachten, ob der Prädiktor zeit- und oder individuenabhängig ist. Die integrierte Intensitätsrate schreibt sich als

$$\Lambda_{t_i}(\boldsymbol{\beta}, a, \boldsymbol{\alpha}_1, \mathbf{u}^{\alpha_1}, \boldsymbol{\alpha}_2, \mathbf{u}^{\alpha_2}, \boldsymbol{\alpha}_3, \mathbf{u}^{\alpha_3}) = \int_0^{t_i} \lambda_{is}(\boldsymbol{\beta}, a, \boldsymbol{\alpha}_1, \mathbf{u}^{\alpha_1}, \boldsymbol{\alpha}_2, \mathbf{u}^{\alpha_2}, \boldsymbol{\alpha}_3, \mathbf{u}^{\alpha_3}) ds$$

oder ausführlicher

$$\begin{aligned} \Lambda_{t_i}(\boldsymbol{\beta}, a, \boldsymbol{\alpha}_1, \mathbf{u}^{\alpha_1}, \boldsymbol{\alpha}_2, \mathbf{u}^{\alpha_2}, \boldsymbol{\alpha}_3, \mathbf{u}^{\alpha_3}) = & \int_0^{t_i} \exp\left(x_{ij}^t \boldsymbol{\beta} + x_{is}^t a \right. \\ & + \sum_{l=0}^g \alpha_{1l} s^l + \sum_{k=0}^K u_k^{\alpha_1} (s - \kappa_k)_+^g \\ & + x_{ij} \sum_{l=0}^g \alpha_{2l} s^l + x_{ij} \sum_{k=0}^K u_k^{\alpha_2} (s - \kappa_k)_+^g \\ & \left. + \sum_{l=0}^g \alpha_{3l} x_{ij}^l + \sum_{k=0}^K u_k^{\alpha_3} (x_{ij} - \kappa_k)_+^g \right) ds \end{aligned}$$

Um die nicht von der Zeit abhängenden Terme vor das Integral ziehen zu können, führt man noch folgende Abkürzung ein

$$\lambda_{is}^*(a, \boldsymbol{\alpha}_1, \mathbf{u}^{\alpha_1}, \boldsymbol{\alpha}_2, \mathbf{u}^{\alpha_2}) = \exp\left(\eta_{is}(a) + \eta_s(\boldsymbol{\alpha}_1, \mathbf{u}^{\alpha_1}) + \eta_{is}(\boldsymbol{\alpha}_2, \mathbf{u}^{\alpha_2})\right),$$

von der insbesondere bei den partiellen Ableitungen Gebrauch gemacht wird. Es gilt

$$\lambda_{is}^*(a, \boldsymbol{\alpha}_1, \mathbf{u}^{\alpha_1}, \boldsymbol{\alpha}_2, \mathbf{u}^{\alpha_2}) = \frac{\lambda_{is}(\boldsymbol{\beta}, a, \boldsymbol{\alpha}_1, \mathbf{u}^{\alpha_1}, \boldsymbol{\alpha}_2, \mathbf{u}^{\alpha_2}, \boldsymbol{\alpha}_3, \mathbf{u}^{\alpha_3})}{\exp(\mathbf{x}_{ij}^t \boldsymbol{\beta}) \exp(\eta_i(\boldsymbol{\alpha}_3, \mathbf{u}^{\alpha_3}))}.$$

4.5.2 Schätzung der Intensitätsrate

Man kann nun entweder mit der Laplace–Approximation arbeiten oder die Schätzfunktionen aus Abschnitt (3.5) verwenden. Hier wird letzterer Weg verfolgt, da die Theorie dazu hinreichend gut ausgebaut ist. Wegen $\lambda(\theta) = \exp(\eta(\theta))$ gilt $\frac{\dot{\lambda}(\theta)}{\lambda(\theta)} = \dot{\eta}(\theta)$ und Gleichung (3.26) schreibt sich für Individuum i unter Beachtung einer nichtinformativen

Rechtszensierung (vgl. 3.1) mit Zensierungsindikatorvariable δ_i und $t_i = T_i \wedge U_i$ als

$$G_i(\theta) = \sum_{\{s \leq t_i, N_{is}^c \neq N_{is-}^c\}} \dot{\eta}_{is}(\theta) - \int_0^{t_i} \dot{\eta}_{is}(\theta) \lambda_{is}(\theta) ds \quad (4.37)$$

$$= \delta_i \cdot \dot{\eta}_{it_i}(\theta) - \int_0^{t_i} \dot{\eta}_{is}(\theta) \lambda_{is}(\theta) ds \quad (4.38)$$

Die Darstellung der Schätzfunktionen berücksichtigt noch nicht die Tatsache, dass auch zufällige Parameter geschätzt werden. Dies geschieht weiter unten mit Hilfe von Straftermen. Im Unterschied zu Abschnitt (4.3.3) ist die Intensitätsrate λ_s zusätzlich abhängig von der Zeit. Dies erfordert die Berechnung des Integrals von λ_s über die Zeit. Im Fall linearer Splines erhält man eine analytische Lösung, der Vektor der partiellen Ableitungen $\dot{\eta}_{is}(\theta)$ ist linear in s , z.B. ergibt sich für $\dot{\eta}_{is}(\boldsymbol{\alpha}_2, \mathbf{u}^{\alpha_2}) = (x_{ij}, x_{ij} \cdot s, x_{ij} \cdot (s - \kappa_1)_+, \dots, x_{ij} \cdot (s - \kappa_K)_+)$ mit $(\frac{d\dot{\eta}_{is}(\theta)}{ds} = \dot{\eta}'_{is}(\theta))$

$$\int_0^{t_i} \dot{\eta}_{is}(\theta) \lambda_{is}(\theta) ds = \left[\frac{(\dot{\eta}_{is}(\theta) \cdot \dot{\eta}'_{is}(\theta) - 1) \lambda_{is}(\theta)}{(\dot{\eta}'_{is}(\theta))^2} \right]_{ug}^{t_i} \quad (4.39)$$

Bei der unteren Integrationsgrenze ug ist zu beachten, dass sie bei den partiellen Ableitungen von $\dot{\eta}_{is}$ nach $\boldsymbol{\alpha}$ Null, bei den Ableitungen nach u_k jeweils κ_k sind. Für $t_i \leq \kappa_k$ ist der Wert des Integrals Null. Bei höheren Splinegraden muß numerisch integriert werden. Schätzungen $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}_1, \hat{\boldsymbol{\alpha}}_2, \hat{\boldsymbol{\alpha}}_3, \hat{a}, \hat{\mathbf{u}}^{\alpha_1}, \hat{\mathbf{u}}^{\alpha_2}, \hat{\mathbf{u}}^{\alpha_3}$ für $\boldsymbol{\beta}, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3, a, \mathbf{u}^{\alpha_1}, \mathbf{u}^{\alpha_2}, \mathbf{u}^{\alpha_3}$ erhält man als Nullstellen der Schätzgleichungen $G(\cdot)$, d.h.

$$G(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}_1, \hat{\boldsymbol{\alpha}}_2, \hat{\boldsymbol{\alpha}}_3, \hat{a}, \hat{\mathbf{u}}^{\alpha_1}, \hat{\mathbf{u}}^{\alpha_2}, \hat{\mathbf{u}}^{\alpha_3}) = \sum_{i=1}^n G_i(\cdot) = 0 \quad (4.40)$$

Es gibt verschiedene Verfahren zum Lösen von nichtlinearen Gleichungen. Sie lassen sich danach unterscheiden, ob man den Gradienten (und evtl. noch die 2. Ableitungen) verwendet oder nicht. Zum ersten Fall gehören z.B. Newton Raphson und Quasi Newton Verfahren; sie haben den Vorteil, schneller zu konvergieren als gradientenfreie Verfahren (z.B. Nelder Mead Verfahren). Sei $\theta = (\boldsymbol{\beta}, a, \theta_1, \theta_2, \theta_3)$ mit den Abkürzungen $\theta_1 = (\boldsymbol{\alpha}_1, \mathbf{u}^{\alpha_1})$, $\theta_2 = (\boldsymbol{\alpha}_2, \mathbf{u}^{\alpha_2})$ und $\theta_3 = (\boldsymbol{\alpha}_3, \mathbf{u}^{\alpha_3})$. Ausgehend vom Startwert θ^0 werden iterativ z.B. im k -ten Schritt nach folgender Vorschrift neue Schätzungen generiert

$$\theta^{k+1} = \theta^k - s^k \cdot d^k \quad , \quad (4.41)$$

wobei s_k die Schrittweite und d_k die Richtung im k -ten Schritt ist. Das Verfahren bricht ab, falls $\theta^{k+1} \approx \theta^k$ bzw. falls $G(\theta^{k+1}) \approx 0$ gilt. In der Literatur Press, Teukolsky, Vetterling und Flannery (1992) wird empfohlen, die Schrittweite adaptiv zu wählen. Zwar wird durch den Gradienten die optimale (infinitesimale) Richtung angegeben, jedoch nicht die Schrittweite, die ja automatisch Eins ist, falls man die Schrittweitenbestimmung unterläßt. Bei der Schrittweitenbestimmung hat man ein eindimensionales Minimierungsproblem zu lösen. Zur Berechnung von d_k wird – je nach Verfahren – neben dem Gradienten noch die Hessematrix benötigt. Neben der bisherigen Unterteilung ist es bei hochdimensionalen Problemen ein weit verbreitetes Vorgehen, blockweise neue Schätzungen zu berechnen. Dies hat den Vorteil, nicht für alle Parameter simultan eine Hessematrix berechnen zu müssen. Im Vergleich zum obigen Algorithmus wird jetzt der k -te Schritt blockweise aufgeteilt. Es bietet sich an, eine Einteilung entsprechend den Parametern der glatten Effekte zu wählen. Sei $\theta^k = (\beta^k, a^k, \theta_1^k, \theta_2^k, \theta_3^k)$ die Schätzung zu Beginn des k -ten Schritts.

1. Sei $\beta_0^k = \beta^k$, man berechnet

$$\beta_{p+1}^k = \beta_p^k - s_p^k \cdot d_p^k \quad , \quad p = 0, 1, \dots$$

bis in der p^* -ten Iteration eines der Abbruchkriterien erfüllt ist und setzt $\beta^{k+1} = \beta_{p^*}^k$ sowie $\theta^k = (\beta^{k+1}, a^k, \theta_1^k, \theta_2^k, \theta_3^k)$,

2. sei $a_0^k = a^k$, man berechnet

$$a_{p+1}^k = a_p^k - s_p^k \cdot d_p^k \quad , \quad p = 0, 1, \dots$$

bis in der p^* -ten Iteration eines der Abbruchkriterien erfüllt ist und setzt $a^{k+1} = a_{p^*}^k$ sowie $\theta^k = (\beta^{k+1}, a^{k+1}, \theta_1^k, \theta_2^k, \theta_3^k)$,

3. sei $\theta_{10}^k = \theta_1^k$, man berechnet

$$\theta_{1p+1}^k = \theta_{1p}^k - s_p^k \cdot d_p^k \quad , \quad p = 0, 1, \dots$$

bis in der p^* -ten Iteration eines der Abbruchkriterien erfüllt ist und setzt $\theta_1^{k+1} = \theta_{1p^*}^k$ sowie $\theta^k = (\beta^{k+1}, a^{k+1}, \theta_1^{k+1}, \theta_2^k, \theta_3^k)$,

4. sei $\theta_{20}^k = \theta_2^k$, man berechnet

$$\theta_{2p+1}^k = \theta_{2p}^k - s_p^k \cdot d_p^k \quad , \quad p = 0, 1, \dots$$

bis in der p^* -ten Iteration eines der Abbruchkriterien erfüllt ist und setzt $\theta_2^{k+1} = \theta_{2p^*}^k$ sowie $\theta^k = (\beta^{k+1}, a^{k+1}, \theta_1^{k+1}, \theta_2^{k+1}, \theta_3^k)$,

5. sei $\theta_{30}^k = \theta_3^k$, man berechnet

$$\theta_{3p+1}^k = \theta_{3p}^k - s_p^k \cdot d_p^k \quad , \quad p = 0, 1, \dots$$

bis in der p^* -ten Iteration eines der Abbruchkriterien erfüllt ist

und setzt $\theta_3^{k+1} = \theta_{3p^*}^k$ sowie $\theta^k = (\boldsymbol{\beta}^{k+1}, a^{k+1}, \theta_1^{k+1}, \theta_2^{k+1}, \theta_3^{k+1})$.

Bei diesem Vorgehen verzichtet man im Vergleich mit der vollständigen (Pseudo-) Hessematrix auf die gemischten zweiten partiellen Ableitungen (z.B. $\frac{\partial G}{\partial \boldsymbol{\beta}^k \partial \alpha_1^k}$). Dafür ist die Dimension in jedem Teilschritt kleiner. Eine Anwendung auf die Schätzfunktionen liefert eine Aufspaltung in marginale Schätzgleichungen $G(\boldsymbol{\beta})$, $G(a)$, $G(\theta_1)$, $G(\theta_2)$, $G(\theta_3)$, für die man im k -ten Schritt jeweils einen neuen Wert erhält. Da – abgesehen von der Schätzfunktion für $\boldsymbol{\beta}$ und a – alle Schätzfunktionen zufällige unabhängige normalverteilte Parameter mit Erwartungswert 0 enthalten, werden die Schätzfunktionen jeweils um einen Strafterm $G_{\sigma_{\alpha_1}^2}$ korrigiert und zwar für $G(\theta_1)$ mit $(\mathbf{u}^{\alpha_1} = (u_1^{\alpha_1}, \dots, u_K^{\alpha_1}))$

$$G_{\sigma_{\alpha_1}^2}(\mathbf{u}^{\alpha_1}) = \frac{1}{\sigma_{\alpha_1}^2} (\mathbf{u}^{\alpha_1})^t \quad , \quad (4.42)$$

der die martingale Struktur der Schätzfunktionen nicht zerstört, d.h. es gilt $\mathbb{E}(G_{\sigma_{\alpha_1}^2}(\theta_1)) = 0$. Ein analoges Vorgehen erfolgt für $G(\mathbf{u}^{\alpha_2}), G(\mathbf{u}^{\alpha_3})$.

Bei den Schätzfunktionen ebenso wie bei den Ableitungen ist zu beachten, dass bezüglich der Zeit integriert werden muß. Zur Vereinfachung der Notation sei $\theta_{-<vek>}$ der Teilvektor von θ ohne den Vektor $<vek>$. Die Berechnungen erfolgen unter der Annahme, dass die Integration nach der Zeit und die Differentiation nach den Parametern vertauschbar ist.

Die marginale Schätzfunktion für $\boldsymbol{\beta}$

Die Ableitung des integrierten Intensitätsprozesses Λ nach $\boldsymbol{\beta}$ pro Individuum (mit $\lambda_{is}^*(\theta)$ als Kurzform für $\lambda_{is}^*(\theta_{-(\boldsymbol{\beta}, \theta_3)})$) lautet

$$\frac{\partial}{\partial \boldsymbol{\beta}} \Lambda_t(\boldsymbol{\beta}, \theta_{-\boldsymbol{\beta}}) = \frac{\partial}{\partial \boldsymbol{\beta}} \eta_i(\boldsymbol{\beta}) \exp(\eta_i(\boldsymbol{\beta})) \exp(\eta_i(\theta_3)) \int_0^{t_i} \lambda_{is}^*(\theta) ds \quad (4.43)$$

$$= \mathbf{x}_i^t \exp(\eta_i(\boldsymbol{\beta}) + \eta_i(\theta_3)) \int_0^{t_i} \lambda_{is}^*(\theta) ds \quad . \quad (4.44)$$

Summiert über alle Individuen erhält man

$$\frac{\partial}{\partial \boldsymbol{\beta}} \Lambda_t(\boldsymbol{\beta}, \theta_{-\boldsymbol{\beta}}) = \sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\beta}} \Lambda_{t_i}(\boldsymbol{\beta}, \theta_{-\boldsymbol{\beta}}) \quad . \quad (4.45)$$

Die marginale Schätzfunktion für $\boldsymbol{\beta}$ lautet demnach ($\boldsymbol{\delta} = (\delta_1, \dots, \delta_n)^t$)

$$G(\boldsymbol{\beta}) = \mathbf{X}^t \boldsymbol{\delta} - \mathbf{X}^t \text{vek} \left(\exp(\eta_i(\boldsymbol{\beta}) + \eta_i(\theta_3)) \int_0^{t_i} \lambda_{is}^*(\theta) ds \right) \quad (4.46)$$

$$= \mathbf{X}^t \left[\boldsymbol{\delta} - \text{vek} \left(\exp(\eta_i(\boldsymbol{\beta}) + \eta_i(\theta_3)) \int_0^{t_i} \lambda_{is}^*(\theta) ds \right) \right] \quad (4.47)$$

mit

$$\text{vek} \left(\exp(\eta_i(\boldsymbol{\beta}) + \eta_i(\theta_3)) \int_0^{t_i} \lambda_{is}^*(\theta) ds \right) = \left(\exp(\eta_1(\boldsymbol{\beta}) + \eta_1(\theta_3)) \int_0^{t_1} \lambda_{1s}^*(\theta) ds, \dots, \exp(\eta_n(\boldsymbol{\beta}) + \eta_n(\theta_3)) \int_0^{t_n} \lambda_{ns}^*(\theta) ds \right)^t .$$

Man beachte die Ähnlichkeit der Schätzfunktion für $\boldsymbol{\beta}$ mit der Schätzfunktion in einem Poisson'schen generalisierten linearen Modell, falls man $\boldsymbol{\delta}$ als abhängige Variable auffaßt, die allerdings nur die Werte 0 und 1 annimmt. Die obigen Darstellungen werden jetzt für die anderen Parameter übernommen.

Die marginale Schätzfunktion für a

$$\frac{\partial}{\partial a} \Lambda_i(a, \theta_{-a}) = \exp(\eta_i(\boldsymbol{\beta}) + \eta_i(\theta_3)) \int_0^{t_i} \frac{\partial}{\partial a} \eta_{is}(a) \lambda_{is}^*(\theta) ds \quad (4.48)$$

mit $\frac{\partial}{\partial a} \eta_{is}(a) = \mathbf{x}_{is}^t$, wobei dies ein Vektor ist. Summiert über die Individuen gilt

$$\frac{\partial}{\partial a} \Lambda_t(a, \theta_{-a}) = \sum_{i=1}^n \frac{\partial}{\partial a} \Lambda_i(a, \theta_{-a}) \quad . \quad (4.49)$$

Die marginale Schätzfunktion $G(\cdot)$ für a lautet ($\mathbf{X}_{as} = (\mathbf{x}_{1s}, \dots, \mathbf{x}_{ns})$)

$$G(a) = \mathbf{X}_{at}^t \boldsymbol{\delta} - \frac{\partial}{\partial a} \Lambda_t(a, \theta_{-a}) \quad (4.50)$$

Da die Kovariablen zeitabhängig sind, erhält man leicht veränderte Terme.

Die marginale Schätzfunktion für θ_1 (Typ 1)

Die partiellen Ableitungen nach den Parametern θ_1 des glatten Effektes für die Baseline lauten

$$\frac{\partial}{\partial \theta_1} \Lambda_{t_i}(\theta_1, \theta_{-\theta_1}) = \exp(\eta_i(\boldsymbol{\beta}) + \eta_i(\theta_3)) \int_0^{t_i} \frac{\partial}{\partial \theta_1} \eta_{is}(\theta_1) \lambda_{is}^*(\theta) ds \quad . \quad (4.51)$$

In Matrixnotation

$$\begin{aligned} \frac{\partial}{\partial \theta_1} \Lambda_t(\theta_1, \theta_{-\theta_1}) &= \begin{pmatrix} \sum_{i=0}^n \exp(\eta_i(\boldsymbol{\beta}) + \eta_i(\theta_3)) \int_0^{t_i} \frac{\partial}{\partial \alpha_{11}} \eta_{is}(\theta_1) \lambda_{is}^*(\theta) ds \\ \vdots \\ \sum_{i=0}^n \exp(\eta_i(\boldsymbol{\beta}) + \eta_i(\theta_3)) \int_0^{t_i} \frac{\partial}{\partial \mathbf{u}_K^{\alpha_1}} \eta_{is}(\theta_1) \lambda_{is}^*(\theta) ds \end{pmatrix} \\ &= \begin{pmatrix} \sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\alpha}_1} \Lambda_{t_i}(\boldsymbol{\alpha}_1, \theta_{-\boldsymbol{\alpha}_1}) \\ \sum_{i=1}^n \frac{\partial}{\partial \mathbf{u}^{\alpha_1}} \Lambda_{t_i}(\mathbf{u}^{\alpha_1}, \theta_{-\mathbf{u}^{\alpha_1}}) \end{pmatrix} \end{aligned}$$

mit

$$\frac{\partial}{\partial \theta_1} \eta_{is}(\theta_1) = \left(\underbrace{1, s^1, \dots, s^g}_{\dot{\eta}_{is}(\boldsymbol{\alpha}_1)}, \underbrace{(s - \kappa_1)_+^g, \dots, (s - \kappa_K)_+^g}_{\dot{\eta}_{is}(\mathbf{u}^{\alpha_1})} \right) \quad . \quad (4.52)$$

Die marginale Schätzfunktion $G(\cdot)$ für θ_1 ergänzt um den Strafterm $G_{\sigma_{\alpha_1}^2}(\mathbf{u}^{\alpha_1})$ lautet $(\dot{\boldsymbol{\eta}}_t(\boldsymbol{\alpha}_1))^t = (\dot{\eta}_{1t_1}, \dots, \dot{\eta}_{nt_n}), \dot{\boldsymbol{\eta}}_t(\mathbf{u}^{\alpha_1})^t$ analog)

$$G(\theta_1) = \begin{pmatrix} G(\boldsymbol{\alpha}_1) \\ G(\mathbf{u}^{\alpha_1}) - G_{\sigma_{\alpha_1}^2}(\mathbf{u}^{\alpha_1}) \end{pmatrix} \quad (4.53)$$

$$= \begin{pmatrix} \dot{\boldsymbol{\eta}}_t(\boldsymbol{\alpha}_1)^t \delta - \frac{\partial}{\partial \boldsymbol{\alpha}_1} \Lambda_t(\boldsymbol{\alpha}_1, \theta_{-\boldsymbol{\alpha}_1}) \\ \dot{\boldsymbol{\eta}}_t(\mathbf{u}^{\alpha_1})^t \delta - \frac{\partial}{\partial \mathbf{u}^{\alpha_1}} \Lambda_t(\theta_1, \theta_{-\mathbf{u}^{\alpha_1}}) - G_{\sigma_{\alpha_1}^2}(\mathbf{u}^{\alpha_1}) \end{pmatrix} \quad (4.54)$$

Die marginale Schätzfunktion für θ_2 (Typ 2)

Die Ableitungen nach den zeitvariierenden Effekten θ_2 lauten

$$\frac{\partial}{\partial \theta_2} \Lambda_{t_i}(\theta_2, \theta_{-\theta_2}) = \exp(\eta_i(\boldsymbol{\beta}) + \eta_i(\theta_3)) \int_0^{t_i} \frac{\partial}{\partial \theta_2} \eta_{is}(\theta_2) \lambda_{is}^*(\theta) ds \quad . \quad (4.55)$$

In Matrixnotation

$$\begin{aligned} \frac{\partial}{\partial \theta_2} \Lambda_t(\theta_2, \theta_{-\theta_2}) &= \begin{pmatrix} \sum_{i=0}^n \exp(\eta_i(\boldsymbol{\beta}) + \eta_i(\theta_3)) \int_0^{t_i} \frac{\partial}{\partial \alpha_{21}} \eta_{is}(\theta_2) \lambda_{is}^*(\theta) ds \\ \vdots \\ \sum_{i=0}^n \exp(\eta_i(\boldsymbol{\beta}) + \eta_i(\theta_3)) \int_0^{t_i} \frac{\partial}{\partial \mathbf{u}_{2K}^\alpha} \eta_{is}(\theta_2) \lambda_{is}^*(\theta) ds \end{pmatrix} \\ &= \begin{pmatrix} \sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\alpha}_2} \Lambda_{t_i}(\boldsymbol{\alpha}_2, \theta_{-\boldsymbol{\alpha}_2}) \\ \sum_{i=1}^n \frac{\partial}{\partial \mathbf{u}^{\alpha_2}} \Lambda_{t_i}(\mathbf{u}^{\alpha_2}, \theta_{-\mathbf{u}^{\alpha_2}}) \end{pmatrix} \end{aligned}$$

und mit

$$\frac{\partial}{\partial \theta_2} \eta_{is}(\theta_2) = x_{ij} \left(\underbrace{1, s^1, \dots, s^g}_{\dot{\eta}_{is}(\boldsymbol{\alpha}_2)}, \underbrace{(s - \kappa_1)_+^g, \dots, (s - \kappa_K)_+^g}_{\dot{\eta}_{is}(\mathbf{u}^{\alpha_2})} \right) \quad (4.56)$$

Die marginale Schätzfunktion $G(\cdot)$ für θ_2 , ergänzt um den Strafterm $G_{\sigma_{\alpha_2}^2}(\mathbf{u}^{\alpha_2})$, sowie mit $(\dot{\boldsymbol{\eta}}_t(\boldsymbol{\alpha}_2))^t = (\dot{\eta}_{1t_1}, \dots, \dot{\eta}_{nt_n})$, $\dot{\boldsymbol{\eta}}_t(\mathbf{u}^{\alpha_2})^t$ analog

$$G(\theta_2) = \begin{pmatrix} G(\boldsymbol{\alpha}_2) \\ G(\mathbf{u}^{\alpha_2}) - G_{\sigma_{\alpha_2}^2}(\mathbf{u}^{\alpha_2}) \end{pmatrix} \quad (4.57)$$

$$= \begin{pmatrix} \dot{\boldsymbol{\eta}}_t(\boldsymbol{\alpha}_2)^t \delta - \frac{\partial}{\partial \boldsymbol{\alpha}_2} \Lambda_t(\boldsymbol{\alpha}_2, \theta_{-\boldsymbol{\alpha}_2}) \\ \dot{\boldsymbol{\eta}}_t(\mathbf{u}^{\alpha_2})^t \delta - \frac{\partial}{\partial \mathbf{u}^{\alpha_2}} \Lambda_t(\theta_2, \theta_{-\mathbf{u}^{\alpha_2}}) - G_{\sigma_{\alpha_2}^2}(\mathbf{u}^{\alpha_2}) \end{pmatrix} \quad (4.58)$$

Die marginale Schätzfunktion für θ_3 (Typ 3)

Abschließend erhält man die Ableitungen für die Parameter bezüglich des glatten Effekts θ_3

$$\frac{\partial}{\partial \theta_3} \Lambda_t(\theta_3, \theta_{-\theta_3}) = \frac{\partial}{\partial \theta_3} \eta_{is}(\theta_3) \exp(\eta_i(\boldsymbol{\beta}) + \eta_i(\theta_3)) \int_0^{t_i} \lambda_{is}^*(\theta) ds, \quad (4.59)$$

wobei

$$\frac{\partial}{\partial \theta_3} \eta_{is}(\theta_3) = \left(\underbrace{1, x_i^1, \dots, x_i^g}_{\mathbf{x}_{i\alpha_3}}, \underbrace{(x_i - \kappa_1)_+^g, \dots, (x_i - \kappa_K)_+^g}_{\mathbf{x}_{iu^{\alpha_3}}} \right) =: \mathbf{x}_{i\theta_3} \quad (4.60)$$

gilt. In Matrixnotation gilt

$$\begin{aligned} \frac{\partial}{\partial \theta_3} \Lambda_t(\theta_3, \theta_{-\theta_3}) &= \begin{pmatrix} \sum_{i=0}^n \exp(\eta_i(\boldsymbol{\beta}) + \eta_i(\theta_3)) 1 \int_0^{t_i} \lambda_{is}^*(\theta) ds \\ \vdots \\ \sum_{i=0}^n \exp(\eta_i(\boldsymbol{\beta}) + \eta_i(\theta_3)) (x_i - \kappa_K)_+^g \int_0^{t_i} \lambda_{is}^*(\theta) ds \end{pmatrix} \\ &= \begin{pmatrix} \sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\alpha}_3} \Lambda_t(\boldsymbol{\alpha}_3, \theta_{-\boldsymbol{\alpha}_3}) \\ \sum_{i=1}^n \frac{\partial}{\partial \mathbf{u}^{\alpha_3}} \Lambda_t(\mathbf{u}^{\alpha_3}, \theta_{-\mathbf{u}^{\alpha_3}}) \end{pmatrix}. \end{aligned}$$

Die marginale Schätzfunktion $G(\cdot)$ für θ_3 (mit $\mathbf{X}_{\alpha_3} = (\mathbf{x}_{1\alpha_3}, \dots, \mathbf{x}_{n\alpha_3})$, $\mathbf{X}_{u^{\alpha_3}}$ analog) und Strafterm $G_{\sigma_{\alpha_3}^2}(\mathbf{u}^{\alpha_3})$

$$G(\theta_2) = \begin{pmatrix} G(\boldsymbol{\alpha}_3) \\ G(\mathbf{u}^{\alpha_3}) - G_{\sigma_{\alpha_3}^2}(\mathbf{u}^{\alpha_3}) \end{pmatrix} \quad (4.61)$$

$$= \begin{pmatrix} \mathbf{X}_{\alpha_3}^t \delta - \frac{\partial}{\partial \boldsymbol{\alpha}_3} \Lambda_t(\boldsymbol{\alpha}_3, \theta_{-\boldsymbol{\alpha}_3}) \\ \mathbf{X}_{u^{\alpha_3}}^t \delta - \frac{\partial}{\partial \mathbf{u}^{\alpha_3}} \Lambda_t(\theta_3, \theta_{-\mathbf{u}^{\alpha_3}}) - G_{\sigma_{\alpha_3}^2}(\mathbf{u}^{\alpha_3}) \end{pmatrix} \quad (4.62)$$

Bei der Lösung der nichtlinearen Gleichungssysteme für die Parameter spielen die zweiten partiellen Ableitungen neben den ersten partiellen Ableitungen eine besondere Rolle.

Hier erhält man als Pendant zu den zweiten partiellen Ableitungen die Martingal Information als Verallgemeinerung der Fisher Information (vgl. Heyde (1997)).

Die Martingal Information für marginale Schätzfunktion

Die Formel für die Martingal Information lautet

$$I_G(\theta) = \sum_{i=1}^n I_{G_i}(\theta) = \sum_{i=1}^n \int_0^{t_i} \left(\frac{\partial}{\partial \theta} \lambda_{is}(\theta) \right)^2 \lambda_{is}(\theta)^{-1} ds \quad (4.63)$$

bzw. im multivariaten Fall als Matrix

$$I_{G_i}(\theta) = \int_0^{t_i} \left(\frac{\partial}{\partial \theta} \lambda_{is}(\theta) \right) \lambda_{is}^{-1}(\theta) \left(\frac{\partial}{\partial \theta} \lambda_{is}(\theta) \right)^t ds \quad . \quad (4.64)$$

Dies läßt sich vereinfachen zu

$$I_{G_i}(\theta) = \int_0^{t_i} \eta_{is}(\theta) \lambda_{is} \eta_{is}(\theta)^t ds \quad . \quad (4.65)$$

In Matrixnotation erhält man für $\boldsymbol{\beta}$ mit

$$I_G(\boldsymbol{\beta}) = \sum_{i=1}^n I_{G_i}(\boldsymbol{\beta}) = \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^t \exp(\eta_i(\boldsymbol{\beta}) + \eta_i(\theta_3)) \int_0^{t_i} \lambda_{is}^*(\theta) ds \quad . \quad (4.66)$$

Es folgen die Matrizen für die anderen Parameter. Für die als Zufallseffekte betrachteten Parameter der jeweiligen glatten Effekte muß noch die Ableitung der Strafterme (vgl. 4.42) $G_{\sigma_{\alpha_1}^2}(\mathbf{u}^{\alpha_1}), \dots, G_{\sigma_{\alpha_3}^2}(\mathbf{u}^{\alpha_3})$ mitberücksichtigt werden. Dies führt zu Diagonalmatrizen

$$D(\sigma_{\alpha_i}^2) = \frac{1}{\sigma_{\alpha_i}^2} \text{Diag}(\underbrace{0, \dots, 0}_{p \text{ mal}}, \underbrace{1, \dots, 1}_{q \text{ mal}}) \quad i = 1, \dots, 3 \quad , \quad (4.67)$$

wobei p die Anzahl der festen, q die Anzahl der zufälligen Parameter bezeichnet. In Matrixnotation gilt für die zweiten Ableitungen bezüglich a

$$I_G(a) = \sum_{i=1}^n I_{G_i}(a) = \sum_{i=1}^n \exp(\eta_i(\boldsymbol{\beta}) + \eta_i(\theta_3)) \int_0^{t_i} \mathbf{x}_{is} \mathbf{x}_{is}^t \lambda_{is}^*(\theta) ds \quad (4.68)$$

bzw. θ_1 (Typ 1)

$$\begin{aligned} I_G(\theta_1) &= \sum_{i=1}^n I_{G_i}(\theta_1) + D(\sigma_{\alpha_1}^2) \\ &= \sum_{i=1}^n \exp(\eta_i(\boldsymbol{\beta}) + \eta_i(\theta_3)) \int_0^{t_i} \eta_{is}(\theta_1) \eta_{is}(\theta_1)^t \lambda_{is}^*(\theta) ds + D(\sigma_{\alpha_1}^2) \end{aligned}$$

bzw. θ_2 (Typ 2)

$$\begin{aligned} I_G(\theta_2) &= \sum_{i=1}^n I_{G_i}(\theta_2) + D(\sigma_{\alpha_2}^2) \\ &= \sum_{i=1}^n \exp(\eta_i(\boldsymbol{\beta}) + \eta_i(\theta_3)) \int_0^{t_i} \eta_{is}(\theta_2) \eta_{is}(\theta_2)^t \lambda_{is}^*(\theta) ds + D(\sigma_{\alpha_2}^2) \end{aligned}$$

und θ_3 (Typ 3)

$$\begin{aligned} I_G(\theta_3) &= \sum_{i=1}^n I_{G_i}(\theta_3) + D(\sigma_{\alpha_3}^2) = \\ &= \sum_{i=1}^n \mathbf{x}_{ij\theta_3} \mathbf{x}_{ij\theta_3}^t \exp(\eta_i(\boldsymbol{\beta}) + \eta_i(\theta_3)) \int_0^{t_i} \lambda_{is}^*(\theta) ds + D(\sigma_{\alpha_3}^2) \end{aligned}$$

Schätzung der Varianzparameter

In Abschnitt 4.3.1 bzw. wurde bereits die Bias Korrektur der Varianzparameter angesprochen. Sie wurde im Kontext von Longitudinaldaten entwickelt (Jørgensen und Knudsen, 2004). Hier soll der Ansatz auf die in diesem Abschnitt vorgestellten Survivalmodelle übertragen werden. Die Modellierung der (für alle Individuen) gemeinsamen Baseline-Hazardrate bzw. der zeitabhängigen Kovariableneffekte mit zufälligen Parametern führt dazu, dass der Response (die individuellen Ereigniszeitpunkte) nur bedingt unabhängig ist. Zur Vereinfachung der Notation wird in diesem Abschnitt angenommen, dass es nur einen zeitabhängigen Kovariableneffekt gibt. Betrachtet man die Zensierungsvariablen δ_i , $i = 1, \dots, n$ als abhängige Variable, so gilt approximativ, dass (siehe Pawitan (2001), bzw. Cai et al. (2002)) die Zufallsgrößen δ_i gegeben den Vektor \mathbf{u}^{α_1} bedingt unabhängig und poissonverteilt sind mit Parameter $\exp(\eta_i(\boldsymbol{\beta}) + \eta_i(\theta_1) + o_i)$ ($= \mathbf{E}(\delta_i | \mathbf{u}^{\alpha_1})$). Der Term $o_i = \ln(\int_0^{t_i} \lambda_{is}^*(\theta) ds)$ wird als Offset bezeichnet, für den keine Parameter geschätzt werden. Da die Zufallseffekte $u_k^{\alpha_1}$ unabhängig normalverteilt sind

mit Varianz $\sigma_{\alpha_1}^2$, folgt, dass (für das Individuum i) $u^{i\alpha_1} = x_{i1} \sum_{k=1}^K u_k^{\alpha_1} (T_i - \kappa_k)_+^g$ normalverteilt ist mit $\mathbf{E}(u^{\alpha_1}) = 0$ und Varianz $\sigma_{u^{i\alpha_1}}^2 = \sigma_{\alpha_1}^2 x_{i1}^2 \sum_{k=1}^K (T_i - \kappa_k)_+^{2g}$. Bezeichne $\eta_{if}(\theta_1)$ den polynomialen Anteil des glatten Prädiktors mit festen Effekten, so lautet der marginale Erwartungswert von δ_i unter Verwendung der Momente der Log-Normalverteilung und der Zerlegung $\eta_i(\theta_1) = \eta_{if}(\theta_1) + u^{i\alpha_1}$

$$\begin{aligned}
 \mathbf{E}(\delta_i) &= \mathbf{E}(\mathbf{E}(\delta_i | u^{\alpha_1})) \\
 &= \mathbf{E}(\exp(\eta_i(\boldsymbol{\beta}) + \eta_i(\theta_1) + o_i)) \\
 &= \exp(\eta_i(\boldsymbol{\beta}) + \eta_{if}(\theta_1) + o_i) \mathbf{E}(\exp(u^{i\alpha_1})) \\
 &= \exp(\eta_i(\boldsymbol{\beta}) + \eta_{if}(\theta_1) + o_i) \exp\left(\frac{\sigma_{u^{i\alpha_1}}^2}{2}\right) \\
 &=: \mu_i
 \end{aligned}$$

bzw. (mit $\mathbf{E}(\mathbf{E}(\delta_i | u^{\alpha_1})) = \mathbf{E}(\mathbb{V}(\delta_i | u^{\alpha_1}))$ aufgrund der Verteilungsannahme)

$$\begin{aligned}
 \mathbb{V}(\delta_i) &= \mathbf{E}(\mathbb{V}(\delta_i | u^{\alpha_1}) + \mathbb{V}(\mathbf{E}(\delta_i | u^{\alpha_1}))) \\
 &= \mu_i + \mathbf{E}\left(\mathbf{E}(\delta_i | u^{\alpha_1}) - \mu_i\right)^2 \\
 &= \mu_i + \exp\left(2(\eta_i(\boldsymbol{\beta}) + \eta_{if}(\theta_1) + o_i)\right) \mathbb{V}(\exp(u^{i\alpha_1})) \\
 &= \mu_i + \exp\left(2(\eta_i(\boldsymbol{\beta}) + \eta_{if}(\theta_1) + o_i)\right) (\exp(2\sigma_{u^{i\alpha_1}}^2) - \exp(\sigma_{u^{i\alpha_1}}^2)) \\
 &= \mu_i + \exp\left(2(\eta_i(\boldsymbol{\beta}) + \eta_{if}(\theta_1) + o_i)\right) \exp(\sigma_{u^{i\alpha_1}}^2) (\exp(\sigma_{u^{i\alpha_1}}^2) - 1) \\
 &= \mu_i + \mu_i^2 (\exp(\sigma_{u^{i\alpha_1}}^2) - 1)
 \end{aligned}$$

Aufgrund der bedingten Unabhängigkeit gilt $\mathbb{C}(\delta_i, \delta_j | u^{\alpha_1}) = 0$, $i \neq j$. Somit berechnet sich die marginale Korrelation als

$$\begin{aligned}
 \mathbb{C}(\delta_i, \delta_j) &= \mathbf{E}(\mathbb{C}(\xi_i, \xi_j | u^{\alpha_1})) + \mathbf{E}\left([\mathbf{E}(\xi_i | u^{\alpha_1}) - \mathbf{E}(\xi_i)][\mathbf{E}(\xi_j | u^{\alpha_1}) - \mathbf{E}(\xi_j)]\right) \\
 &= 0 + \mathbf{E}[\mathbf{E}(\delta_i | u^{\alpha_1}) \mathbf{E}(\delta_j | u^{\alpha_1})] - \mathbf{E}(\delta_i) \mathbf{E}(\delta_j) \\
 &= \mathbf{E}[\exp(\eta_i(\boldsymbol{\beta}) + \eta_i(\theta_1) + o_i) \exp(\eta_j(\boldsymbol{\beta}) + \eta_j(\theta_1) + o_j)] - \mu_i \mu_j \\
 &= \exp(\eta_i(\boldsymbol{\beta}) + \eta_{if}(\theta_1) + o_i + \eta_j(\boldsymbol{\beta}) + \eta_{jf}(\theta_1) + o_j) \mathbf{E}(\exp(u^{i\alpha_1} u^{j\alpha_1})) - \mu_i \mu_j \\
 &= \exp(\eta_i(\boldsymbol{\beta}) + \eta_{if}(\theta_1) + o_i + \eta_j(\boldsymbol{\beta}) + \eta_{jf}(\theta_1) + o_j) (\exp(\sigma_{u^{i\alpha_1}}^2 \sigma_{u^{j\alpha_1}}^2)) - \mu_i \mu_j \\
 &= \mu_i \mu_j [\exp\left(\frac{1}{2} \sigma_{u^{i\alpha_1}}^2 \sigma_{u^{j\alpha_1}}^2\right) - 1]
 \end{aligned}$$

Damit wird die Mittelwert- bzw. die Varianz- bzw. die Korrelationsstruktur zur Verfügung gestellt um mit dem (pseudo) poissonverteilten Response Pearsonsche Schätzfunktionen aufzubauen (s. 4.3.1).

Methodik zur Nullstellenbestimmung der Schätzfunktionen

Die Berechnung der Martingal Information ist aufgrund der – außer im Fall linearer Splines – numerischen Integration durchaus aufwendig, insbesondere bei den Parametervektoren θ_1 und θ_2 , bei denen über alle Matrizenkomponenten integriert werden muß. Bei den Martingal Informationen bezüglich θ_1 und θ_2 könnten zudem numerische Instabilitäten entstehen, da aufgrund der Multiplikation der Terme $\eta_{is}(\cdot)$ und $\eta_{is}(\cdot)^t$ Potenzen der Ordnung $2g$ auftreten, d.h. bei kubischen Splines erhält man Potenzen bis zum sechsten Grad. Ein Verfahren, das möglicherweise nicht ganz so schnell konvergiert, dafür aber den Ruf größerer Robustheit besitzt, ist das Quasi Newton Verfahren, besser die Klasse der Quasi Newton Verfahren (Stoer, 1999). Sie unterscheiden sich dadurch vom Newton Verfahren, dass man zwar ebenso die Information des Gradienten über die Abstiegsrichtung ausnutzt, aber anstelle der Hessematrix H eine einfache Approximation wählt. Der Algorithmus

$$\theta^{k+1} = \theta^k - s^k \cdot d^k \quad (4.69)$$

mit $d^k = H_k^{-1}G^k$ läßt sich umformen zu

$$H_k(\theta^{k+1} - \theta^k) = -s^k G^k \quad . \quad (4.70)$$

Anstelle der inversen Hessematrix muß nun ein Gleichungssystem mit den Unbekannten $\Delta^{k+1} = \theta^{k+1} - \theta^k$ gelöst werden. Bei den Quasi Newton Verfahren verwendet man anstelle von H_k eine Approximation B_k . Für $k = 0$ setzt man entweder $B^0 = I$, d.h. die Einheitsmatrix, dann entspricht der erste Schritt dem Gradientenverfahren oder aber man verwendet eine Differenzenapproximation ∇G basierend auf dem Gradienten $G(\cdot)$ mit Komponenten l und Einheitsvektoren e_l gemäß

$$\nabla G_l = \frac{1}{h}(G(\theta + he_l) - G(\theta)) \quad . \quad (4.71)$$

Da die Hessematrix symmetrisch ist, sollte eine symmetrisierte Form der Differenzenapproximation verwendet werden (Gill, Murray und Wright, 1993). Bei der Berechnung

von B_{k+1} wird eine iterative Formel verwendet

$$B_{k+1} = B_k + U_k \quad , \quad (4.72)$$

wobei

$$U_k = \frac{1}{c^k} (y^k - B_k \Delta^k) (y^k - B_k \Delta^k)^t \quad , \quad (4.73)$$

mit $y^k = G(\theta^{k+1}) - G(\theta^k)$ und $c^k = (y^k - B_k \Delta^k)^t \Delta^k$ gilt. Die Matrix U_k hat den Rang 1. Man spricht daher auch vom symmetrischen Rang 1 Verfahren. Die iterative Berechnung ist nützlich bei der Lösung des Gleichungssystems

$$B_k \Delta^k = -s^k G^k \quad . \quad (4.74)$$

Da die Matrix B_k symmetrisch ist, läßt sich die Cholesky Zerlegung verwenden

$$B_k = L_k D_k L_k^t \quad (4.75)$$

$$= L_k D_k^{\frac{1}{2}} D_k^{\frac{1}{2}} L_k^t \quad (4.76)$$

$$= \bar{L}_k \bar{L}_k^t \quad (4.77)$$

mit \bar{L}_k (bzw. L_k und Diagonalmatrix D_k) als untere Dreiecksmatrix (die Diagonalelemente von L_k haben alle den Wert Eins). Setzt man $u^k = \frac{1}{\sqrt{c^k}} (y^k - B_k \Delta^k)$, dann gilt

$$B_{k+1} = L_k D_k L_k^t + u^k (u^k)^t \quad (4.78)$$

$$= L_k (D_k + p^k (p^k)^t) L_k^t \quad (4.79)$$

$$= \hat{L}_k \hat{D}_k \hat{L}_k^t \quad (4.80)$$

mit $L_k p^k = u^k$. Diese Berechnung ist effizienter als eine (vollständige) Neuberechnung der Cholesky Zerlegung (Gill et al., 1993, S. 42), da die Elemente von \hat{D}_k und \hat{L}_k in einem Schritt aus den Elementen von D_k und L_k berechnet werden können. In der Nähe der Nullstelle kann B als Näherung der Hessematrix aufgefaßt werden. Eine weitere Einstellgröße des Algorithmus ist die Schrittlänge s^k . Da die Lösung eines nichtlinearen Gleichungssystems $G(\theta)$ (theoretisch) gleichbedeutend ist mit der Minimierung der Funktion $g(\theta) = \frac{1}{2} G^t(\theta) G(\theta)$, läßt sich die Wahl von s^k anhand der Funktion $g(\theta)$ motivieren. Offensichtlich soll s^k so gewählt werden, dass $g(\theta^{k+1}) < g(\theta^k)$ gilt. Formuliert

man dies als eindimensionales Optimierungsproblem, so muß man die Funktion

$$m(s^k) = g(\theta^k - s^k \Delta^k) \quad (4.81)$$

minimieren. In der Literatur wird eine approximative Lösung vorgeschlagen, nämlich $m(s^k)$ durch ein kubisches Polynom zu ersetzen (Press et al., 1992). Die adaptive Wahl der Schrittweite dürfte insbesondere zu Beginn des Algorithmus wichtig sein. Nahe der Nullstelle ist zu vermuten, dass die quadratische Taylor Approximation, die dem Lösungsvorgehen zugrunde liegt, gut paßt. Das Vorgehen läßt sich folgendermaßen zusammenfassen:

1. Im $(k = 0)$ –ten Schritt werden verwendet:

Startwert: θ^0 ,

Gradient: $G(\theta^0)$,

Quasi Hessematrix:

$$B_0 = I \text{ bzw. } B_0 = \nabla G(\theta^0),$$

$$B_0 = L_0 L_0^t,$$

Δ^0 ist Lösung von:

$$L_0 L_0^t \Delta^0 = -G(\theta^0),$$

Durchführung der Schrittweitenbestimmung s_0 ($s_0 \leq 1$)

mit θ^0, Δ^0 und

Funktion: $g(\theta^0)$,

Gradient:

$$\nabla g(\theta^0) \approx B_0 G(\theta^0),$$

Resultat: s_0 ,

berechne θ^1 ;

2. im $(k + 1)$ –ten Schritt werden verwendet:

Wert θ^k ,

Gradient $G(\theta^k)$,

Quasi Hessematrix:

$$B_{k+1} = B_k + u^k (u^k)^t$$

$$B_{k+1} = L_{k+1} L_{k+1}^t,$$

Δ_{k+1} ist Lösung von:

$$L_{k+1} L_{k+1}^t \Delta_{k+1} = -G(\theta^k),$$

Durchführung der Schrittweitenbestimmung s_{k+1} ($s_{k+1} \leq 1$)

mit $\theta^k, g(\theta^k), \Delta_{k+1}$,

Funktion: $g(\theta^k)$,

Gradient:

$$\nabla g(\theta^k) \approx B_{k+1} G(\theta^k),$$

Resultat: s_{k+1} ,

Resultat: θ^{k+1} ;

Der Algorithmus endet, wenn $G(\theta) \approx 0$ gilt. Hierzu noch ein Zitat (Press et al., 1992) (in runden Klammern stehen die oben verwendeten Symbole):

Note that this method essentially minimizes $f(g)$ by taking [...] steps designed to bring $F(G)$ to zero. This is *not* equivalent to minimizing $f(g)$ directly by taking [...] steps to bring $\nabla f(\nabla g)$ to zero.

Hier handelt es sich bei den Gradienten um die jeweiligen Schätzfunktionen $G(\cdot)$ und bei den Hesse Matrizen H um die jeweiligen Martingal Informationen.

Es wird ein hybrides Vorgehen verfolgt. Beim blockweisen iterativen Vorgehen wird für β und θ_3 die Martingal Information verwendet, dies entspricht dem Newton Verfahren, für den zeitlichen Effekt (θ_1) und den Effekt der zeitvariierenden Koeffizienten (θ_2) werden Lösungen gemäß den Quasi Newton Verfahren berechnet.

Methodik zur numerischen Integration der Schätzfunktionen

Es existiert eine Vielzahl von Verfahren (Stoer, 1999), bei denen das Integral $\int f(x) dx$ möglichst gut durch eine endliche Summe $\sum_{i=1}^n w_i f(x_i)$ approximiert werden soll. Sie unterscheiden sich durch die Wahl der Stützstellen x_i , die Gewichte w_i und den Integrationsbereich. Üblicherweise werden sie so gewählt, dass 'einfache' Funktionen (meist Polynome) exakt integriert werden können. Da die Funktionsauswertungen rechenintensiv sind, ist eine geringe Anzahl n von Stützstellen vorteilhaft. Im Rahmen dieser Arbeit handelt es sich bei der zu integrierenden Funktion um $\lambda_{is}^*(\cdot)$ bzw. den Vektor $\hat{\eta}_{is}(\theta_{(\cdot)}) \lambda_{is}^*(\theta)$. Die Funktionen sind pro Individuum verschieden, ebenso ist der Integrationsbereich individuenabhängig $[0, t_i]$. Der hier vorgeschlagene Weg folgt (Mason und Handscomb, 2003) sowie (Press et al., 1992, S. 192). Er basiert auf der Verwendung von Chebychev Polynomen T_n der ersten Art⁴. Auf dem Bereich $[-1, 1]$ bilden die Chebychev Polynome bezüglich einer Gewichtsfunktion w ein Orthogonalsystem bezüglich der quadratischen Norm, d.h. definiert man als inneres Produkt $\langle f, g \rangle := \int_{-1}^1 w(x) f(x) g(x) dx$, so gilt $\langle T_m, T_n \rangle = 0$, $m \neq n$. Weiter existiert für quadratintegrier-

⁴dieser Zusatz wird im folgenden stets vorausgesetzt

bare Funktionen eine Reihenentwicklung, d.h.

$$f(x) \approx \sum_{i=0}^{\infty} c_i T_i(x) \quad \text{mit} \quad c_i = \frac{\langle f, T_i \rangle}{\langle T_i, T_i \rangle} .$$

Die Koeffizienten c_i lassen sich im allgemeinen wegen der dafür nötigen Integration nicht exakt berechnen. Zur numerischen Integration kann man ein einfaches Integrationsverfahren verwenden, z.B. die Sehnen-Trapezregel mit den Stützstellen $x_i = \cos\left(\frac{\pi(i-0.5)}{n}\right)$, $i = 1, \dots, n$. Dies führt zu

$$c_i = \frac{2}{n} \sum_{k=1}^n f(x_k) T_i(x_k) . \quad (4.82)$$

Bemerkenswerterweise bleibt die Orthogonalität an den Stützstellen erhalten, d.h. es gilt $\sum_{i=1}^n T_k(x_i) T_l(x_i) = 0$, $k, l < n, k \neq l$. Für das praktische Rechnen wird man nur eine endliche Anzahl von Chebychev Polynomen verwenden. Die Funktion f wird in den Stützstellen durch eine endliche Reihe interpoliert. Eine weitere sehr angenehme Eigenschaft der gewählten Reihendarstellung ist, dass die Chebychev Polynome ein Verhalten besitzen, das den Minimax Polynomen ähnelt, d.h. dass der maximale Abstand zu der interpolierten Funktion minimal wird. Für diese theoretisch schönen Eigenschaften hält sich der dafür benötigte algorithmische Aufwand in Grenzen. Zuerst muß der Integrationsbereich von $[0, t_i]$ auf $[-1, 1]$ transformiert werden. Danach wird eine Approximation mittels Chebychev Polynomen

$$T_n(x) = \cos(n \arccos x) \quad (4.83)$$

durchgeführt. Es existiert auch eine rekursive Darstellung der T_n mit Polynomen, die beim praktischen Rechnen verwendet wird, wobei auf die numerische Stabilität des Rekursionsverfahren geachtet werden muß (Press et al., 1992, S. 193). Die Stützstellen x_i entsprechen den Nullstellen des Chebychev Polynoms. Die Approximation für f lautet

$$f(x) \approx \sum_{i=1}^n c_i T_i(x) - 0.5c_0 \quad (4.84)$$

mit Koeffizienten c_i , die gemäß folgender Formel berechnet werden⁵

$$c_i = \frac{2}{n} \sum_{k=1}^n f(x_k) T_i(x_k) \quad (4.85)$$

$$= \frac{2}{n} \sum_{k=1}^n f \left(\cos \left(\frac{\pi(k-0.5)}{n} \right) \right) \cos \left(\frac{i\pi(k-0.5)}{n} \right) \quad . \quad (4.86)$$

Typischerweise werden die Koeffizienten c_n schnell klein. Da die Polynome nur Werte zwischen $[-1, 1]$ annehmen, wird der Approximationsfehler durch den Term $c_n T_n(x)$ bestimmt – man beachte die Orthogonalitätseigenschaft der Polynome. Ein weiterer Vorteil dieses Vorgehens ist, dass eine Formel existiert, die es ermöglicht, die Koeffizienten c_i passend zur Funktion f umzurechnen in Koeffizienten C_i , die die Stammfunktion von f approximieren

$$C_i = \frac{c_{i-1} - c_{i+1}}{2i} \quad , i > 0 \quad . \quad (4.87)$$

Mit den Koeffizienten C_i kann nun das Integral für beliebiges $0 < y < t_i$ berechnet werden (C_0 fungiert als Normierungskonstante)

$$\int_0^y f(x) dx \approx \sum_{i=1}^n C_i T_i(y) - 0.5C_0 \quad (4.88)$$

bzw.

$$\int_y^{t_i} f(x) dx = \int_0^{t_i} f(x) dx - \int_0^y f(x) dx \approx \sum_{i=1}^n C_i (T_i(t_i) - T_i(y)) \quad . \quad (4.89)$$

Betrachtet man das Integral der Funktion $f(x)$ als Stammfunktion, so erhält man eine Approximation durch Chebychev Polynome. Eine wiederholte Anwendung der Approximation liefert

$$C_i^1 = \frac{C_{i-1} - C_{i+1}}{2i} \quad , i > 0 \quad , \quad (4.90)$$

die Koeffizienten der Polynomdarstellung für das Integral der Stammfunktion. Auf diese Weise erhält man iterierte Integralapproximationen, die für die partielle Integration (siehe unten) benötigt werden. Die Approximation ist relativ effizient, man braucht nur für die Koeffizienten c_i aufwendige Funktionsauswertungen von $f(x)$ vornehmen. Bei

⁵Die Formel in Press et al. (1992) enthält in der verwendeten Auflage einen kleinen Fehler

den weiteren Koeffizienten C_i und C_i^1 , etc. genügt die einfache Update Formel. Die numerische Stabilität wird im folgenden Text noch analysiert.

Betrachtet man (z.B.) von der Schätzfunktion $G_i(\theta_1)$ den Term, der integriert werden muß und setzt man die untere Integrationsgrenze ein

$$\begin{pmatrix} \int_0^{t_i} 1 \lambda_{is}^*(\theta) ds \\ \int_0^{t_i} s^1 \lambda_{is}^*(\theta) ds \\ \vdots \\ \int_0^{t_i} s^g \lambda_{is}^*(\theta) ds \\ \int_0^{t_i} (s - \kappa_1)_+^g \lambda_{is}^*(\theta) ds \\ \vdots \\ \int_0^{t_i} (s - \kappa_K)_+^g \lambda_{is}^*(\theta) ds \end{pmatrix} = \begin{pmatrix} \int_0^{t_i} 1 \lambda_{is}^*(\theta) ds \\ \int_0^{t_i} s^1 \lambda_{is}^*(\theta) ds \\ \vdots \\ \int_0^{t_i} s^g \lambda_{is}^*(\theta) ds \\ \int_{\kappa_1}^{t_i} (s - \kappa_1)^g \lambda_{is}^*(\theta) ds \\ \vdots \\ \int_{\kappa_K}^{t_i} (s - \kappa_K)^g \lambda_{is}^*(\theta) ds \end{pmatrix}, \quad (4.91)$$

so erkennt man, dass nach dem Ausmultiplizieren von $(s - \kappa_k)^g$ die Integrale als Summen geschrieben werden können. Sei $g = 3$, dann gilt für $\kappa = 1, \dots, K$

$$\begin{aligned} \int_{\kappa_k}^{t_i} (s - \kappa_k)^3 \lambda_{is}^*(\theta) ds &= \int_{\kappa_k}^{t_i} s^3 \lambda_{is}^*(\theta) ds - 3\kappa_k \int_{\kappa_k}^{t_i} s^2 \lambda_{is}^*(\theta) ds \\ &\quad + 3\kappa_k^2 \int_{\kappa_k}^{t_i} s \lambda_{is}^*(\theta) ds - \kappa_k^3 \int_{\kappa_k}^{t_i} \lambda_{is}^*(\theta) ds, \end{aligned}$$

wobei die Integrale nur für $t_i > \kappa_k$ berechnet werden müssen. Anderenfalls sind sie Null. Schreibt man zur Abkürzung der Notation für das (un-) bestimmte Integral

$$\int_u^{t_i} \lambda_{is}^*(\theta) ds = I_i^0(u), \quad I_{is}^0(u) = \int_u^s \lambda_{ir}^*(\theta) dr \quad 0 \leq u \leq s \leq t_i$$

bzw. rekursiv ($n \geq 0$)

$$\int_u^{t_i} I_{is}^n(u) ds = I_i^{n+1}(u), \quad I_{is}^{n+1}(u) = \int_u^s I_{ir}^n(u) dr \quad 0 \leq u \leq s \leq t_i,$$

so gilt für die Integrale mit Hilfe der partiellen Integration

- $\int_0^{t_i} s \lambda_{is}^*(\theta) ds = t_i I_i^0(0) - I_i^1(0) =: \Delta I_i^0(0)$,
- $\int_0^{t_i} s^2 \lambda_{is}^*(\theta) ds = t_i^2 I_i^0(0) - 2t_i I_i^1(0) + 2I_i^2(0) =: \Delta I_i^1(0)$,
- $\int_0^{t_i} s^3 \lambda_{is}^*(\theta) ds = t_i^3 I_i^0(0) - 3t_i^2 I_i^1(0) + 6t_i I_i^2(0) - 6I_i^3(0) =: \Delta I_i^2(0)$,

sowie für $\kappa_k, k = 1, \dots, K$

- $\int_{\kappa_k}^{t_i} s \lambda_{is}^*(\theta) ds = t_i I_i^0(0) - \kappa_k I_{i\kappa_k}^0(0) - I_i^1(\kappa_k) =: \Delta I_i^0(\kappa_k)$,
- $\int_{\kappa_k}^{t_i} s^2 \lambda_{is}^*(\theta) ds = t_i^2 I_i^0(0) - \kappa_k^2 I_{i\kappa_k}^0(0) - 2t_i I_i^1(0) + 2\kappa_k I_{i\kappa_k}^1(0) + 2I_i^2(\kappa_k)$
 $=: \Delta I_i^1(\kappa_k)$,
- $\int_{\kappa_k}^{t_i} s^3 \lambda_{is}^*(\theta) ds = t_i^3 I_i^0(0) - \kappa_k^3 I_{i\kappa_k}^0(0) - 2t_i^2 I_i^1(0) + 2\kappa_k^2 I_{i\kappa_k}^1(0)$
 $+ 6t_i I_i^2(0) - 6\kappa_k I_{i\kappa_k}^2(0) - 6I_i^3(\kappa_k)$
 $=: \Delta I_i^2(\kappa_k)$,

und

$$\int_{\kappa_k}^{t_i} (s - \kappa_k)^3 \lambda_{is}^*(\theta) ds = \Delta I_i^2(\kappa_k) - 3\kappa_k \Delta I_i^1(\kappa_k) + 3\kappa_k^2 \Delta I_i^0(\kappa_k) - \kappa_k^3 I_i^0(\kappa_k) .$$

Im Fall von $G(\theta_2)$ gelten analoge Aussagen, es müssen nur die Terme vor den Integralen (rechte Seite) modifiziert werden. Zusätzlich muß noch für $G(a)$ die Integration der Funktion $x_{ijs} \lambda_{is}^*(\theta)$ durchgeführt werden.

Zusammenfassend bleibt festzustellen, dass man pro Individuum und pro zeitvariierendem Effekt nur die Integrale $I_i^0(u), I_i^1(u), I_i^2(u), I_i^3(u)$ für $u = 0$ bzw. $\kappa_k, k = 1, \dots, K$ approximieren muß. Das Vorgehen der zugehörigen Chebychev Approximation lautet (pro Individuum und zeitvariierendem Effekt)

1. berechne $\lambda_{is}^*(\theta)$ an den Integrationsstützstellen,
2. berechne die Koeffizienten c_i zur Approximation von $\lambda_{is}^*(\theta)$,
3. berechne die Koeffizienten C_i zur Approximation $\hat{I}_i^0(u)$ für $I_i^0(u)$ für den Bereich $[0, t_i]$, ‘nützliches Nebenprodukt’, die Approximationen $\hat{I}_i^0(\kappa_k)$ für $I_i^0(\kappa_k) k = 1, \dots, K$
4. berechne die Koeffizienten C_i^1 (aus C_i) zur Approximation $\hat{I}_i^1(u)$ für $I_i^1(u)$ sowie die Approximationen $\hat{I}_i^1(\kappa_k)$,
5. berechne die Koeffizienten C_i^2 (aus C_i^1) zur Approximation $\hat{I}_i^2(u)$ für $I_i^2(u)$ sowie die Approximationen $\hat{I}_i^2(\kappa_k)$,
6. berechne die Koeffizienten C_i^3 (aus C_i^2) zur Approximation $\hat{I}_i^3(u)$ für $I_i^3(u)$ sowie die Approximationen $\hat{I}_i^3(\kappa_k)$.

Die Eleganz der Chebychev Approximation liegt nun darin, dass man nur einmal die Funktionsauswertungen vornehmen muß, die Integrale $I_i^0(u), \dots, I_i^3(u), u = 0, \kappa_1, \dots, \kappa_K$ lassen sich sehr effizient ohne weitere Funktionsauswertungen berechnen. Entscheidend

ist natürlich die Anzahl der nötigen Integrationsstützstellen und die Stabilität der iterativen Integration. Dies wird anhand dreier Testfunktionen erörtert.

1. $f(x) = 1.2 \cos(x) + x$, Integrationsbereich $[0, 10]$,
2. $g(x) = 0.2 \exp(2x)$, Integrationsbereich $[0, 1]$,
3. $h(x) = \exp\left(0.25 \sum_{i=0}^3 x^i + 0.25 \sum_{j=1}^8 (x - g_j)_+^3\right)$,

wobei im letzten Fall ein äquidistantes Gitter (mit Gitterpunkten g_j) auf $[0, 1]$ angenommen wird. Für beide Funktionen werden die Integrale $I^j(\kappa_k)$, $j = 0, \dots, 3, k = 1, \dots, 5$ approximiert an den Stellen

1. $\kappa_k = 1.5(k - 1) + 1$, $k = 1, \dots, 5$,
2. $\kappa_k = 0.15(k - 1) + 0.1$, $k = 1, \dots, 5$,

d.h. die Integrationsbereiche entsprechen jeweils den Intervallen $[0, \kappa_k]$. In beiden Fällen lassen sich die Integrale exakt berechnen, so dass sich in Abhängigkeit von der Anzahl der Chebychev Polynome der absolute sowie der relative Fehler für die Integrale berechnen läßt. Der Tabelle entnimmt man die Fehler im Fall von f (E-7 = 10^{-7} , etc.). Offensichtlich wird die Approximation immer genauer, je mehr Polynome man verwendet. Bereits bei 30 Polynomen ist die Approximation praktisch exakt. Ebenfalls ersichtlich ist, dass die Approximation der iterierten Integrale bei wenigen Polynomen ($n = 15$) etwas schlechter ist.

		Anzahl der Chebychev Polynome n		
		15	20	30
absoluter Fehler	$\hat{I}_i^0(u)$	E-6 / E-7	E-10 / E-11	E-14 / E-15
	$\hat{I}_i^1(u)$	E-4 / E-6	E-9 / E-10	E-14 / E-15
	$\hat{I}_i^2(u)$	E-4 / E-6	E-8 / E-11	E-13 / E-14
	$\hat{I}_i^3(u)$	E-4 / E-5	E-8 / E-10	E-13 / E-15
relativer Fehler	$\hat{I}_i^0(u)$	E-6 / E-8	E-11 / E-12	E-15 / E-15
	$\hat{I}_i^1(u)$	E-7 / E-7	E-10 / E-11	E-14 / E-15
	$\hat{I}_i^2(u)$	E-5 / E-6	E-10 / E-10	E-14 / E-16
	$\hat{I}_i^3(u)$	E-4 / E-6	E-8 / E-10	E-13 / E-16

Detailliertere Angaben entnimmt man den Tabellen im Anhang (B.1, B.2, B.3). In übereinstimmender Weise zum Verhältnis der Größe des Fehler zur Anzahl der Polynome ist die grafische Darstellung der Koeffizienten c_i , $i = 1, \dots, 30$. Man sieht, dass die Koeffizienten selbst auf der logarithmischen Skala betragsmäßig sehr schnell gegen Null

gehen (siehe Grafiken (4.1), (4.2)). Da die Chebychev Polynome nur Werte aus dem Bereich $[-1, 1]$ annehmen, ist das jeweilige Größenverhältnis der Koeffizienten zueinander entscheidend für den Beitrag der korrespondierenden Polynome. Das Symbol \diamond bedeutet, dass der Koeffizient negativ ist, es wurde daher dessen Betrag gezeichnet. Analog

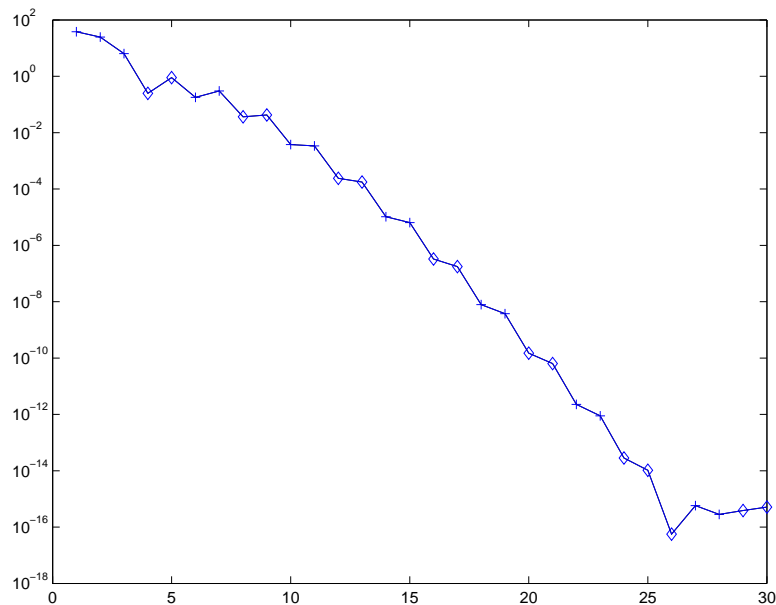


Abbildung 4.1: Beträge der Chebychev Koeffizienten für erste Stammfunktion von f , Darstellung auf logarithmischer Skala

zum geschilderten Fall kann man auch für die iterierten Integrale die Koeffizienten grafisch darstellen. Es ist ersichtlich, dass sich ein qualitativ gleiches Verhalten zeigt, die Koeffizienten werden extrem schnell klein.

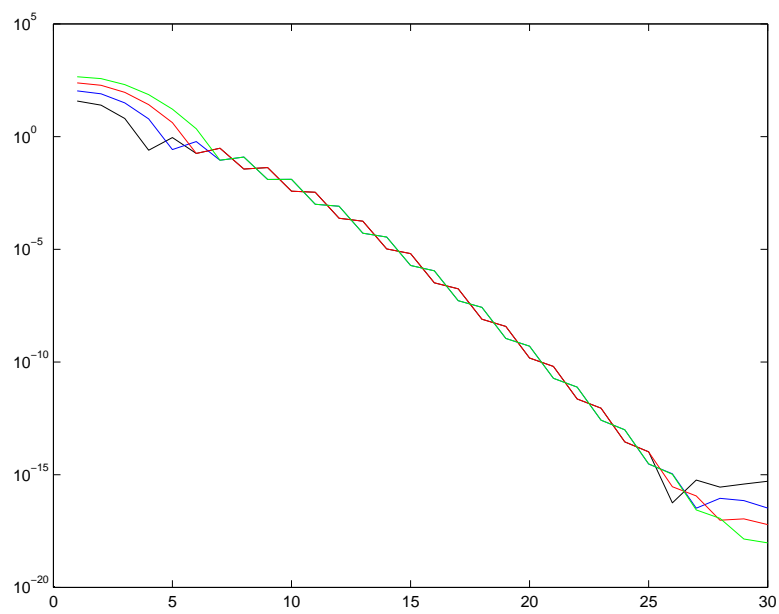


Abbildung 4.2: Beträge der Chebychev Koeffizienten für iterierte Stammfunktionen von f , Darstellung auf logarithmischer Skala

Eine analoge Analyse erfolgt für die Funktion $g(x)$. Betrachtet man die Tabelle sowie die Grafiken (siehe (4.3), (4.4)), so sieht man ein ähnliches Verhalten wie für die Funktion $f(x)$.

		Anzahl der Chebychev Polynome		
		15	20	30
absoluter Fehler	$\hat{I}_i^0(u)$	E-16 / E-16	E-16 / E-17	E-16 / E-18
	$\hat{I}_i^1(u)$	E-16 / E-16	E-17 / E-18	E-17 / E-18
	$\hat{I}_i^2(u)$	E-16 / E-17	E-18 / E-18	E-18 / E-19
	$\hat{I}_i^3(u)$	E-16 / E-16	E-18 / E-19	E-18 / E-19
relativer Fehler	$\hat{I}_i^0(u)$	E-14 / E-15	E-15 / E-16	E-15 / E-16
	$\hat{I}_i^1(u)$	E-13 / E-15	E-15 / E-16	E-15 / E-16
	$\hat{I}_i^2(u)$	E-12 / E-14	E-13 / E-16	E-14 / E-16
	$\hat{I}_i^3(u)$	E-11 / E-14	E-13 / E-16	E-13 / E-16

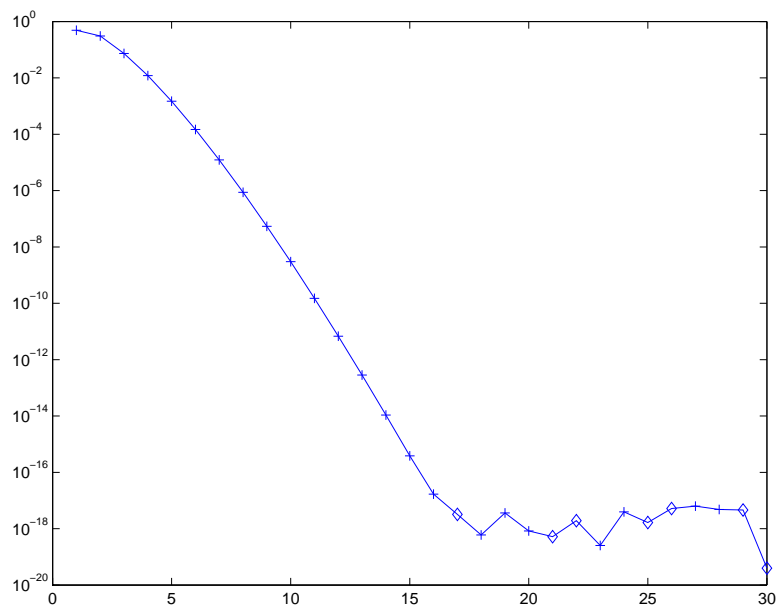


Abbildung 4.3: Beträge der Chebychev Koeffizienten für erste Stammfunktion von g , Darstellung auf logarithmischer Skala

Details zu den Fehlern findet man im Anhang (B.4, B.5, B.6). Im Gegensatz zu einfachen Verfahren (z.B. Sehnentrapezregel) bedarf es nur weniger Chebychev Polynome / Integrationsstützstellen um eine ausreichende Genauigkeit zu erhalten. Die Chebychev Polynome erweisen sich vermutlich aufgrund ihrer Orthogonalitätseigenschaft als ro-

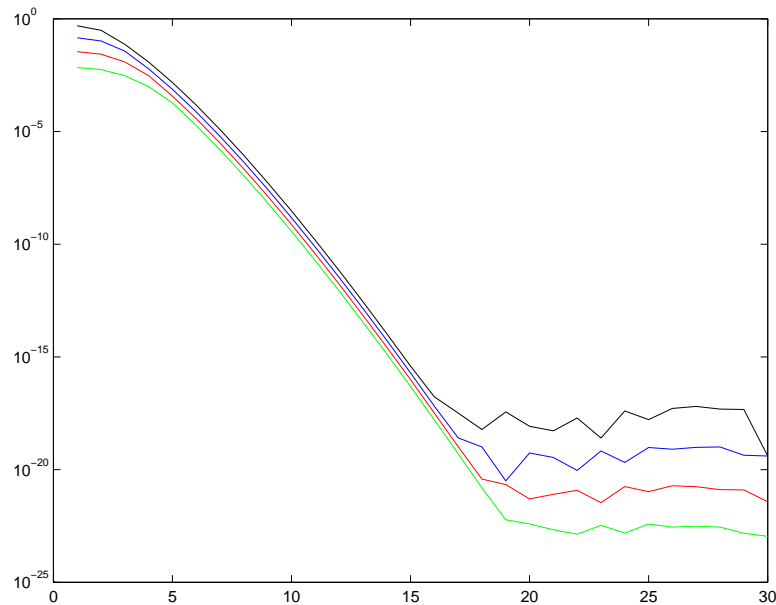


Abbildung 4.4: Beträge der Chebyshev Koeffizienten für iterierte Stammfunktionen von g , Darstellung auf logarithmischer Skala

bust und effizient. Im Fall der dritten Funktion

$$h(x) = \exp\left(0.25 \sum_{i=0}^3 x^i + 0.25 \sum_{j=1}^8 (x - g_j)_+^3\right)$$

existiert kein exakter Wert für die Integration, es erfolgt daher ein Vergleich zu einer adaptiven Integrationsroutine in Matlab, sowie ein Vergleich zur Integration mittels der Trapezregel. Die Ergebnisse sind im Anhang (B.7) zu finden. Zu den jeweiligen Integrationsbereichen wurde der Wert von Matlab angegeben sowie bei den alternativen Verfahren (Gauß–Chebychev, Trapezregel) der Wert und der absolute (relative) Fehler – im Vergleich zu Matlab. Es zeigt sich, dass die Gauß–Chebychev Integration (mit 40 bzw. 50 Basispolynomen) ziemlich gute Ergebnisse liefert, die durchgehend besser sind als bei der Trapezregel (mit 100, 500, 1000 Stützstellen). Sie sind allerdings schlechter als bei den beiden vorhergehenden Funktionen. Eine Ursache hierfür könnten Überlagerungseffekte sein (Mason und Handscomb, 2003), die darauf beruhen, dass Polynome verschiedenen Grades dieselben Nullstellen haben. Betrachtet man die Entwicklung der Koeffizienten der Chebyshev Polynome, so erkennt man, dass der exponentielle Abfall der Koeffizienten ab etwa dem fünfzehnten Polynom stark verflacht und die Koeffizienten betragsmäßig relativ ähnlich sind (siehe Grafiken (4.5), (4.6)).

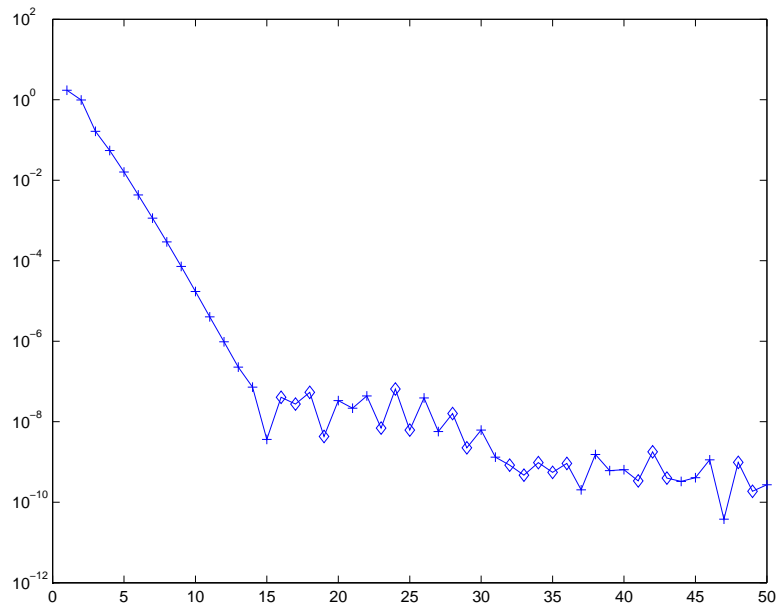


Abbildung 4.5: Beträge der Chebyshev Koeffizienten für erste Stammfunktion von h , Darstellung auf logarithmischer Skala

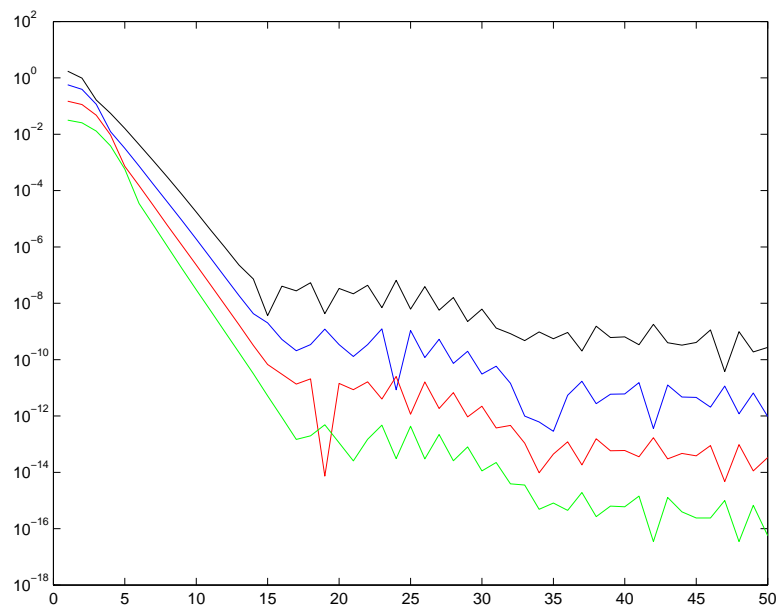


Abbildung 4.6: Beträge der Chebyshev Koeffizienten für iterierte Stammfunktionen von h , Darstellung auf logarithmischer Skala

4.6 Schätzung im bayesianischen Kontext

4.6.1 Darstellung der a posteriori Verteilung

Einige grundlegende Aspekte bayesianischer Techniken zur Modellierung wurden in dem Abschnitt 3.6 vorgestellt. Eine Anwendung dieser Methoden erfolgt nun in diesem Abschnitt. Für die a posteriori Verteilung bzw. Dichte $f_{post}(\theta|T, \gamma)$ der jetzt als zufällig betrachteten Parameter $\theta = (\boldsymbol{\beta}, a, \theta_1, \theta_2, \theta_3)$ benötigt man den Likelihoodprozeß $L_t(\theta, T)$ mit Beobachtungen $T = (t_1, \dots, t_n)$, die a priori Verteilungen bzw. Dichten $f_{prio}(\theta|\gamma)$ der Parameter mit Hyperparametern γ sowie Verteilungen bzw. Dichten $f_{hyp}(\gamma)$, für die Hyperparameter γ der a priori Verteilungen. Die Modellierung der Intensitätsrate wurde bereits in den vorherigen Abschnitten behandelt. Im Gegensatz zum Ansatz über gemischte Modelle werden nun alle Parameter als Zufallsgrößen betrachtet. Mit dieser Notation gilt

$$f_{post}(\theta|T, \gamma) \propto L_t(\theta, T) \cdot f_{prio}(\theta|\gamma) \cdot f_{hyp}(\gamma) \quad . \quad (4.92)$$

Die Likelihood schreibt sich (vgl. (2.40) und (3.1) als

$$L_t(\theta, T) = \prod_{i=1}^n \lambda_{t_i}^{\delta_i} \exp \left\{ - \int_0^{t_i} \lambda_s ds \right\} . \quad (4.93)$$

Als a priori Dichte wird für die Parametervektoren $\boldsymbol{\beta}$ und a eine nichtinformative Verteilung gewählt, d.h.

$$f_{prio}(\cdot) \propto const \quad , \quad const \in \mathbb{R}^+ \quad . \quad (4.94)$$

Im Fall von $\theta_1 = (\boldsymbol{\alpha}_1, \mathbf{u}^{\alpha_1})$ (Typ 1) lautet der glatte Effekt ($r = 1$)

$$\eta_s(\boldsymbol{\alpha}_r, \mathbf{u}^{\alpha_r}) = \sum_{l=0}^g \alpha_{r,l} s^l + \sum_{k=0}^K u_k^{\alpha_r} (s - \kappa_k)_+ \quad ,$$

im Fall von $\theta_2 = (\boldsymbol{\alpha}_2, \mathbf{u}^{\alpha_2})$ wird lediglich eine Kovariable multiplikativ aufgenommen. Für die Komponenten $\alpha_{r,l}$ wird jeweils eine nichtinformative a priori Dichte gewählt,

wogegen für die Parameter $u_k^{\alpha_r}$ eine Normalverteilung gewählt wird

$$f_{prio}(u_k^{\alpha_r}) \propto \exp(-0.5\gamma_r) \cdot \exp\left(-\frac{(u_k^{\alpha_r})^2}{2\exp(\gamma_r)}\right), \quad k = 1, \dots, K, \quad (4.95)$$

mit $\exp(\gamma_r)$ als Varianz und der Hyperparameter γ_r hat folgende Verteilung

$$f_{hyp}(\gamma_r) \propto \exp\left(-\frac{(\gamma_r - \mu_{hyp})^2}{2\sigma_{hyp}^2}\right), \quad (4.96)$$

d.h., der Varianzparameter $\exp(\gamma_r)$ ist jeweils log-normalverteilt. Der Parameter σ_{hyp}^2 sollte entsprechend dem Vorwissen gewählt werden, bei ungenauem oder nicht vorhandenem Vorwissen groß genug, z.B. $\sigma_{hyp}^2 = 10$. Damit wird sichergestellt, dass $\exp(-\gamma_r)$ hinreichend große und kleine Werte annehmen kann. Wählt man $\mu_{hyp} = 0$, so streut γ_r um den Wert Null, negative Werte von μ_{hyp} resultieren dementsprechend zu kleineren Werten der Varianz $\exp(\gamma_r)$.

Ein analoges Vorgehen erfolgt im Fall von $\theta_3 = (\boldsymbol{\alpha}_3, \mathbf{u}^{\alpha_3})$, bei dem anstelle eines zeitlichen glatten Effekts ein räumlicher glatter Effekt geschätzt wird. Mit Hilfe der hybriden MCMC Verfahren werden Realisierungen von Zufallsgrößen aus den Verteilungen der (Hyper-) Parameter simuliert. Damit läßt sich die a posteriori Dichte darstellen als

$$\begin{aligned} f_{post}(\theta|T, \gamma) &\propto L_t(\theta, T) \cdot f_{prio}(\boldsymbol{\beta}) \cdot f_{prio}(a) \prod_{|gE|} f_{prio}(\boldsymbol{\alpha}_{(\cdot)}) \\ &\quad \cdot \prod_{|gE|} f_{prio}(u_r^{\alpha_r}) \cdot \prod_{|gE|} f_{hyp}(\gamma) \\ &\propto \prod_{\{T_n \leq t\}} \lambda_{t_i}^{\delta_i} \exp\left\{-\int_0^{t_i} \lambda_s ds\right\} \\ &\quad \cdot \prod_{|gE|} \exp(-0.5\gamma_r) \cdot \exp\left(-\frac{(u_r^{\alpha_r})^2}{2\exp(\gamma_r)}\right) \\ &\quad \cdot \prod_{|gE|} \exp\left(-\frac{(\gamma_r - \mu_{hyp})^2}{2\sigma_{hyp}^2}\right), \end{aligned} \quad (4.97)$$

wobei $|gE|$ jeweils die Anzahl der glatten Effekte bezeichnet. Bei Verwendung der hybriden Monte Carlo Verfahren benötigt man die (logarithmierte) a posteriori Dichte nur für den Akzeptanzschritt (vgl. 3.42). Zur Generierung neuer Werte benötigt man den Gradienten der logarithmierten a posteriori Dichte. Mit den oben gewählten a priori Ver-

teilungen gilt, dass die Gradienten (blockweise pro Effekt) fast identisch sind mit den marginalen Schätzgleichungen aus Kapitel (4.5). Anstelle der Strafterme werden nun Terme der a priori Dichte verwendet. Zusätzlich erhält man noch einen Gradienten für die Hyperparameter. Somit ergibt sich für

- β

$$\frac{\partial}{\partial \beta} \ln f_{post}(\theta|T, \gamma) \propto G(\beta) \quad (4.98)$$

- a

$$\frac{\partial}{\partial a} \ln f_{post}(\theta|T, \gamma) \propto G(a) \quad (4.99)$$

- $\theta_1 = (\alpha_1, \mathbf{u}^{\alpha_1})$ (Typ 1)

$$\frac{\partial}{\partial \alpha_1} \ln f_{post}(\theta|T, \gamma) \propto G(\alpha_1) \quad (4.100)$$

$$\frac{\partial}{\partial \mathbf{u}^{\alpha_1}} \ln f_{post}(\theta|T, \gamma) \propto G(\mathbf{u}^{\alpha_1}) + \frac{\partial}{\partial \mathbf{u}^{\alpha_1}} \ln f_{prio}(\mathbf{u}^{\alpha_1}) \quad (4.101)$$

- $\theta_2 = (\alpha_2, \mathbf{u}^{\alpha_2})$ (Typ 2)

$$\frac{\partial}{\partial \alpha_2} \ln f_{post}(\theta|T, \gamma) \propto G(\alpha_2) \quad (4.102)$$

$$\frac{\partial}{\partial \mathbf{u}^{\alpha_2}} \ln f_{post}(\theta|T, \gamma) \propto G(\mathbf{u}^{\alpha_2}) + \frac{\partial}{\partial \mathbf{u}^{\alpha_2}} \ln f_{prio}(\mathbf{u}^{\alpha_2}) \quad (4.103)$$

- $\theta_3 = (\alpha_3, \mathbf{u}^{\alpha_3})$ (Typ 3)

$$\frac{\partial}{\partial \alpha_3} \ln f_{post}(\theta|T, \gamma) \propto G(\alpha_3) \quad (4.104)$$

$$\frac{\partial}{\partial \mathbf{u}^{\alpha_3}} \ln f_{post}(\theta|T, \gamma) \propto G(\mathbf{u}^{\alpha_3}) + \frac{\partial}{\partial \mathbf{u}^{\alpha_3}} \ln f_{prio}(\mathbf{u}^{\alpha_3}) \quad (4.105)$$

- Hyperparameter γ_r (pro glattem Effekt r)

$$\frac{\partial}{\partial \gamma_r} \ln f_{post}(\theta|T, \gamma_r) = \frac{\partial}{\partial \gamma_r} \ln f_{prio}(\theta|\gamma_r) + \frac{\partial}{\partial \gamma_r} \ln f_{hyp}(\gamma_r) \quad (4.106)$$

Die Ableitung läßt sich schreiben als

$$\begin{aligned} \frac{\partial}{\partial \gamma_r} \ln f_{post}(\theta|T, \gamma) &= \frac{\partial}{\partial \gamma_r} \ln f_{prio}(\theta|\gamma_r) + \frac{\partial}{\partial \gamma_r} \ln f_{hyp}(\gamma_r) \\ &= \sum_{k=1}^K [-0.5 + 0.5(u_k^{\alpha_r})^2 \exp(-\gamma_r)] - \frac{(\gamma_r - \mu_{hyp})}{\sigma_{hyp}^2}, \end{aligned} \quad (4.107)$$

wobei die Summation ($k = 1, \dots, K$) über alle Knotenpunkte erfolgt.

4.6.2 Modifikation des Leapfrog Algorithmus

Im Abschnitt (3.6.4) wurde eine Verallgemeinerung vorgeschlagen (Formel 3.44), nämlich die Hessematrix zu verwenden. Im Abschnitt (4.5.2) erfolgte eine blockweise Berechnung. Eine Alternative dazu ist die inkrementelle beobachtete Informationsmatrix $J_t(\theta)$ (Rao, 1999). Bezeichne $I_t(\theta) = \mathbb{E}_\theta(\dot{l}_t(\theta)\dot{l}_t^T(\theta))$, die erwartete Information und $\dot{l}_t(\theta)$ den Scoreprozeß. Dann gilt unter sehr allgemeinen Voraussetzungen (Rao, 1999)

$$J_t(\theta) = \langle \dot{l}^{(c)}(\theta) \rangle_t + \sum_{s \leq t} [\Delta \dot{l}_s(\theta)][\Delta \dot{l}_s(\theta)]^T \quad (4.108)$$

sowie

$$I_t(\theta) = \mathbb{E}_\theta(J_t(\theta)) \quad . \quad (4.109)$$

Wegen $\Delta \dot{l}_s(\theta) = \dot{l}_s(\theta) - \dot{l}_s(\theta)_{s-}$ wird nur über die Sprungzeitpunkte summiert. Der Term $\dot{l}^{(c)}(\theta)_t$ bezeichnet die stetige Komponente des Scoreprozesses. Es gilt $\dot{l}_t^{(c)}(\theta) = \int_0^t \frac{\partial}{\partial \theta} \lambda_s(\theta) ds$, d.h. $\dot{l}_t^{(c)}(\theta)$ hängt vom Zufall nur insofern ab, da der Parameter θ ein Zufallsvektor (Bayes) ist. Somit kann $\dot{l}_t^{(c)}(\theta)$ als vorhersagbar angenommen werden und es gilt $\langle \dot{l}^{(c)}(\theta) \rangle_t = 0$. (s.a. die Definition der vorhersagbaren quadratischen Variation (2.26)). Es folgt

$$J_t(\theta) = \sum_{s \leq t} [\Delta \dot{l}_s(\theta)][\Delta \dot{l}_s(\theta)]^T = \int_0^t \left(\frac{\partial}{\partial \theta} \ln(\lambda_s(\theta)) \right) \left(\frac{\partial}{\partial \theta} \ln(\lambda_s(\theta)) \right)^T dN_s^c \quad . \quad (4.110)$$

Wegen $\lambda_s(\theta) = \exp(\eta_s(\theta))$ bestehen die Einträge der Matrix $J_t(\theta)$ aus den Kovariablen – bei den Ableitungen nach den festen Parametern, aus den Ereigniszeitpunkten und ihren Quadraten – bei der Ableitung nach der Baseline-Hazard sowie im Fall zeitabhängiger

Kovariableneffekte aus den Produkten von Ereigniszeitpunkten und den korrespondierenden Kovariablen (jeweils summiert über die Anzahl der Individuen, die noch unter Risiko stehen).

Beispiel

Sei $\lambda_{s_i}(\theta) = \exp(x_i\beta + \alpha_0 + \alpha_1 s + \alpha_2 s^2 + u^1(\kappa_1 - s)_+^2)$, $i = 1, \dots, n$ mit Ereigniszeitpunkten T_i (bzw. $(T_i)_+^2 := (\kappa_1 - T_i)_+^2$) und Indikatorvariable C_i für den Zensierungsmechanismus, dann hat $J_t(\theta)$ die Gestalt $(\theta = (\beta, \alpha_0, \alpha_1, \alpha_2, u^1))^6$

$$\begin{bmatrix} \sum x_i^2 \mathbf{1}_{C_i} & \sum x_i \mathbf{1}_{C_i} & \sum x_i T_i \mathbf{1}_{C_i} & \sum x_i T_i^2 \mathbf{1}_{C_i} & \sum x_i (T_i)_+^2 \mathbf{1}_{C_i} \\ \sum x_i \mathbf{1}_{C_i} & \sum \mathbf{1}_{C_i} & \sum T_i \mathbf{1}_{C_i} & \sum T_i^2 \mathbf{1}_{C_i} & \sum (T_i)_+^2 \mathbf{1}_{C_i} \\ \sum x_i T_i \mathbf{1}_{C_i} & \sum T_i \mathbf{1}_{C_i} & \sum T_i^2 \mathbf{1}_{C_i} & \sum T_i^3 \mathbf{1}_{C_i} & \sum T_i (T_i)_+^2 \mathbf{1}_{C_i} \\ \sum x_i T_i^2 \mathbf{1}_{C_i} & \sum T_i^2 \mathbf{1}_{C_i} & \sum T_i^3 \mathbf{1}_{C_i} & \sum T_i^4 \mathbf{1}_{C_i} & \sum T_i^2 (T_i)_+^2 \mathbf{1}_{C_i} \\ \sum x_i (T_i)_+^2 \mathbf{1}_{C_i} & \sum (T_i)_+^2 \mathbf{1}_{C_i} & \sum T_i (T_i)_+^2 \mathbf{1}_{C_i} & \sum T_i^2 (T_i)_+^2 \mathbf{1}_{C_i} & \sum (T_i)_+^4 \mathbf{1}_{C_i} \end{bmatrix} .$$

Die Matrix $J_t(\theta)$ enthält keine zu schätzenden Parameter und ist – im Gegensatz zur Martingal Information – entsprechend einfach zu berechnen. Es wird daher vorgeschlagen $J_t(\theta)$ zu invertieren und die Hauptdiagonale als Ersatz für M^{-1} zu verwenden. Die Matrix $J_t(\theta)$ muß – je nach Art und Umfang der Daten – nicht positiv definit sein. Man kann sich überlegen, dass numerische Instabilitäten auftreten bei starker Korrelation der Kovariablen oder aber –insbesondere bei starker Zensierung– bei einem relativ feinen Splinegitter, da die meisten Terme am rechten Rand des Gitters Null sind.

⁶aus Platzgründen wurde anstelle der korrekten Form $\sum_{i=1}^n$ nur \sum verwendet

5 Implementierung der Methoden und Anwendung

In diesem Kapitel wird die praktische Umsetzung der bisher behandelten Modelle beschrieben. Die Anwendung der Methoden erfolgt zur Kontrolle an simulierten Datensätzen sowie anhand echter Daten. Zuerst werden Bayes Methoden verwendet. Es folgen Verfahren, die auf Schätzgleichungen beruhen. Vorweg sollen im folgenden Abschnitt zwei Beispiele Einblick in die Problematik liefern.

5.1 Einführende Beispiele

Der einfachste Fall der Schätzung eines Intensitätsprozesses ist eine Stichprobe mit unabhängigen und identisch verteilten Zufallsgrößen, die eine Lebensdauerverteilung besitzen. Man denke etwa an eine Exponentialverteilung oder an eine Weibullverteilung. In diesem Spezialfall entspricht der Intensitätsprozeß der Hazardrate der entsprechenden Lebensdauerverteilung. Im Weibullmodell lautet die Hazardrate

$$\lambda(t) = \lambda\alpha(\lambda t)^{\alpha-1} \quad , \quad t, \lambda, \alpha > 0 \quad , \quad (5.1)$$

die für $\alpha = 1$ in den Spezialfall der Exponentialverteilung übergeht.

5.1.1 Ein naives Beispiel

Im Fall der Exponentialverteilung erhält man als Maximum Likelihood Schätzwert für den unbekannt Parameter λ – je nach Wahl der Parametrisierung – das arithmetische Mittel \bar{x} oder dessen Kehrwert. Im folgenden sei T exponentialverteilt mit Parametrisierung $\mathbb{E}(T) = \frac{1}{\lambda}$. Verwendet man zur Berechnung der Hazardrate ein Modell, das eine glatte Hazardrate schätzt, so wird man nicht das arithmetische Mittel der Realisationen $t_i, i = 1, \dots, n$ von T verwenden, sondern versuchen, die metrische Struktur der Daten zu erfassen. Ein naiver Schätzer der Hazardrate könnte nun so konstruiert werden, dass man die Daten der Größe nach sortiert, in (z.B.) Quartile einteilt und je Quartil

eine Schätzung der Hazardrate durchführt. In Ermangelung eines speziellen Verfahrens kann man nun je Quartil die Hazardrate schätzen über den Kehrwert des arithmetischen Mittels. Dann erhält man folgende Grafik, die anhand eines simulierten Datensatzes erstellt wurde. Der Datensatz besteht aus $n = 200$ exponentialverteilten Realisationen mit Parameter $\lambda = 1$. Die Grafik enthält die Schätzung der Hazardrate für alle Daten.

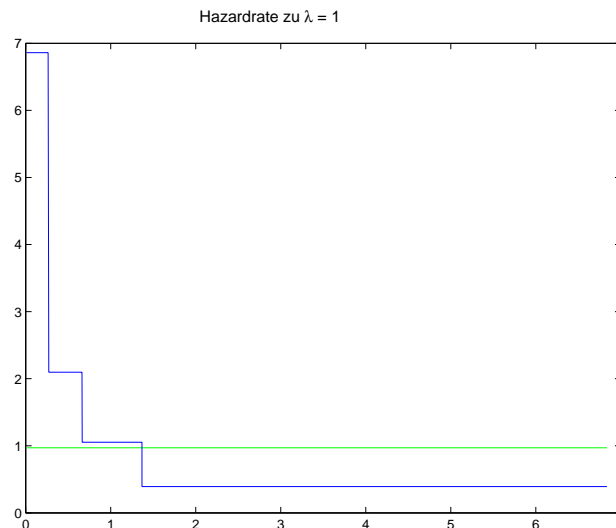


Abbildung 5.1: ‘quartilsweise’ Schätzung der Hazardrate

Dies entspricht der horizontalen Linie nahe dem (theoretisch exakten) Wert Eins, sowie die entsprechenden Schätzer pro Quartil. Die Schätzungen pro Quartil sind plausibel. Betrachtet man die ersten 25% der Daten, so sind die Werte klein und der dazu korrespondierende Schätzwert für λ ist entsprechend groß. In jedem weiteren Quartil hat man dann größere Realisationen, so dass die entsprechenden Schätzwerte für λ immer kleiner werden. Es ist offensichtlich, dass die Hazardrate anhand der Quartile eine sehr schlechte Annäherung an die wahre Hazardrate darstellt. Betrachtet man nochmals das Vorgehen der quartilsweisen Schätzung, so erkennt man, dass das Einteilen der Daten in Quartile zu einem großen Bias führt. Dahinter verbirgt sich eine Zensierungsproblematik. Da die Daten der Größe nach sortiert wurden, werden bei der Schätzung innerhalb des ersten Quartils nur die kleinsten Werte berücksichtigt, alle anderen nicht. Die Zensierung ist hier fatalerweise unmittelbar an die Größe der Realisationen geknüpft und daher liegt keine nichtinformative Zensierung vor, sondern eine (höchst) informative Zensierung. Unabhängig von den Methoden – Bayes oder Schätzgleichungen – wird im Rahmen dieser Arbeit die Modellierung eines zeitabhängigen Intensitätsprozesses $\lambda(t)$ angestrebt. Dazu wird natürlich nicht in expliziter Weise eine Sortierung der Daten durchgeführt, dies geschieht allerdings implizit, da die metrische Struktur der Daten,

z.B. in Form von Differenzen zwischen aufeinanderfolgenden Lebensdauern wichtig ist. Dieses einfache Beispiel lehrt nun zwei Dinge. Erstens kann die Schätzung nicht ausschließlich lokal durchgeführt werden, d.h. dass zur Schätzung des Intensitätsprozesses zu einem Zeitpunkt t nur benachbarte Beobachtungen herangezogen werden und zweitens ist die Schätzung einer flexiblen Intensitätsrate kein triviales Unterfangen. Zu dem ersten Punkt sei entschärfend angemerkt, dass bei dem verwendeten ad hoc Verfahren nicht berücksichtigt wird, dass die Realisationen einem Prozeß mit gleicher Intensität entspringen. Es gibt sehr wohl Methoden, die ebenso auf Intervallen beruhen, aber die Struktur des Punktprozesses berücksichtigen (Gamerman, 1991).

5.1.2 Beispiel zur komplexeren Schätzung

Es wurde eine unabhängige, identisch verteilte Stichprobe von $n = 200$ exponentialverteilten Zufallsgrößen mit Parameter $\lambda = 1$ erzeugt. In diesem Fall entspricht die Konstante λ der Hazardrate. Als Maximum Likelihood Schätzwert erhält man $\hat{\lambda} = 1.0578$. Die Zufallsgröße läßt sich auch als weibullverteilt auffassen, wobei beide Parameter den Wert Eins annehmen. Schätzt man beide Parameter zu den gleichen Daten, so erhält man die Werte $\hat{\lambda} = 1.0206$ und $\hat{\alpha} = 1.1091$. In der Grafik (5.2) ist die theoretische Hazardrate, sowie die geschätzte Hazardrate des Exponentialmodells sowie die geschätzte Hazardrate des Weibullmodells dargestellt. Abgesehen von dem Spezialfall, dass im Weibullmodell als Schätzung für α genau der theoretische Wert Eins erreicht wird, hat die Hazardrate eine verfälschte Form. Dieses Problem mag künstlich erscheinen, denn es sollte bekannt sein, ob die Zufallsvariablen exponentialverteilt sind oder nicht. Geht man aber davon aus, dass dies nicht genau bekannt ist und sucht sich eine größere Klasse von Verteilungen, so erhält man fast sicher nicht die wahre Form der Hazardrate. Ganz unabhängig von den später verwendeten Methoden kann also schon in einfachen Beispielen gezeigt werden, dass die Schätzungen in einem komplexeren Modell nicht unbedingt die wahre Situation eines einfachen Modells widerspiegeln. Vermeiden läßt sich dies wohl am ehesten durch genaue Analyse des Vorwissens, in diesem Fall die Annahme, dass die Zufallsgrößen exponentialverteilt sind oder nicht. In diesem Zusammenhang drängt sich geradezu die Frage auf, ob nicht geeignete Verfahren der Modellwahl dieses Problem befriedigend lösen, vor allem im Hinblick auf die semiparametrische Modellierung. Sei I die wahre Parametermenge und I_n die Menge der geschätzten Parameter anhand einer Stichprobe vom Umfang n . Eine natürliche Forderung wäre, dass das Verfahren zur Modellwahl konsistent ist, d.h. $I_n \rightarrow I$ (in Wahrscheinlichkeit), (vgl. Bunea

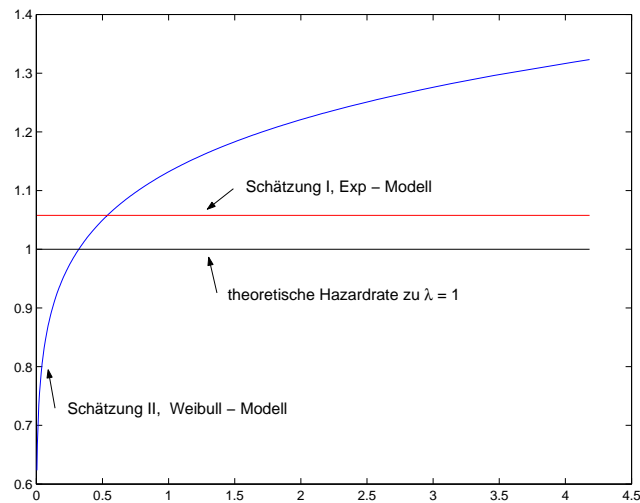


Abbildung 5.2: Hazardrate, theoretisch und zwei Schätzungen

(2004)). Eine gewisse Popularität zur Zeit haben Schwellenwertverfahren erlangt, bei denen Parametern der Wert Null zugewiesen wird, sobald sie betragsmäßig unterhalb einer Schranke liegen, siehe etwa Tibshirani (1996). Dieses (und andere) Vorgehen wird in (Leeb und Pötscher, 2005) kritisiert, in Analogie zur Kritik von Le Cam (Le Cam, 1953) am Hodges Schätzer. Die Kritik beruht darauf, dass die Konsistenz nicht gleichmäßig ist (auch nicht lokal), mit der Folge, dass es keine Gewähr gibt, dass die Verteilung der geschätzten Parameter (als Funktion der zufälligen Realisationen) auf Basis der endlichen Stichprobe wesentliche Charakteristika mit der entsprechenden asymptotischen Verteilung gemeinsam hat. In der Veröffentlichung Ishwaran und Rao (2005) wird ein Vorgehen zur Modellwahl vorgeschlagen, dass die Kritikpunkte von Leeb und Pötscher (2005) berücksichtigt.

5.2 Bayesianische Verfahren

5.2.1 Simulationstudien

Die ersten Beispiele zur Schätzung der Hazardrate liefern Datensätze mit unabhängig, identisch exponential- bzw. weibullverteilten Zufallsgrößen. Danach werden via Cox-Modell Datensätze mit Kovariablen betrachtet. Anschließend wird noch ein Datensatz diskutiert, bei dem ein Kovariableneffekt zeitabhängig ist. Hauptaugenmerk ist die semiparametrische Schätzung der Hazardrate $\lambda(t)$, $t \in [0, T]$. Die Hazardrate wird zu

jedem Parametervektor der MCMC Simulationen auf einem hinreichend feinen, gleichabständigen zeitlichen Gitter (50 Werte) berechnet. Zu den Zeitpunkten lassen sich die typischen statistischen Kenngrößen berechnen, z.B. Minimum, Maximum, Median oder Quantile. Mittels linearer Interpolation lassen sich damit die Hazardrate sowie die Konfidenzbänder für den gesamten Zeitbereich visuell darstellen. Jede der entsprechenden Grafiken enthält die theoretische Hazardrate (schwarze Linie) sowie die Schätzungen. Im einzelnen: Das Minimum bzw. Maximum (blaue, kleine Punkte), das 5%-Quantil sowie das 95% -Quantil (blaue gestrichelte Linien) sowie die Medianschätzung (schwarze Punkte).

Simulation mit weibullverteilten Zufallsgrößen

Da die Exponentialverteilung ein Spezialfall der Weibullverteilung ist, werden die Simulationen zu beiden Verteilungen gemeinsam abgehandelt. Die Simulation der Datensätze erfolgte mit Matlab (Version 13) oder mit R (Version 1.9.1). Es wurden jeweils Datensätze mit Stichprobenumfang $n = 200, 500, \text{ bzw. } 1000$ erzeugt, deren Ereigniszeitpunkte mit t_i bezeichnet werden. Der erste Datensatz beinhaltet exponentialverteilte Zufallsvariablen mit Parameter $\lambda = 1$, der zweite weibullverteilte mit Parametern $\alpha = 2$ und $\lambda = 1$. Dies entspricht einer Hazardrate von $\lambda(t) = 2t$. Die Zensierung erfolgt durch weitere n exponential- bzw. weibullverteilte Zufallsvariablen mit Ereigniszeitpunkten u_i , wobei der Skalenparameter variiert wurde, so dass verschiedene Zensierungsszenarien erreicht wurden. Dann wird ein paarweiser Vergleich durchgeführt, falls $t_i < u_i$, so nimmt man an, dass das Ereignis beobachtet wird, anderenfalls geht man davon aus, dass das Ereignis zensiert ist und die Ereigniszeit u_i beträgt. Es wird also stets der kleinere von beiden Werten gewählt. Dieses Vorgehen entspricht dem Fall der nichtinformativen Rechtszensierung (vgl. Abschnitt (3.1)), den Beitrag zum Likelihoodprozeß pro Individuum findet man in der Formel (3.1). Hierbei ist die Anzahl der zensierten Daten zufällig. Die folgenden Ausführungen der Ceteris Paribus Analyse beziehen sich auf den Datensatz mit exponentialverteilten Zufallsvariablen, gefolgt von ergänzenden Aussagen zum Datensatz mit weibullverteilten Zufallsgrößen.

Ceteris Paribus Analyse der Simulationen

Im Rahmen der Simulationsstudien soll der Einfluß der ‘Einstellgrößen’ betrachtet werden. Die Einstellgrößen betreffen die Modellkomplexität, die Datensituation und die

Parameter des hybriden Monte Carlo Verfahrens. Eine Liste könnte folgendermaßen aussehen:

- Wahl der Parametrisierung,
- Transformation des Zeitbereichs,
- Wahl der Schrittweite τ bzw. Anzahl der Leapfrog Schritte (vgl. Abschnitt (3.5.4)),
- Verwendung von Polynomen zur Modellierung der Hazardrate,
- Verwendung von Splines, Anzahl der Knotenpunkte,
- Grad des Polynoms bzw. Splines,
- Datensätze mit verschieden starker Zensierung,
- Wahl der a priori bzw. Parameter.

Die Beurteilung des Einflusses erfolgt anhand der Samplingpfade und der geschätzten Hazardrate. Die Akzeptanzwahrscheinlichkeit ist bei den hybriden Monte Carlo Verfahren kein Kriterium, da sie eigentlich immer über 90% liegt oder Null ist. Letzteres tritt auf, falls die Schrittweite τ zu groß gewählt wird. Dementsprechend groß ist der Diskretisierungsfehler des Hamilton Systems, so dass die neuen Werte im Akzeptanzschritt abgelehnt werden. Die Samplingpfade lassen sich grafisch darstellen. Sie sollten nach der Burn in Phase keine Muster mehr aufweisen. Im Idealfall sind die Samplingpfade nicht autokorreliert. Ein oft angewandter Trick zur Reduzierung der Autokorrelation ist das Ausdünnen der per MCMC Verfahren vorgeschlagenen Werte, d.h. man verwendet nur jeden k -ten Wert für die weiteren Analysen. Zusätzlich kann man sich einen Einblick verschaffen, indem man die a posteriori Dichte anhand der simulierten Werte mittels eines Kerndichteschätzers grafisch darstellt. Weiter lassen sich die Quartile anhand der simulierten Werte angeben, natürlich auch der empirische Mittelwert und die empirische Varianz. Die empirischen Mittelwerte lassen sich zur Schätzung der Hazardrate verwenden. Bei der Varianz muß berücksichtigt werden, dass eine naive Schätzung nicht die Autokorrelation berücksichtigt. Die für die Analyse verwendete Software R stützt sich daher auf die Spektraldichte, deren Wert an der Stelle Null eine Schätzung der Varianz darstellt (Brockwell und Davis, 1996).

Wahl der Parametrisierung

Hat die Hazardrate folgende Form

$$h_{it}(\theta) = h_t^0(\gamma) \cdot \exp(\beta^t x_{it}), \quad (5.2)$$

so muß sichergestellt werden, dass h_t^0 positiv ist, da die Hazardrate bzw. der Intensitätsprozeß strikt positiv ist. Bei den MCMC Simulationen kann es nun passieren, dass die neu vorgeschlagenen Werte zu einer Hazardrate führen, die negativ ist. Als ad-hoc Methode wird man einfach neue Werte simulieren. Dieses Phänomen tritt insbesondere in der Burn in Phase auf, wenn die stationäre Verteilung noch nicht erreicht wurde. Eine wichtige Rolle spielt hierbei die Diskretisierungsschrittweite τ . Bei großem τ werden entsprechend große Schritte in der ‘Energiewelt’ gemacht. Man vermeidet diese Problematik, indem man die Hazardrate strikt als Exponentialfunktion schreibt oder eine a priori Verteilung verwendet, die der Positivitätsbedingung genügt.

Transformation des Zeitbereichs

Dividiert man die Ereigniszeitpunkte durch den maximalen Wert, so erhält man Lebensdauern auf $[0, 1]$. Dadurch verschiebt sich natürlich die geschätzte Hazardrate – bei exponentialverteilten Zufallsgrößen um eine Konstante. Eine Transformation ist unvermeidlich bei großen Lebensdauern, da über die exp- Funktion numerisch integriert werden muß. Andere Transformationen sind genauso möglich, beispielsweise, wenn die Skala der Ereigniszeit in Tagen gemessen wird. In diesem Fall bietet sich ein ‘Einheitenwechsel’ auf Monat oder Jahr an.

Wahl der Schrittweite bzw. Anzahl der Leapfrog Schritte

Die Anzahl der Leapfrog Schritte wird zufällig bestimmt anhand der diskreten Gleichverteilung auf $[1, l_{\text{max}}]$. Zusammen mit der Diskretisierungsschrittweite τ bestimmt sie die zurückgelegte Wegstrecke auf dem Energiepotential $H(x, p)$. Mit zunehmender Größe von τ steigt der Approximierungsfehler des Leapfrog Differenzenschemas. Ab einem gewissen Wert fällt daher die Akzeptanzrate stark ab. Wird hingegen τ sowie

l_{smax} sehr klein gewählt bewegt man sich nur wenig auf dem Energiepotential bis zum Akzeptanzschritt. Entsprechend ineffizient ist das hybride Monte Carlo Verfahren, da die Autokorrelation bei den Parametern sehr hoch ist. Ein guter Kompromiß scheinen Werte von 0.01 bzw. von 0.005 für τ und 100 für l_{smax} zu sein. Diese Werte müssen gegebenenfalls modifiziert werden – in Abhängigkeit von den Kovariablen. Insbesondere falls die Kovariablen nicht standardisiert sind, da dann die Ableitungen wesentlich größer sind und τ kleiner gewählt werden muß.

Verwendung von Polynomen, Splines, Grad der Basisfunktionen

Die Darstellung der Hazardrate erfolgt entsprechend dem oben geschilderten Verfahren. Die Abbildungen (5.3), (5.5) und (5.6) modellieren jeweils die Hazardrate als Polynom vom Grad 1 bzw. 2 sowie als Spline vom Grad 1 mit neun Gitterpunkten, die datengeteuert quartilsweise gewählt wurden. Alle Grafiken haben gemeinsam, dass am rechten Rand der Bereich zwischen minimaler und maximaler Hazardrate trompetenförmig auseinanderdriftet. Bei Polynomen (Splines analog) höheren Grades werden zudem von dem Maximum der Hazardrate am rechten Rand sehr große Werte angenommen, daher muß immer die Skala der y-Achse bei der grafischen Interpretation mitberücksichtigt werden. Bei Verwendung eines Polynoms vom Grad Eins hat man im Mittel einen leichten Abwärts- oder Aufwärtstrend (vgl. Abbildung (5.7) und (5.8)), wobei allerdings bei hinreichend großen Datensätzen ($n = 1000$) die Konfidenzstreifen entsprechend nahe bei dem theoretisch exakten Wert liegen. Betrachtet man die Samplingpfade (Abbildung (5.4)), z.B. bei einem Polynom ersten Grades, so sieht man, dass die aposteriori Dichte des zweiten Parameters, der für die Steigung verantwortlich ist, nicht symmetrisch um den Wert Null verteilt ist. Die Splines scheinen eine etwas bessere Anpassung zu haben, da die Steigung der Funktion eher flacher ist, als bei einer rein polynomialen Modellierung (vgl. (5.9), (5.10)).

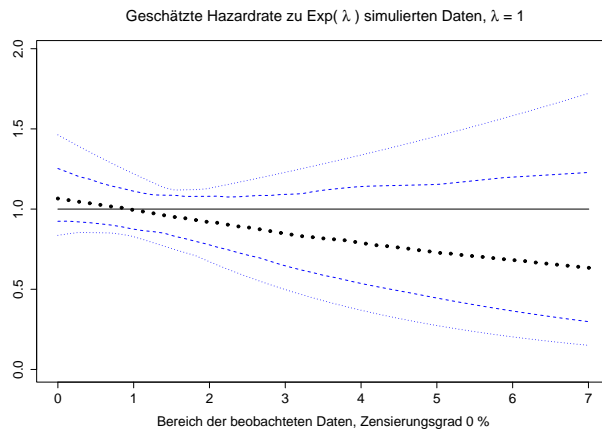


Abbildung 5.3: Polynom Grad 1, $n = 200$

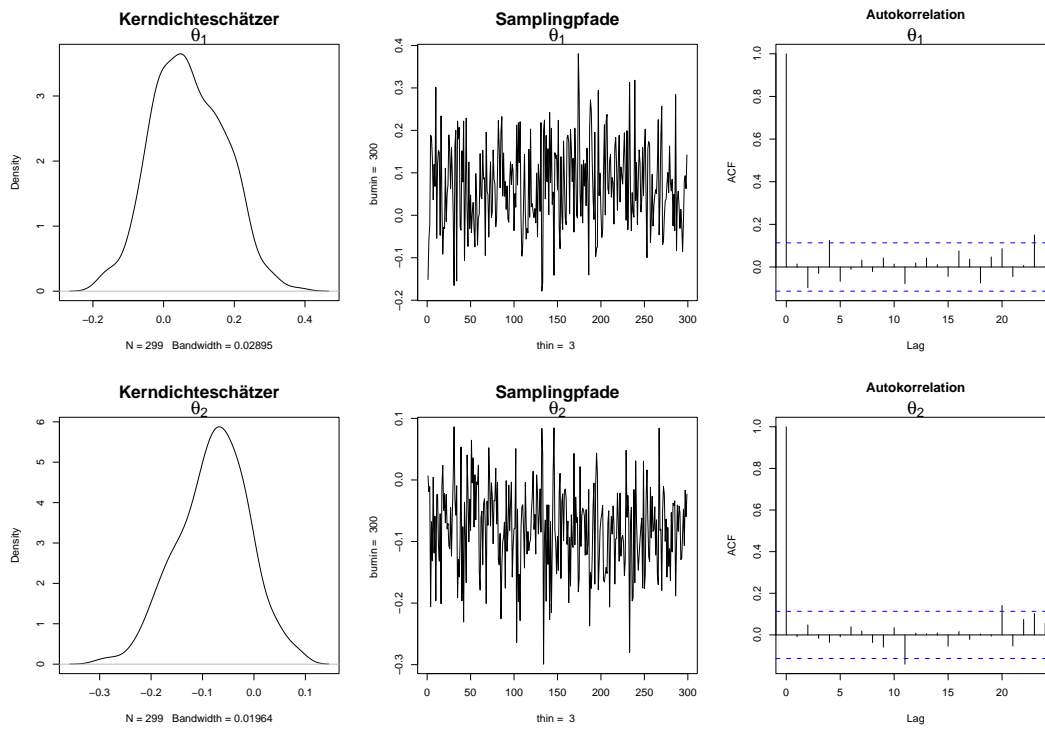


Abbildung 5.4: Polynom Grad 1, Samplingpfade der Hazardrate, $n = 200$

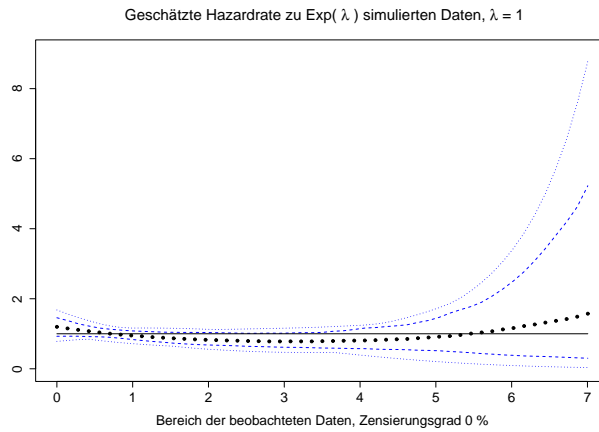


Abbildung 5.5: Polynom Grad 2, $n=200$

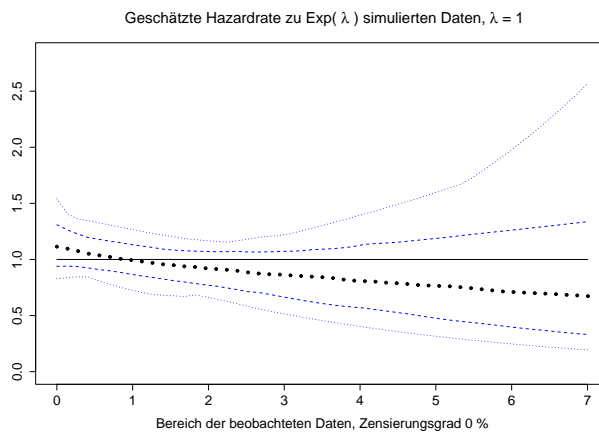


Abbildung 5.6: Spline Grad 1 mit 9 Gitterpunkten, $n = 200, \mu = -10$

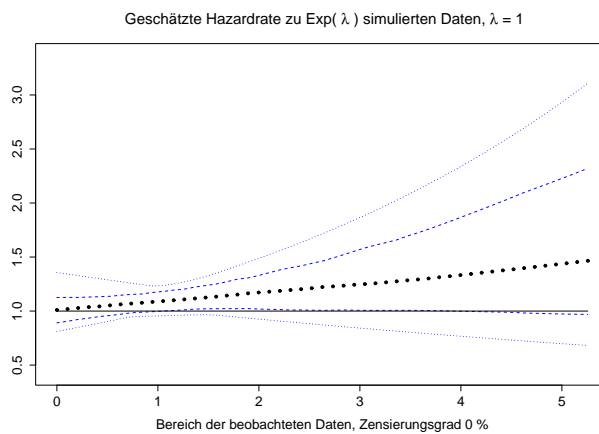


Abbildung 5.7: Polynom Grad 1, $n = 500$

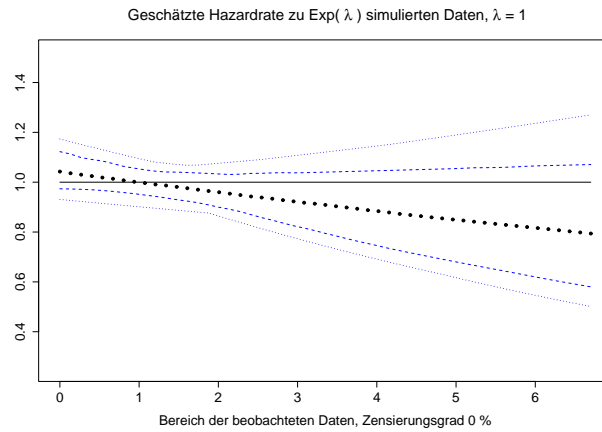


Abbildung 5.8: Polynom Grad 1, $n = 1000$

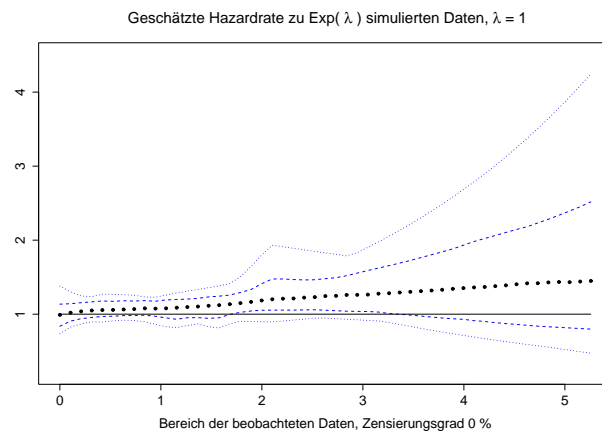


Abbildung 5.9: Spline Grad 1 mit 9 Gitterpunkten, $n = 500$

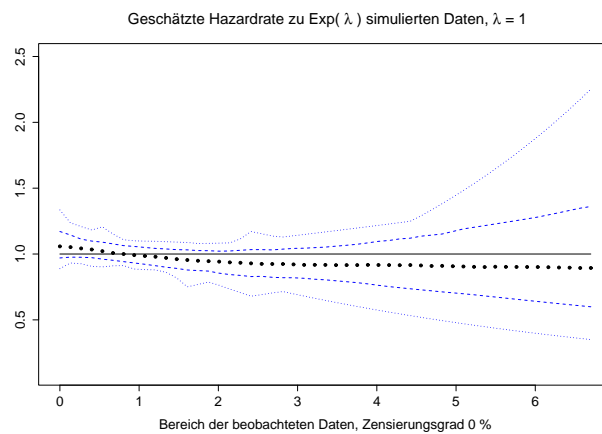


Abbildung 5.10: Spline Grad 1 mit 9 Gitterpunkten, $n = 1000$

Betrachtet man die Simulationen mit weibullverteilten Daten, so ergibt sich am rechten Rand – ganz analog – wieder ein großer Konfidenzbereich, insbesondere bei einem hohen Polynomgrad (siehe Abbildung (5.13)). Bei einem Polynomgrad Eins wird die Hazardrate überschätzt (vgl. Abbildung (5.11)), die Schätzung nimmt einen exponentiellen Verlauf an, bei einem Polynom vom Grad Zwei ist das Ergebnis auch am rechten Rand recht gut (Abbildung (5.12)). Die Verwendung von Splines führt zu einer Verbes-

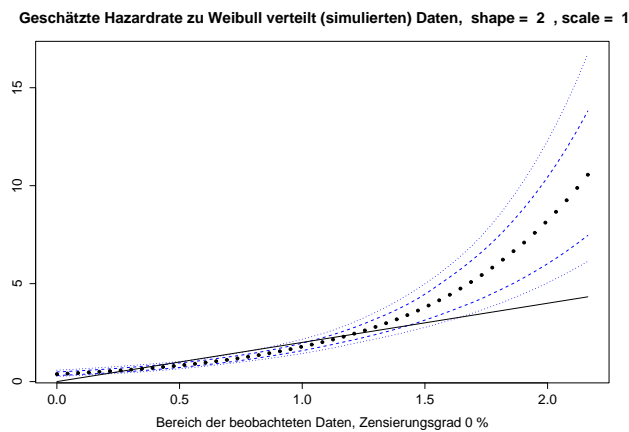


Abbildung 5.11: Polynom Grad 1, $n = 200$

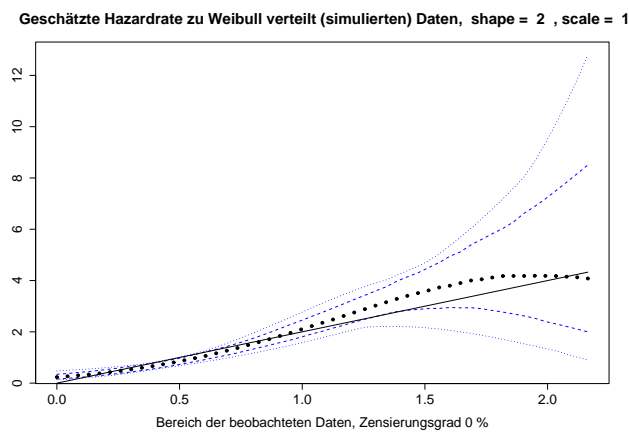
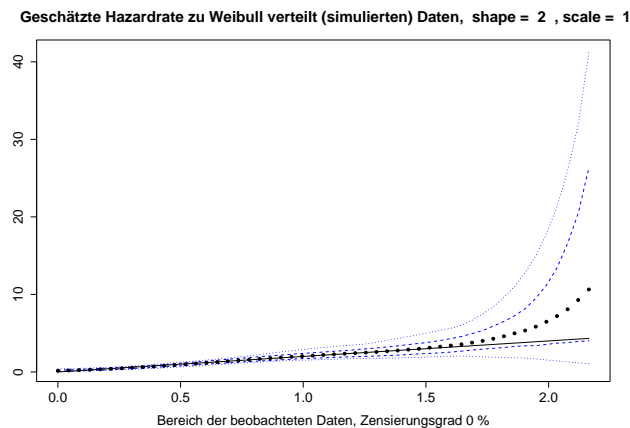
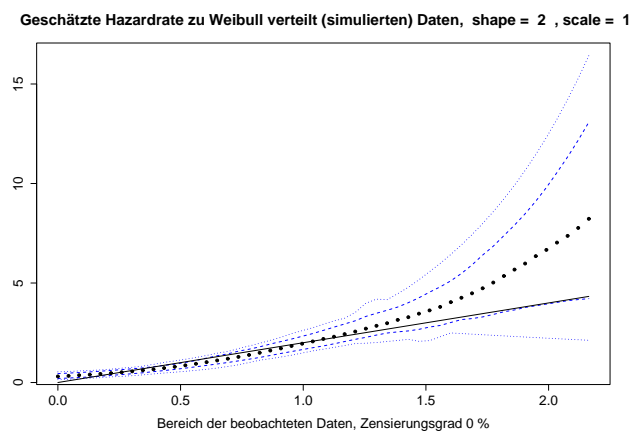


Abbildung 5.12: Polynom Grad 2, $n = 200$

serung (siehe Abbildung (5.14)), insbesondere wenn der Datensatz vergrößert wird (vgl. Abbildung (5.15)).

Abbildung 5.13: Polynom Grad 3, $n = 200$ Abbildung 5.14: Spline Grad 1, 9 Gitterpunkte, $n = 200$

Datensätze mit verschieden starker Zensierung

Es wurden Datensätze erzeugt, bei denen die Anzahl der zensierten Daten bis zu 74.5% betrug. Es ist zu beobachten, dass der maximale Ereigniszeitpunkt mit zunehmender Zensierung sinkt. Dies ist normal, da die Wahrscheinlichkeit einer Zensierung steigt, je größer die Ereigniszeit ist. Bei entsprechend starker Zensierung passiert dies häufiger. Die Güte der Schätzung ist – wie nicht anders zu erwarten war – abhängig vom Grad der Zensierung. Bei einer Zensierung im Bereich bis zu ca. 20% kann die Schätzung noch relativ ähnlich zur Berechnung mit vollständigem Datensatz sein, bei einer stärkeren Zensierung sind teilweise deutliche Unterschiede auszumachen.

Bei starker Zensierung kommt es insbesondere zu einer Verschlechterung der Schätzun-

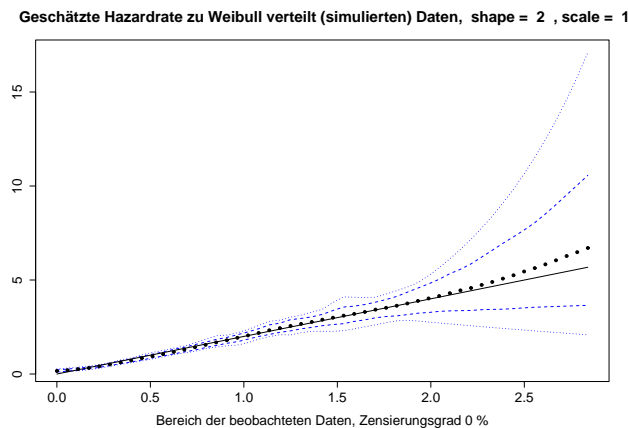


Abbildung 5.15: Spline Grad 1, 9 Gitterpunkte, $n = 1000$

gen am rechten Rand - in Ermangelung einer hinreichenden Anzahl von unzensierten Beobachtungszeitpunkten.

Bei der Analyse ist zu beachten, dass die Ergebnisse von dem Zensierungsmechanismus abhängen (vgl. Abschnitt (5.2.1)). Man gelangt zu den Lebensdauerdaten, indem man jeweils das Minimum von zwei unabhängigen Lebensdauerverteilungen betrachtet. Wählt man z.B. eine bimodale Verteilung (via Mischungen) für die zensierten Daten, kann dies zu anderen Resultaten führen.

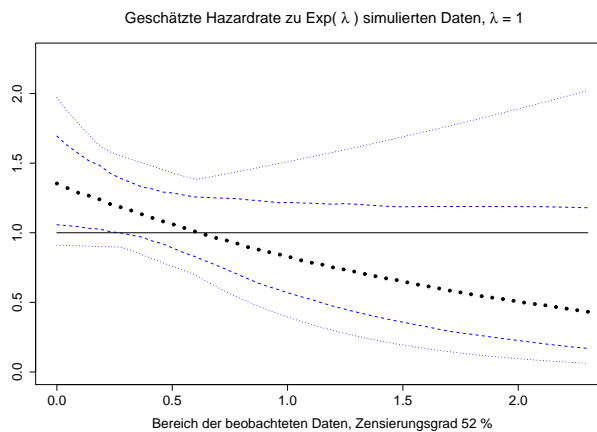


Abbildung 5.16: Polynom Grad 1, $n = 200$

Im Fall der exponentialverteilten Daten führt die Zensierung zu einer stärkeren Steigung (vgl. Abbildung (5.16)), zudem wird der Konfidenzbereich breiter, insbesondere am rechten Rand (siehe Grafik (5.17)). Die Splines zeigen ein ähnliches Verhalten, mit

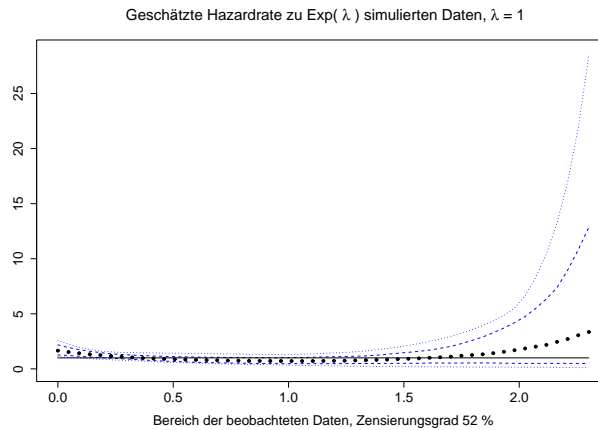


Abbildung 5.17: Polynom Grad 2

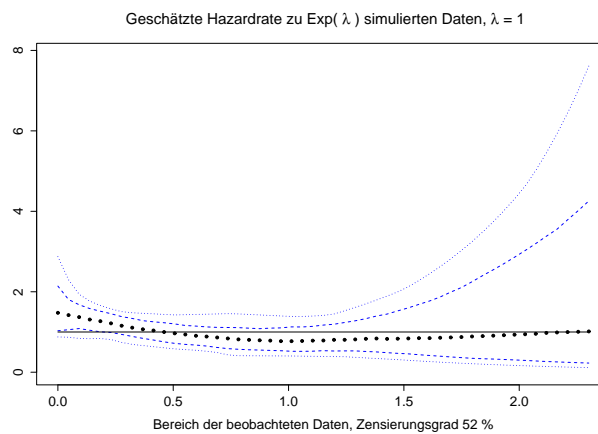


Abbildung 5.18: Spline Grad 1, 9 Gitterpunkte

Ausnahme des Falles $\mu = 5$. Dort wird auch bei einer Zensierung von 52% ein gutes Ergebnis erzielt (Abbildung (5.18))(siehe unten zum Vergleich Abbildung (5.23)) Ergebnis ohne Zensierung). Für die Situation mit weibullverteilten Daten ergibt sich ein verändertes Bild —bei einer Zensierung von 74.5%. Bei Polynomen vom Grad 2 oder höher wird die Hazardrate deutlich falsch geschätzt (siehe Abbildung (5.20) bzw. Abbildung (5.21)), wogegen die Schätzung beim Polynomgrad Eins gut ist (Abbildung (5.19)).

Geschätzte Hazardrate zu Weibull verteilt (simulierten) Daten, shape = 2 , scale = 1

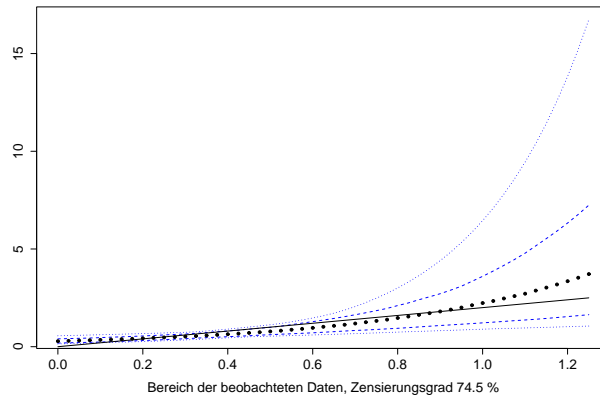


Abbildung 5.19: Polynom Grad 1, $n = 200$

Geschätzte Hazardrate zu Weibull verteilt (simulierten) Daten, shape = 2 , scale = 1

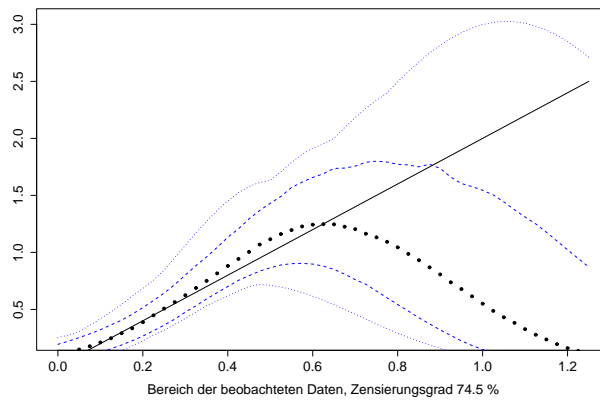


Abbildung 5.20: Polynom Grad 2

Geschätzte Hazardrate zu Weibull verteilt (simulierten) Daten, shape = 2 , scale = 1

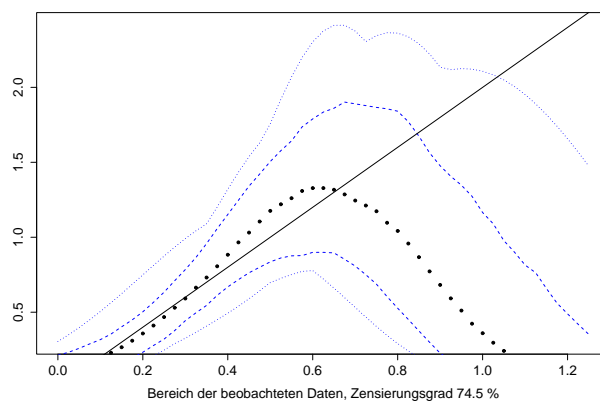


Abbildung 5.21: Polynom Grad 3

Wahl der a priori Parameter

Die Modellierung mit Splines führt (je nach Anzahl der Gitterpunkte) zu einer verschiedenen starken lokalen Anpassung. Die Stärke der Anpassung ist abhängig von der Wahl der Hyperparameter der a priori Verteilung und deren Einfluß auf die Varianz der a priori Verteilung (vgl. Formel (4.29) und Abschnitt (4.5.1)). Wählt man etwa den Hyperparameter μ_0 der a priori Verteilung der zufälligen Parameter $u_{\kappa_k}^\alpha$ betreffend der Monome $(t - \kappa_k)_+^g$ sehr klein, so liegen die Werte von $u_{\kappa_k}^\alpha$ nahe bei Null (vgl. die unteren beiden Zeilen der Grafik (5.24)), $\mu_0 = -10, \sigma_0^2 = 10$.

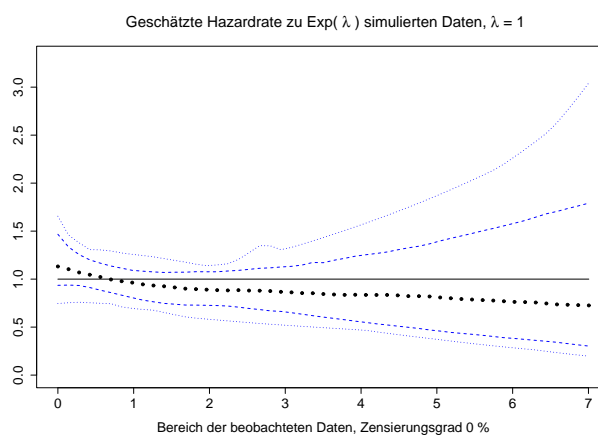


Abbildung 5.22: Spline Grad 1, 9 Gitterpunkte, $n = 200, \mu = -6$

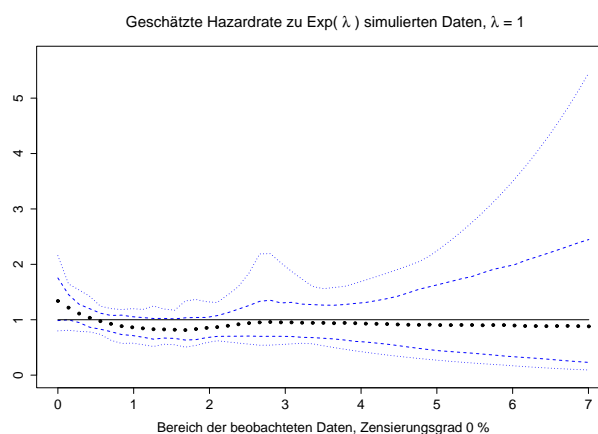


Abbildung 5.23: Spline Grad 1, 9 Gitterpunkte, $n = 200, \mu = +5$

Die korrespondierende Hazardrate befindet sich in Grafik (5.6). Sie zeigt einen Verlauf,

der einer rein polynomialen Modellierung gleicht. Wählt man einen größeren Wert für μ_0 , so ändert sich der Verlauf der Hazardrate (siehe (5.22), mit $\mu = -6$). Die Grafiken (5.23) und (5.25) basieren auf dem Wert $\mu_0 = 5$. Die Hazardrate im Bereich $[0, 2.5]$ einen leicht parabelförmigen Verlauf, bei sonst guter mittlerer Anpassung. Die Samplingpfade weisen eine höhere Autokorrelation auf, dies ist möglicherweise ein Hinweis auf die Überparametrisierung bei der Schätzung einer Konstanten.

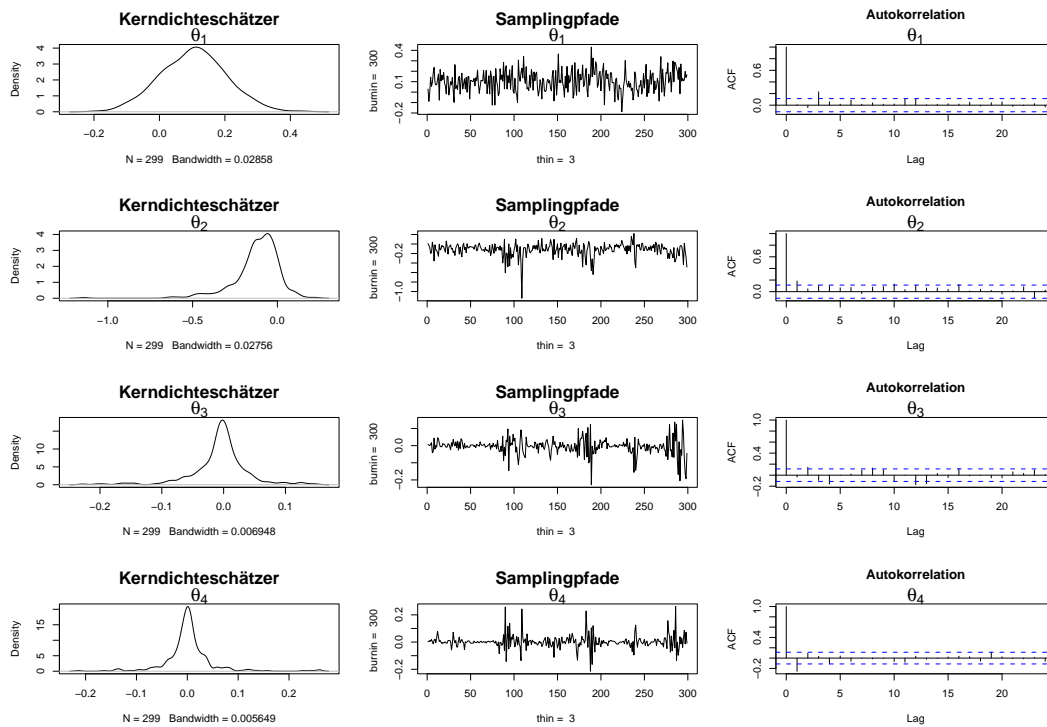


Abbildung 5.24: Spline Grad 1, $\mu_0 = -10$

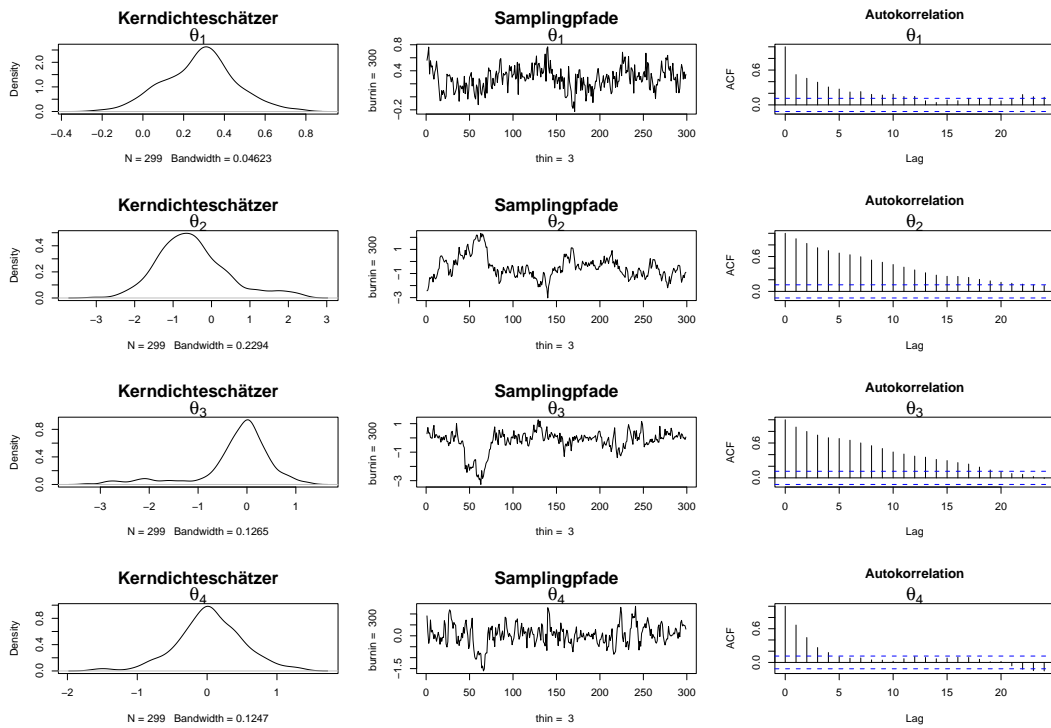


Abbildung 5.25: Spline Grad 1, $\mu_0 = +5$

Simulation mit Kovariablen

Die Darstellung der Simulation von Datensätzen folgt Bender, Augustin und Blettner (2005). Ergänzt man die Hazardrate um Kovariablen, so gilt $\lambda(t|x) = \lambda_0(t) \exp(x'\beta)$. Sei $\Lambda_0(t) = \int_0^t \lambda_0(s) ds$, so gilt für die Verteilungsfunktion $F(\cdot|x)$ des Cox-Modells mit Kovariablen x

$$F(t|x) = 1 - \exp(-\Lambda_0(t) \exp(x'\beta)) \quad . \quad (5.3)$$

Sei η gleichverteilt auf $(0, 1)$, dann läßt sich via Inversionsmethode zeigen, dass sich die Lebensdauern $T \sim F(\cdot|x)$ eines Cox-Modells simulieren lassen gemäß

$$T = \Lambda_0^{-1}(-\ln(\eta) \exp(-x'\beta)) \quad . \quad (5.4)$$

Entsprechend gilt für das Cox-Exponential (λ) Modell

$$T = -\frac{\ln(\eta)}{\lambda \exp(x'\beta)} \quad , \quad (5.5)$$

bzw.

$$T = \left(-\frac{\ln(\eta)}{\lambda \exp(x'\beta)} \right)^{1/\alpha} \quad (5.6)$$

für das Cox-Weibull Modell (vgl. Formel (5.1)). Es wurden Daten zu zwei Modellen simuliert, entsprechend dem oben geschilderten Vorgehen, also für den Cox-Exponentialfall (mit Parameter $\lambda = 1$) und für den Cox-Weibull Fall mit den Parametern $\alpha = 2$ und $\lambda = 1$. Es wurden zwei Kovariablen aufgenommen, die Kovariablen wurden simuliert als Realisationen von normalverteilten Zufallsgrößen mit Erwartungswert $\mu_1 = 1$ und $\sigma_1^2 = 1$ bzw. $\mu_2 = -0.5$ und $\sigma_2^2 = 2$ mit festen Effekten $\beta_1 = 0.5$ und $\beta_2 = -0.5$. Zusätzlich wurden Datensätze mit verschieden starker Zensierung betrachtet. Als Stichprobenumfang wurde $n = 200$ bzw. $n = 500$ gewählt. Die Verteilung der Lebensdauern ist linkssteil, wie man dem folgenden Boxplot (Grafik (5.26)) entnehmen kann. Während die maximale Lebensdauer (im unzensierten Fall) 25.091 beträgt, ist der Wert für das 75% Quartil 1.006. Für den größten Teil des zeitlichen Bereichs steht demnach nur eine geringe Anzahl von Daten zur Verfügung. Im Vergleich zu dem Fall ohne Kovariablen ist der Zeitbereich wesentlich größer. Die Hazardrate im Fall ohne Zensierung bei einem Polynom vom Grad Eins ist in Abbildung (5.27) dargestellt. Die Samplingpfade der Parameter sind in den Grafiken (5.28) bzw. (5.29). Die Tabelle (5.1) enthält einige

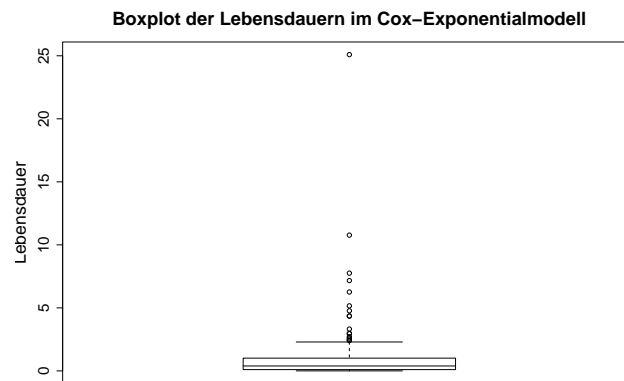
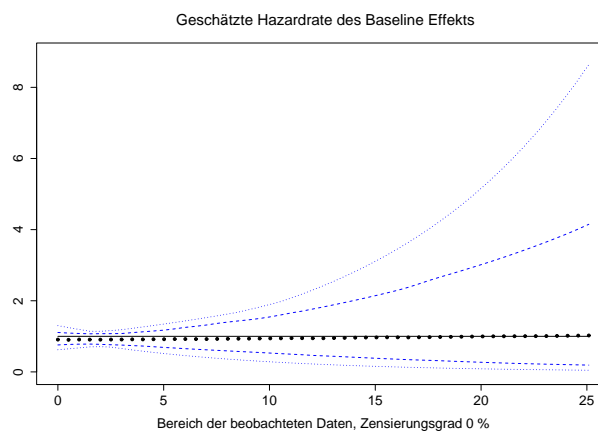
Abbildung 5.26: Boxplot für Datensatz mit $n = 200$ 

Abbildung 5.27: Polynom vom Grad 1

deskriptive Maßzahlen. Die Ergebnisse passen gut zu den theoretischen Werten zum einen bei den festen Parametern, die zu den Kovariablen korrespondieren, zum anderen führen die Schätzungen der Parameter des Polynoms zu einer Hazardrate, die nahe an der theoretischen Gerade verläuft. Insgesamt kann man sehr zufrieden sein, die Ergebnisse sind eigentlich besser als im Fall ohne Kovariablen. Bei starker Zensierung kann es zu einer Verschlechterung der Schätzungen kommen, die sich neben den Parametern, die für die Schätzung der Hazardrate benötigt werden, auch bei den Kovariableneffekten niederschlagen kann. Verwendet man anstelle des Polynoms einen Spline zur Schätzung der Hazardrate erhält man ähnliche Ergebnisse.

Bei zunehmender Zensierung verbreitern sich die Konfidenzstreifen der geschätzten Hazardrate (vgl. Grafik (5.30)). Zudem steigt die Autokorrelation in den Samplingpfaden.

Tabelle 5.1: Parameterschätzungen für ein Polynom ersten Grades

	\bar{x}	x_{Median}	s^2	x_0	$x_{0.05}$	$x_{0.95}$	x_1
$\hat{\beta}_1$	0.480	0.480	0.005	0.269	0.367	0.598	0.693
$\hat{\beta}_2$	-0.509	-0.510	0.002	-0.617	-0.569	-0.441	-0.388
$\hat{\alpha}_0$	-0.097	-0.096	0.014	-0.474	-0.277	0.103	0.263
$\hat{\alpha}_1$	0.001	0.005	0.002	-0.120	-0.066	0.063	0.101

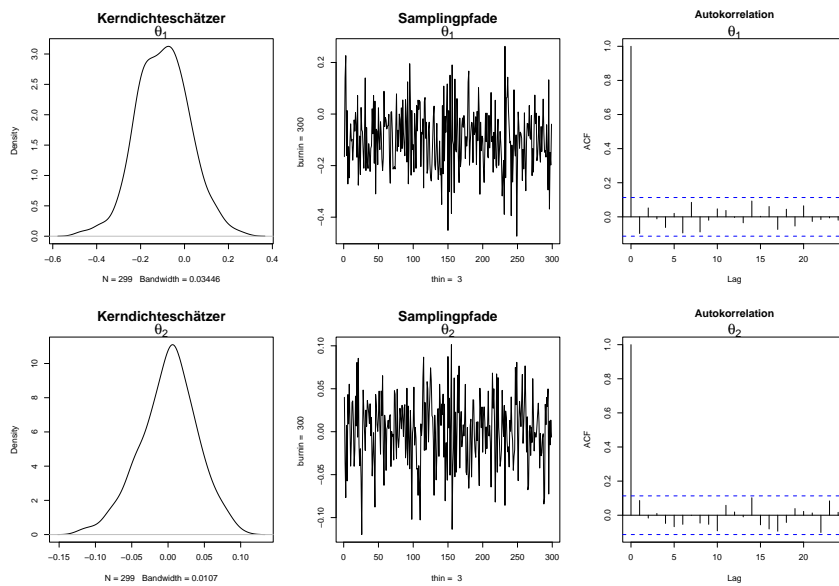


Abbildung 5.28: Polynom Grad 1

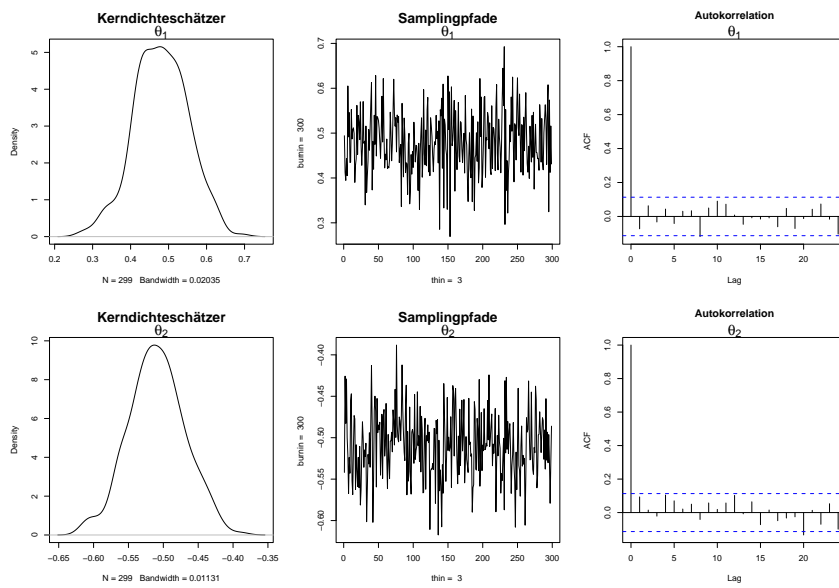


Abbildung 5.29: Polynom Grad 1

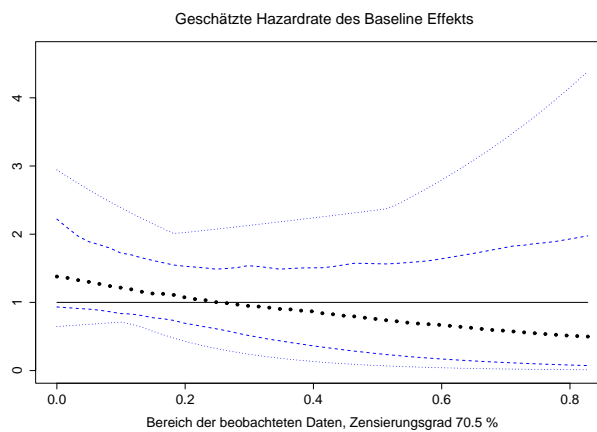


Abbildung 5.30: Polynom vom Grad 1

Die Ergebnisse für den Datensatz im Cox-Weibull Modell sind insgesamt ebenfalls relativ zufriedenstellend. Bedauerlich ist allerdings der trompetenförmige Verlauf der Konfidenzstreifen. Entsprechend dem Fall ohne Kovariablen ist der Verlauf der Hazardrate am rechten Rand günstiger bei Verwendung eines linearen Splines (hier mit neun Gitterpunkten) als bei einem Polynom (vgl. Grafik (5.32) bei $n = 500$). Die Hazardrate für das Splinemodell ist in Grafik (5.31) bei einem Datenumfang von $n = 200$ wiedergegeben. Die dazu korrespondierende Tabelle (C.1) der Parameterschätzungen findet sich im Anhang.

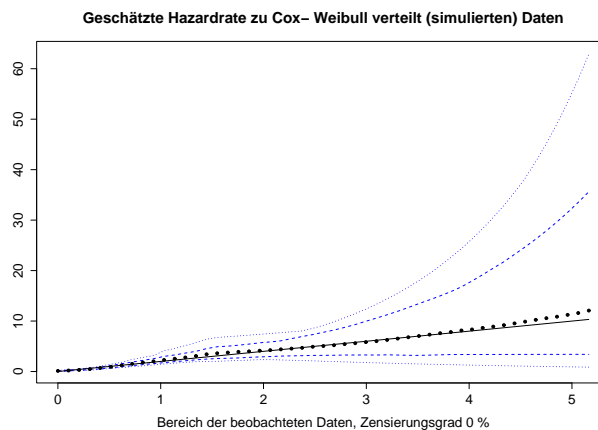


Abbildung 5.31: Spline vom Grad 1

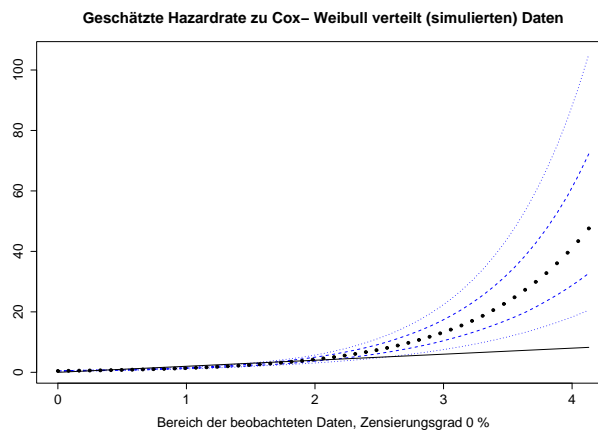


Abbildung 5.32: Polynom vom Grad 1

Simulation mit Kovariablen und zeitabhängigen Effekten

Die Formel (5.7) liefert ein Vorgehen zur Simulation von Datensätzen mit zeitabhängigen Effekten. Sei $\lambda(t|x) = \lambda_0(t) \exp(x'_1\beta_1 + x'_2\beta_{2t})$, dann gilt

$$F(t|x) = 1 - \exp[-\Lambda(t) \exp(x'_1\beta_1)] \quad , \quad (5.7)$$

wobei in Erweiterung zum bisherigen Fall für Λ gilt:

$$\Lambda(t) = \int_0^t \lambda_0(s) \exp(x'_2\beta_{2s}) ds \quad . \quad (5.8)$$

Sei T entsprechend $F(.|x)$ verteilt, dann folgt aus der Inversionsmethode, dass

$$U = \exp[-\Lambda(t) \exp(x'_1\beta_1)] \quad . \quad (5.9)$$

gleichverteilt auf $[0, 1]$ ist. Zufallsvariablen, die gemäß $F(.|x)$ verteilt sind, erhält man durch

$$T = \Lambda^{-1}[-\ln(U) \exp(-x'_1\beta_1)] \quad . \quad (5.10)$$

Die Inversion ist gewährleistet, da die Hazardrate strikt positiv ist. Im allgemeinen läßt sich die kumulierte Hazardrate nur numerisch bestimmen. Daher gestaltet sich die Inversion schwieriger, sie ist nicht in geschlossener Form möglich. Wählt man in Anlehnung an den exponentialverteilten Fall $\lambda_0(t) = \lambda$ sowie für β_{2t}

$$\beta_{2t} = \begin{cases} (1 - 0.5t)\beta_2 & , \text{ für } t \in [0, 2] \\ 0 & , \text{ sonst } \end{cases} \quad , \quad (5.11)$$

(oder einen anderen beliebigen linearen Zusammenhang) so gilt für die kumulierte Hazardrate $\Lambda(t) = \lambda \exp(x'_1\beta_1)\Lambda_0(t)$ für $0 \leq t \leq 2$

$$\begin{aligned} \Lambda_0(t) &= \int_0^t \exp(x_2\beta_{2s}) ds \\ &= \exp(\beta_2 x_2) \int_0^t \exp(-0.5x_2\beta_2 s) ds \\ &= \exp(\beta_2 x_2) [(-0.5x_2\beta_2)^{-1} \exp(-0.5x_2\beta_2 s)]_0^t \\ &= \frac{2 \exp(\beta_2 x_2)}{(x_2\beta_2)} (1 - \exp(-0.5t\beta_2 x_2)) \end{aligned}$$

bzw. für $t \geq 2$

$$\begin{aligned}\Lambda_0(t) &= \Lambda_0(2) + \int_2^t 1 ds \\ &= \Lambda_0(2) + t - 2 \quad .\end{aligned}$$

Die Umkehrfunktion $\Lambda_0^{-1}(t)$ wird stückweise definiert, für $t \in [0, \Lambda_0(2)]$

$$\Lambda_0^{-1}(t) = -\frac{2}{\beta_2 x_2} \ln \left(1 - \frac{t x_2 \beta_2}{2 \exp(x_2 \beta_2)} \right) \quad ,$$

wobei die Einschränkung des Definitionsbereichs (in Abhängigkeit von λ und $x_2 \beta_2$) die Positivität des Arguments der Logarithmusfunktion garantiert sowie für $t > \Lambda_0(2)$

$$\Lambda_0^{-1}(t) = t - \Lambda_0(2) + 2 \quad .$$

Die Daten werden analog zu Formel (5.10) gemäß

$$T = \Lambda_0^{-1} \left[-\ln(U) \frac{1}{\lambda} \exp(-x'_1 \beta_1) \right] \quad ,$$

simuliert, wobei jeweils die Fallunterscheidung beachtet werden muß.

Zur Überprüfung wird ein zweigleisiges Vorgehen eingeschlagen. Zum einen wird ein Datensatz mit einem zeitlich festen Kovariableneffekt erzeugt, bei dem der zeitlich feste Effekt als zeitveränderlicher Effekt modelliert wird. Der geschätzte Effekt sollte dann idealerweise konstant sein. Zum anderen wird ein Datensatz mit einem zeitabhängigen Effekt erzeugt, der entsprechend modelliert wird. Wie bisher wird der Baseline-Effekt $h_0(t)$ grafisch auf der linearen Skala dargestellt, also in der Form $\exp(h_0(t))$, der Effekt des zeitvariierenden Koeffizienten β_i auf der logarithmischen Skala, also in der Form $\ln(\exp(\beta_i \cdot (\cdot)))$. Der erste Fall kann als der einfachere angesehen werden, da es möglich sein sollte, eine konstante Funktion zu schätzen. Der für die Simulation verwendete Wert β beträgt 0.5. Der Umfang des Datensatzes betrug $n = 500$. In Grafik (5.33) sieht man die Schätzung des konstanten Effekts anhand eines Splines vom Grad Eins sowie in Grafik (5.34). Die Schätzung der Baseline-Hazardrate, der schwarze Strich, entspricht dem exakten Verlauf.

Bei einem Zensierungsgrad von 64.5% erhält man für den zeitvariierenden Effekt einen relativ ähnlichen Verlauf im Vergleich zum nichtzensierten Fall. Bei der Schätzung der Baseline-Hazardrate führt die Zensierung zu einer Verfälschung in Form eines monoton

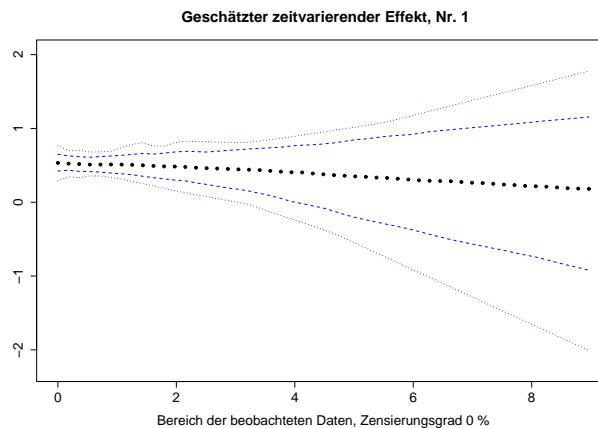


Abbildung 5.33: Spline vom Grad 1, 15 Gitterpunkte, ohne Zensierung

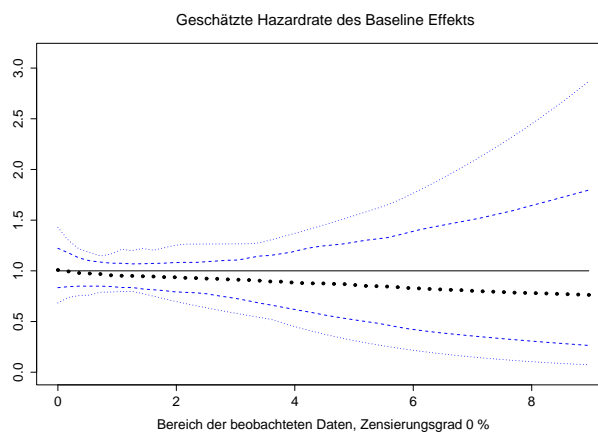


Abbildung 5.34: Spline vom Grad 1, 15 Gitterpunkte, ohne Zensierung

fallenden Verlaufs (Grafik (5.35)).

Bei der Schätzung zu simulierten Daten mit einem zeitvariierenden Effekt gemäß (5.11) treten mehr Probleme auf. Die guten Schätzungen in den Grafiken (5.39) und (5.38) erhält man bei einem Datensatz mit $n = 2500$ ohne Zensierung. Bei einer Zensierung von 35.4% erhält man zwar für die Baseline-Hazardrate gute Resultate, schwieriger ist die Schätzung des zeitvariierenden Effekts. Es gelingt nicht, die geknickte Darstellung zu finden. Vielmehr schätzt man eine linear fallende Funktion. Dieses Verhalten ist auch typisch für kleinere Datensätze, etwa mit $n = 500$ oder $n = 1000$ und bei einer geringen Anzahl von Gitterpunkten. Eine Schätzung basierend auf einem Spline vom Grad 1 mit 15 Gitterpunkten findet man in Grafik (5.36).

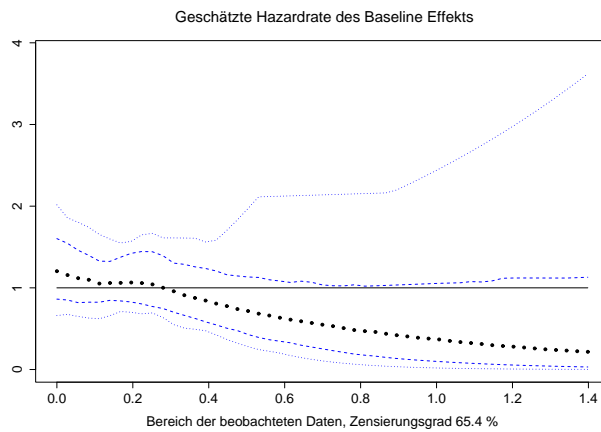


Abbildung 5.35: Spline vom Grad 1, 15 Gitterpunkte

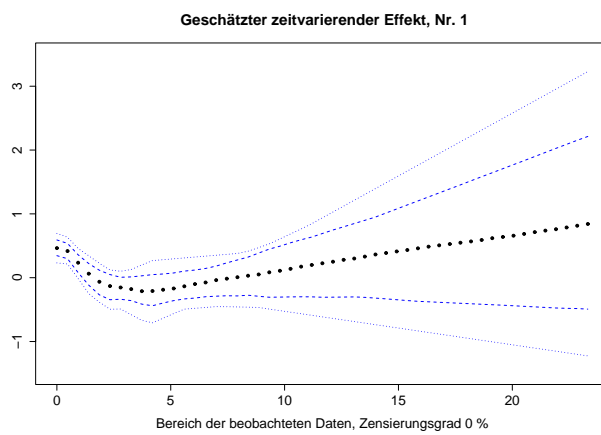


Abbildung 5.36: Spline vom Grad 1, 15 Gitterpunkte, ohne Zensierung

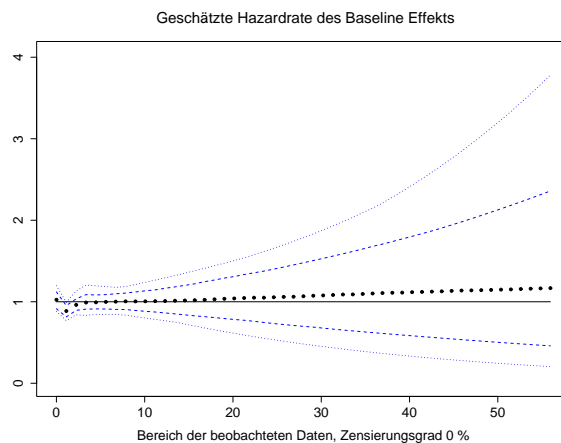


Abbildung 5.37: Spline vom Grad 1, 7 Gitterpunkte, ohne Zensierung

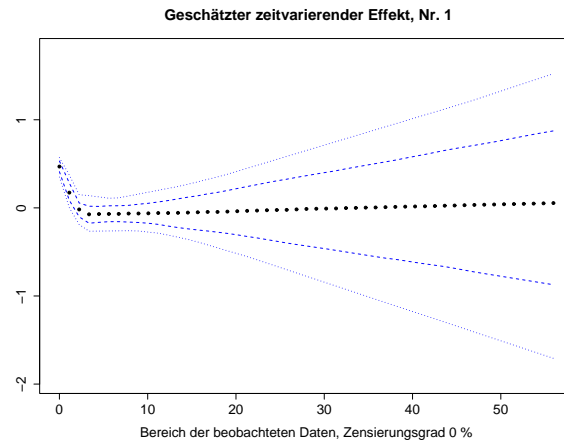


Abbildung 5.38: Spline vom Grad 1, 7 Gitterpunkte, ohne Zensierung

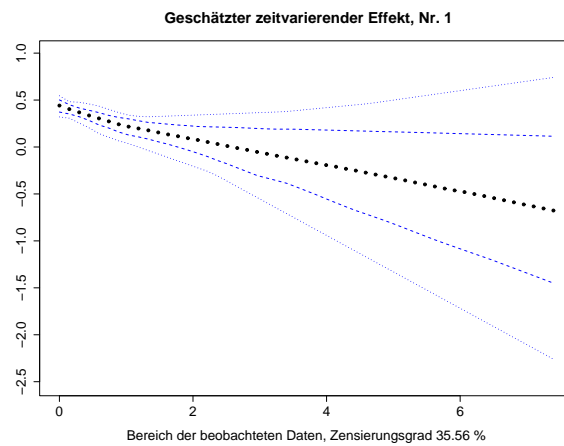


Abbildung 5.39: Spline vom Grad 1, 7 Gitterpunkte

Abschließende Bemerkungen

Das hybride Monte Carlo Verfahren hat sich als zuverlässig erwiesen. Die Qualität der Ergebnisse hängt von der Datensituation ab. Die etwas enttäuschenden Ergebnisse im exponential — bzw. weibullverteilten Fall sollten vielleicht im Vergleich zu (5.1.2) betrachtet werden. Der Anteil der zensierten Daten spielt bei den Schätzungen eine Rolle, als Faustregel könnte man sagen, dass eine Zensierung bis ca. 15–20% – je nach Komplexität der Hazardrate – zu vernachlässigbaren Verlusten führt. Da die unabhängige Rechtszensierung dazu führt, dass das zu betrachtende Zeitintervall kleiner wird – große Lebensdauern werden eher zensiert – kann eine leichte Zensierung sogar die Schätzung verbessern. Allerdings kann man sich bei (starker) Zensierung nicht sicher sein, ob der zeitvariierende Effekt richtig geschätzt wird. Betrachtet man den Fall des zeitabhängigen Kovariableneffekts (siehe Grafik (5.39)), so würde man wohl fälschlicherweise einen linear abfallenden Effekt vermuten. Die Aussage wird allerdings relativiert durch die Konfidenzstreifen, die den Wert Null enthalten. Die Konfidenzstreifen sollten stets mitberücksichtigt werden.

Das typische Vorgehen soll kurz zusammengefaßt werden. Je nach Größe der Lebensdauern sollte abgewägt werden, ob eine Transformation des Zeitbereichs sinnvoll ist. Bei der Wahl der Schrittweite τ sollte mit einem kleinen Wert begonnen werden, etwa $\tau = 0.0001$. Weisen die Sampling Pfade eine (hohe) Autokorrelation auf, so sollte τ vergrößert werden, analog kann eine Erhöhung der Anzahl der Leapfrog–Schritte die Korrelation zwischen den einzelnen Parametern reduzieren. Als Startwert könnte man etwa einen Wert von 100 wählen, Werte oberhalb von 200 haben eine eher geringe Wirkung. Generell gilt, dass die Schrittweite und die Anzahl der Leapfrog–Schritte zueinander passen müssen. Bei großem τ und einer großen Anzahl von Leapfrog–Schritten kann der Approximationsfehler des Leapfrog Schemas zu einer geringen Akzeptanzwahrscheinlichkeit führen. Um einen visuellen Eindruck von dem möglichen zeitvariierendem Effekt zu bekommen, sollte man zuerst mit Polynomen (Grad Eins bis Drei) arbeiten. Danach kann man Verfeinerungen mit Splines vornehmen. In den durchgeführten Simulationen zeigte sich, dass bei Splines im Vergleich zu Polynomen der rechte Rand des Zeitbereichs günstiger dargestellt wird. Ein Nachteil bei Splines kann sein, dass sie für die Hazardrate kaum gebraucht werden, da die Hazardrate hinreichend gut durch Polynome dargestellt werden kann. In diesem Fall nehmen die zufälligen Parameter u^α sehr kleine Werte nahe bei Null an. Die Monte Carlo Pfade der korrespondierenden (Hyper-) Parameter weisen dann eine nur langsam abfallende Autokorrelation auf, offenbar wegen des schwachen Einflusses auf die Daten. Eine längere Burn in Phase muß in Kauf genommen werden.

5.2.2 Ergebnisse zu Datensatz PBC

Der Datensatz wurde in Kapitel (1) vorgestellt. Es werden dieselben Variablen wie im Cox-Modell (vgl. Abschnitt (1.2)) gewählt. In einem ersten Schritt wurde angenommen, dass alle Parametereffekte zeitunabhängig sind, danach werden Modelle mit zeitabhängigen Effekten berechnet.

In den Bemerkungen am Ende des Abschnitts (3.6.4) sowie im Abschnitt (4.6.2) wurde erwähnt, dass man bei der kinetischen Energie $k(\mathbf{p})$ anstelle eines Vektors \mathbf{p} mit i. i. d. standardnormalverteilten Zufallsgrößen eine Modifikation vornehmen kann. Diese Modifikation läßt sich einerseits im Sinne der stochastischen Modellierung inhaltlich als eine approximative Korrektur zweiter Ordnung (via Korrespondenz zur Informationsmatrix) interpretieren, andererseits als eine Korrektur des Hamilton Systems, da die Parameter aufgrund der jeweiligen Zugehörigkeit zu den korrespondierenden Kovariablen nicht als gleichwertig angesehen werden können. Anschaulich gesprochen haben die Parameter (= Teilchen des Systems) eine unterschiedliche Masse und es wird versucht, dies über die Korrektur miteinzubeziehen¹.

Zum leichteren Verständnis der Modelle wird die Notation zur Darstellung der Intensitätsrate kurz in einer informellen Weise wiederholt und zwar in Gestalt der Prädiktoren, die exponentiell in die Intensitätsrate eingehen. Mit $\mathbf{x}_i^t \boldsymbol{\beta}$ wird stets der Prädiktor der festen Effekte bezeichnet. Der Prädiktor der Baseline-Hazard ist $\eta_s^{bs}(\theta_{bs})$, die individuenspezifischen zeitabhängigen Prädiktoren lauten entsprechend $\eta_{is}^{edema}(\theta_{edema})$, $\eta_{is}^{\log(bili)}(\theta_{\log(bili)})$ und $\eta_{is}^{\log(protime)}(\theta_{\log(protime)})$. Entsprechend des jeweiligen Modells werden für die zeitabhängigen Effekte Polynome oder Splines verwendet. Alle Kovariablen, die ohne zeitabhängigen Effekt modelliert werden, sind im linearen Prädiktor enthalten.

Modell mit festen Parametereffekten

Bei einem Modell mit festen Parametereffekten und einer Baseline-Hazardrate, die mit einem Polynom vom Grad 1 modelliert wird, lautet die Intensitätsrate (vgl. Abschnitt 4.5.1) pro Individuum i

$$\lambda_{is} = \exp(\eta_s^{bs}(\theta_{bs}) + \mathbf{x}_i^t \boldsymbol{\beta}) \quad (5.12)$$

Die Ergebnisse der Simulationen befinden sich in Tabelle (D.3), s. S. 156, die a posteriori Schätzungen sind – abgesehen vom Effekt für $\log(protime)$ – denen des Cox-Modells (vgl. (1.2)) sehr ähnlich. Das gleiche Modell wurde anschließend mit zentrierten Kovariablen gerechnet (siehe (D.4)). Dies vermindert die Autokorrelation der Samp-

¹Auf diese Korrektur bezieht sich im folgenden der Hinweis ‘Modifiziertes Verfahren’

lingpfade. Zudem verschiebt sich der Parameter $\hat{\alpha}_0$, der die Konstante der Hazardrate modelliert, entsprechend der multiplikativen Struktur der Hazardrate. Die Ergebnisse dieser modifizierten Vorgehensweise finden sich in den Tabellen (D.1) (Kovariablen nicht zentriert) bzw. (D.2) (Kovariablen zentriert). Vergleicht man die Samplingpfade der Parameter betreffend die Baseline-Hazardrate (Grafik (D.1)) bzw. (Grafik (D.2)), so erkennt man, dass bei der modifizierten Version die Samplingpfade kaum autokorreliert sind.

Modell mit zeitabhängigen Parametereffekten

Die Ergebnisse in Abschnitt 1.2 legen nahe, eine Modellierung der Kovariablen edema , $\log(\text{bili})$, $\log(\text{protime})$ mit zeitabhängigen Parametereffekten durchzuführen. Der Grafik (A.1) bzw. der Tabelle (A.1) kann man entnehmen, dass die durchschnittliche Überlebenszeit bei der Variable edema stark von den Ausprägungen (0, 0.5, 1.0) abhängt. So gibt es nur zwei Individuen außerhalb des Zeitbereichs $[0, 2.658]$ ohne Zensurierung mit der Ausprägung $\text{edema} = 1$, nämlich $T_1 = 3.926$ und $T_2 = 9.385$. Dies hat zur Folge, dass das modifizierte Vorgehen bei der Variable edema scheitert, falls man Splines vom Grad 1 verwendet, da die Matrix J_t aus Abschnitt (4.6.2) nicht den (numerisch) vollen Rang hat. Bei Splines vom Grad 2 mit einer geringen Anzahl von Gitterpunkten (z.B. 4) ist kein Rangverlust zu beobachten. Die prinzipielle Frage, ob man einen zeitvariierenden Effekt für eine Kovariable mit wenigen Ausprägungen, die zeitlich sehr ungleich verteilt sind, schätzen kann, bleibt davon unberührt. Möglicherweise ist für solche Fälle eine Stratifizierung der Intensitätsrate günstiger.

Modell mit zeitabhängigen Parametereffekten für $\log(\text{bili})$, $\log(\text{protime})$

Es wurden 2 Modelle gerechnet, zum einen ein Modell, bei dem die zeitvariierenden Effekte mit Polynomen vom Grad 3 dargestellt werden (modifiziertes Vorgehen), zum anderen mit Splines vom Grad 2 mit 4 Gitterpunkten. Die Intensitätsrate lautet

$$\lambda_{is} = \exp(\eta_s^{bs}(\theta_{bs}) + \mathbf{x}_i^t \boldsymbol{\beta} + \eta_{is}^{\log(\text{bili})}(\theta_{\log(\text{bili})}) + \eta_{is}^{\log(\text{protime})}(\theta_{\log(\text{protime})})) \quad (5.13)$$

Eine Tabelle der Ergebnisse für den polynomialen Fall befindet sich im Anhang (siehe Tabelle (D.5), S. 159), ebenso wie die Grafiken der Samplingpfade (siehe Abbildungen (D.6)–(D.9)) bzw. der geschätzten zeitvariierenden Effekte (Abbildungen (D.3)–(D.5), s. S. 160 bzw. S. 161). Die Ergebnisse im Falle der Splines finden sich ebenfalls im Anhang (vgl. Tabelle D.6, S. 164 bzw. Grafiken D.10 – D.12, S. 165 ff)

Modell mit zeitabhängigen Parametereffekten für $\log(\text{bili})$, edema und $\log(\text{protime})$

Es wurde ein Modell gerechnet mit zeitvariierenden Effekten für $\log(\text{bili})$, edema , $\log(\text{protime})$

$$\lambda_{is} = \exp(\eta_s^{bs}(\theta_{bs}) + \mathbf{x}_i^t \boldsymbol{\beta} + \eta_{is}^{\log(\text{bili})}(\theta_{\log(\text{bili})}) + \eta_{is}^{\log(\text{protime})}(\theta_{\log(\text{protime})}) + \eta_{is}^{\text{edema}}(\theta_{\text{edema}})) \quad (5.14)$$

Für die zeitvariierenden Effekte wurde ein Polynom vom Grad 3 verwendet (modifiziertes Vorgehen). Eine Tabelle zu den Ergebnissen findet sich in (D.7). Die Grafiken ((5.40) – (5.43)) beschreiben den funktionalen Verlauf. Die korrespondierenden Samplingpfade befinden sich in (D.13)–(D.16).

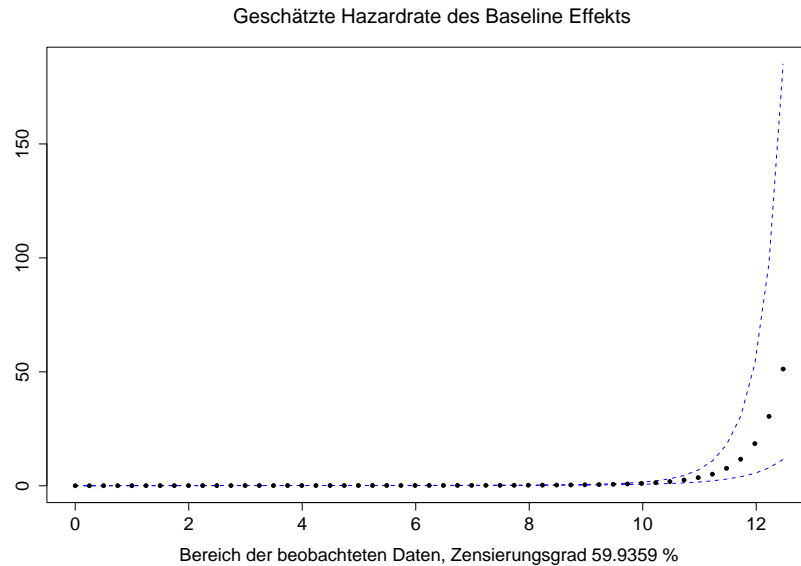


Abbildung 5.40: Baseline-Hazardrate, Polynom Grad 3

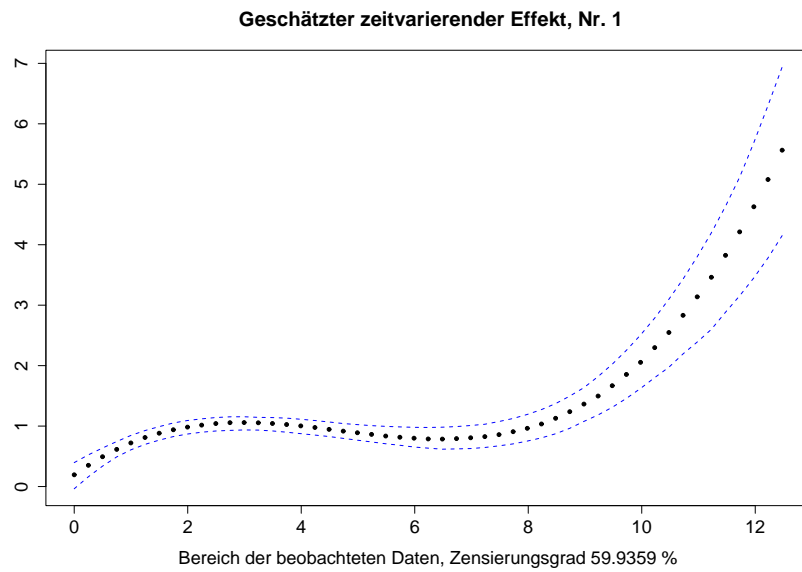


Abbildung 5.41: zeitvariierender Effekt für $\log(\text{bili})$, Polynom Grad 3

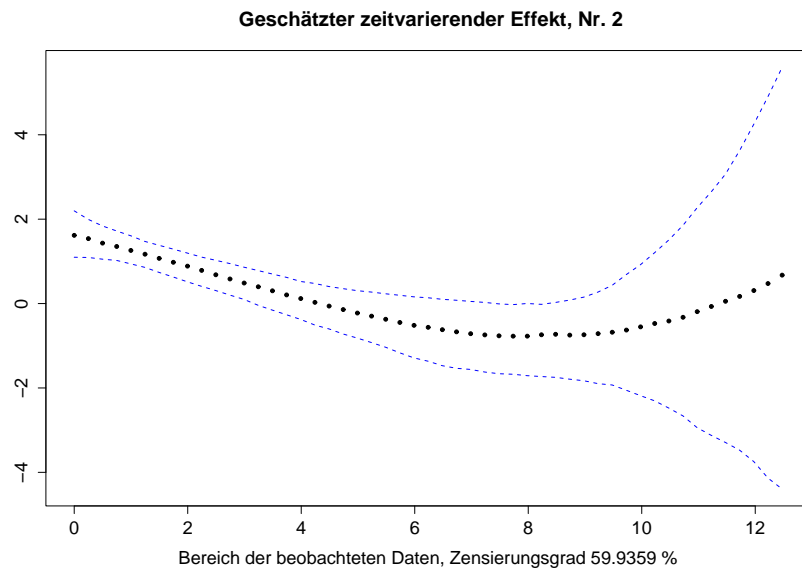


Abbildung 5.42: zeitvariierender Effekt für edema, Polynom Grad 3

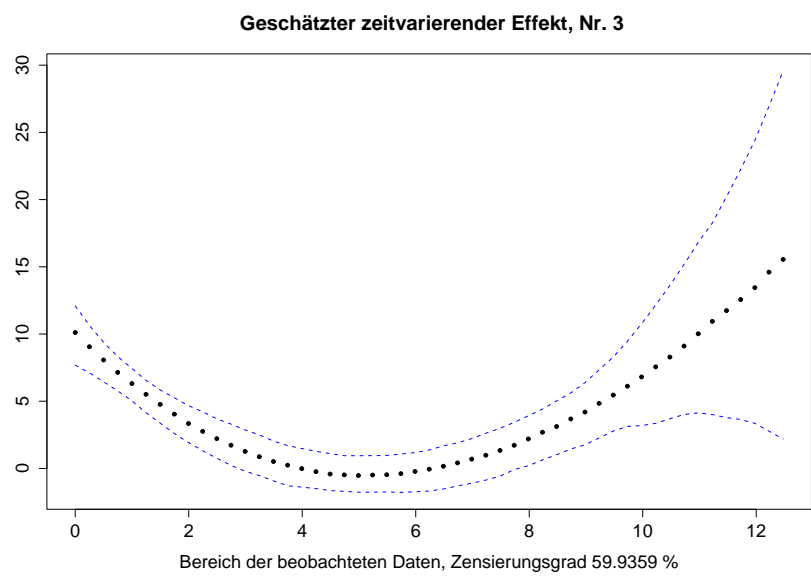


Abbildung 5.43: zeitvariierender Effekt für $\log(\text{prot ime})$, Polynom Grad 3

Alternativ wurde ein Modell mit Splines vom Grad 2 mit 4 Gitterpunkten gerechnet. Die Ergebnisse finden sich in Tabelle (D.8), die Grafiken mit den zeitvariierenden Effekten in (5.44)–(5.47). Exemplarisch sind die Samplingpfade für die Baseline Hazardrate (siehe (D.18),(D.19)) sowie für den zeitvariierenden Effekt für $\log(\text{bili})$ (siehe (D.20), (D.21)) im Anhang zu finden.

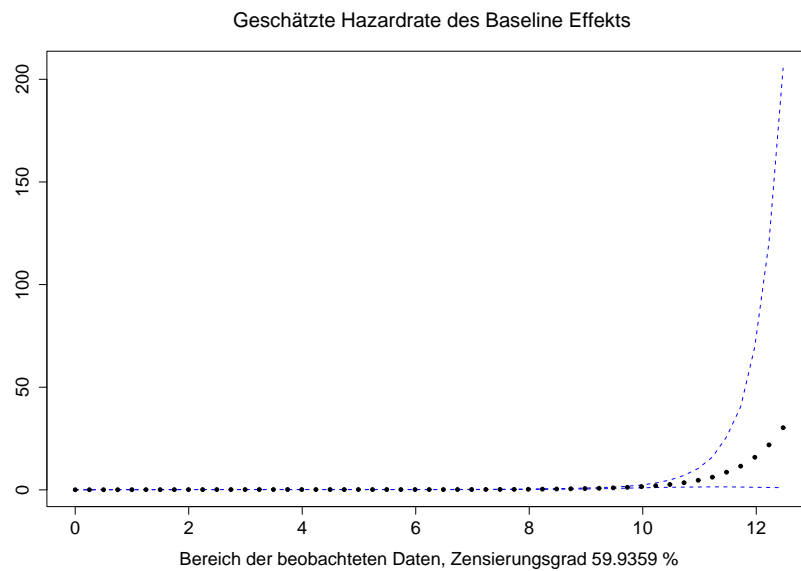


Abbildung 5.44: Baseline-Hazardrate, Spline Grad 2, 4 Gitterpunkte

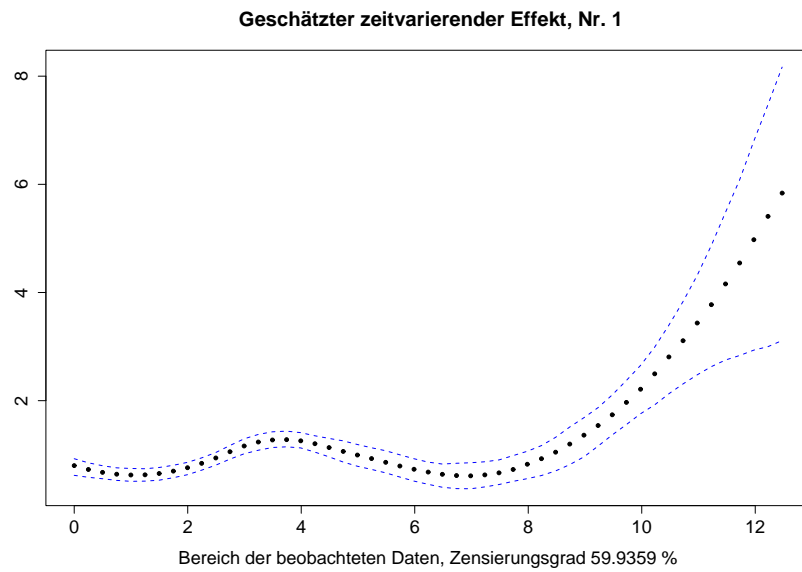


Abbildung 5.45: zeitvarierender Effekt für $\log(\text{bili})$, Spline Grad 2, 4 Gitterpunkte

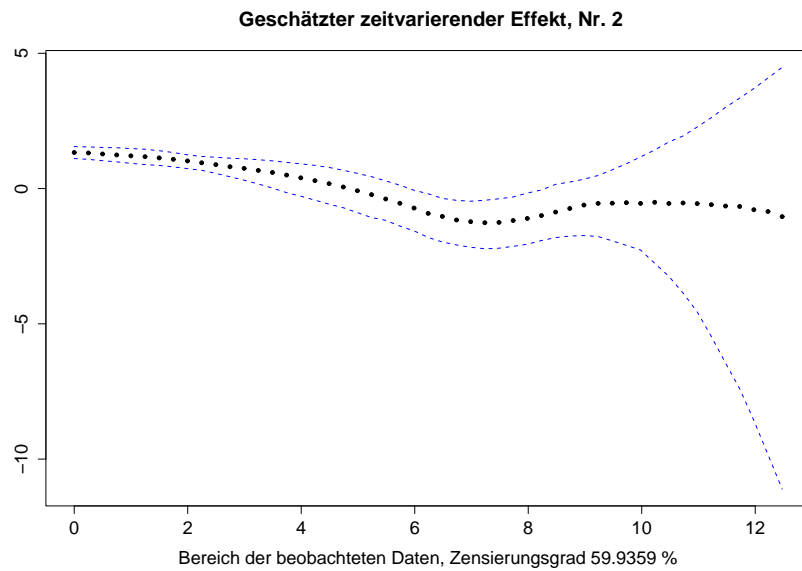


Abbildung 5.46: zeitvarierender Effekt für edema, Spline Grad 2, 4 Gitterpunkte

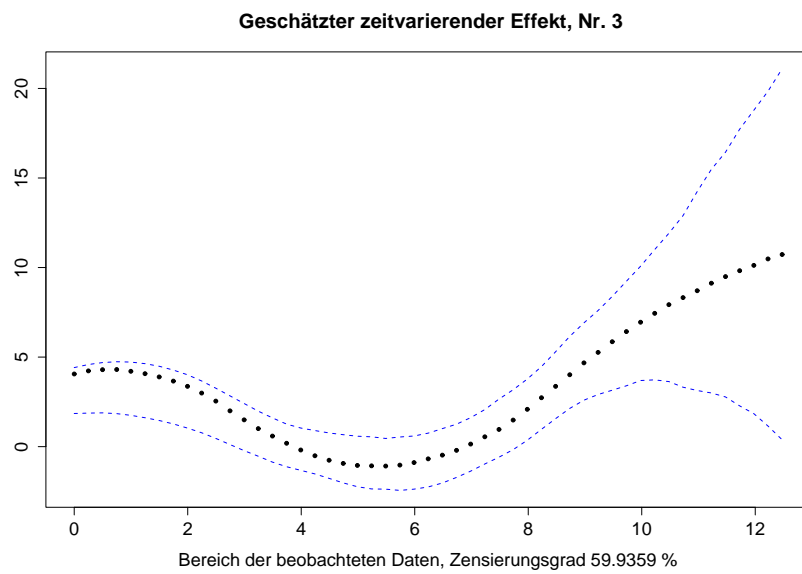


Abbildung 5.47: zeitverändernder Effekt für $\log(\text{protime})$, Spline Grad 2, 4 Gitterpunkte

Fazit

Es wurde bereits erwähnt, dass die Bayes Schätzungen unter der Annahme von zeitlich konstanten Kovariableneffekten denen im Cox-Modell sehr ähnlich sind. Bemerkenswerter ist, dass der funktionale Verlauf der zeitabhängigen Parametereffekte für die Variablen $\log(\text{bili})$, edema , $\log(\text{protime})$ im Bayes Modell stark dem mit `lowess` modellierten Verläufen anhand der partiellen skalierten Schoenfeldresiduen gleicht. Dies ist ein für die Praxis erfreuliches Ergebnis. Zwar können die Aussagen dieses Beispiels nicht verallgemeinert werden, aber es erscheint durchaus sinnvoll im ersten Schritt ein Cox-Modell zu berechnen, selbst wenn die Proportional Hazard Annahme verletzt ist, um danach anhand der Schoenfeldresiduen den glatten zeitlichen Verlauf des Parametereffekts zu schätzen.

Im Unterschied zu den Schätzungen anhand der Residuen ist die Variation der Funktionen im Bayes Modell am rechten Rand ausgeprägter, zudem weitet sich dort der Konfidenzbereich trompetenförmig.

Die semiparametrische Modellierung basiert auf Splines und ist so konzipiert, dass die Glattheit der Splines datengesteuert wird. Je nach Komplexität des unbekanntem zeitvariierenden Effekts können Polynome vom Grad 2 oder 3 ausreichend sein. Zudem sollte bei Polynomen die Korrektur (vgl 4.6) der Schrittweite des Leapfrog Verfahrens immer funktionieren, was bei Splines nicht unbedingt gewährleistet ist.

5.3 Verfahren mit Schätzfunktionen

Es werden einige Resultate – allerdings auch einige Probleme – zusammengetragen. Die Resultate beziehen sich auf den Datensatz PBC. Es wurde ein Modell mit festen Kovariableneffekten gerechnet, die Baseline-Hazardrate wurde mit Polynomen geschätzt. Die angesprochenen Probleme sind numerischer Natur. Sie haben ihre Ursache in der Form der Schätzfunktionen sowie der Modellierung der zeitvariierenden Effekte. Auf eine Modellierung mit Splines, dem eigentlichen Ziel und eine Schätzung via gemischte Modelle wurde verzichtet. Bei den – im Vergleich zur Verwendung von Polynomen – komplexeren Modellen ist zu erwarten, dass noch größere numerische Schwierigkeiten auftreten.

5.3.1 Modell mit zeitunabhängigen Kovariableneffekten

Gewissermaßen als Einstieg in die Verfahren, die auf Schätzfunktionen basieren, wird ein Modell berechnet, das konstante Kovariableneffekte annimmt. Die Baseline-Hazardrate wird als einziger zeitabhängiger Effekt modelliert. Es wird Bezug genommen auf die Darstellung in Abschnitt (4.5.2). Der Beitrag zur Schätzfunktion für die festen Kovariableneffekte pro Individuum i lautet (vgl. Gleichung 4.46) und Zensierungsindikator δ_i

$$\mathbf{x}_i^t \left[\delta_i - \left(\exp(\eta_i(\boldsymbol{\beta}) + \eta_i(\theta_3)) \int_0^{t_i} \lambda_{i3}^*(\theta) ds \right) \right] . \quad (5.15)$$

Die Schätzungen $\hat{\boldsymbol{\beta}}$ erhält man als Nullstellen der Schätzfunktion. Analog erhält man die Schätzungen für die Baseline Hazardrate. Entsprechend Abschnitt (4.5.2) werden die Schätzungen blockweise iterativ durchgeführt. Es zeigt sich, dass zwar die Schätzfunktion für die festen Effekte bzw. für die Baseline Hazard jeweils Null ist, dass aber pro Schätzfunktion (mindestens) ein Schätzwert divergiert, einmal nach $+\infty$, einmal nach $-\infty$. Zur Vermeidung dieser Konvergenzprobleme wird vorgeschlagen, die Kovariablen \mathbf{x}_i zu zentrieren, was auch zum gewünschten Erfolg führt. Eine Tabelle der Ergebnisse findet sich im Anhang (D.9). Neben den Schätzungen aus dem Cox-Modell finden sich die Ergebnisse der Schätzfunktionen, wobei im ersten Fall die Baseline-Hazardrate mit einem Polynom ersten Grades, im zweiten Fall mit einem Polynom zweiten Grades geschätzt wurde. Für die fixen Effekte ist der Standardfehler beim Cox-Modell größer, teilweise um fast 100%. Auffallend ist ebenfalls die geringe Standardabweichung der Parameter der Baseline-Hazardrate.

5.3.2 Modell mit zeitabhängigen Kovariableneffekten

In einem weiteren Schritt wurde versucht, Kovariableneffekte zeitabhängig zu modellieren. Hierfür können allerdings keine Ergebnisse dargestellt werden, da die Verfahren zum Bestimmen der Nullstellen wegen numerischer Schwierigkeiten abgebrochen wurden. Als eine mögliche Ursache der Schwierigkeiten könnte die Approximation der Hessematrix durch die Gradienten beim Quasi Newton Verfahren in Betracht kommen, auf die daher verzichtet wurde. Statt dessen wurde die Approximation der Hessematrix über finite Differenzen durchgehend verwendet, die bisher nur bei der Initialisierung vorgeschlagen wurde (siehe Formel (4.71)). Der Algorithmus dazu findet sich in Press et al. (1992). Man braucht dazu zwar mehr Funktionsauswertungen, bekommt aber bei vorsichtiger Wahl von h eine im allgemeinen gute Näherung der Hessematrix. Dies wurde auch überprüft, indem alle Elemente der Matrix – als Funktionen der Zeit – numerisch integriert wurden. Dies führte leider ebenfalls nicht zum Erfolg. Ursache war, dass die Hessematrix nicht positiv definit war, bzw. dass bei der Schrittweitenbestimmung (s. Seite 79) keine Lösung gefunden wurde. Eine mögliche Fehlerquelle könnte natürlich auch die numerische Integration darstellen. Die numerische Integration wird besser, je mehr Chebychev Polynome verwendet werden. Unabhängig von der verwendeten Zahl der Polynome konnte keine Lösung gefunden werden.

Im folgenden wird versucht, eine Ursache der Probleme zu finden. Der Prädiktor für die j -te Kovariable und das i -te Individuum sei $\eta_{ijs} = x_{ij}f_j(s)$ mit $f_j(s)$ als zeitvariierendem Effekt. Zur Vereinfachung der Notation wird angenommen, dass $f_j(s)$ ein Polynom ist mit $f_j(s) = \alpha_0 + \alpha_1 s + \alpha_2 s^2$. Die Aussagen sind allerdings auch für Splines gültig. Der Prädiktor modelliert den Intensitätsprozeß λ_{is} , d.h. es gilt

$$\lambda_{is} = \exp(\dots + \eta_{ijs} + \dots) \quad . \quad (5.16)$$

Die Schätzfunktionen enthalten die Ableitungen nach den Parametern. Betrachtet man die partiellen Ableitungen für $(\alpha_0, \alpha_1, \alpha_2)$, so gilt

$$\lambda'_{is} = x_{ij}\lambda_{is} \cdot (1, s, s^2)^t \quad . \quad (5.17)$$

Für die Hessematrix benötigt man die zweiten Ableitungen

$$\lambda''_{is} = x_{ij}^2\lambda_{is} \cdot (1, s, s^2)^t \cdot (1, s, s^2) \quad . \quad (5.18)$$

Das äußere Produkt der beiden Vektoren führt zu einer Matrix mit Rang 1, die folgende Gestalt hat

$$(1, s, s^2)^t \cdot (1, s, s^2) = \begin{pmatrix} 1 & s & s^2 \\ s & s^2 & s^3 \\ s^2 & s^3 & s^4 \end{pmatrix} . \quad (5.19)$$

Die elementweise Integration über die Lebensdauer von $[0, T_i]$ für das Individuum i ändert nichts daran, dass die Matrix den Rang 1 hat. Um eine Matrix mit vollem Rang r zu erhalten, muß man mindestens die Matrizen von r Individuen aufsummieren, d.h. je komplexer der zeitvariierende Effekt modelliert wird, desto mehr Individuen braucht man um eine Hessematrix mit vollem Rang zu erhalten. Ein weiterer Aspekt ist die zeitliche Struktur der Lebensdauern. Da in der Matrix Polynome auftauchen, wird die Summe der individuellen Hessematrizen stark beeinflusst durch die Hessematrizen der Individuen mit den größten Lebensdauern, da ihre Werte am größten sind. Insbesondere dominieren sie möglicherweise die rechte Spalte mit den höchsten Potenzen. Die Kondition der Hessematrix, also das Verhältnis aus dem größtem und dem kleinstem Eigenwert über alle Individuen ist für den Datensatz PBC aber sehr ungünstig. Dies ist nicht weiter überraschend angesichts der Struktur der einzelnen Matrizen pro Individuum. Das Versagen der Verfahren zur Bestimmung der Nullstellen erscheint unter diesem Aspekt plausibel. Im vorherigen Abschnitt wurde die Baseline-Hazardrate mit Polynomen geschätzt. Setzt man $x_{ij} = 1$, so erkennt man, dass dort die gleichen potentiellen Gefahren lauern wie bei zeitvariierenden Effekten. Komplexere Modelle mit zeitvariierenden Effekten scheinen einen größeren Datensatz zu erfordern. Der geringe Standardfehler korrespondiert zu entsprechend großen Werten in der Hessematrix. Es ist zu vermuten, dass bei den Matrizen, die fast einen Rangverlust haben, die Varianz entsprechend unterschätzt wird.

5.3.3 Lokale Modellierung und numerische Kondition

Den obigen Ausführungen ist zu entnehmen, dass die Kondition der Hessematrix von Beobachtungen mit großen Lebensdauern dominiert wird. Im Abschnitt 5.1.1 wurde bereits auf eine Veröffentlichung (Gameran, 1991) hingewiesen, die ein Verfahren vorschlägt, dass auf der Intervallbildung beruht. Dabei wird für jedes Intervall eine – sozusagen lokale – Intensitätsrate geschätzt. In Anlehnung an dieses Vorgehen wird vorgeschlagen, den Bereich der Lebensdauern in zwei Intervalle zu partitionieren. Der

Trennpunkt zwischen den Intervallen sei t_{part} . Für jedes Individuum i gilt bezüglich seiner Lebensdauer T_i

$$T_i = \min(T_i, t_{part}) + (T_i - t_{part})_+ \quad (5.20)$$

Betrachtet man den Zeitbereich von $[0, t_{part}]$, so kann man für diesen Bereich die Intensitätsrate mit den rechtszensierten Lebensdauern $\min(T_i, t_{part})$ schätzen. Dies führt bei der Schätzfunktion (5.15) zu einem verkürzten Integrationsbereich, der auch dazu führen sollte, dass sich die Kondition der Hessematrix verbessert, da die dominanten Lebensdauern gestutzt werden. Diesem Zweck entsprechend sollte der Wert von t_{part} gewählt werden – abgesehen davon, dass man sich einen möglichst großen Wert für t_{part} wünscht. Zur Schätzung der Intensitätsrate rechts von t_{part} muß allerdings gewährleistet sein, dass mindestens eine nicht zensierte Beobachtung größer als t_{part} ist. Denkbar ist auch, dass man für den ‘Restbereich’ $[t_{part}, T_{max}]$ eine vereinfachte Schätzung vornimmt, etwa durch parametrische Verteilungsannahmen, so dass sich die Intensitätsrate entsprechend einfach schätzen läßt.

6 Zusammenfassung und Ausblick

On ne finit pas un œuvre, on l'abandonne (Gustave Flaubert)

Als Einstieg in die semiparametrischen Lebensdauermodelle wurde ein Cox-Modell gerechnet. Der Datensatz motivierte die Fragestellung nach zeitvariierenden Kovariableneffekten. Dazu wurde als theoretischer Unterbau der Likelihoodprozeß dargestellt. In engem Zusammenhang dazu steht die Doob-Meyer Zerlegung sowie der Kompensator. Läßt sich für die Lebensdauerdaten jeweils der genaue Eintritt des Ereignisses bestimmen, kann der Likelihoodprozeß als Radon-Nikodym Dichte bezüglich eines homogenen Poissonprozesses dargestellt werden und der Kompensator läßt sich als Integral des Intensitätsprozesses darstellen.

Dieser Intensitätsprozeß steht im Mittelpunkt der Modellierung, die zum einen auf Schätzfunktionen basierte, zum anderen auf Bayes Methoden. Die Modellierung der zeitvariierenden Effekte basierte auf Splines. Im Falle der Schätzfunktionen wurde dargestellt wie man mit einem Ansatz über gemischte Modelle die sonst üblichen Verfahren der Kreuzvalidierung vermeidet. Dabei werden die Parameter bezüglich der gestutzten Polynome als Zufallsvariablen aufgefaßt. Parameterschätzungen erhält man als Nullstellen der Schätzfunktionen.

Die bayesianischen Schätzungen wurden mit Markov Chain Monte Carlo Verfahren realisiert. Dabei wurde auf das Hamilton Monte Carlo Verfahren zurückgegriffen, dessen Idee darin besteht neue Werte dadurch vorzuschlagen, dass man auf einem Energieniveau entsprechend den diskretisierten Hamilton'schen Bewegungsgleichungen eine 'gewisse' Wegstrecke zurücklegt. Das Energieniveau setzt sich zusammen aus der potentiellen Energie, was der Likelihood entspricht und der kinetischen Energie. Es wurden Vorschläge gemacht wie man die Fischer Information oder Elemente davon in die kinetische Energie miteinfließen lassen kann, mit dem Ziel die Qualität der diskretisierten Bewegungsgleichungen zu erhöhen.

Unabhängig von den gewählten Methoden muß der Intensitätsprozeß numerisch integriert werden; dazu wurden Chebychev Polynome verwendet. Die numerische Integration wurde im Vorfeld mit Testfunktionen überprüft und erwies sich – wie auch bei den

späteren Berechnungen – als zuverlässig.

Die bayesianischen Verfahren wurden in umfangreichen Simulationsstudien getestet. Eine Besonderheit der Simulationsstudien bzw. Datensituation bei Lebensdauern, die noch erwähnt werden sollte, sind die zensierten Daten. Die Schätzungen verschlechtern sich bei hoher Zensierung, womit leider in der Praxis zu rechnen ist. Generell läßt sich sagen, dass je komplizierter der zu schätzende Intensitätsprozeß ist, desto mehr Daten werden benötigt. Bei eher einfachen zeitvariierenden Effekten kann es ausreichend sein, Polynome zur Modellierung zu verwenden. Wie bereits erwähnt wurde, weisen die aus den Schoenfeldresiduen geschätzten zeitvariierenden Effekte eine bemerkenswerte Ähnlichkeit auf. Leider gilt dies nicht für den rechten Rand, hier weisen die bayesianischen Schätzungen einen sich trompetenförmig vergrößernden Konfidenzbereich auf, der auf Basis der Schoenfeldresiduen nicht zu beobachten ist, nicht zuletzt wegen der Tatsache, dass die Schoenfeldresiduen beschränkt sind. Ob die Ähnlichkeit allgemeiner Natur ist bedarf einer weiteren Forschung.

Die Ergebnisse bei den Schätzfunktionen sind dagegen bescheidener, die numerischen Probleme und ihre Ursachen wurden diskutiert. Ein möglicherweise interessanter Ausweg könnte die am Ende angesprochene lokale Modellierung sein, bei der die Daten künstlich zensiert werden.

Von einem generellen Interesse könnte auch die Beziehung zu intervallzensierten Daten sein. Man kann sich vorstellen, dass für die Dauer einer Studie jeweils der exakte Ereigniszeitpunkt vorhanden ist. Angenommen, danach werden zusätzlich Daten erhoben, z.B. durch quartalsweise Arztbesuche, dann wäre es natürlich interessant die Daten kombiniert zu analysieren. Sie passen allerdings nicht in das hier vorgestellte Konzept, da ja nach Ende der Studie die Daten nur quartalsweise vorliegen und somit vorher-sagbar sind. Allerdings läßt sich der Likelihoodprozeß verallgemeinern zum Doleans Exponential (Rao, 1999), der zusätzlich noch Sprungkomponenten enthält, nämlich die Quartalszeitpunkte. Ob allerdings eine gemeinsame statistische Inferenz möglich ist, kann und soll an dieser Stelle nicht beantwortet werden.

Zum Schluß soll noch angemerkt werden, dass die verwendeten Programme für die Schätzungen in Java geschrieben wurden. Die Grafiken sowie die Tabellen wurden mit der freien Statistiksoftware *R* geschrieben.

A Ergebnisse zum Cox Modell

Tabelle A.1: Tabellen zum Datensatz PBC
N=312

	N	time	status
age			
[26.3,42.3)	78	5.49	0.24
[42.3,49.8)	78	6.35	0.33
[49.8,56.8)	78	5.71	0.47
[56.8,78.4]	78	4.43	0.55
edema			
0	263	5.91	0.34
0.5	29	4.36	0.59
1	20	1.69	0.95
log(prottime)			
[2.20,2.31)	89	5.41	0.18
[2.31,2.37)	85	6.54	0.28
[2.37,2.42)	61	5.83	0.51
[2.42,2.84]	77	4.16	0.70
log(bili)			
[-1.204,-0.105)	90	6.95	0.16
[-0.105, 0.336)	66	6.51	0.21
[0.336, 1.253)	78	5.28	0.50
[1.253, 3.332]	78	3.16	0.74
log(albumin)			
[0.673,1.20)	81	3.34	0.68
[1.200,1.27)	76	5.52	0.39
[1.270,1.34)	81	6.32	0.22
[1.340,1.53]	74	6.91	0.30
Overall			
	312	5.49	0.40

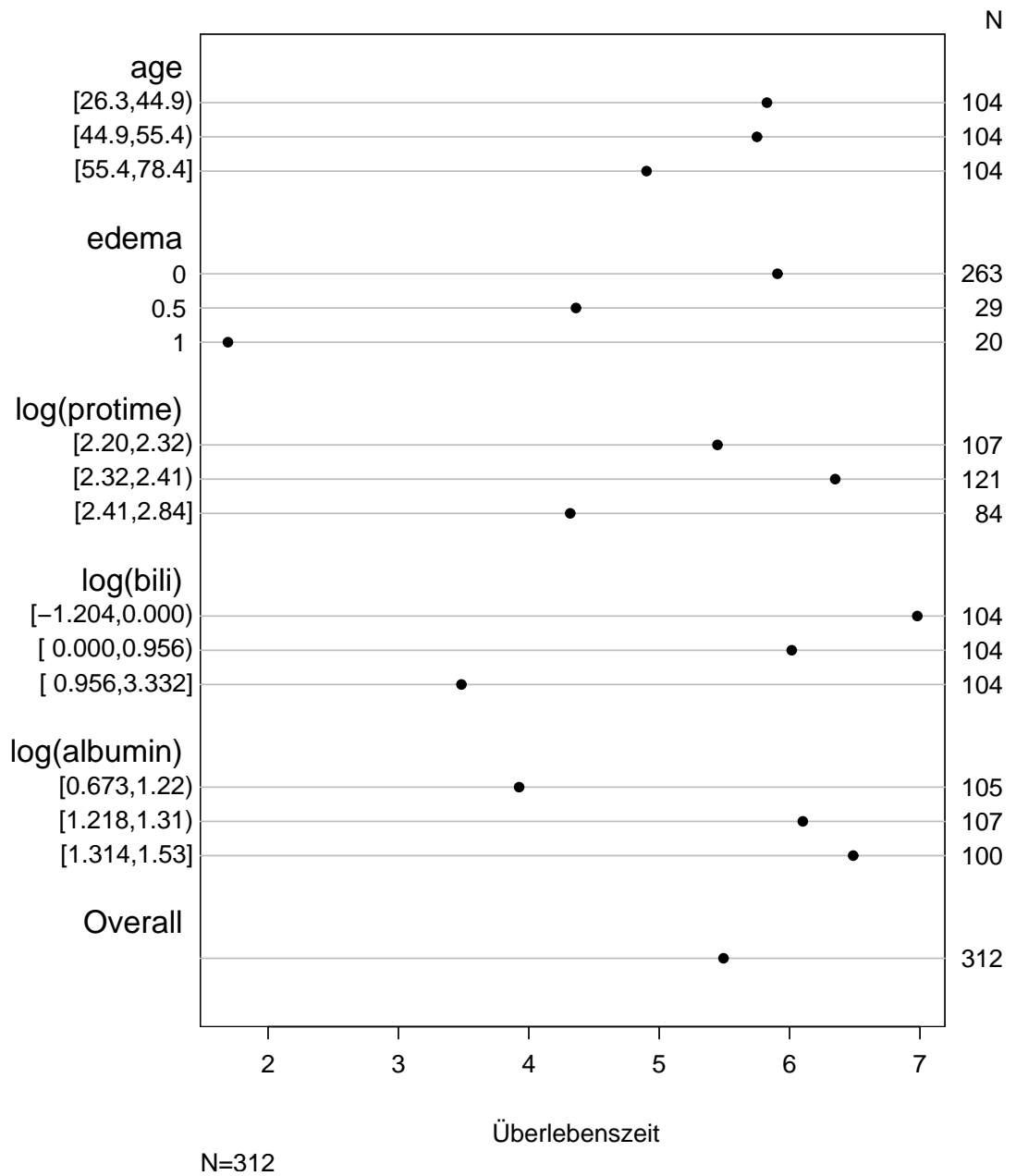


Abbildung A.1: durchschnittliche Überlebenszeiten in Abhängigkeit von Kovariablen

B Ergebnisse zur Numerischen Integration

Tabelle B.1: Fehler bei der Numerischen Integration eines exponentiellen Splines vom Grad 1

rechte Integrationsgrenze	0.1	0.25	0.4	0.55	0.7
absoluter Fehler					
Integral 1. Ordnung	0	$-3.469447e - 16$	$-1.110223e - 16$	$-3.885781e - 16$	$-5.551115e - 16$
Integral 2. Ordnung	0	$-5.551115e - 17$	$-1.110223e - 16$	$-1.526557e - 16$	$-1.804112e - 16$
Integral 3. Ordnung	0	$-6.938894e - 18$	$-2.428613e - 17$	$-3.989864e - 17$	$-4.857226e - 17$
Integral 4. Ordnung	0	$2.168404e - 18$	$-8.673617e - 19$	$-3.469447e - 18$	$-6.071532e - 18$
relativer Fehler					
Integral 1. Ordnung	0	$-5.348132e - 15$	$-9.059045e - 16$	$-1.938852e - 15$	$-1.816940e - 15$
Integral 2. Ordnung	0	$-7.465126e - 15$	$-5.217938e - 15$	$-3.376718e - 15$	$-2.179933e - 15$
Integral 3. Ordnung	0	$-1.170071e - 14$	$-9.204440e - 15$	$-5.334649e - 15$	$-2.877504e - 15$
Integral 4. Ordnung	0	$6.006791e - 14$	$-3.433805e - 15$	$-3.589113e - 15$	$-2.229449e - 15$

Tabelle B.2: Fehler bei der Numerischen Integration eines exponentiellen Splines vom Grad 1

rechte Integrationsgrenze	0.1	0.25	0.4	0.55	0.7
absoluter Fehler					
Integral 1. Ordnung	0	$-4.163336e - 17$	$-1.526557e - 16$	$-2.220446e - 16$	$-4.440892e - 16$
Integral 2. Ordnung	0	$0.000000e + 00$	$-6.938894e - 18$	$-2.081668e - 17$	$-5.551115e - 17$
Integral 3. Ordnung	0	$-6.938894e - 18$	$-6.938894e - 18$	$-3.469447e - 18$	$0.000000e + 00$
Integral 4. Ordnung	0	$-2.168404e - 18$	$-1.734723e - 18$	$8.673617e - 19$	$2.168404e - 18$
relativer Fehler					
Integral 1. Ordnung	0	$-6.417758e - 16$	$-1.245619e - 15$	$-1.107915e - 15$	$-1.453552e - 15$
Integral 2. Ordnung	0	$0.000000e + 00$	$-3.261211e - 16$	$-4.604615e - 16$	$-6.707486e - 16$
Integral 3. Ordnung	0	$-1.170071e - 14$	$-2.629840e - 15$	$-4.638825e - 16$	$0.000000e + 00$
Integral 4. Ordnung	0	$-6.006791e - 14$	$-6.867610e - 15$	$8.972782e - 16$	$7.962318e - 16$

Tabelle B.3: Fehler bei der Numerischen Integration eines exponentiellen Splines vom Grad 1

rechte Integrationsgrenze	0.1	0.25	0.4	0.55	0.7
absoluter Fehler					
Integral 1. Ordnung	0	$-9.714451e - 17$	$-1.387779e - 16$	$-1.665335e - 16$	$-2.220446e - 16$
Integral 2. Ordnung	0	$-2.775558e - 17$	$-3.469447e - 17$	$-4.163336e - 17$	$-5.551115e - 17$
Integral 3. Ordnung	0	$-6.938894e - 18$	$-6.938894e - 18$	$-8.673617e - 18$	$-3.469447e - 18$
Integral 4. Ordnung	0	$-1.301043e - 18$	$-1.301043e - 18$	$8.673617e - 19$	$1.301043e - 18$
relativer Fehler					
Integral 1. Ordnung	0	$-1.497477e - 15$	$-1.132381e - 15$	$-8.309364e - 16$	$-7.267760e - 16$
Integral 2. Ordnung	0	$-3.732563e - 15$	$-1.630606e - 15$	$-9.209230e - 16$	$-6.707486e - 16$
Integral 3. Ordnung	0	$-1.170071e - 14$	$-2.629840e - 15$	$-1.159706e - 15$	$-2.055360e - 16$
Integral 4. Ordnung	0	$-3.604074e - 14$	$-5.150707e - 15$	$8.972782e - 16$	$4.777391e - 16$

Tabelle B.4: Fehler bei der Numerischen Integration eines Polynoms

rechte Integrationsgrenze	1	2.5	4	5.5	7
absoluter Fehler					
Integral 1. Ordnung	$-7.400e-07$	$-6.922170e-07$	$-8.997604e-09$	$-5.811782e-08$	$-7.209528e-07$
Integral 2. Ordnung	$1.720e-06$	$1.288530e-06$	$-1.254959e-07$	$-2.080270e-06$	$-1.395873e-06$
Integral 3. Ordnung	$1.570e-06$	$2.164301e-06$	$1.302250e-06$	$3.653573e-07$	$-6.896336e-07$
Integral 4. Ordnung	$-1.270e-06$	$9.709375e-07$	$4.529010e-06$	$7.804269e-06$	$6.389645e-06$
relativer Fehler					
Integral 1. Ordnung	$-4.900e-07$	$-1.801163e-07$	$-1.268727e-09$	$-4.070345e-09$	$-2.850925e-08$
Integral 2. Ordnung	$2.390e-06$	$2.703850e-07$	$-9.919814e-09$	$-7.408695e-08$	$-2.429211e-08$
Integral 3. Ordnung	$6.780e-06$	$5.536092e-07$	$7.952755e-08$	$8.016748e-09$	$-6.406062e-09$
Integral 4. Ordnung	$-2.248e-05$	$4.041481e-07$	$2.804521e-07$	$1.306356e-07$	$3.777211e-08$

Tabelle B.5: Fehler bei der Numerischen Integration eines Polynoms

rechte Integrationsgrenze	1	2.5	4	5.5	7
absoluter Fehler					
Integral 1. Ordnung	0	$9.758239e - 11$	$1.069758e - 10$	$9.511147e - 11$	$9.258017e - 11$
Integral 2. Ordnung	0	$1.760814e - 10$	$2.742002e - 10$	$3.633254e - 10$	$4.585132e - 10$
Integral 3. Ordnung	0	$1.492770e - 10$	$4.793534e - 10$	$9.752412e - 10$	$1.605144e - 09$
Integral 4. Ordnung	0	$5.498713e - 11$	$5.672049e - 10$	$1.704834e - 09$	$3.667424e - 09$
relativer Fehler					
Integral 1. Ordnung	0	$2.539114e - 11$	$1.508435e - 11$	$6.661236e - 12$	$3.660976e - 12$
Integral 2. Ordnung	0	$3.694889e - 11$	$2.167413e - 11$	$1.293951e - 11$	$7.979419e - 12$
Integral 3. Ordnung	0	$3.818376e - 11$	$2.927380e - 11$	$2.139895e - 11$	$1.491032e - 11$
Integral 4. Ordnung	0	$2.288813e - 11$	$3.512331e - 11$	$2.853719e - 11$	$2.167982e - 11$

Tabelle B.6: Fehler bei der Numerischen Integration eines Polynoms

rechte Integrationsgrenze	1	2.5	4	5.5	7
absoluter Fehler					
Integral 1. Ordnung	0	-1.065814e - 14	-1.421085e - 14	-1.953993e - 14	-3.552714e - 14
Integral 2. Ordnung	0	1.776357e - 15	-2.664535e - 14	-4.973799e - 14	-9.947598e - 14
Integral 3. Ordnung	0	2.042810e - 14	7.105427e - 15	-6.394885e - 14	-1.563194e - 13
Integral 4. Ordnung	0	-8.881784e - 15	-7.105427e - 15	-7.105427e - 14	-1.989520e - 13
relativer Fehler					
Integral 1. Ordnung	0	-2.773271e - 15	-2.003833e - 15	-1.368500e - 15	-1.404880e - 15
Integral 2. Ordnung	0	3.727505e - 16	-2.106179e - 15	-1.771374e - 15	-1.731162e - 15
Integral 3. Ordnung	0	5.225331e - 15	4.339237e - 16	-1.403179e - 15	-1.452063e - 15
Integral 4. Ordnung	0	-3.697001e - 15	-4.399928e - 16	-1.189377e - 15	-1.176096e - 15

Tabelle B.7: Numerische Integration eines exponentiellen Splines

rechte Integrationsgrenze	0.1	0.25	0.4	0.55	0.7
Matlab					
Matlab	0.13014655	3.336994e - 01	5.528325e - 01	7.992022e - 01	1.095562e + 00
Gauß-Cebychev, 40					
Wert	0.13014655	3.336993e - 01	5.528325e - 01	7.992022e - 01	1.095562e + 00
absoluter Fehler	0.00000000	-4.819857e - 09	-2.507390e - 09	-6.721134e - 09	-1.696771e - 08
relativer Fehler	0.00000000	-1.444371e - 08	-4.535533e - 09	-8.409804e - 09	-1.548767e - 08
Gauß-Cebychev, 50					
Wert	0.13014655	3.336994e - 01	5.528325e - 01	7.992022e - 01	1.095562e + 00
absoluter Fehler	0.00000000	2.750654e - 09	-2.800213e - 09	-2.446843e - 09	-7.614064e - 09
relativer Fehler	0.00000000	8.242912e - 09	-5.065211e - 09	-3.061607e - 09	-6.949915e - 09
Trapez, 100					
Wert	0.13052168	3.278292e - 01	5.405701e - 01	7.803591e - 01	1.069898e + 00
absoluter Fehler	-0.00037513	5.870118e - 03	1.226246e - 02	1.884306e - 02	2.566451e - 02
relativer Fehler	-0.00288235	1.759104e - 02	2.218115e - 02	2.357733e - 02	2.342589e - 02
Trapez, 500					
Wert	0.12889773	3.312079e - 01	5.490730e - 01	7.941444e - 01	1.089173e + 00
absoluter Fehler	0.00124882	2.491440e - 03	3.759584e - 03	5.057795e - 03	6.389588e - 03
relativer Fehler	0.00959549	7.466123e - 03	6.800584e - 03	6.328555e - 03	5.832246e - 03
Trapez, 1000					
Wert	0.12952245	3.324543e - 01	5.509541e - 01	7.966757e - 01	1.092372e + 00
absoluter Fehler	0.00062410	1.245010e - 03	1.878446e - 03	2.526453e - 03	3.190361e - 03
relativer Fehler	-0.00288235	1.759104e - 02	2.218115e - 02	2.357733e - 02	2.342589e - 02

C Ergebnisse der Simulationsstudien

Tabelle C.1: Werte für ein Cox Weibull Modell mit Spline Grad 1

	\bar{x}	x_{Median}	s^2	x_0	$x_{0.05}$	$x_{0.95}$	x_1
fixe Effekte							
$\hat{\beta}_1$	0.426	0.430	0.007	0.158	0.301	0.574	0.682
$\hat{\beta}_2$	-0.545	-0.543	0.002	-0.681	-0.626	-0.471	-0.424
zeitvariierender Effekt Nr. 1							
$\hat{\alpha}_0$	-2.122	-2.118	0.072	-2.961	-2.568	-1.703	-1.418
$\hat{\alpha}_1$	4.685	4.625	0.705	2.898	3.358	6.079	7.258
\hat{u}_0^α	-0.474	-0.479	0.565	-2.882	-1.671	0.696	1.912
\hat{u}_1^α	-0.555	-0.570	0.563	-3.247	-1.635	0.670	1.845
\hat{u}_2^α	-0.391	-0.407	0.453	-2.902	-1.467	0.722	2.586
\hat{u}_3^α	-0.550	-0.467	0.570	-2.712	-1.913	0.772	1.320
\hat{u}_4^α	-0.464	-0.423	0.481	-2.454	-1.613	0.639	1.610
\hat{u}_5^α	-0.474	-0.425	0.581	-3.056	-1.844	0.727	1.609
\hat{u}_6^α	-0.575	-0.555	0.448	-2.389	-1.697	0.457	1.543
\hat{u}_7^α	-0.243	-0.309	0.510	-2.081	-1.505	0.956	1.901
\hat{u}_8^α	-0.640	-0.643	0.235	-2.081	-1.491	0.070	0.954
Hyperparameter							
$\ln(\hat{\sigma}_1^2)$	-0.474	-0.460	0.461	-2.168	-1.528	0.731	1.816

	\bar{x}	x_{Median}	s^2	x_0	$x_{0.05}$	$x_{0.95}$	x_1
feste Effekte							
$\hat{\beta}_1$	0.495	0.495	0.000	0.438	0.458	0.531	0.564
$\hat{\beta}_2$	-0.509	-0.510	0.000	-0.555	-0.529	-0.489	-0.472
zeitvariierender Effekt Nr. 1							
$\hat{\alpha}_0$	0.020	0.024	0.004	-0.125	-0.089	0.115	0.179
$\hat{\alpha}_1$	0.053	0.007	0.055	-0.345	-0.250	0.537	0.794
\hat{u}_0^α	-0.220	-0.239	0.084	-0.725	-0.618	0.285	0.412
\hat{u}_1^α	0.102	0.102	0.092	-0.618	-0.404	0.602	0.715
\hat{u}_2^α	0.237	0.220	0.041	-0.247	-0.090	0.581	0.676
\hat{u}_3^α	-0.384	-0.269	0.172	-1.180	-1.042	0.158	0.367
\hat{u}_4^α	0.006	0.073	0.066	-0.513	-0.396	0.349	0.441
\hat{u}_5^α	0.291	0.300	0.017	-0.024	0.077	0.494	0.617
\hat{u}_6^α	-0.083	-0.080	0.003	-0.230	-0.177	0.002	0.096
zeitvariierender Effekt Nr. 2							
$\hat{\alpha}_0$	0.469	0.469	0.001	0.350	0.411	0.532	0.570
$\hat{\alpha}_1$	-0.417	-0.428	0.034	-0.834	-0.697	-0.083	0.037
\hat{u}_0^α	-0.104	-0.074	0.033	-0.536	-0.439	0.160	0.355
\hat{u}_1^α	0.027	0.038	0.049	-0.503	-0.345	0.356	0.464
\hat{u}_2^α	0.485	0.476	0.028	0.181	0.267	0.794	0.944
\hat{u}_3^α	-0.035	-0.031	0.023	-0.461	-0.286	0.219	0.284
\hat{u}_4^α	-0.306	-0.312	0.081	-0.942	-0.767	0.161	0.293
\hat{u}_5^α	0.179	0.182	0.054	-0.352	-0.230	0.531	0.666
\hat{u}_6^α	0.173	0.170	0.006	-0.027	0.051	0.297	0.351
Hyperparameter							
$\ln(\hat{\sigma}_1^2)$	-1.684	-1.586	0.186	-2.730	-2.516	-1.129	-0.963
$\ln(\hat{\sigma}_2^2)$	-1.996	-2.024	0.061	-2.791	-2.417	-1.568	-1.379

D Ergebnisse der Analyse des Datensatzes

D.1 Modell mit flexibler Baseline-Hazardrate

Anmerkung zur Notation:

$\hat{\beta}_1$ ist der geschätzte Effekt für die Kovariable age

$\hat{\beta}_2$ ist der geschätzte Effekt für die Kovariable log(albumin)

$\hat{\beta}_3$ ist der geschätzte Effekt für die Kovariable log(bili)

$\hat{\beta}_4$ ist der geschätzte Effekt für die Kovariable edema

$\hat{\beta}_5$ ist der geschätzte Effekt für die Kovariable log(protime)

Tabelle D.1: Parameterschätzungen mit Polynom vom Grad 1, modifiziertes Verfahren

	\bar{x}	x_{Median}	s^2	x_0	$x_{0.05}$	$x_{0.95}$	x_1
feste Effekte							
$\hat{\beta}_1$	0.032	0.032	0.000	0.023	0.025	0.039	0.042
$\hat{\beta}_2$	-3.003	-2.962	0.116	-3.833	-3.546	-2.467	-1.985
$\hat{\beta}_3$	0.886	0.883	0.002	0.769	0.811	0.968	1.031
$\hat{\beta}_4$	0.739	0.727	0.022	0.409	0.502	0.985	1.176
$\hat{\beta}_5$	2.860	2.892	0.248	1.584	2.003	3.682	4.056
Effekte Baseline							
$\hat{\alpha}_0$	-8.957	-8.881	1.561	-11.966	-10.974	-6.889	-5.477
$\hat{\alpha}_1$	0.222	0.223	0.000	0.179	0.190	0.249	0.271

Tabelle D.2: Parameterschätzungen mit Polynom vom Grad 1, modifiziertes Verfahren, Kovariablen zentriert

	\bar{x}	x_{Median}	s^2	x_0	$x_{0.05}$	$x_{0.95}$	x_1
feste Effekte							
$\hat{\beta}_1$	0.033	0.033	0.000	0.022	0.026	0.041	0.044
$\hat{\beta}_2$	-2.954	-2.951	0.128	-3.966	-3.524	-2.347	-1.768
$\hat{\beta}_3$	0.894	0.893	0.003	0.734	0.813	0.981	1.020
$\hat{\beta}_4$	0.729	0.735	0.021	0.337	0.493	0.960	1.187
$\hat{\beta}_5$	2.831	2.852	0.257	0.999	1.963	3.619	4.033
Effekte Baseline							
$\hat{\alpha}_0$	-2.645	-2.646	0.007	-2.871	-2.777	-2.506	-2.434
$\hat{\alpha}_1$	0.222	0.223	0.000	0.158	0.192	0.248	0.272

Tabelle D.3: Parameterschätzungen mit Polynom vom Grad 1

	\bar{x}	x_{Median}	s^2	x_0	$x_{0.05}$	$x_{0.95}$	x_1
feste Effekte							
$\hat{\beta}_1$	0.033	0.032	0.000	0.023	0.026	0.040	0.045
$\hat{\beta}_2$	-3.030	-3.017	0.111	-3.839	-3.550	-2.479	-1.823
$\hat{\beta}_3$	0.888	0.892	0.002	0.726	0.814	0.965	0.999
$\hat{\beta}_4$	0.720	0.728	0.023	0.330	0.457	0.959	1.054
$\hat{\beta}_5$	2.760	2.804	0.223	1.125	1.978	3.465	3.860
Effekte Baseline							
$\hat{\alpha}_0$	-8.696	-8.977	1.307	-11.122	-10.049	-6.900	-4.384
$\hat{\alpha}_1$	0.221	0.222	0.000	0.167	0.190	0.250	0.267

Tabelle D.4: Parameterschätzungen mit Polynom vom Grad 1, Kovariablen zentriert

	\bar{x}	x_{Median}	s^2	x_0	$x_{0.05}$	$x_{0.95}$	x_1
feste Effekte							
$\hat{\beta}_1$	0.033	0.033	0.000	0.019	0.025	0.040	0.045
$\hat{\beta}_2$	-2.977	-2.981	0.152	-4.011	-3.575	-2.346	-1.659
$\hat{\beta}_3$	0.893	0.892	0.002	0.760	0.819	0.978	1.038
$\hat{\beta}_4$	0.748	0.753	0.022	0.268	0.501	0.963	1.152
$\hat{\beta}_5$	2.708	2.750	0.272	1.416	1.887	3.526	3.947
Effekte Baseline							
$\hat{\alpha}_0$	-2.652	-2.649	0.007	-2.917	-2.783	-2.510	-2.435
$\hat{\alpha}_1$	0.222	0.223	0.000	0.177	0.193	0.248	0.279

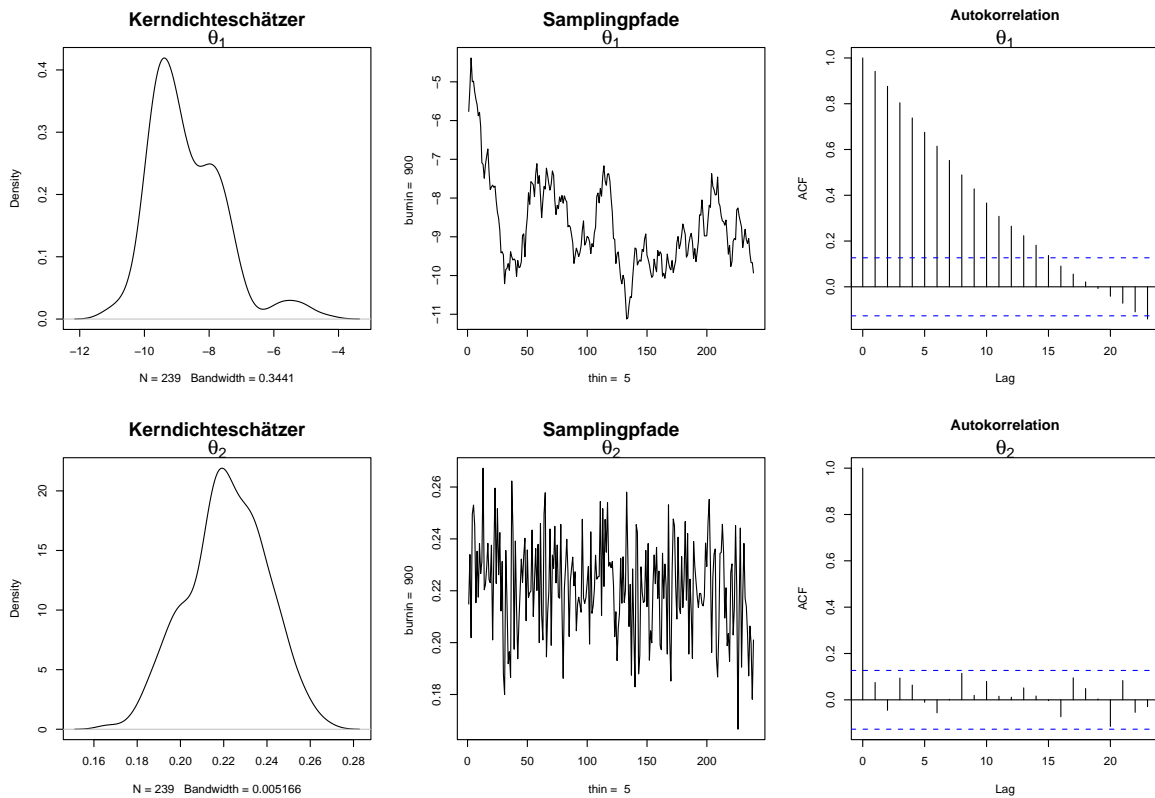


Abbildung D.1: Samplingpfade und geschätzte Dichte der Parameter der Hazardrate

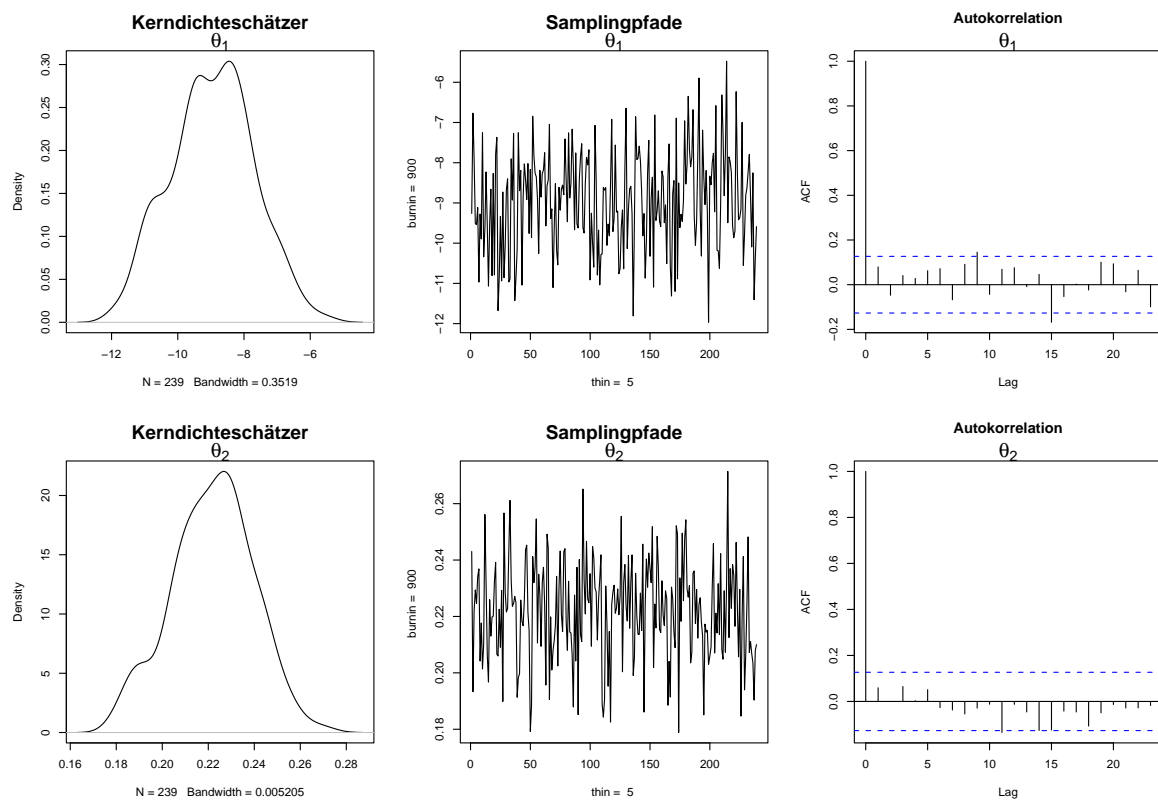


Abbildung D.2: Samplungpfade und geschätzte Dichte der Parameter der Hazardrate, modifiziertes Verfahren

D.2 Modell mit flexibler Baseline-Hazardrate, sowie glatten Kovariableneffekten für $\log(\text{bili})$, $\log(\text{protime})$

Anmerkung zur Notation:

$\hat{\beta}_1$ ist der geschätzte Effekt für die Kovariable $\log(\text{albumin})$

$\hat{\beta}_2$ ist der geschätzte Effekt für die Kovariable age

$\hat{\beta}_3$ ist der geschätzte Effekt für die Kovariable edema

D.2.1 Modellierung mit Polynomen vom Grad 3

Tabelle D.5: Parameterschätzungen (Modell: Polynom, Grad 3, modifiziertes Verfahren)

	\bar{x}	x_{Median}	s^2	x_0	$x_{0.05}$	$x_{0.95}$	x_1
feste Effekte							
$\hat{\beta}_1$	-3.214	-3.215	0.109	-4.180	-3.698	-2.695	-2.348
$\hat{\beta}_2$	0.034	0.034	0.000	0.021	0.027	0.041	0.046
$\hat{\beta}_3$	0.793	0.793	0.023	0.220	0.571	1.051	1.259
Effekte Baseline							
$\hat{\alpha}_0$	-3.479	-3.467	0.041	-4.043	-3.838	-3.205	-3.062
$\hat{\alpha}_1$	0.715	0.694	0.023	0.370	0.494	1.007	1.127
$\hat{\alpha}_2$	-0.130	-0.125	0.001	-0.232	-0.202	-0.081	-0.053
$\hat{\alpha}_3$	0.009	0.009	0.000	0.004	0.006	0.014	0.016
Effekte $\log(\text{protime})$							
$\hat{\alpha}_0$	10.965	10.943	1.252	8.240	9.148	12.671	14.041
$\hat{\alpha}_1$	-5.005	-4.880	1.188	-8.268	-6.841	-3.415	-2.185
$\hat{\alpha}_2$	0.625	0.608	0.070	0.002	0.239	1.065	1.361
$\hat{\alpha}_3$	-0.019	-0.017	0.000	-0.073	-0.048	0.009	0.026
Effekte $\log(\text{bili})$							
$\hat{\alpha}_0$	0.315	0.312	0.015	-0.031	0.094	0.509	0.654
$\hat{\alpha}_1$	0.602	0.598	0.014	0.292	0.414	0.805	0.970
$\hat{\alpha}_2$	-0.153	-0.151	0.001	-0.260	-0.202	-0.103	-0.070
$\hat{\alpha}_3$	0.011	0.011	0.000	0.005	0.007	0.015	0.019

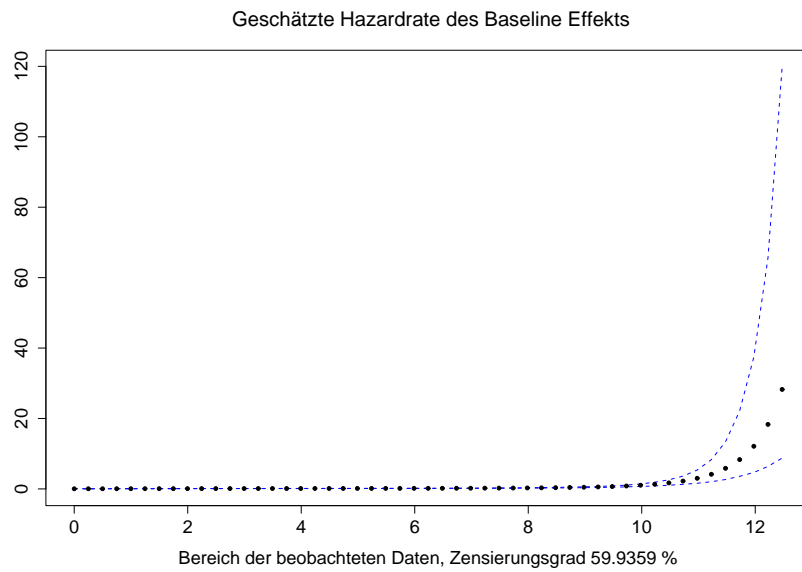


Abbildung D.3: Baseline-Hazardrate, Polynom, Grad 3

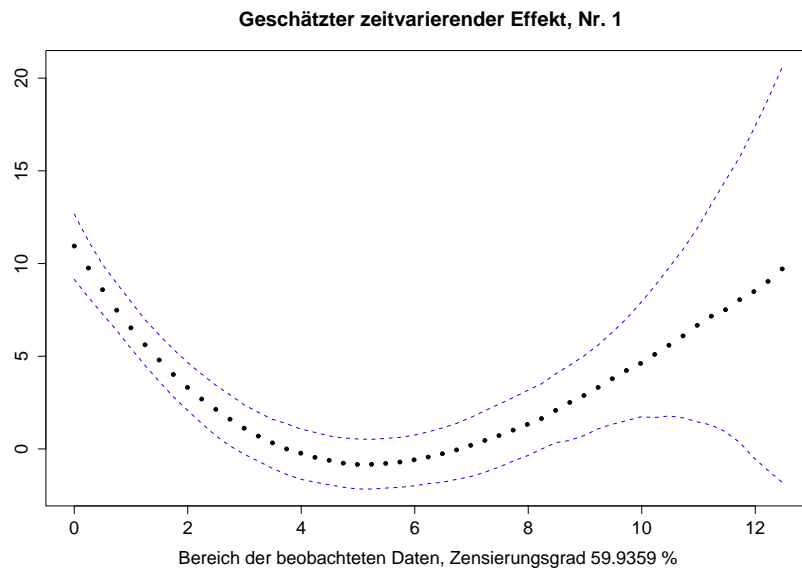


Abbildung D.4: zeitvariiender Effekt für $\log(\text{protime})$, Polynom, Grad 3

Geschätzter zeitvariierender Effekt, Nr. 2

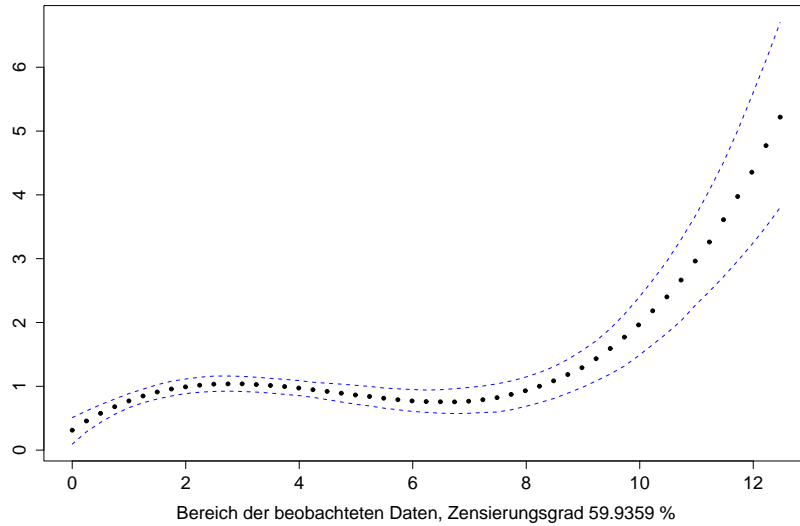


Abbildung D.5: zeitvariierender Effekt für $\log(bili)$, Polynom, Grad 3

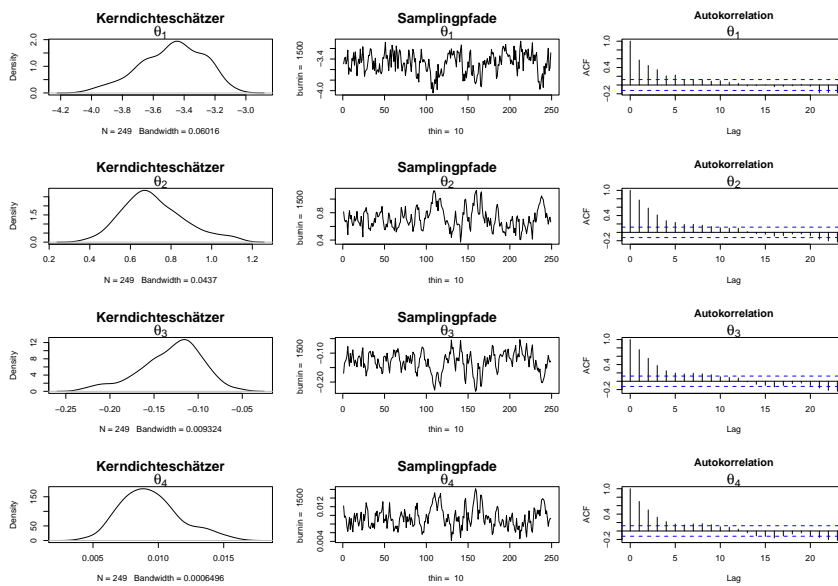


Abbildung D.6: Samplingpfade und geschätzte Dichte der Parameter der Hazardrate, Polynom, Grad 3, modifiziertes Verfahren

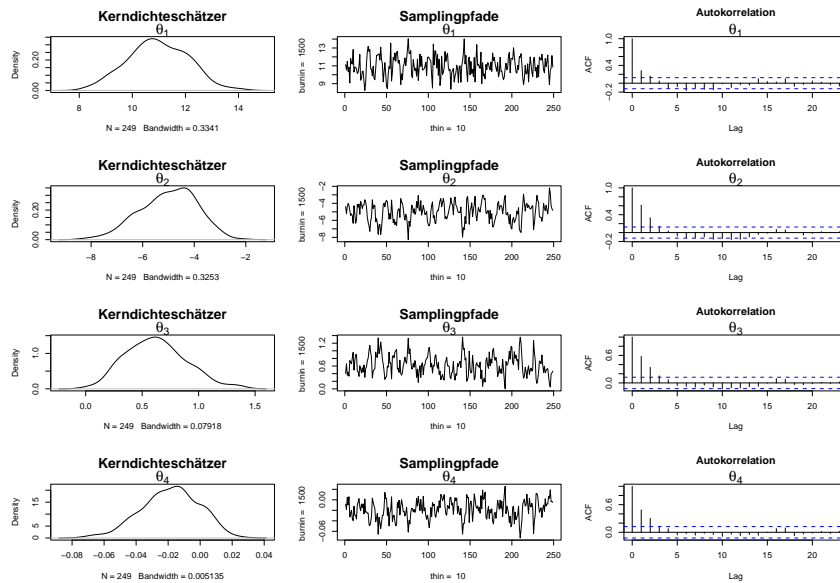


Abbildung D.7: Samplingpfade und geschätzte Dichte der Parameter für $\log(\text{bili})$, Polynom Grad 3, modifiziertes Verfahren

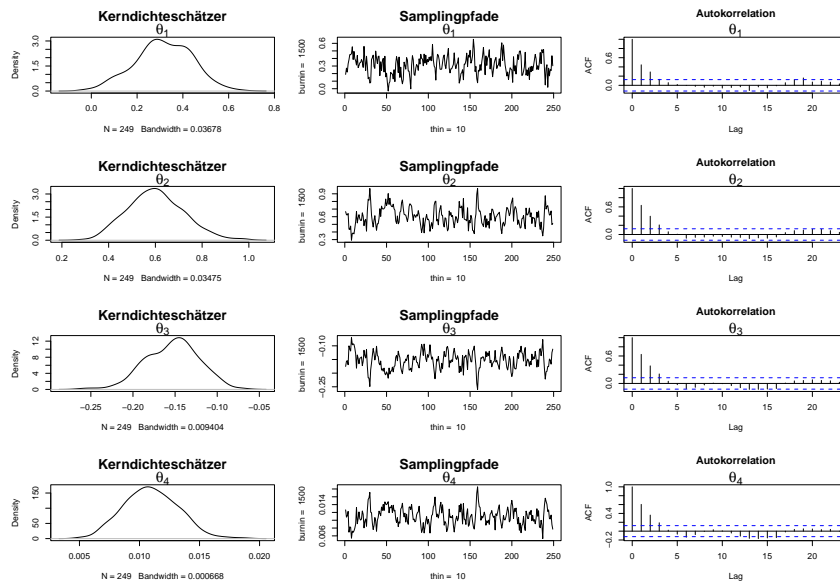


Abbildung D.8: Samplingpfade und geschätzte Dichte der Parameter für $\log(\text{prot ime})$, Polynom Grad 3, modifiziertes Verfahren

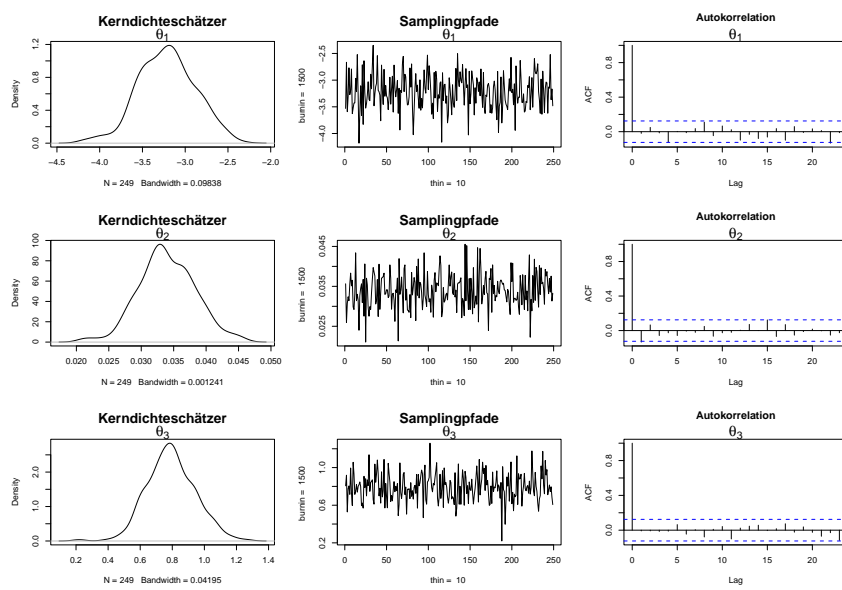


Abbildung D.9: Samplingpfade und geschätzte Dichte der Parameter für age, edema und log(albumin), Polynom Grad 3, modifiziertes Verfahren

D.2.2 Modellierung mit Splines vom Grad 2

Tabelle D.6: Parameterschätzungen Modell: Spline, Grad 2, 4 Gitterpunktemodifiziertes Verfahren)

	\bar{x}	x_{Median}	s^2	x_0	$x_{0.05}$	$x_{0.95}$	x_1
feste Effekte							
$\hat{\beta}_2$	-3.234	-3.218	0.094	-3.954	-3.702	-2.738	-2.476
$\hat{\beta}_1$	0.033	0.033	0.000	0.023	0.027	0.039	0.047
$\hat{\beta}_3$	0.673	0.677	0.021	0.371	0.438	0.905	1.100
Effekt Baseline							
$\hat{\alpha}_0$	-2.681	-2.681	0.007	-2.873	-2.814	-2.551	-2.506
$\hat{\alpha}_1$	0.271	0.266	0.002	0.199	0.212	0.351	0.370
$\hat{\alpha}_2$	-0.011	-0.012	0.000	-0.035	-0.028	0.005	0.015
\hat{u}_0^α	-0.013	-0.008	0.001	-0.102	-0.076	0.031	0.063
\hat{u}_1^α	0.044	0.035	0.002	-0.031	-0.003	0.122	0.225
\hat{u}_2^α	0.038	0.029	0.002	-0.092	-0.012	0.124	0.190
\hat{u}_3^α	-0.007	0.000	0.002	-0.204	-0.079	0.046	0.096
Effekte log(bili)							
$\hat{\alpha}_0$	0.511	0.509	0.008	0.258	0.367	0.647	0.724
$\hat{\alpha}_1$	0.215	0.209	0.004	0.084	0.113	0.329	0.363
$\hat{\alpha}_2$	-0.015	-0.013	0.000	-0.050	-0.038	0.002	0.009
\hat{u}_0^α	-0.049	-0.052	0.001	-0.131	-0.109	0.007	0.034
\hat{u}_1^α	0.059	0.052	0.002	-0.037	0.001	0.132	0.188
\hat{u}_2^α	0.073	0.073	0.002	-0.046	0.003	0.149	0.234
\hat{u}_3^α	0.026	0.020	0.003	-0.139	-0.067	0.122	0.193
Effekte log(protime)							
$\hat{\alpha}_0$	9.320	9.313	0.770	7.712	7.865	10.816	11.521
$\hat{\alpha}_1$	-3.587	-3.506	0.249	-4.599	-4.372	-2.799	-2.512
$\hat{\alpha}_2$	0.315	0.310	0.003	0.163	0.218	0.409	0.445
\hat{u}_0^α	0.001	-0.001	0.001	-0.134	-0.036	0.045	0.154
\hat{u}_1^α	0.001	0.001	0.001	-0.166	-0.047	0.058	0.129
\hat{u}_2^α	-0.009	-0.003	0.002	-0.181	-0.076	0.037	0.165
\hat{u}_3^α	-0.007	-0.003	0.004	-0.245	-0.112	0.073	0.236
Hyperparameter							
$\ln \hat{\sigma}_1^2$	-5.821	-5.830	2.509	-11.819	-7.966	-3.192	-2.110
$\ln \hat{\sigma}_2^2$	-5.207	-4.976	1.371	-9.012	-7.596	-3.568	-2.381
$\ln \hat{\sigma}_3^2$	-6.799	-6.871	2.377	-9.463	-9.118	-4.292	-3.656

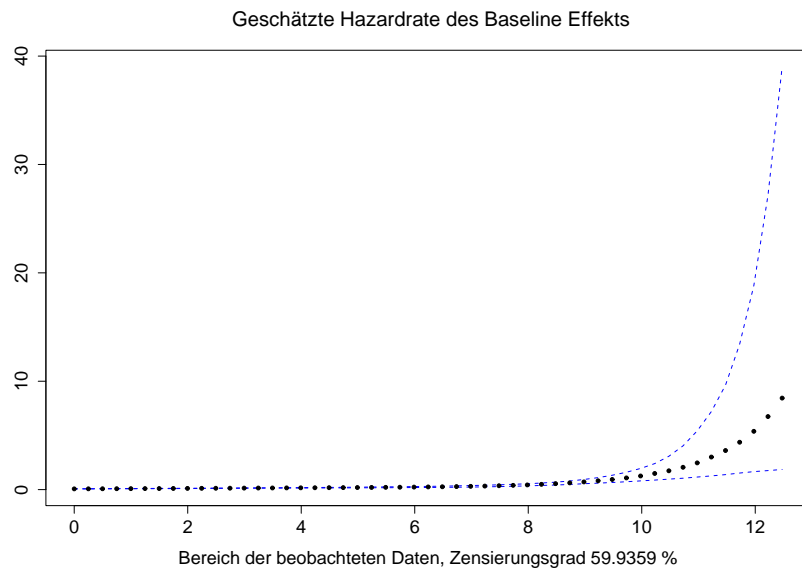


Abbildung D.10: Baseline-Hazardrate, Spline, Grad 2

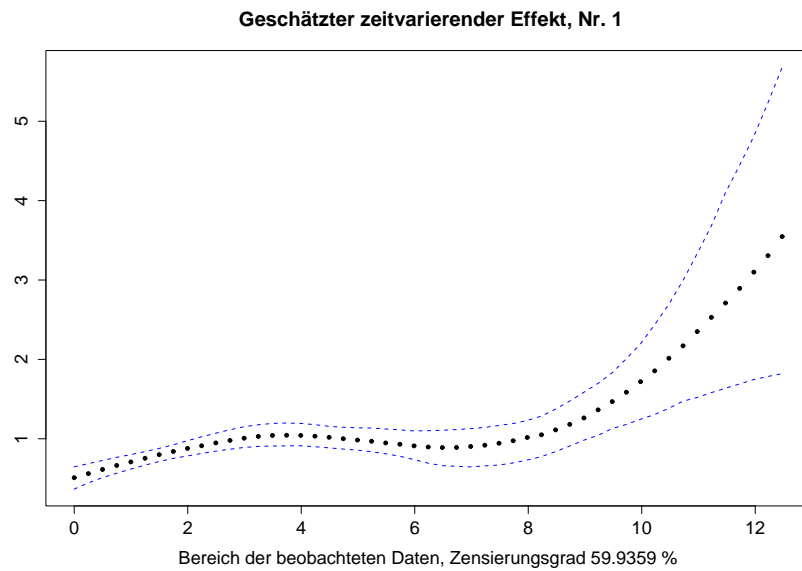


Abbildung D.11: zeitvariierender Effekt für $\log(\text{bili})$, Spline, Grad 2

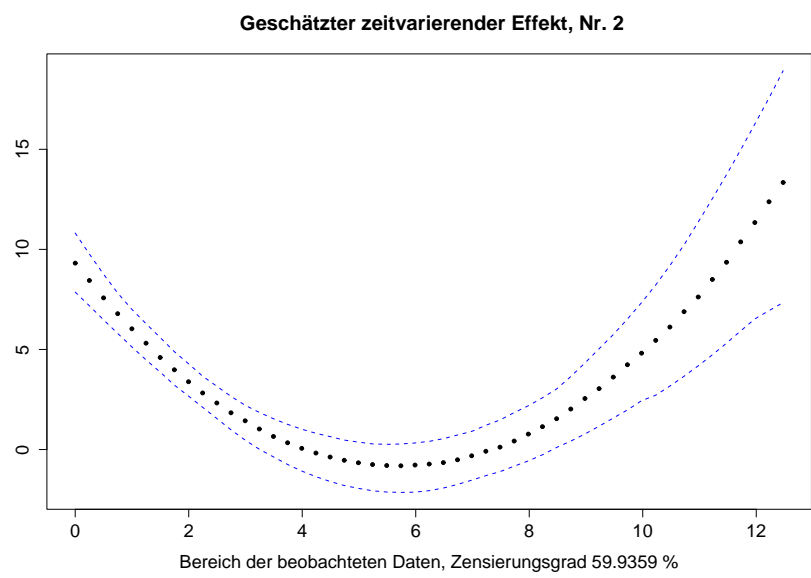


Abbildung D.12: zeitvariierender Effekt für $\log(\text{prot ime})$, Spline, Grad 2

D.3 Modell mit flexibler Baseline-Hazardrate, sowie glatten Kovariableneffekten für $\log(\text{bili})$, edema , $\log(\text{protime})$

Anmerkung zur Notation:

$\hat{\beta}_1$ ist der geschätzte Effekt für die Kovariable $\log(\text{albumin})$

$\hat{\beta}_2$ ist der geschätzte Effekt für die Kovariable age

D.3.1 Modellierung mit Polynomen vom Grad 3

Tabelle D.7: Parameterschätzungen (Modell: Polynom, Grad 3, modifiziertes Verfahren)

	\bar{x}	x_{Median}	s^2	x_0	$x_{0.05}$	$x_{0.95}$	x_1
feste Effekte							
$\hat{\beta}_1$	-3.072	-3.061	0.150	-4.192	-3.671	-2.494	-1.772
$\hat{\beta}_2$	0.033	0.033	0.000	0.020	0.026	0.040	0.045
Effekte Baseline							
$\hat{\alpha}_0$	-3.510	-3.522	0.034	-3.963	-3.798	-3.219	-3.100
$\hat{\alpha}_1$	0.818	0.823	0.017	0.511	0.598	1.046	1.131
$\hat{\alpha}_2$	-0.161	-0.160	0.001	-0.237	-0.214	-0.109	-0.094
$\hat{\alpha}_3$	0.011	0.012	0.000	0.006	0.008	0.015	0.017
Effekte $\log(\text{bili})$							
$\hat{\alpha}_0$	0.195	0.194	0.018	-0.165	-0.037	0.396	0.566
$\hat{\alpha}_1$	0.683	0.681	0.014	0.391	0.491	0.868	0.943
$\hat{\alpha}_2$	-0.168	-0.169	0.001	-0.241	-0.217	-0.119	-0.087
$\hat{\alpha}_3$	0.012	0.012	0.000	0.006	0.008	0.015	0.017
Effekte edema							
$\hat{\alpha}_0$	1.623	1.617	0.116	0.737	1.098	2.197	2.657
$\hat{\alpha}_1$	-0.337	-0.314	0.097	-1.233	-0.842	0.127	0.371
$\hat{\alpha}_2$	-0.026	-0.036	0.007	-0.251	-0.151	0.119	0.205
$\hat{\alpha}_3$	0.004	0.004	0.000	-0.015	-0.007	0.013	0.022
Effekte $\log(\text{protime})$							
$\hat{\alpha}_0$	10.019	10.113	1.834	5.568	7.681	12.087	12.994
$\hat{\alpha}_1$	-4.241	-4.346	2.039	-8.213	-6.407	-1.935	0.099
$\hat{\alpha}_2$	0.463	0.481	0.123	-0.586	-0.136	1.041	1.530
$\hat{\alpha}_3$	-0.007	-0.007	0.001	-0.077	-0.047	0.032	0.067

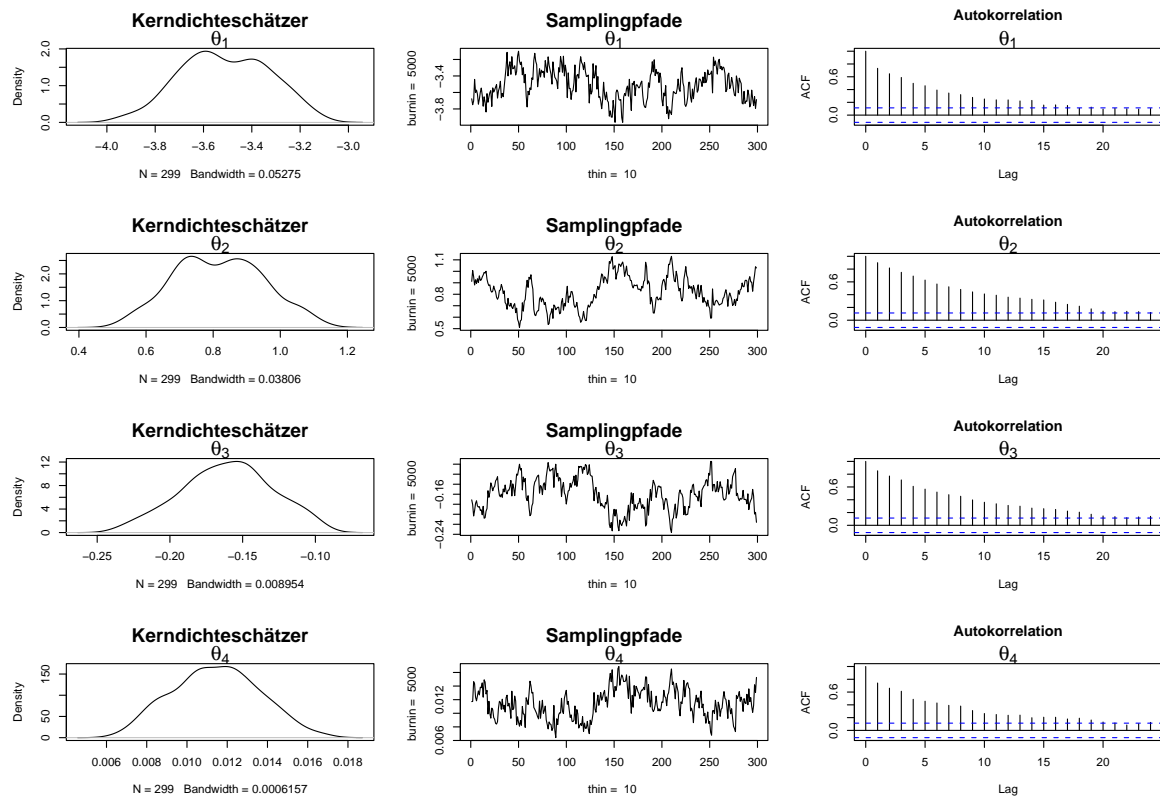


Abbildung D.13: Samplungpfade und geschätzte Dichte der Parameter der Hazardrate, Polynom, Grad 3, modifiziertes Verfahren

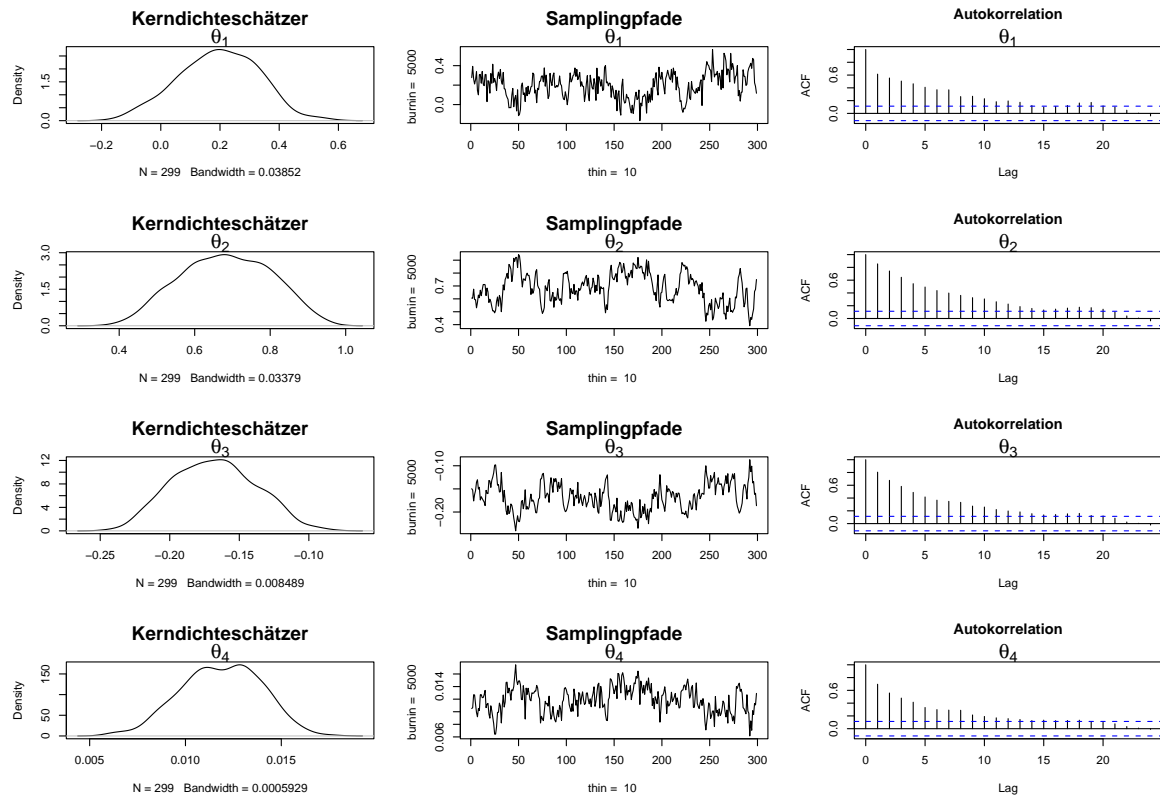


Abbildung D.14: Samplungpfade und geschätzte Dichte der Parameter für $\log(\text{bili})$, Polynom, Grad 3, modifiziertes Verfahren

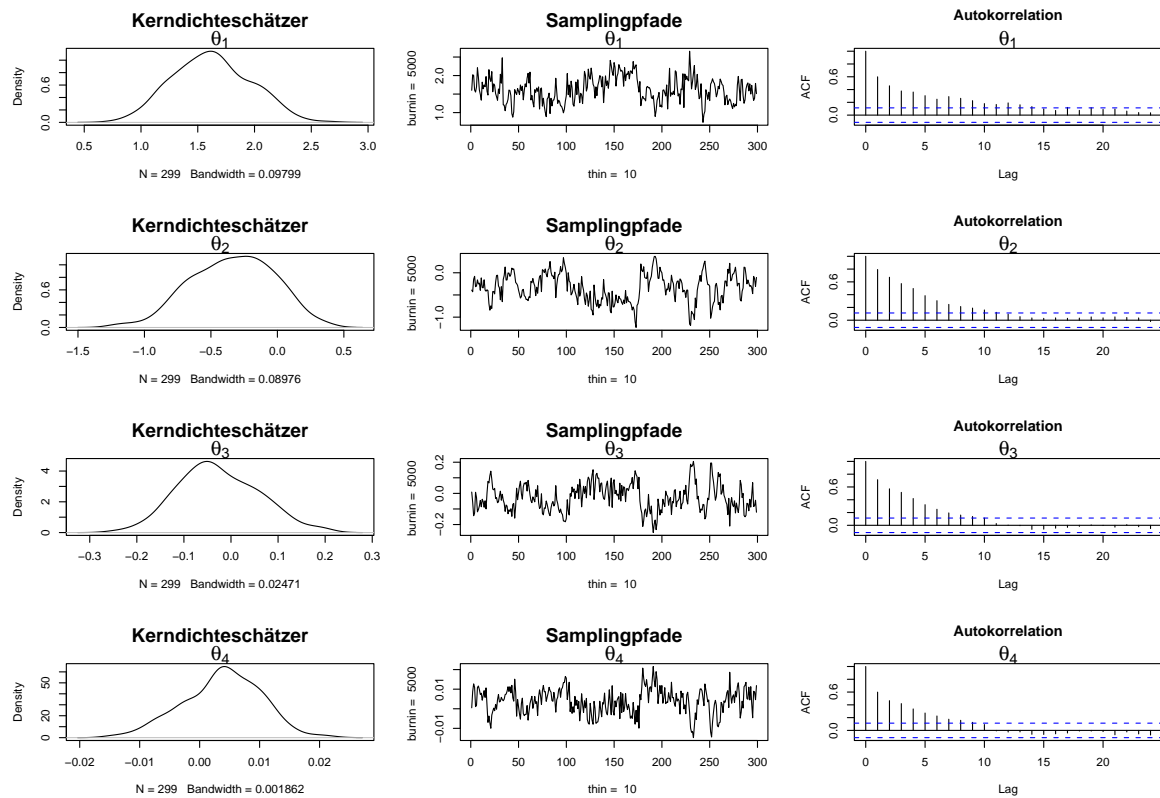


Abbildung D.15: Samplungspfade und geschätzte Dichte der Parameter für edema, Polynom, Grad 3, modifiziertes Verfahren

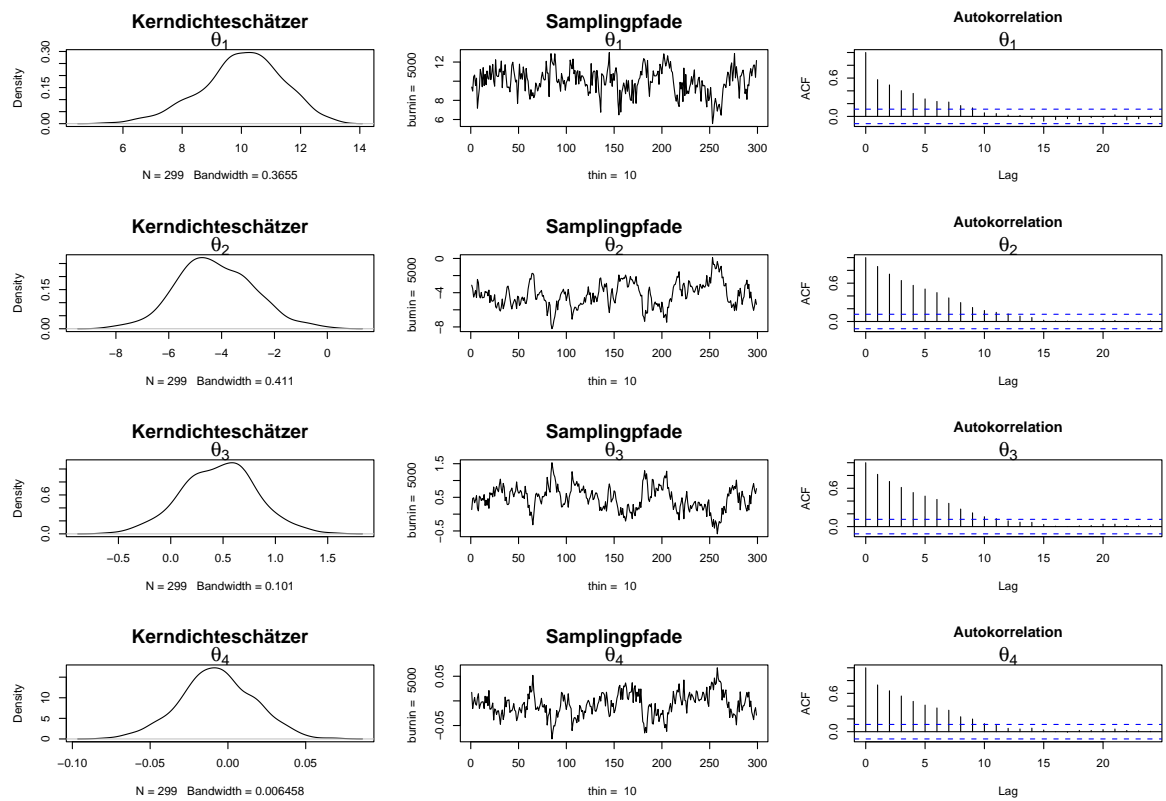


Abbildung D.16: Samplingpfade und geschätzte Dichte der Parameter für $\log(\text{protime})$, Polynom, Grad 3, modifiziertes Verfahren

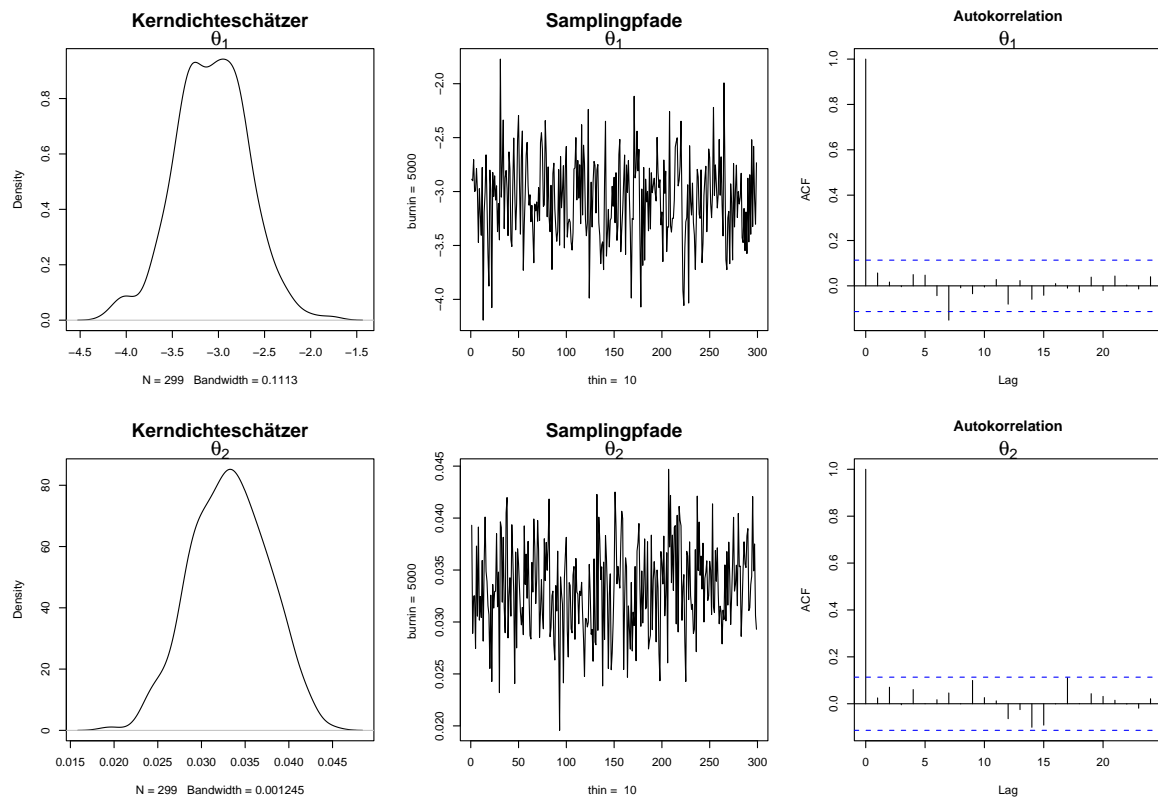


Abbildung D.17: Samplingspfade und geschätzte Dichte der Parameter für age und log(albumin), Polynom, Grad 3, modifiziertes Verfahren

D.3.2 Modellierung mit Splines vom Grad 2

Tabelle D.8: Parameterschätzungen (Modell: Spline Grad 2, 4 Gitterpunkte)

	\bar{x}	x_{Median}	s^2	x_0	$x_{0.05}$	$x_{0.95}$	x_1
festе Effekte							
$\hat{\beta}_1$	-3.074	-3.074	0.012	-3.355	-3.269	-2.873	-2.650
$\hat{\beta}_2$	0.033	0.033	0.000	0.022	0.026	0.041	0.047
Effekte Baseline							
$\hat{\alpha}_0$	-3.036	-3.051	0.017	-3.417	-3.217	-2.808	-2.739
$\hat{\alpha}_1$	0.655	0.628	0.030	0.334	0.423	0.967	1.038
$\hat{\alpha}_2$	-0.089	-0.087	0.002	-0.189	-0.173	-0.019	-0.002
\hat{u}_0^α	-0.001	0.007	0.006	-0.224	-0.128	0.126	0.155
\hat{u}_1^α	0.130	0.128	0.005	-0.054	0.005	0.254	0.354
\hat{u}_2^α	0.126	0.121	0.009	-0.085	-0.027	0.293	0.350
\hat{u}_3^α	-0.107	-0.099	0.014	-0.396	-0.283	0.100	0.221
Effekte log(bili)							
$\hat{\alpha}_0$	0.795	0.799	0.008	0.523	0.619	0.926	0.985
$\hat{\alpha}_1$	-0.322	-0.329	0.008	-0.508	-0.453	-0.168	-0.050
$\hat{\alpha}_2$	0.151	0.154	0.001	0.078	0.110	0.187	0.197
\hat{u}_0^α	-0.420	-0.422	0.004	-0.555	-0.519	-0.316	-0.270
\hat{u}_1^α	0.280	0.293	0.011	0.048	0.089	0.422	0.513
\hat{u}_2^α	0.150	0.161	0.012	-0.131	-0.031	0.325	0.395
\hat{u}_3^α	0.008	-0.018	0.029	-0.342	-0.277	0.290	0.407
Effekte edema							
$\hat{\alpha}_0$	1.320	1.334	0.021	1.037	1.109	1.552	1.603
$\hat{\alpha}_1$	-0.080	-0.074	0.017	-0.419	-0.313	0.119	0.204
$\hat{\alpha}_2$	-0.039	-0.038	0.004	-0.183	-0.143	0.062	0.105
\hat{u}_0^α	-0.013	0.009	0.021	-0.377	-0.314	0.182	0.236
\hat{u}_1^α	-0.023	-0.052	0.019	-0.307	-0.249	0.217	0.272
\hat{u}_2^α	0.382	0.431	0.039	-0.146	-0.021	0.618	0.729
\hat{u}_3^α	-0.482	-0.467	0.030	-0.955	-0.792	-0.213	-0.163
Effekte log(protime)							
$\hat{\alpha}_0$	3.686	4.057	0.716	1.646	1.851	4.414	4.535
$\hat{\alpha}_1$	0.606	0.668	0.069	0.117	0.171	0.967	1.051
$\hat{\alpha}_2$	-0.471	-0.500	0.016	-0.663	-0.623	-0.237	-0.198
\hat{u}_0^α	0.787	0.816	0.052	0.224	0.294	1.105	1.161
\hat{u}_1^α	0.114	0.126	0.062	-0.352	-0.260	0.512	0.555
\hat{u}_2^α	0.086	0.146	0.063	-0.373	-0.305	0.451	0.507
\hat{u}_3^α	-0.716	-0.726	0.021	-1.096	-0.960	-0.475	-0.391
Hyperparameter							
$\ln(\hat{\sigma}_1^2)$	-3.116	-3.012	0.241	-3.927	-3.812	-2.490	-2.253
$\ln(\hat{\sigma}_2^2)$	-0.552	-0.700	0.218	-1.385	-1.156	0.296	0.464
$\ln(\hat{\sigma}_3^2)$	0.387	0.479	0.280	-0.730	-0.512	1.155	1.301
$\ln(\hat{\sigma}_4^2)$	0.318	0.279	0.101	-0.477	-0.247	0.851	1.202

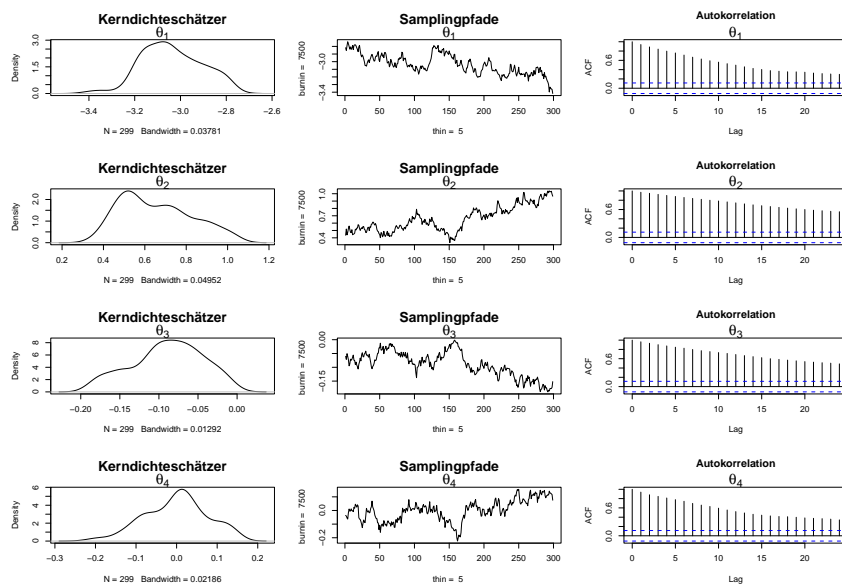


Abbildung D.18: Samplingpfade und geschätzte Dichte der Parameter der Baseline-Hazardrate, Spline Grad 2, 4 Gitterpunkte

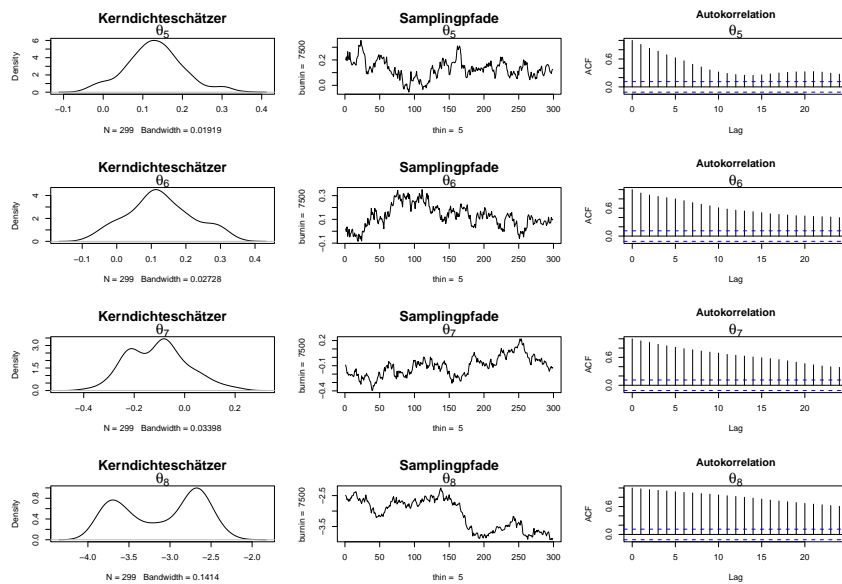


Abbildung D.19: Samplingpfade und geschätzte Dichte der Parameter der Baseline-Hazardrate, Spline Grad 2, 4 Gitterpunkte

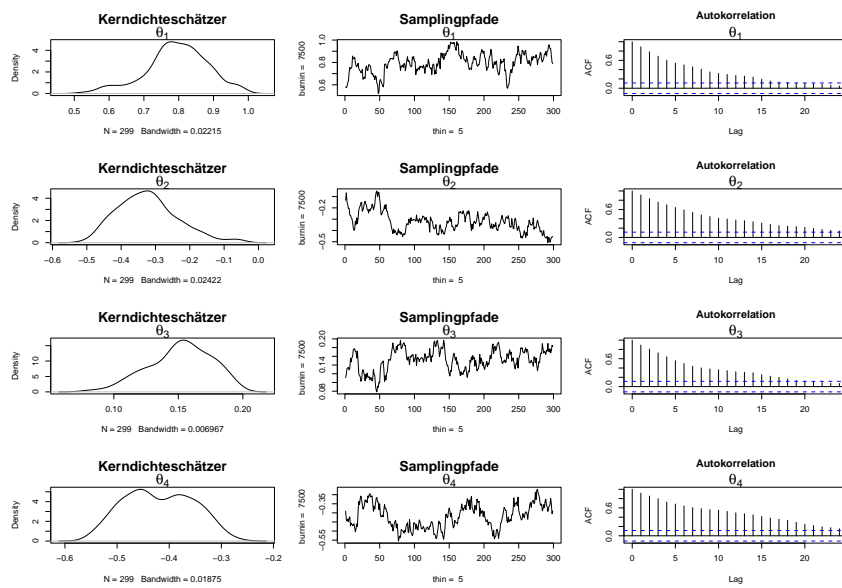


Abbildung D.20: Samplingpfade und geschätzte Dichte der Parameter für $\log(\text{bili})$, Spline Grad 2, 4 Gitterpunkte

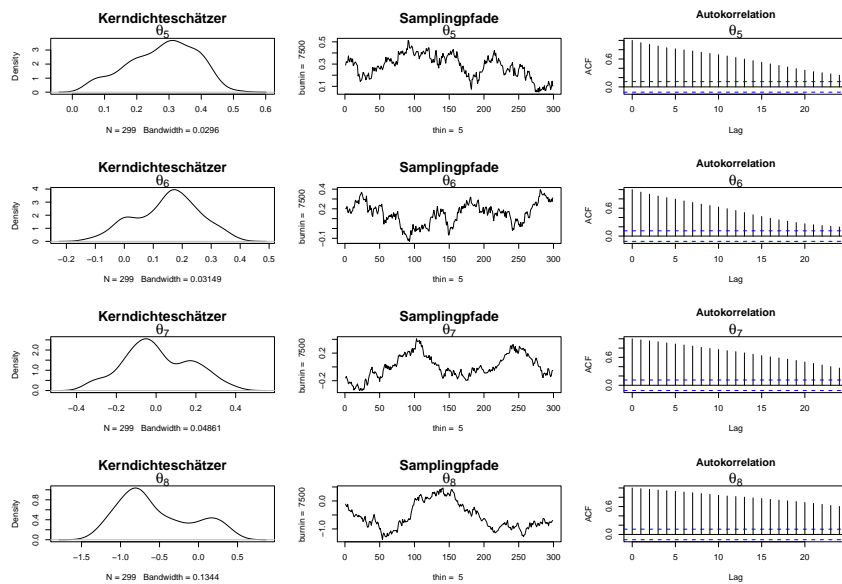


Abbildung D.21: Samplingpfade und geschätzte Dichte der Parameter für $\log(\text{bili})$, Spline Grad 2, 4 Gitterpunkte

D.4 Ergebnisse der Schätzfunktionen

Tabelle D.9: Parameterschätzungen

	$\hat{\cdot}$	Standardfehler
feste Effekte Cox Modell		
$\hat{\beta}_1$	0.0332	0.0087
$\hat{\beta}_2$	-3.0599	0.7240
$\hat{\beta}_3$	0.8801	0.0987
$\hat{\beta}_4$	0.7856	0.2990
$\hat{\beta}_5$	3.0140	1.0240
feste Effekte Schätzfunktionen, Modell 2		
$\hat{\beta}_1$	0.0409	0.0055
$\hat{\beta}_2$	-3.8681	0.4417
$\hat{\beta}_3$	1.1963	0.0575
$\hat{\beta}_4$	0.8213	0.1817
$\hat{\beta}_5$	3.936	0.6374
Effekte Baseline Hazardrate		
$\hat{\alpha}_0$	-2.7387	0.1051
$\hat{\alpha}_1$	0.4361	0.0161
feste Effekte Schätzfunktionen, Modell 1		
$\hat{\beta}_1$	0.0397	0.0055
$\hat{\beta}_2$	-4.2572	0.4417
$\hat{\beta}_3$	1.2277	0.0575
$\hat{\beta}_4$	0.9929	0.1817
$\hat{\beta}_5$	3.7516	0.6374
Effekte Baseline Hazardrate		
$\hat{\alpha}_0$	-3.4934	0.1751
$\hat{\alpha}_1$	0.7994	0.0693
$\hat{\alpha}_2$	-0.0303	0.0060

Anmerkung zur Notation:

$\hat{\beta}_1$ ist der geschätzte Effekt für die Kovariable age

$\hat{\beta}_2$ ist der geschätzte Effekt für die Kovariable $\log(\text{albumin})$

$\hat{\beta}_3$ ist der geschätzte Effekt für die Kovariable $\log(\text{bili})$

$\hat{\beta}_4$ ist der geschätzte Effekt für die Kovariable edema

$\hat{\beta}_5$ ist der geschätzte Effekt für die Kovariable $\log(\text{protime})$

Literaturverzeichnis

- Aalen, O. O. (1978). Nonparametric inference for a family of counting processes, *The Annals of Statistics* **6**: 701–726.
- Andersen, P. K., Borgan, O., Gill, R. D. und Keiding, N. (1993). *Statistical Models Based on Counting Processes*, 1. Auflage, Springer-Verlag, New York.
- Aven, T. (1986). Bayesian inference in a parametric process model, *Scandinavian Journal of Statistics* **13**: 87–97.
- Barndorff-Nielsen, O. (1983). On a formula for the distribution of the maximum likelihood estimator, *Biometrika* **70**: 343–365.
- Bender, R., Augustin, T. und Blettner, M. (2005). Generating survival times to simulate cox proportional hazards models, *Statistics in Medicine* **24**: 1713–1723.
- Besag, J. E., Green, P. J., Higdon, D. und Mengersen, K. (1995). Bayesian computation and stochastic systems (with discussion), *Statistical Science* **10**: 3–66.
- Blossfeld, H.-P., Hamerle, A. und Mayer, K. U. (1989). *Event History Analysis. Statistical Theory and Applications in the Social Sciences*, Lawrence Erlbaum.
- Booth, J. G., Hobert, J. P. und Jank, W. (2001). A survey of monte carlo algorithms for maximizing the likelihood of a two-stage hierarchical model, *Statistical Modelling* **1**: 333–349.
- Borgan, O. (1984). Maximum likelihood estimation in parametric counting process models, with application to censored failure time data, *Scandinavian Journal of Statistics* **11**: 1–16.
- Breiman, L. (1992). *Probability*, Society for Industrial and Applied Mathematics.
- Breslow, N. E. und Clayton, D. G. (1993). Approximate inference in generalized linear mixed models, *Journal of the American Statistical Association* **88**: 9–25.
- Brockwell, T. J. und Davis, R. J. (1996). *Time Series*, 1. Auflage, Springer, New York.

- Buja, A., Hastie, T. J. und Tibshirani, R. J. (1989). Linear smoothing and additive models, *Annals of Statistics* **17**: 453–510.
- Bunea, F. (2004). Consistent covariate selection and post model selection inference in semiparametric regression, *Annals of Statistics* **32**: 898–927.
- Cai, T., Hyndman, R. J. und Wand, M. P. (2002). Mixed model-based hazard estimation, *Journal of Computational & Graphical Statistics* **11**: 784–798.
- Chang, I. S. und Hsiung, C. A. (1990). Finite sample optimality of maximum partial likelihood estimation in cox’s model for counting processes, *Journal of Statistical Planning and Inference* **25**: 35–42.
- Choo, K. (2000). *Learning hyperparameters for neural networks models using hamilton dynamics*, Master’s thesis, University of Toronto.
- Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*, 2. Auflage, Springer-Verlag, New York.
- Courant, R. und Hilbert, D. (1993). *Methoden der mathematischen Physik*, 4. Auflage, Springer-Verlag, Berlin.
- Cox, D. R. (1972). Regression models and life tables (with discussion), *J. R. Statist. Soc B* **34**: 187–220.
- Dalgaard, P. (2002). *Introductory Statistics with R*, Springer-Verlag, New York.
- Deuffhard, P. und Hohmann, A. (1991). *Numerische Mathematik*, 1. Auflage, Walter de Gruyter, Berlin.
- Dickinson, E. R., Grambsch, P. M., Fleming, T. R., Fisher, L. D. und Langworthy, A. (1989). Prognosis in primary biliary cirrhosis, *Hepatology* **10**: 1–7.
- Doob, J. L. (1953). *Stochastic Processes*, 1. Auflage, Wiley & Son, New York.
- Dothan, M. U. (1990). *Prices in Financial Markets*, 1. Auflage, Oxford, Oxford.
- Duane, S., Kennedy, A. D., Pendelton, B. J. und Roweth, D. (1987). Hybrid monte carlo, *Physics Letters B* **195(2)**: 216–222.
- Eilers, P. H. C. und Marx, B. D. (1996). Flexible smoothing with b-splines and penalties, *Statistical Science* **11**: 89–121.
- Eilers, P. H. C. und Marx, B. D. (2004). Splines, knots and penalties. submitted.

- Fahrmeir, L. und Tutz, G. (1994). *Multivariate Statistical Modelling Based on Generalized Linear Models*, 1. Auflage, Springer Verlag, New York.
- Fischer, A. (2003). *An Uncoupling–Coupling Method for Markov Chain Monte Carlo Simulations with an application to Biomolecules*, PhD thesis, Freie Universität Berlin.
- Fleming, T. R. und Harrington, D. P. (1991). *Counting Process and Survival Analysis*, 1. Auflage, Wiley, New York.
- Forster, O. (1993). *Analysis 2*, 5. Auflage, Vieweg, Braunschweig.
- Fox, J. (2002). Cox proportional-hazards regression for survival data, <http://cran.at.r-project.org/doc/contrib/Fox-Companion/appendix.html>. Appendix, An R and S-PLUS Companion to Applied Regression.
- Gamerman, D. (1991). Dynamic bayesian models for survival data, *Applied Statistics* **40**: 63–79.
- Garcke, J., Griebel, M. und Thess, M. (2001). Data mining with sparse grids, *Computing* **67(3)**: 225–253.
- Geman, S. und Geman, D. (1984). Stochastic relaxation, gibbs distributions and the bayesian restoration of images, *IEEE Transaction on Pattern Analysis and Machine Intelligence* **6**: 721–741.
- Gerstner, T. und Griebel, M. (1998). Numerical integration using sparse grids, *Numer. Algorithms* **18**: 209–232.
- Gill, P. E., Murray, W. und Wright, M. H. (1993). *Practical Optimization*, 10. Auflage, Academic Press, London.
- Gill, R. (1984). Understanding cox’s regression model: A martingal approach, *Journal of the American Statistical Association* **79**: 441–447.
- Grambsch, P. M. und Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals, *Biometrika* **81**: 515–526.
- Grambsch, P. M. und Therneau, T. M. (2000). *Survival Analysis*, 1. Auflage, Springer-Verlag, New York.
- Gray, R. J. (1992). Flexible methods for analyzing survival data using splines, with application to breast cancer prognosis, *Journal of the American Statistical Association* **87**: 942–951.

- Gustafson, P., MacNab, Y. und Wen, S. (2004). On the value of derivative evaluations and random walk suppression in markoc chain monte carlo algorithms, *Statistics and Computing* **14**: 23–38.
- Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems, *Journal of the American Statistical Association* **72**: 320–340.
- Hastie, T. J. und Tibshirani, R. J. (1990). *Generalized additive models*, 1. Auflage, Chapman and Hall, New York.
- Hastie, T. und Tibshirani, R. (1993). Varying-coefficients models, *J. R. Statist. Soc B* **55**: 757–796.
- Hennerfeind, A., Brezger, A. und Fahrmeir, L. (2005). Geoadditive survival models, *Technical report*, Ludwig–Maximilians–Universität München. SFB 386 Discussion Paper 414.
- Heumann, C. (2003). Monte carlo methods for missing data in generalized linear models and generalized linear mixed models, Habilitation. Institut für Statistik, Ludwig–Maximilians–Universität München.
- Heyde, C. (1997). *Quasi-Likelihood and its application*, 1. Auflage, Springer-Verlag, New York.
- Hjort, N. L. (1986). Bayes estimators and asymptotic efficiency in parametric counting process models, *Scandinavian Journal of Statistics* **13**: 63–85.
- Hämmerlin, G. und Hoffmann, K. H. (1994). *Numerische Mathematik*, 4. Auflage, Springer-Verlag, Berlin.
- Hougaard, P. (2000). *Analysis of multivariate Survival data*, Springer-Verlag, New York.
- Hutton, J. E. und Nelson, P. I. (1986). Quasi-likelihood estimation for semimartingals, *Stochast. Process. Appl* **22**: 245–257.
- Ishwaran, H. (2000). Inference for the random effects in bayesian generalized linear mixed models, *ASA Proceedings of the Bayesian Statistical Science Section*.
- Ishwaran, H. und Rao, J. S. (2005). Spike and slab variable selection: Frequentist and bayesian strategies, *The Annals of Statistics* **33**: 730–773.
- Jørgensen, B. und Knudsen, S. J. (2004). Parameter orthogonality and bias adjustment for estimating functions, *Scandinavian Journal of Statistics* **31**: 93–115.

- Karr, A. F. (1991). *Point Processes and Their Statistical Inference*, 2. Auflage, Dekker, New York.
- Klein, J. und Moeschberger, M. (2003). *Survival analysis*, 2. Auflage, Springer-Verlag, New York.
- Kneib, T. und Fahrmeir, L. (2004). A mixed model approach for structured hazard regression, *Technical report*, Ludwig–Maximilians–Universität München. SFB 386 Discussion Paper 400.
- Knorr-Held, L. (1999). Conditional prior proposals in dynamic models, *Scandinavian Journal of Statistics* **26**: 129–144.
- Kopp, P. E. (1984). *Martingales and stochastic integrals*, Cambridge University Press, Cambridge.
- Kauermann, G. (2005). Penalised Spline Fitting in Multivariable Survival Models with Varying Coefficients, *Computational Statistics and Data Analysis* **49**: 169–186.
- Kvaløy, J. T. und Reiersølmoen Neef, L. (2004). Tests for the proportional intensity assumption based on the score process, *Life Time Analysis* **10**: 139–157.
- Laird, N. M. und Ware, J. M. (1982). Random effects models for longitudinal data, *Biometrics* **38**: 963–974.
- Lasota, A. und Mackey, M. (1985). *Probabilistic Properties of Deterministic Systems*, 1. Auflage, Cambridge University Press, London.
- Last, G. und Brandt, A. (1995). *Marked Point Process on the Real Line*, 1. Auflage, Springer-Verlag, New York.
- Le Cam, L. (1953). On some asymptotic properties of maximum likelihood estimates and related bayes estimates, *Univ. Calif. Publ. Statist.* **1**.
- Leeb, H. und Pötscher, B. M. (2005). Model selection and inference: Facts and fiction, *Econometric Theory* **21**: 21–59.
- Lin, X. und Zhang, D. (1999). Inference in generalized additive mixed models by using smoothing splines, *J.R.Statist. Soc. B* **61**: 381–400.
- Lipster, R. S. und Shiryaev, A. N. (1977). *Statistics of Random Processes I, General Theory*, 1. Auflage, Springer-Verlag, New York.

- Lipster, R. S. und Shiryaev, A. N. (1978). *Statistics of Random Processes II, General Theory*, 1. Auflage, Springer-Verlag, New York.
- Liu, J. S. (2001). *Monte Carlo Strategies in Scientific Computing*, 1. Auflage, Springer-Verlag, New York.
- Liu, Q. und Pierce, D. A. (1994). A note on gauss–hermite quadratur, *Biometrika* **81**: 624–629.
- MacKay, D. J. C. (2003). *Information Theory, Inference and Learning Algorithms*, 6. Auflage, Cambridge University Press, Cambridge.
- Martinussen, T., Scheike, T. H. und Skovgaard, I. M. (2002). Efficient estimation of fixed and time-varying covariate effects in multiplicative intensity models, *Scandinavian Journal of Statistics* **29**: 57–74.
- Mason, J. C. und Handscomb, D. C. (2003). *Chebyshev Polynomials*, 1. Auflage, Chapman & Hall/CRC, Boca Raton.
- McCulloch, C. E., Casella, G. und Searle, S. R. (1992). *Variance Components*, 1. Auflage, Wiley, New York.
- McCulloch, C. E. und Searle, S. R. (2001). *Generalized, Linear and Mixed Models*, 1. Auflage, Wiley, New York.
- Mebane, W. R. und Wand, J. (1999). Poisson-normal dynamic generalized linear mixed models of u.s. house campaign contributions, *Working Papers*, Washington University in St. Louis.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. und Teller, E. (1953). Equations of state calculations by fast computing machines, *Journal of Chemical Physics* **21(6)**.
- Neal, R. M. (1993). *Bayesian Learning for Neuronal Networks*, 1. Auflage, Springer-Verlag, New York.
- Patterson, H. D. und Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal, *Biometrika* **58**: 545–554.
- Pawitan, Y. (2001). *Statistical Modelling and Inference Using Likelihood*, Clarendon Press, Oxford.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T. und Flannery, B. P. (1992). *Numerical Recipes in C*, 1. Auflage, Cambridge University Press, Cambridge.

- Rao, B. L. S. (1999). *Semimartingales and their Statistical Inference*, 1. Auflage, Chapman and Hall-CRC, Boca Raton.
- Ripatti, S. und Palmgren, J. (2000). Estimation of multivariate frailty models using penalized partial likelihood, *Biometrics* **56**: 1016–1022.
- Schoenfeld, D. (1982). Partial residuals for the proportional hazards regression model, *Biometrika* **69**: 239–241.
- Shiryayev, A. N. und Prokhorov, R. J. (2000). *Stochastic Calculus*, 1. Auflage, Springer, New York.
- Sirjaev, A. N. (1988). *Wahrscheinlichkeit*, 1. Auflage, VEB Deutscher Verlag der Wissenschaften, Berlin.
- Snyder, D. (1975). *Random Point Processes*, 1. Auflage, Wiley, New York.
- Stoer, J. (1999). *Numerische Mathematik I*, 8. Auflage, Springer-Verlag, Berlin.
- Therneau, T. M., Grambsch, P. M. und Fleming, T. R. (1990). Martingale-based residuals for survival models, *Biometrika* **77**: 147–160.
- Tian, L., Zucker, D. und Wei, L. J. (2005). On the cox model with time-varying regression coefficients, *Journal of the American Statistical Association* **100**: 172–183.
- Tibshirani, R. J. (1996). Regression shrinkage and selection via the lasso, *J. Royal. Statist. Soc B.* **58**: 267–288.
- Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion), *The Annals of Statistics* **22**: 1701–1762.
- Tierney, L. und Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities, *Journal of the American Statistical Association* **81**: 82–86.
- Tong, H. (1990). *Non-linear Time Series*, Oxford University Press, Oxford.
- Verbyla, A. P., Cullis, B. R., Kenward, M. G. und Welham, S. J. (1999). The analysis of designed experiments and longitudinal data by using smoothing splines, *Appl. Statist.* **48**: 269–311.
- Vermunt, J. K. (1996). *Log-linear event history analysis - a general approach with missing data, unobserved heterogeneity and latent variables*, Tilburg University Press, Tilburg.

- Wand, M. P. (2003). Smoothing and mixed models, *Computational Statistics* **18**: 223–249.
- Winkler, G. (1995). *Image Analysis, Random Fields and Dynamic Monte Carlo Methods: A Mathematical Introduction*, 1. Auflage, Springer-Verlag, New York.
- Winnett, A. und Sasieni, P. (2001). Miscellanea a note on scaled schoenfeld residuals for the proportional hazards model, *Biometrika* **88**: 565–571.