

DISSERTATION

SCHÄTZUNG VON MARGINALEN MODELLEN MIT PROPORTIONALEN HASARDFUNKTIONEN FÜR GRUPPEN VON KORRELIERTEN ZÄHLPROZESSEN

ausgeführt zum Zwecke der Erlangung des akademischen
Grades
eines Doktors der technischen Wissenschaften

eingereicht an der an der Technischen Universität Graz
Technisch-Naturwissenschaftliche Fakultät

von Mag. rer. nat. Franz Quehenberger

Institut für Medizinische Informatik, Statistik und
Dokumentation
Engelgasse 13, A-8010 Graz
franz.quehenberger@kfunigraz.ac.at

Graz, im August 1999

Kurzfassung

Abhängigkeiten zwischen Überlebensdaten können entstehen, wenn die beobachteten Einheiten zu Haufen gruppiert sind. Die multivariate Modellierung solcher Abhängigkeiten ist kaum anwendbar, während das Cox-Modell ein flexibles Instrument zur Modellierung der marginalen Überlebensverteilung darstellt. Die Parameterschätzung mittels der Cox'schen Likelihood ist auch bei gehäuften Beobachtungen konsistent (Lee et al., 1992). Liang et al. (1993) betrachten eine Pseudo-likelihood, die eine U -Statistik ist. Diese beiden Parameterschätzer werden verallgemeinert auf Zählprozesse mit proportionalen Hasardfunktionen. Asymptotische Eigenschaften der Schätzung von Erwartungswert und Varianz der Modellparameter werden für den Fall nicht identisch verteilter Beobachtungen bewiesen. Für den Logrank-Test stimmen diese Schätzer asymptotisch überein. Zwei Simulationsstudien haben zwar keine wesentlichen Unterschiede zwischen den Schätzern ergeben, die U -Statistiken sind aber empfindlicher bei einer Ausdünnung der Risikomenge. Die Schätzmethoden werden auf einen Datensatz aus der Zahnmedizin angewendet, wobei die Liegedauer von Teleskopkronen—gruppiert nach Patienten—modelliert wird. Zur Modellvalidierung wird die Methode der Schoenfeld-Residuen adaptiert.

Abstract

Dependencies between survival data can arise if the observed units are clustered. Modelling multivariate dependencies of this kind is hardly feasible, whereas the Cox model represents a flexible tool for modelling the marginal survival distribution. Parameter estimation by the Cox likelihood is consistent even if the data are correlated (Lee et al., 1992)). Liang et al. (1993) proposed a pseudolikelihood which is a U -statistic. The two parameter estimators are generalised to counting processes with proportional hazard functions. Asymptotic properties of expectation and variance of model parameters are proved for non identically distributed data. For the logrank test these estimators are asymptotically equivalent. Two simulation studies showed essentially no differences between the estimators, but the U -statistics were more sensitive against thinning of the risk set. The estimators are applied to data from dental medicine, whereby the durability of telescopic crowns clustered by patient is modelled. The method of Schoenfeld residuals is adapted for correlated data.

Zusammenfassung

Wenn die Beobachtungen zu Haufen gruppiert sind, dann bedeutet vollständige statistische Modellierung, daß die multivariate Verteilung vollständig spezifiziert und geschätzt werden muß. Dieser Ansatz ist speziell in der Überlebensanalyse oft nicht durchführbar. Beim marginalen Ansatz wird nur ein Modell für die Einzelbeobachtung angegeben. Die Abhängigkeit zwischen den Beobachtungen werden erst bei der Schätzung der Parameter berücksichtigt.

Hier wird die Schätzung des marginalen Modells für Zählprozesse mit proportionalen Hasardraten behandelt. Im ersten Kapitel werden die Grundlagen der Theorie der Zählprozesse vorgestellt. Im zweiten Kapitel wird die Modellbildung mit Zählprozessen behandelt. Neben Zensierung und Filterung werden der marginale Ansatz und der Ansatz der bedingten Unabhängigkeit (Gebrechlichkeitsmodell) bei gehäuften Daten beschrieben und durch Beispiele illustriert.

In Abschnitt 3 werden die benötigten Ergebnisse aus der Theorie der U -Statistiken für nicht identisch verteilte Stichproben bereitgestellt. Es sind dies vor allem der zentrale Grenzwertsatz, der Jackknife-Schätzer für die Varianz und ein Satz über Kerne mit einem Parameter, der aus der Stichprobe geschätzt werden muß.

In Kapitel 4 werden in Verallgemeinerung von Lee et al. (1992) die Eigenschaften der Cox-Likelihood bei der Schätzung des marginalen Modells der proportionalen Hasardfunktionen für nicht identisch verteilte Daten gehäufte Daten behandelt.

Kapitel 5 beinhaltet die Schätzung des marginalen Modells der proportionalen Hasardraten mittels Schätzfunktionen, die zugleich U -Statistiken sind—ein Ansatz, der auf Liang et al. (1993) zurückgeht. Zuletzt wird gezeigt, daß im Fall des Logrank-Tests die marginale Schätzung mittels Cox-Likelihood und mittels U -Statistik asymptotisch übereinstimmen.

In Kapitel 6 werden Simulationen zum Verhalten der Schätzer des marginalen Modells für paarweise abhängige Überlebensdaten beschrieben. Bei der Schätzung mit der Cox'schen Likelihood ohne Berücksichtigung der Abhängigkeit der Daten ist der Schätzer der Varianz des Regressionsparameters stark verzerrt. U -Statistiken sind empfindlich gegen Ausdünnung der Risikomenge. Ansonsten stimmen die Ergebnisse gut überein.

In Kapitel 7 werden die Schätzer des marginalen Modells auf die Daten aus einer retrospektiven Kohortenstudie über den Einfluß von Lage, Prothesenart und Hygiene auf die Liegedauer von Teleskopkronen. Die Liegedauer der Teleskopkronen im selben Patienten sind korreliert. Zur Modellvalidierung wird die Methode der Schoenfeld-Residuen für abhängige Daten adaptiert.

Eine weitere Zusammenfassung wird in (Quehenberger, 1997) gegeben.

Danksagung

Mein Dank gilt vor allem meiner Familie für das Verständnis, das sie der zeitaufwendigen Arbeit entgegengebracht hat.

Darüberhinaus bedanke ich mich bei Prof. Dr. G. Gell dafür, daß er mir die Dissertation ermöglichte.

Prof. Dr. E. Stadlober und Prof. Dr. W. Müller danke ich für die Betreuung der Arbeit.

Schließlich bedanke ich mich bei Dr. H. Jerebic für die Überlassung eines Datensatzes.

Inhaltsverzeichnis

1	Stochastische Prozesse	1
1.1	Stoppzeiten	2
1.2	Martingale	4
1.3	Vorhersagbare und optionale Variation	6
1.4	Stochastische Integration	7
1.5	Zählprozesse	8
1.5.1	Der Einsprung-Zählprozeß	9
1.5.2	Markierte Punkt-Prozesse	11
1.5.3	Der Erneuerungssatz	12
1.5.4	Produkte von Zählprozessen	12
1.5.5	Starten, Stoppen und Filtern	13
1.5.6	Stochastische Transformation der Zeit	14
1.6	Schwache Konvergenz	15
1.6.1	Straffheit	17
1.6.2	Empirische Prozesse	18
1.7	Grenzwertsätze und Ungleichungen	21
1.7.1	Der zentrale Grenzwertsatz	21
1.7.2	Lenglart'sche Ungleichung	21
1.7.3	Funktionen, Reihen und Konvergenz in Wahrscheinlichkeit	22
1.8	Das Produkt-Integral	24
1.8.1	Markov-Prozesse	26
1.9	Likelihoods für Zählprozesse	27
1.10	Partielle Likelihoods für Zählprozesse	31

2	Modelle für Zählprozesse	35
2.1	Vollständig beobachtete Prozesse	36
2.1.1	Modelle von Markov-Prozessen	37
2.1.2	Das Regressionsmodell für das relative Risiko	39
2.2	Rechts-Zensierung	40
2.2.1	Einleitung	40
2.2.2	Unabhängige Rechts-Zensierung	43
2.2.3	Die Konstruktion von Likelihoods	46
2.2.4	Nicht-informative Rechts-Zensierung	50
2.2.5	Links-Stutzen	51
2.3	Verallgemeinerte Zensierung, Filterung und Stutzen	54
2.4	Teilweise spezifizierte Modelle	55
2.5	Gebrechlichkeits-Modelle	57
2.6	Marginale Modelle	62
3	<i>U</i>-Statistiken	69
3.1	Definitionen	69
3.2	Der Jackknife-Schätzer für die Varianz	72
3.3	<i>U</i> -Statistiken mit geschätztem Kern	77
4	Schätzung mit der Cox'schen Likelihood	87
4.1	Modellformulierung	87
4.2	Die partielle Likelihood von Cox	89
4.3	Asymptotische Eigenschaften	91
4.4	Das Cox-Modell für Haufendaten	99
4.4.1	Konsistenz von $\hat{\beta}$	101
4.4.2	Asymptotische Verteilung von $\hat{\beta}$	105
5	Schätzung durch <i>U</i>-Statistiken	115
5.1	Der Schätzer	115
5.2	Asymptotische Verteilung	117
5.3	Gewichtsfunktionen	125

5.4	Der Schätzer von Liang et al. (1993)	127
5.5	Der Logrank-Test als U -Statistik	129
5.5.1	k -Stichproben Tests	129
5.5.2	Das Cox-Modell für k Stichproben	132
5.5.3	Der Pseudolikelihoodschätzer	132
6	Simulationen	135
6.1	Bivariate Überlebensfunktionen	135
6.1.1	Positiv stabile Gebrechlichkeits-Verteilungen	137
6.1.2	Die Clayton-Familie	138
6.1.3	Transformation der Normalverteilung	141
6.2	Schema der Simulationsstudien	142
6.3	Die Schätzer	144
6.4	Ein Clayton-Modell	145
6.4.1	Plan der Simulationsstudie	145
6.4.2	Ergebnisse	147
6.4.3	Vergleich mit dem Ergebnis von (Liang et al. 1993)	151
6.5	Vergleich der Gütefunktionen	157
6.5.1	Plan der Simulationsstudie	157
6.5.2	Ergebnisse	158
6.5.3	Zusammenfassung der Simulationsergebnisse	174
7	Ein Beispiel aus der Zahnmedizin	175
7.1	Beschreibung der Studie	175
7.2	Das Modell	176
7.3	Ergebnisse	178
7.3.1	Der Breslow-Schätzer	178
7.3.2	Überprüfung der Annahme proportionaler Hasardraten	178
7.3.3	Das Modell der proportionalen Hasardraten	180

Kapitel 1

Stochastische Prozesse

Diese Einführung basiert auf Andersen, Borgan, Gill und Keiding (1993, p. 59 ff). Ab nun wird unter \mathcal{T} entweder das abgeschlossene Intervall $[0, \tau]$ für eine positive reelle Zahl τ oder das halboffene Intervall $[0, \tau)$ für ein positives τ oder $\tau = \infty$ verstanden. Die konkrete Definition wird nur dann angegeben, wenn die Gültigkeit der Aussage davon abhängt.

Unter einem *stochastischen Prozeß* X wird, falls nicht anders angegeben, eine Familie von reellen Zufallsvariablen mit Indexmenge \mathcal{T} verstanden. Für ein Element von X sind die Schreibweisen X_t und $X(t)$ üblich, während $X(t, \omega)$ eine Realisation der Zufallsvariablen $X(t) : \Omega \rightarrow R$ bezeichnet. Die Abbildung $X(\cdot, \omega) : t \rightarrow X(t, \omega)$ heißt *Pfad des Prozesses* X .

Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Eine *Filtrierung* (\mathcal{F}_t) ist ein System von Teil- σ -Algebren von \mathcal{F} mit Indexmenge \mathcal{T} . Eine *aufsteigende* Filtrierung (\mathcal{F}_t) , d.h. $s < t \rightarrow \mathcal{F}_s \subseteq \mathcal{F}_t$, wird als *Geschichte* bezeichnet. Eine Geschichte ist *rechtsseitig stetig*, wenn $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$, und *vollständig*, wenn sie die P -Nullmengen enthält, d.h. wenn $A \subseteq B \in \mathcal{F}$ und $P(B) = 0$, dann $A \in \mathcal{F}_0$. Man sagt, *eine Filtrierung erfülle die üblichen Bedingungen*, wenn sie *aufsteigend*, *rechtsseitig stetig* und *vollständig* ist.

Die Vollständigkeit kann bei leichter Umformulierung der Ergebnisse bei den üblichen Bedingungen weggelassen werden (Andersen et al. 1993).

Man sagt, ein stochastischer Prozeß X ist *an eine Filtrierung* (\mathcal{F}_t) *angepaßt*, wenn X_t für alle $t \in \mathcal{T}$ \mathcal{F}_t -meßbar ist. Er ist *càdlàg*¹, wenn die Pfade rechtsseitig stetig sind und die linksseitigen Grenzwerte existieren, beziehungsweise *càglàd*, wenn die Pfade linksseitig stetig sind und die rechtsseitigen Grenzwerte

¹Abkürzung von „Continue à droite, limites à gauche“ (frz.)

te existieren. $D(\mathcal{T})$, die Menge der càdlàg Funktionen, wird als *Skorochod-Raum* bezeichnet.

Ein stochastischer Prozeß X erzeugt eine Geschichte (\mathcal{F}_t) durch $\mathcal{F}_t = \sigma\{X_s : s \leq t\}$, die von den Zufallsvariablen von 0 bis t erzeugte σ -Algebra. Ein Ergebnis von Courrège und Priouret (1965) besagt, daß *reelle Sprung-Prozesse*, d.h. Prozesse, deren Pfade càdlàg-Treppenfunktionen sind, rechtsseitig stetige Geschichten erzeugen. Erweiterungen mit Nullmengen ändern die Aussage nicht, sodaß reelle Sprung-Prozesse die üblichen Bedingungen erfüllen. Mit \mathcal{F}_{t-} wird die von $X(s)$, $0 \leq s < t$ erzeugte σ -Algebra bezeichnet.

1.1 Stoppzeiten

Eine *Stoppzeit* T ist eine $\bar{\mathcal{T}}$ -Zufallsvariable mit $\{T \leq t\} \in \mathcal{F}_t$ für alle $t \in \mathcal{T}$. Jede feste Zeit h ist eine Stoppzeit. Für einen càdlàg, angepaßten Prozeß ist

$$T = \inf \{t \in \mathcal{T} : |X(t)| \geq c\}$$

eine Stoppzeit. Zu einer Stoppzeit T kann man die σ -Algebren \mathcal{F}_T und \mathcal{F}_{T-} definieren, die von den Ereignissen erzeugt werden, die bis zum Zeitpunkt T (einschließlich) beziehungsweise vor dem Zeitpunkt T eingetreten sind:

$$\begin{aligned} \mathcal{F}_T &:= \{A \in \mathcal{F} \mid A \cap \{T \leq t\} \in \mathcal{F}_t \forall t \geq 0\} \\ \mathcal{F}_{T-} &:= \{A \in \mathcal{F} \mid A \cap \{T < t\} \in \mathcal{F}_t \forall t \geq 0\} \end{aligned}$$

\mathcal{F}_T kann charakterisiert werden als diejenige σ -Algebra, die von T und von allen Zufallsvariablen $X(T)$ erzeugt wird, wobei X ein beliebiger càdlàg angepaßter Prozeß ist. Für die Charakterisierung von \mathcal{F}_{T-} hat man $X(T)$ durch $X(T-)$ zu ersetzen.

Stoppzeiten und die noch zu definierenden Zählprozesse hängen eng zusammen. Der *Einsprung-Zählprozeß* (siehe 1.5.1) wird folgendermaßen definiert: Sei die reelle Zufallsvariable T die Zeit des Eintretens eines zufälligen Ereignisses. Der Indikator-Prozeß $(I(T \leq t))$ ist ein càdlàg Prozeß, der bis zur Zeit T Null ist, dann auf 1 springt, und auf diesem Wert verharrt. Man kann zeigen, daß der Prozeß $(I(T \leq t))$ an eine Filtrierung (\mathcal{F}_t) , die die üblichen Bedingungen erfüllt, genau dann angepaßt ist, wenn T eine Stoppzeit ist. Insbesondere ist $(I(T \leq t))$ an die Filtrierung (\mathcal{N}_t) angepaßt, die durch $\mathcal{N}_t = \sigma\{T \wedge t, I(T \leq t)\}$ definiert ist².

²Schreibweise: $x \wedge y$ bezeichnet das Minimum von x und y .

Im allgemeinen ist $X(T)$ keine Zufallsvariable, d.h. die Funktion ist nicht meßbar bezüglich \mathcal{F}_T . $X(T)$ ist nur im Fall $\mathbb{P}(t < \infty) = 1$ fast sicher definiert. Für càdlàg-Prozesse ist $X(T)$ eine \mathcal{F}_T -meßbare Zufallsvariable.

Der *gestoppte Prozeß* für eine Stoppzeit T ist definiert als $X^T(t) = X(t \wedge T)$.

Man sagt, eine Folge (T_n) von Stoppzeiten ist *lokalisierend*, wenn sie monoton wachsend ist und für alle $t \in \mathcal{T}$ die Bedingung $\lim_{n \rightarrow \infty} P(T_n \geq t) = 1$ erfüllt. Ein Prozeß hat eine bestimmte Eigenschaft *lokal*, wenn eine lokalisierende Folge (T_n) von Stoppzeiten existiert, sodaß der Prozeß $I(T_n > 0)X^{T_n}$ diese Eigenschaft hat.

Ein Beispiel für eine solche lokale Eigenschaft ist folgende Definition: Ein stochastischer Prozeß X ist *lokal beschränkt*, falls eine lokalisierende Folge T_n und Konstanten c_n existieren, sodaß für jedes n gilt: $\sup_{t \leq T_n} |X(t)| \leq c_n$ f.s. auf $\{T_n > 0\}$.

Eine äquivalente Formulierung ist: Ein Prozeß X ist *lokal beschränkt*, wenn eine nicht fallende Folge $T_n; n = 1, 2, \dots$ von Stoppzeiten mit $P(T_n \geq t) \rightarrow 1$ für $n \rightarrow \infty$ und $t \in \mathcal{T}$ existiert, sodaß X^{T_n} beschränkt ist auf $\{T_n > 0\}$.

Jeder linksseitig stetige angepaßte Prozeß ist lokal beschränkt. Als lokalisierende Folge von Stoppzeiten kann man hier $T_n = \inf \{t : X(t) > n\}$ wählen. Nur die Pfade von $I(T_n > 0)X^{T_n}$ sind beschränkt. Würde in der Definition von lokalisierenden Eigenschaften die Beschränktheit von X^{T_n} gefordert, so wären nur diejenigen linksseitig stetigen angepaßten Prozesse lokal beschränkt, deren Anfangswert beschränkt ist.

In dieser Arbeit sind Integrale über stochastische Prozesse pfadweise Lebesgue-Stieltjes-Integrale. Wir nehmen an, daß Y ein Prozeß mit lokal beschränkter Variation ist, d.h. $\int_{[0,t]} |dY(s)| < \infty$ für fast alle $\omega \in \Omega$ und für alle $t \in \mathcal{T}$. Weiters sei $\int_{[0,t]} |X(s)| |dY(s)| < \infty$ für fast alle $\omega \in \Omega$ und für alle $t \in \mathcal{T}$. Dann ist $\int X dY$ definiert als der stochastische Prozeß

$$t \rightarrow \int_{[0,t]} X(s) dY(s).$$

Es sei vereinbart, daß $Y(0-) = 0$, so daß für $\int_0^0 = X(0)Y(0)$ gilt. Bei Integralen bezüglich des Lebesgue-Maßes ist die integrierende Funktion $f(x) = x$. Man schreibt vereinfachend $\int X dx$ und $\int X$ für $t \rightarrow \int_0^t X(s) ds$. Zu einem càdlàg Prozeß X definiert man die *linksseitig stetige Modifikation* $X_-(t) = X(t-)$ und den *Sprungprozeß* $\Delta X = X - X_-$.

Zwei Prozesse werden als *ununterscheidbar* bezeichnet, wenn $P(X(t) = Y(t)) \forall t \in \mathcal{T} = 1$.

1.2 Martingale

Ein *Martingal* ist ein (càdlàg) angepaßter Prozeß M , der *integrierbar* ist, d.h.

$$E(|M(t)|) < \infty \quad \text{für alle } t \in \mathcal{T}$$

und die *Martingalbedingung* erfüllt:

$$E(M(t)|\mathcal{F}_s) = M(s) \quad \text{für alle } s \leq t \quad (1.2.1)$$

M ist ein *Submartingal*, wenn

$$E(M(t)|\mathcal{F}_s) \geq M(s) \quad \text{für alle } s \leq t \quad (1.2.2)$$

M ist ein *Supermartingal*, wenn $-M$ ein Submartingal ist. Ein Martingal heißt *quadratisch integrierbar*, wenn

$$\sup_{t \in \mathcal{T}} E(M(t)^2) < \infty. \quad (1.2.3)$$

Ein quadratisch integrierbares Martingal kann für $\tau \notin \mathcal{T}$ durch den punktweisen Grenzwert, der fast sicher existiert, auf $\bar{\mathcal{T}}$ fortgesetzt werden. Das Resultat ist ein quadratisch integrierbares Martingal auf $\bar{\mathcal{T}}$.

Ein Martingal auf $\bar{\mathcal{T}}$ ist automatisch *gleichmäßig integrierbar*, in dem starken Sinn, daß die Menge der $M(T)$, wobei T eine beliebige Stoppzeit ist, gleichmäßig integrierbar ist:

$$\lim_{c \rightarrow \infty} \sup_T E[|M(T)|I(|M(T)| > c)] = 0 \quad (1.2.4)$$

Ein gleichmäßig integrierbares Martingal auf $\bar{\mathcal{T}}$ erfüllt Doob's optionalen Stichprobenziehungssatz, d.h. in der Martingalbedingung (1.2.1) können $s \leq t$ durch Stoppzeiten $S \leq T$ ersetzt werden.

Die Integrierbarkeitsbedingungen, die Martingale oder quadratisch integrierbare Martingale erfüllen müssen, können durch Lokalisierung umgangen werden; man arbeitet mit *lokalen Martingalen* und *lokal quadratisch integrierbaren Martingalen*, die mittels lokalisierender Folgen von Stoppzeiten definiert werden.

Sei M ein lokales (quadratisch integrierbares) Martingal und T eine Stoppzeit. Dann ist der Prozeß $I(T > 0)M^T$ nicht notwendigerweise ein lokales (quadratisch integrierbares) Martingal.

Ein *vorhersagbarer Prozeß* H ist in gewisser Weise das Gegenteil eines Martingales. Ein solcher Prozeß ist als Funktion von (t, ω) auf $\mathcal{T} \times \Omega$ meßbar

bezüglich der σ -Algebra auf $\mathcal{T} \times \Omega$, die durch alle angepaßten Prozesse, deren sämtliche Pfade linksseitig stetig sind, erzeugt wird. Ein stochastischer Prozeß H ist genau dann vorhersagbar, wenn $H(T)$ für jede Stoppzeit T \mathcal{F}_{T-} -meßbar ist. Das heißt, daß der Wert eines solchen Prozesses bei t zur Zeit $t-$ festgelegt ist.

Linksseitig stetige angepaßte Prozesse sind aufgrund der Definition vorhersagbar. Jede deterministische, meßbare Funktion ist vorhersagbar (dazu beachte man, daß die linksseitig stetigen Treppenfunktionen auf \mathcal{T} die Borelmengen auf \mathcal{T} erzeugen). Daher gibt es auch rechtsseitig stetige Prozesse, die vorhersagbar sind. Man kann zeigen, daß vorhersagbare càdlàg Prozesse lokal beschränkt sind, da man den Prozeß gerade dann stoppen kann, bevor er einen bestimmten Wert überschreitet.

Ein Prozeß, der sowohl ein lokales Martingal als auch ein vorhersagbarer Prozeß von endlicher Variation ist, ist konstant. Das ist der Grund, warum nicht konstante Martingale mit stetigen Pfaden von unbeschränkter Variation auf jedem Intervall sind. Ein Beispiel dafür ist die Brown'sche Bewegung.

Die *Doob-Meyer Zerlegung* wird in ihrer ursprünglichen Form zunächst auf einer Klasse von Submartingalen angegeben und dann durch Lokalisierung auf eine größere Klasse von Prozessen erweitert:

Man sagt, ein càdlàg Prozeß X ist von der Klasse D , wenn die Familie der Zufallsvariablen $X(T)$, wobei T eine beliebige Stoppzeit ist, *gleichmäßig* integrierbar ist.

Sei X ein Submartingal der Klasse D . Dann existiert ein càdlàg, *nichtfallender und vorhersagbarer* Prozeß \tilde{X} , sodaß $M = X - \tilde{X}$ ein gleichmäßig integrierbares Martingal ist mit $M(0) = 0$. Darüberhinaus ist \tilde{X} integrierbar. Wenn $\tau \notin \mathcal{T}$ gilt auch $E(\tilde{X}(\tau)) < \infty$.

Diese Zerlegung ist notwendigerweise eindeutig, weil sonst $X = \tilde{X} + M = \tilde{X}' + M'$ und $\tilde{X} - \tilde{X}' = M - M'$. Das wäre aber ein vorhersagbares Martingal von endlicher Variation, das bei zum Zeitpunkt Null den Wert Null annimmt. Daher ist es konstant 0 auf \mathcal{T} .

Durch Lokalisierung kann man auf die Integrierbarkeitsbedingungen verzichten. Unter dem *Kompensator* eines càdlàg angepaßten Prozesses X versteht man einen vorhersagbaren und càdlàg Prozeß \tilde{X} von endlicher Variation, sodaß $M = X - \tilde{X}$ ein lokales Martingal mit $M(0) = 0$ ist. Existiert ein Kompensator dann ist er eindeutig. Mittels der Doob-Meyer-Zerlegung zeigt man, daß X genau dann einen Kompensator hat, wenn X als Differenz von 2 lokalen Submartingalen dargestellt werden kann.

Ein Prozeß mit Kompensator heißt *spezielles Semimartingal*. Alle lokalen

Sub- und Supermartingale haben Kompensatoren. Das sind insbesondere nichtfallende nichtnegative lokal integrierbare càdlàg-Prozesse, da sie trivialerweise lokale Submartingale sind. Nach der Jensen'schen Ungleichung ist das Quadrat eines quadratisch integrierbaren Martingals ein Submartingal. Daher ist das Quadrat eines lokal quadratisch integrierbaren Martingals ein Submartingal und besitzt einen nichtfallenden Kompensator. Das Produkt MM' von lokal quadratisch integrierbaren Martingalen kann als Differenz von 2 Quadraten von lokal quadratisch integrierbaren Martingalen angeschrieben werden [$MM' = 1/4(M + M')^2 - 1/4(M - M')^2$] und hat einen Kompensator von endlicher Variation.

1.3 Vorhersagbare und optionale Variation

Für lokal quadratisch integrierbare Martingale M und M' ist M^2 ein lokales Submartingal und MM' die Differenz von zwei lokalen Submartingalen (siehe 1.2). Die Kompensatoren dieser Prozesse sind die *vorhersagbare Variation* $\langle M \rangle = \langle M, M \rangle$ und die *vorhersagbare Kovariation* $\langle M, M' \rangle$ ³. Prozesse mit $\langle M, M' \rangle = 0$ werden als *orthogonal* bezeichnet. Die vorhersagbare Variation $\langle M \rangle(t)$ kann als Grenzwert in Wahrscheinlichkeit des Netzes der endlichen Tupel $E = (t_1, \dots, t_n)$ der Bauart $0 = t_0 < \dots < t_n = t$, deren Feinheit gegen 0 geht, von $\sum_E \text{Var}(M(t_i) - M(t_{i-1}) | \mathcal{F}_{t_{i-1}})$ charakterisiert werden. Die *optionale Variation* $[M]$ erhält man als Grenzwert in der Wahrscheinlichkeit, wenn man den bedingten Erwartungswert wegläßt, d.h. als Grenzwert von $\sum_E (M(t_i) - M(t_{i-1}))^2$. Analog ist die optionale Kovariation $[M, M']$ definiert⁴. Wenn M und M' stetig sind, dann fallen vorhersagbare und optionale Variation zusammen. Für lokale Martingale mit endlicher Variation gibt es die explizite Darstellung

$$\begin{aligned} [M](t) &= \sum_{s \in [0, t]} \Delta M(s)^2 = M(t)^2 - 2 \int_0^t M(s-) dM(s) \\ [M, M'](t) &= \sum_{s \in [0, t]} \Delta M(s) \Delta M'(s) \\ &= M(t)M'(t) - \int_0^t M(s-) dM'(s) - \int_0^t M'(s-) dM(s). \end{aligned}$$

³In Širjaev (1988) findet man dafür die Bezeichnung *quadratische und gemeinsame Charakteristik* der quadratisch integrierbaren Martingale

⁴In Širjaev (1988) findet man dafür die Bezeichnungen *quadratische und gemeinsame Variation* der quadratisch integrierbaren Martingale

Der Prozeß $[M]$ ist nichtfallend, càdlàg und $M^2 - [M]$ ist ein lokales Martingal. Wenn $[M, M']$ sogar lokal integrierbar ist, dann ist das lokale Martingal M und M' quadratisch integrierbar und $\langle M, M' \rangle$ ist der Kompensator von $[M, M']$. Wenn daher M und M' lokal quadratisch integrierbare Martingale von endlicher Variation ohne gemeinsame Sprünge sind, also $\Delta M \Delta M' = 0$, dann ist $[M, M'] = 0$ und daher auch $\langle M, M' \rangle = 0$, d.h. M und M' sind orthogonal und MM' ist ein lokales Martingal.

Mittels der optionalen und quadratischen Variation eines Martingals kann überprüft werden, ob ein quadratisch integrierbares Martingal durch eine Stoppzeit T lokalisiert wird. Wenn das lokale Martingal 0 ist zur Zeit 0, dann kann die Bedingung $I(T > 0)$ weggelassen werden. Für ein lokal integrierbares Martingal M und eine Stoppzeit T ist M^T genau dann ein quadratisch integrierbares Martingal, wenn

$$\mathbb{E} \langle M \rangle (T) < \infty$$

und genau dann, wenn

$$\mathbb{E} [M] (T) < \infty.$$

1.4 Stochastische Integration

Satz 1.4.1. *Sei M ein lokal quadratisch integrierbares Martingal von endlicher Variation, H ein vorhersagbarer Prozeß und $\int H^2 d[M]$ lokal integrierbar oder $\int H^2 d\langle M \rangle$ ist lokal endlich (das ist stets erfüllt, wenn H lokal beschränkt ist).*

Dann ist $\int H dM$ ein lokal quadratisch integrierbares Martingal und

$$\begin{aligned} \left[\int H dM \right] &= \int H^2 d[M] \\ \left\langle \int H dM \right\rangle &= \int H^2 d\langle M \rangle. \end{aligned} \quad (1.4.1)$$

Hinreichend für lokale Beschränktheit von H sind Linksstetigkeit oder Rechtsstetigkeit. Eine ähnliche Formel gilt für die vorhersagbare Variation matrixwertiger Prozesse:

$$\begin{aligned} \left\langle \int H dM, \int K dM' \right\rangle &= \int H d\langle M, M' \rangle K^\top \\ \left[\int H dM, \int K dM' \right] &= \int H d[M, M'] K^\top \end{aligned} \quad (1.4.2)$$

1.5 Zählprozesse

Ein multivariater Zählprozeß $N = (N_1, \dots, N_n)$ ist ein Vektor von angepaßten càdlàg Prozessen mit $N_i(0) = 0$ und mit stückweise konstanten, nichtfallenden Pfaden mit Sprunghöhen $+1$. Zwei Prozesse springen jeweils mit Wahrscheinlichkeit 0 zugleich. $N_i(t)$ sei fast sicher endlich für $i = 1, \dots, n$ und für alle $t \in \mathcal{T}$. Wenn $\tau \notin \mathcal{T}$, dann soll auch $N_i(\tau) = \lim_{t \uparrow \tau} N_i(t) = \infty$ zulässig sein.

Ein Zählprozeß N ist ein rechtsseitig stetiger Sprungprozeß, weshalb die von N erzeugte Filtrierung rechtsseitig stetig ist, sodaß N eine Geschichte erzeugt.

Da die Komponenten von N niemals gleichzeitig springen, ist auch $N = \sum_{i=1}^n N_i$ ein Zählprozeß. Zählprozesse sind *lokale Submartingale*, da sie *angepaßt, càdlàg, nichtfallend* und *lokal beschränkt* sind (für eine lokalisierende Folge T_n von Stoppzeiten, die den Sprüngen von N zugeordnet sind, sodaß $N(T_n) = n$ und $0 \leq N_i^{T_n} \leq n$, siehe Andersen et al. (1993, p73)). Daher haben sie aufgrund der *Doob-Meyer-Zerlegung* eindeutig bestimmte Kompensatoren Λ_i , die zu einem Vektor Λ zusammengefaßt werden können. $\Lambda_i(\cdot)$ ist ein vorhersagbarer Prozeß mit $\Lambda_i(0) = 0$ und der charakterisierenden Eigenschaft, daß $M_i = N_i - \Lambda_i$ ein lokales Martingal ist. Die Sprunghöhen von $\Delta\Lambda_i$ sind durch 1 beschränkt. Für den Kompensator Λ von N gilt $\Lambda = \sum_{i=1}^n \Lambda_i$. Wenn Λ *absolut stetig* ist, dann existiert ein *Intensitätsprozeß* λ_i mit

$$\Lambda_i(t) = \int_0^t \lambda_i(u) du. \quad (1.5.1)$$

In Hinblick auf die Anwendungen ist es sinnvoll, die Vorhersagbarkeit des Intensitätsprozesses zu fordern. Der Intensitätsprozeß ist im Allgemeinen nicht eindeutig. Wenn ein càglàd Intensitätsprozeß existiert, dann ist er eindeutig.

Da N durch die oben definierte lokalisierende Folge T_n lokal beschränkt ist und Λ als vorhersagbarer càdlàg Prozeß lokal beschränkt ist, ist M_i durch die Folge T_n ein *lokal quadratisch integrierbares Martingal mit lokal beschränkter Variation*. Es gilt

$$\begin{aligned} \langle M_i \rangle &= \Lambda_i - \int \Delta\Lambda_i d\Lambda_i \\ \langle M_i, M_{i'} \rangle &= - \int \Delta\Lambda_i d\Lambda_{i'} \quad \text{für } i \neq i' \end{aligned} \quad (1.5.2)$$

Wenn Λ stetig ist, dann folgt daraus

$$\begin{aligned} \langle M_i \rangle &= \Lambda_i \\ \langle M_i, M_{i'} \rangle &= 0 \quad \text{für } i \neq i' \end{aligned} \quad (1.5.3)$$

Die Martingale aus Zerlegungen von Zählprozessen mit stetigen Kompensatoren, insbesondere diejenigen, für die ein Intensitätsprozeß existiert, sind orthogonal.

Sei N ein n -variater Zählprozeß mit Intensität λ , Kompensator $\Lambda = \int \lambda$ und Doob-Meyer-Zerlegung $N = M + \Lambda$. Sei H ein $q \times p$ -Matrix von lokal beschränkten vorhersagbaren Prozessen. Sei $\int H dM$ der q -Vektor der zeilenweisen Summen der Integrale über die Komponenten von H bezüglich der Komponenten von M . Das Ergebnis ist ein Vektor von lokal quadratisch integrierbaren Martingalen von endlicher Variation. Es gilt aufgrund von (1.4.2)

$$\begin{aligned} \langle M \rangle &= \text{diag} \int \lambda \\ \left\langle \int H dM \right\rangle &= \int H \text{diag} \lambda H^\top \end{aligned} \quad (1.5.4)$$

1.5.1 Der Einsprung-Zählprozeß

Ein wichtiges Beispiel eines Zählprozesses ist der *Einsprung-Prozeß*. Sei T eine nicht-negative Zufallsvariable mit absolut stetiger Verteilungsfunktion F , Überlebensfunktion $S = 1 - F$, Dichte f und Hasard-Rate $\alpha = f/S$. Sei τ_F die Obergrenze des Trägers von S . Sei $\tau = \infty$ und $\mathcal{T} = [0, \infty]$, auch wenn $\tau_F < \infty$. Trotz $\int_0^{\tau_F} \alpha(t) dt = \infty$ gilt dann für $t < \tau_F$

$$\int_0^t \alpha(u) du = -\log(1 - F(t)) < \infty.$$

Proposition 1.5.1. *Sei ein univariater Zählprozeß definiert durch*

$$N(t) = I(T \leq t).$$

Sei (\mathcal{N}_t) die durch N erzeugte Filtrierung und sei der linksseitig stetige, vorhersagbare Prozeß Y definiert durch

$$Y(t) = I(T \geq t).$$

Dann ist der Kompensator von N gegeben durch

$$\Lambda(t) = \int_0^t Y(s)\alpha(s) ds.$$

N besitzt daher den Intensitätsprozeß

$$\lambda(t) = Y(t)\alpha(t).$$

Beweis. Zunächst stellt man fest, daß Λ vorhersagbar ist (stetig und angepaßt). Daher muß nur gezeigt werden, daß $M = N - \Lambda$ ein lokales Martingal ist.

Es genügt, wenn man

$$\mathbb{E} (M(\infty)|\mathcal{N}_t) = M(t)$$

zeigt. Dann gilt nämlich für $s < t$:

$$\begin{aligned} \mathbb{E} (M(t)|\mathcal{N}_s) &= \mathbb{E} (\mathbb{E} (M(\infty)|\mathcal{N}_t) |\mathcal{N}_s) \\ &= \mathbb{E} (M(\infty)|\mathcal{N}_s) \\ &= M(s). \end{aligned}$$

Sei nun zunächst $t = 0$. Da \mathcal{N}_0 trivial ist, muß gezeigt werden, daß $\mathbb{E}M(\infty) = 0$. $N(\infty) = 1$ und

$$\begin{aligned} \mathbb{E} \Lambda(\infty) &= \mathbb{E} \int_0^{\tau_F} Y(t)\alpha(t) dt \\ &= \int_0^{\tau_F} \mathbb{P}(t \geq t)\alpha(t) dt \\ &= \int_0^{\tau_F} S(t) \frac{f(t)}{S(t)} dt = \int_0^{\tau_F} f(t) dt = 1. \end{aligned}$$

Wenn man den Erwartungswert bezüglich $\mathcal{N}_t = \sigma(T \wedge t, I(T \leq t))$ berechnet, kann man 2 Fälle unterscheiden, daß $T = s \leq t$ ist für ein $s \leq t$, und daß $T > t$ ist. Im ersten Fall ist $M(\infty) = M(t) = M(s)$ und die Behauptung gilt. Im zweiten Fall ist $M(\infty) - M(t) = \int_t^{\tau_F} I(T \geq s)\alpha(s) ds$ und man muß zeigen, daß der Erwartungswert des Integrals, gegeben $T > t$, null ist. Aber gegeben $T > t$ hat T die Hasardrate $\alpha I_{[t, \tau_F]}$. Das ist aber der analoge Fall zu $t = 0$, nur die Hasardrate ist eine andere. \square

Dieser Beweis stammt aus (Andersen et al. 1993, p. 76 f). Einen längeren Beweis findet man in (Fleming und Harrington 1991, p. 27 ff).

Wenn T keine Dichte hat, kann die *kumulative Hasardfunktion* $A(t) = \int_0^t (1 - F(s-))^{-1} dF(s)$ definiert werden. Der Kompensator von N ist dann der Prozeß $\Lambda = \int Y dA$, wobei Y definiert ist wie oben. Der obige Beweis kann beinahe unverändert übernommen werden. N besitzt daher einen Intensitätsprozeß genau dann, wenn die Verteilung von T eine Dichte besitzt.

Wenn zur Zeit des Sprunges T weitere Information über den Typ des Ereignisses in Form einer Marke J , einer Zufallsvariablen mit Werten aus $\{1, \dots, k\}$, verfügbar ist, dann kann N als k -variater Zählprozeß modelliert werden durch $N_h(t) = I(T \leq t, J = h)$, $t \in [0, \infty)$, $h \in \{1, \dots, k\}$. Die Filtrierung sei selbsterzeugt durch $\mathcal{N}_t = \sigma\{T \wedge t, I(T \leq t), J \cdot I(T \leq t)\}$. Angenommen die Überlebensfunktion von T ist S und die gemeinsame Dichte von (T, J) ist $f_h(t)$ bezüglich Lebesgue-Maß mal Zählmaß. Sei $\alpha_h = f_h/S$. Dann hat N_h den Kompensator $\Lambda_h = \int Y_h \alpha_h$, wobei $Y(t) = I(T \geq t)$. Dieses Ergebnis wird durch die Jacod'sche Formel (1.9.1) verallgemeinert.

1.5.2 Markierte Punkt-Prozesse

Unter Umständen gibt es mehr als eine endliche Anzahl von Typen von Ereignissen. Trotzdem gibt es in endlicher Zeit nur endlich viele Ereignisse. Die Ereignisse können sogar aus einem Kontinuum stammen, beispielsweise ein Meßwert, der bei einem bestimmten Ereignis gemessen wird. Statt des Typs eines Ereignisses ordnet man diesem eine Borel-Menge zu. Der Zählprozeß entspricht dann einem Zählmaß auf der Punktmenge von Paaren aus Sprungzeit und Sprungmarke $(T_1, J_1), (T_2, J_2)$ etc. N und Λ werden dann zu Maßen μ und ν auf dem Produktraum aus Zeit und Typ (beziehungsweise Marke).

Es folgt eine formalere Definition: Sei (E, \mathcal{E}) ein Meßraum von Marken. Bei einem gewöhnlichen Zählprozeß ist $E = \{1, \dots, k\}$ und \mathcal{E} die Potenzmenge. N ist ein *markierter Punkt-Prozeß* bezüglich einer Filtrierung und einem Raum von Marken, wenn N ein (zufälliges) Zählmaß auf dem Produktraum $(\mathcal{T} \times E, \mathcal{B}(\mathcal{T}) \otimes \mathcal{E})$ ist, wobei \mathcal{B} eine Borel- σ -Algebra bezeichnet und N für jedes $A \in \mathcal{E}$ die Bedingung erfüllt, daß

$$N_A(t) = N([0, t] \times A)$$

ein Zählprozeß ist.

Für disjunkte Mengen A_1, \dots, A_k aus \mathcal{E} ist der Prozeß $(N_{A_1}, \dots, N_{A_k})$ ein gewöhnlicher (multivariater) Zählprozeß und hat daher einen Kompensator $(\Lambda_{A_1}, \dots, \Lambda_{A_k})$. Diese Kompensatoren können analog zu N als *vorhersagbares Maß* auf $(\mathcal{T} \times E, \mathcal{B}(\mathcal{T}) \otimes \mathcal{E})$ gewonnen werden:

$$\Lambda_A(t) = \Lambda([0, t] \times A)$$

Wenn (T_n, J_n) Punkte auf dem markierten Punkt-Prozeß sind, dann ist T_n eine Stoppzeit und J_n ist \mathcal{F}_{T_n} -meßbar mit Werten in (E, \mathcal{E}) für $n = 1, 2, \dots$

1.5.3 Der Erneuerungssatz

Seien N ein Zählprozeß, der an zwei Filtrierungen (\mathcal{F}_t) und (\mathcal{G}_t) mit $\mathcal{F}_t \subseteq \mathcal{G}_t$ für alle t angepaßt ist. Man sagt, die Filtrierungen sind *geschachtelt*. Der Intensitätsprozeß bezüglich der größeren Filtrierung (\mathcal{G}_t) sei λ . Im allgemeinen wird die Intensität bezüglich der kleineren Filtrierung (\mathcal{F}_t) von λ verschieden sein. Nach dem *Erneuerungssatz* gibt es einen (\mathcal{F}_t) -vorhersagbaren Prozeß $\tilde{\lambda}$, sodaß für alle t

$$\tilde{\lambda}(t) = \mathbb{E} (\lambda(t) | \mathcal{F}_{t-}), \quad (1.5.5)$$

wobei $\tilde{\lambda}$ der (\mathcal{F}_t) -Intensitätsprozeß von N ist.

1.5.4 Produkte von Zählprozessen

Produktkonstruktionen benötigt man, um unabhängige Komponenten zusammenzufassen. Seien zwei gefilterte Wahrscheinlichkeitsräume gegeben durch $(\Omega^{(i)}, \mathcal{F}^{(i)}, (\mathcal{F}_t^{(i)}, t \in \mathcal{T}), \mathbb{P}^{(i)})$, $i = 1, 2$. Auf diesen Räumen sind jeweils multivariate Zählprozesse $N^{(i)}$ mit Intensitätsprozeß $\lambda^{(i)}$ gegeben. Dann kann auf natürliche Weise das Produkt der filtrierte Wahrscheinlichkeitsräume angegeben werden als

$$\begin{aligned} \Omega &= \Omega^{(1)} \times \Omega^{(2)}, & \mathcal{F} &= \mathcal{F}^{(1)} \otimes \mathcal{F}^{(2)}, \\ \mathcal{F}_t &= \mathcal{F}_t^{(1)} \otimes \mathcal{F}_t^{(2)}, & P &= P^{(1)} \otimes P^{(2)}. \end{aligned}$$

Eine hinreichende Bedingung dafür, daß (\mathcal{F}_t) rechts-stetig ist, ist $\mathcal{F}_t^{(i)} = \mathcal{F}_0^{(i)} \vee \sigma \{N^{(i)}(s) : s \leq t\}$. Dann ist nämlich

$$\mathcal{F}_t = \mathcal{F}_0^{(1)} \otimes \mathcal{F}_0^{(2)} \vee \sigma \{N^{(1)}(s), N^{(2)}(s) : s \leq t\}$$

offensichtlich rechtsseitig stetig. Die komponentenweisen Prozesse $\lambda_h^{(i)}$ sind als Prozesse auf Ω mit der Wahrscheinlichkeit \mathbb{P} vorhersagbar bezüglich (\mathcal{F}_t) und die Prozesse $N_h^{(i)} - \int \lambda_h^{(i)}$ sind lokale Martingale. Daher ändern sich auch die Intensitätsprozesse der $N^{(i)}$ auf dem Produktraum nicht. Aufgrund der Stetigkeit der Kompensatoren hat $(N^{(1)}, N^{(2)})$ fast sicher keine gemeinsamen Sprünge und ist daher ein multivariater Zählprozeß.

Auch für *bedingt unabhängige* Komponenten ist eine Produktkonstruktion möglich. In diesem Fall ist auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ der Zählprozeß $N(t) = (N^{(1)}(t), N^{(2)}(t))$ definiert. Die Filtrierungen $\mathcal{F}_t^{(1)}$ und $\mathcal{F}_t^{(2)}$ sollen *bedingt unabhängig* sein bezüglich einer σ -Algebra $\mathcal{A} \subseteq \mathcal{F}$ (Die σ -Algebra \mathcal{F} ist die von der Vereinigung der \mathcal{F}_t erzeugte σ -Algebra). Das soll bedeuten, daß für alle $A \in \mathcal{F}_t^{(1)}$, $B \in \mathcal{F}_t^{(2)}$, $C \in \mathcal{A}$ mit $P(C) > 0$ gilt:

$$P(A \cap B | C) = P(A | C)P(B | C)$$

Diese Konstruktion wird oft angewendet, wenn $\mathcal{F}_t^{(i)} = \sigma \{N^{(i)}(s) : s \leq t\}$ ist und die σ -Algebra \mathcal{A} die Ereignisse modelliert, die „zur Zeit 0“ eintreten.

Weiters seien die Kompensatoren von $N^{(1)}$ und $N^{(2)}$ bezüglich $(\mathcal{F}_t^{(1)})$ beziehungsweise $(\mathcal{F}_t^{(2)})$ gegeben durch $\Lambda^{(1)}$ und $\Lambda^{(2)}$. Nun definiert man

$$\mathcal{F}_t = \mathcal{A} \vee \mathcal{F}_t^{(1)} \vee \mathcal{F}_t^{(2)}.$$

Wenn die $\mathcal{F}_0^{(i)}$ trivial sind, dann ist $\mathcal{F}_0 = \mathcal{A}$. Man kann leicht direkt überprüfen, daß jedes $(\mathcal{A} \vee \mathcal{F}_t^{(1)})$ -Martingal auch ein \mathcal{F}_t -Martingal ist. Auf diesen Schritt kommt es an, wenn man zeigen will, daß der (\mathcal{F}_t) -Kompensator von N genau aus den $(\mathcal{A} \vee \mathcal{F}_t^{(i)})$ -Kompensatoren von $N^{(i)}(t)$ besteht.

1.5.5 Starten, Stoppen und Filtern

Der gestoppte Prozeß wurde auf Seite 3 eingeführt. Der gestoppte Zählprozeß hat eine Darstellung mittels stochastischer Integration: Sei N ein multivariater Zählprozeß mit Kompensator Λ bezüglich (\mathcal{F}_t) und sei $M = N - \Lambda$. Sei T eine Stoppzeit von (\mathcal{F}_t) . Die Indikatorfunktion $I(t \leq T)$ ist linksseitig stetig und lokal beschränkt. Es gilt $N^T = \int I(t \leq T) dN(t)$ und $\Lambda^T = \int I(T \leq t) d\Lambda(t)$. Wegen $M^T = \int I(T \leq t) dM(t)$ ist M^T wiederum ein lokales Martingal. N^T ist wiederum ein multivariater Zählprozeß. Da das Integral eines vorhersagbaren Prozesses mit einem anderen vorhersagbaren Prozeß wiederum vorhersagbar ist, ist Λ^T ein vorhersagbarer Prozeß, der wegen der Eindeutigkeit des Kompensators der Kompensator von N^T ist. Wenn Λ eine Intensität λ hat, dann ist $\lambda(t)I(t \leq T)$ der Intensitätsprozeß von N^T .

Nun kann man fortfahren, indem man die Filtrierung reduziert von (\mathcal{F}_t) zu $(\mathcal{F}_{t \wedge T})$. Da sowohl N^T als auch Λ^T an diese kleinere Filtrierung angepaßt

sind, bleibt $\lambda(t)I(T \leq t)$ der Intensitätsprozeß von N^T bezüglich dieser Filtrierung (Erneuerungssatz 1.5.3).

Die Verallgemeinerung dieser Konstruktion ist die *Filterung* eines Zählprozesses: Sei C ein vorhersagbarer reellwertiger Prozeß, der nur die Werte 0 und 1 annimmt. Sei $\tilde{N} = \int C dN$. Dann ist \tilde{N} ein Zählprozeß mit Intensität λC .

Sei N ein multivariater Zählprozeß auf einem Raum $(\Omega, \mathcal{F}, \mathbb{P})$ mit Kompensator Λ und Intensität λ bezüglich einer Filtrierung (\mathcal{F}_t) . Beim *Starten* eines Zählprozesses bezüglich einer (\mathcal{F}_t) -Stoppzeit V soll sich die Intensität nicht verändern. Der bei V gestartete Prozeß ist definiert durch

$${}_V N(t) = N(t) - N(t \wedge V).$$

Er wird auch als der bei V *links-gestutzter* Zählprozeß bezeichnet.

Proposition 1.5.2. *Sei $A \in \mathcal{F}_V$ mit $\mathbb{P}(A) > 0$. Der links-gestutzte Zählprozeß ${}_V N$ hat Intensität*

$${}_V(t)\lambda = \lambda(t)I(V < t < \infty),$$

bezüglich der Filtrierung ${}_V \mathcal{F}_t$, definiert durch

$${}_V \mathcal{F}_t = \mathcal{F}_t \vee \mathcal{F}_V,$$

und der bedingten Wahrscheinlichkeit \mathbb{P}^A , definiert durch

$$\mathbb{P}^A(F) = \mathbb{P}(F \cap A) / \mathbb{P}(A), \quad F \in \mathcal{F}.$$

1.5.6 Stochastische Transformation der Zeit

Sei N ein univariater Zählprozeß mit stetigem Kompensator Λ mit $\Lambda(\tau) = \infty$ fast sicher. Sei Λ^{-1} die rechtsseitig stetige inverse Funktion auf $[0, \infty)$. Dann definiert man den zeittransformierten Zählprozeß durch

$$\tilde{N} = N \circ \Lambda^{-1}. \tag{1.5.6}$$

\tilde{N} hat Intensität 1 auf $[0, \infty)$ und stimmt somit in Verteilung mit dem Poisson-Prozeß überein.

Wenn die Bedingung $\Lambda(\tau) = \infty$ nicht fast sicher erfüllt ist, kann an \tilde{N} ab $\Lambda(\tau)$ noch ein unabhängiger Poisson-Prozeß mit Intensität 1 angehängt werden. \tilde{N} kann dann als Poisson-Prozeß interpretiert werden, der zur zufälligen Zeit $\Lambda(\tau)$ gestoppt wurde. $\Lambda(\tau)$ ist eine Stoppzeit bezüglich der transformierten Filtrierung $(\mathcal{F}_{\Lambda^{-1}(u)})_{u \in \mathbb{R}_0^+}$. Eine offensichtliche Folgerung dieser Konstruktion ist:

Folgerung 1.5.3. *Sei N ein Zählprozeß auf \mathcal{T} mit absolut stetigem und beschränktem Kompensator. Dann existiert $\mathbb{E} N(\tau)^s$ für $s \geq 1$.*

Beweis. Sei L_0 mit $\Lambda(\tau) \leq L_0$ eine obere Schranke von $\Lambda(\tau)$.

Sei $\tilde{N}(s) = N \circ \Lambda^{-1}(s)$ für $s \leq \Lambda(\tau)$. Außerhalb dieses Bereiches ist \tilde{N} durch einen unabhängigen Poisson-Prozeß gegeben. Dann ist \tilde{N} in Verteilung äquivalent zu einem Poisson-Prozeß mit Intensität 1.

$\tilde{N}(L_0)$ ist Poisson-verteilt mit Parameter L_0 . Folglich existiert $\mathbb{E} \tilde{N}(L_0)^s$ für $s \geq 1$. Da $N(\tau)$ meßbar ist und wegen $N(\tau) = \tilde{N}(\Lambda^{-1}(\tau)) \leq \tilde{N}(L_0)$ existiert auch $\mathbb{E} N(\tau)^s$ für $s \geq 1$. \square

1.6 Schwache Konvergenz

Siehe Pollard (1984). Der Skorochod-Raum $D(\mathcal{T})$ (das sind die càdlàg Funktionen), ausgestattet mit der Metrik der gleichmäßigen Konvergenz auf kompakten Mengen, ist nicht separabel. Das war ein Argument für die Einführung der Skorochod-Metrik, die eine schwächere Topologie auf $D(\mathcal{T})$ erzeugt, so daß $D(\mathcal{T})$ separabel ist. Schwache Konvergenz ist die Verallgemeinerung der Konvergenz in Verteilung von \mathbb{R} auf metrische Räume. Deren Definition ist von der Metrik abhängig. In diesem Abschnitt werden wichtige Resultate über schwache Konvergenz behandelt.

Eine \mathcal{E}/\mathcal{A} -meßbare Abbildung X von einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{E}, \mathbb{P})$ in eine Menge \mathcal{X} mit σ -Algebra \mathcal{A} wird als *Zufallselement* bezeichnet. Wenn \mathcal{X} ein metrischer Raum ist, dann soll $\mathcal{C}(\mathcal{X}, \mathcal{A})$ für eine σ -Algebra $\mathcal{A} \subset \mathcal{B}(\mathcal{X})$ die Menge der beschränkten, stetigen $\mathcal{A}/\mathcal{B}(\mathbb{R})$ -meßbaren Abbildungen bezeichnen.⁵ Man definiert: Eine Folge X_n in \mathcal{X} *konvergiert in Verteilung* gegen X , wenn $\mathbb{E} f(X_n) \rightarrow \mathbb{E} f(X)$ für jedes $f \in \mathcal{C}(\mathcal{X}, \mathcal{A})$. Man schreibt $X_n \xrightarrow{\mathcal{D}} X$.

In der benötigten Anwendung ist $\mathcal{X} = \mathcal{D}(\mathcal{T})$, ausgestattet mit der Metrik der gleichmäßigen Konvergenz auf kompakten Mengen. \mathcal{A} wird durch die Menge der abgeschlossenen Kugeln

$$\{X \in D(\mathcal{T}) : d_c(X, X_0) \leq \delta\} = B(X_0, \delta)$$

⁵Für einen metrischen Raum \mathcal{Y} bezeichnet $\mathcal{B}(\mathcal{Y})$ die σ -Algebra der Borelmengen auf \mathcal{Y} , d.h. die durch die offenen Mengen auf \mathcal{Y} erzeugte σ -Algebra.

erzeugt und ist echt kleiner als die Borelmengen $\mathcal{B}(D(\mathcal{T}))$. Die durch die Projektionen $\pi_t : X \rightarrow X(t)$ induzierte σ -Algebra \mathcal{P} auf $D(\mathcal{T})$ stimmt mit der durch die δ -Kugeln erzeugten σ -Algebra überein. ⁶

Ein Punkt X von \mathcal{X} heißt *vollständig regulär* bezüglich der Metrik d und der σ -Algebra \mathcal{A} , wenn es zu jeder Umgebung V von X eine \mathcal{A} -meßbare gleichmäßig stetige Funktion g gibt mit $g(X) = 1$ und $g \leq I_V$.

Die Funktionen in $D(\mathcal{T})$ sind vollständig regulär bezüglich der von den Projektionen erzeugten σ -Algebra. Für ein $\delta > 0$ und $V = B(X_0, \delta)$ hat $g(y) = [1 - \delta^{-1}d(x, y)]^+$ diese Eigenschaft.

Lemma 1.6.1. *Sei h eine beschränkte, \mathcal{A} -meßbare reell-wertige Funktion auf \mathcal{X} . Wenn h an jedem Punkt einer separablen \mathcal{A} -meßbaren Menge C von vollständig regulären Punkten stetig ist, dann folgt aus $X_n \xrightarrow{\mathcal{D}} X$ und $\mathbb{P}(X \in C) = 1$, daß $\mathbb{E} h(X_n) \rightarrow \mathbb{E} h(X)$.*

Wenn $\mathbb{E} f(X_n) \rightarrow \mathbb{E} f(X)$ für jede beschränkte, gleichmäßig stetige, \mathcal{A} -meßbare Funktion f und X fast sicher in einer separablen Menge von vollständig regulären Punkten liegt, dann $X_n \xrightarrow{\mathcal{D}} X$.

Folgerung 1.6.2. *Sei \mathcal{X} ein metrischer Raum mit Metrik d und σ -Algebra \mathcal{A} und \mathcal{Y} ein metrischer Raum mit Metrik e und σ -Algebra \mathcal{B} . Sei $\mathcal{X} \otimes \mathcal{Y}$ ausgestattet mit σ -Algebra $\mathcal{A} \otimes \mathcal{B}$ und Metrik*

$$s[(x, y), (x', y')] = \max[d(x, x'), e(y, y')]$$

⁶Das ist Gegenstand von (Pollard 1984, Problem IV.4). Sei \mathcal{B}_0 die durch die abgeschlossenen Kugeln $B(x, \delta)$ erzeugte σ -Algebra auf $D(\mathcal{T})$. Sei \mathcal{P} die durch die Projektionen π_t von $D(\mathcal{T})$ auf \mathbb{R} induzierte σ -Algebra auf $D(\mathcal{T})$. Zur Vereinfachung des Beweises wird angenommen, daß \mathcal{T} kompakt ist. Dann ist die Metrik d_c durch die Supremumsnorm gegeben. Eine Verallgemeinerung auf nicht kompaktes \mathcal{T} ist möglich.

1. Für $\mathcal{P} \subset \mathcal{B}_0$ ist zu zeigen, daß jede Projektion π_t $\mathcal{B}_0/\mathcal{B}$ -meßbar ist. Dafür genügt, daß $\pi_t^{-1}((\alpha, \infty)) \in \mathcal{B}_0$. Es ist aber $\pi_t^{-1}((\alpha, \infty)) = \bigcup_{n=1}^{\infty} B(x_n, n)$ mit $x_n(t) = (n + n^{-1})I_{[t, t+n^{-1})} + \alpha$. Das gilt, weil für ein $x \in D(\mathcal{T})$ mit $x(t) > \alpha$ ein n_0 existiert, sodaß wegen der Rechts-Stetigkeit $x([t, t + n^{-1})) \subseteq (\alpha + n_0^{-1}, \alpha + 2n_0 + n_0^{-1})$ und $d_c(x, x_{n_0}) < n_0$. Daher ist $x \in B(x_{n_0}, n_0)$.
2. Umgekehrt ist jedes $B(x, r)$ in \mathcal{P} , denn $B(x, r) = \bigcap_{t \in \mathbb{Q} \cap \mathcal{T}} \{z : |\pi_t x - \pi_t z| \leq r\} = B_0$. Zum Beweis dieser Behauptung bemerkt man, daß die Bedingung $|\pi_t x - \pi_t z| \leq r$ auf rationalen Zahlen für jedes $z \in B(x, r)$ erfüllt ist. Angenommen es gibt ein $z \in B_0$, sodaß $|z(s) - x(s)| > r$ für ein $s \in \mathcal{T} \setminus \mathbb{Q}$. Dann gibt es ein $\delta > 0$, sodaß $\sup_{0 \leq u-s < \delta} |z(u) - x(u)| > r$. Das Intervall $[s, s + \delta]$ enthält jedoch mindestens eine rationale Zahl, was in Widerspruch zur Konstruktion von B_0 steht. \square

Sei $X_n \xrightarrow{\mathcal{D}} X$ in \mathcal{X} . Wenn X fast sicher in einer separablen Menge von vollständig regulären Punkten liegt und $Y_n \xrightarrow{\mathcal{P}} y_0$ in \mathcal{Y} für einen vollständig regulären Punkt y_0 , dann $(X_n, Y_n) \xrightarrow{\mathcal{D}} (X, y_0)$ im Produktraum $\mathcal{X} \otimes \mathcal{Y}$.

Satz 1.6.3. (Satz von der stetigen Abbildung) Sei H eine \mathcal{A}/\mathcal{A}' -meßbare Abbildung von \mathcal{X} in einen metrischen Raum X' . Wenn H stetig ist auf einer separablen, \mathcal{A} -meßbaren Menge C von vollständig regulären Punkten, dann folgt aus $X_n \xrightarrow{\mathcal{D}} X$ und $X \in C$ fast sicher, daß $HX_n \xrightarrow{\mathcal{D}} HX$.

Bemerkung: Das in Lehrbüchern oft zitierte Lemma von Slutsky ist die Anwendung von Folgerung 1.6.2 und Satz 1.6.3 auf die stetigen Funktionen Multiplikation, Addition und Division.

Satz 1.6.4. (Darstellungssatz) Sei $\{P_n\}$ eine Folge von Wahrscheinlichkeitsmaßen auf einem metrischen Raum. Wenn P_n schwach gegen P konvergiert und P gleich 0 ist außerhalb einer separablen Menge von regulären Punkten, dann existieren Zufallselemente $\{X_n\}$ und X mit Verteilungen $\{P_n\}$ und P , sodaß X_n fast sicher gegen X konvergiert.

$C(\mathcal{T})$, der Raum der stetigen Funktionen auf \mathcal{T} , ist ein separabler, abgeschlossener, meßbarer Teilraum von $D(\mathcal{T})$. Auf $C(\mathcal{T})$ stimmt die Borel- σ -Algebra mit der durch die Projektionen induzierten σ -Algebra überein.

1.6.1 Straffheit

In der Analysis ist Kompaktheit ein wertvolles Hilfsmittel, um Konvergenzkriterien zu erhalten. Für schwache Konvergenz wird der verwandte Begriff der Straffheit verwendet. Die folgende Definition stammt von Pollard (1984, p. 81)

Definition 1.6.5. Sei \mathcal{E} ein metrischer Raum. Sei \mathcal{A} eine Teil- σ -Algebra der Borel- σ -Algebra $\mathcal{B}(\mathcal{E})$. Ein Wahrscheinlichkeitsmaß P ist straff, wenn für jedes $\epsilon > 0$ eine kompakte Menge K_ϵ von vollständig regulären Punkten existiert, sodaß $P(K_\epsilon) > 1 - \epsilon$.

Eine Folge (P_n) von Wahrscheinlichkeitsmaßen auf $(\mathcal{X}, \mathcal{A})$ ist gleichmäßig straff, wenn für jedes $\epsilon > 0$ eine kompakte Menge K_ϵ von vollständig regulären Punkten existiert, so daß $\liminf P_n(G) > 1 - \epsilon$ für jede \mathcal{A} -meßbare offene Menge G , die K_ϵ enthält.

Ein Zufallselement X von $(\mathcal{X}, \mathcal{A})$ ist straff, wenn das induzierte Maß P_X , definiert durch $P_X(A) = P(X \in A)$ für $A \in \mathcal{A}$, straff ist. Eine Folge (X_n)

von Zufallselementen von $(\mathcal{X}, \mathcal{A})$ ist gleichmäßig straff, wenn die induzierten Wahrscheinlichkeitsmaße, die durch $P_n(A) = P(X_n \in A)$ für $A \in \mathcal{A}$ definiert sind, gleichmäßig straff sind.

Man kann zeigen, daß jede kompakte Teilmenge von vollständig regulären Punkten ein abzählbarer Durchschnitt von offenen, \mathcal{A} -meßbaren Mengen ist. Daher wird der erste Teil der Definition, die \mathcal{A} -Meßbarkeit von K_ϵ , nicht benötigt.

Folgerung 1.6.6. Sei \mathcal{E} ein metrischer Raum. Sei \mathcal{A} eine Teil- σ -Algebra der Borel- σ -Algebra $\mathcal{B}(\mathcal{E})$.

Sei $X \in \mathcal{X}$ straff und $P(X \in C) = 1$ für eine separable Menge C von vollständig regulären Punkten. Sei (X_n) eine Folge in \mathcal{X} , die schwach gegen ein straffes Zufallselement X konvergiert. Dann ist (X_n) gleichmäßig straff.

Beweis. Sei $\epsilon > 0$ beliebig gewählt. Dann gibt es eine kompakte Menge K_ϵ von vollständig regulären Punkten, so daß $P(X \in K_\epsilon) > 1 - \epsilon$. Sei G eine offene Menge, die K_ϵ enthält.

Wegen Satz 1.6.4 gibt es eine Folge (\tilde{X}_n) von Zufallselementen, die die selben Wahrscheinlichkeiten wie (X_n) induzieren, und die fast sicher gegen X konvergieren. Wenn $X(\omega) \in G$, dann ist für genügend großes n $\tilde{X}_n(\omega) \in G$, das heißt

$$\liminf I \left\{ \tilde{X}_n \in G \right\} \geq I \{X \in G\} \quad \text{fast sicher}$$

Aufgrund des Lemmas von Fatou gilt

$$\begin{aligned} \liminf P_n(G) &= \liminf P(X_n \in G) \\ &\geq P(\liminf \{X_n \in G\}) \\ &\geq P(X \in G) \\ &= P(G) \end{aligned}$$

□

1.6.2 Empirische Prozesse

Sei X_{n1}, \dots, X_{nn} eine Dreieckschema von zeilenweise unabhängigen Zufallsvariablen mit Verteilungsfunktionen F_{n1}, \dots, F_{nn} . Die empirische Verteilungsfunktion ist

$$\mathbb{F}_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{\{X_{ni} \leq x\}} \quad \text{für } x \in \mathbb{R}. \quad (1.6.1)$$

Die *mittlere Verteilungsfunktion* ist

$$\bar{F}_n(x) = \frac{1}{n} \sum_{i=1}^n F_{ni}(x) \quad \text{für } x \in \mathbb{R}. \quad (1.6.2)$$

Der *empirische Prozeß* ist definiert durch

$$\mathbb{E}_n(x) = n^{-1/2} \sum_{i=1}^n \mathbb{F}_n(x) - \bar{F}_n(x) \quad \text{für } x \in \mathbb{R}. \quad (1.6.3)$$

Die folgende Verallgemeinerung des starken Gesetzes der großen Zahl für unabhängige und gleichverteilte Zufallsvariable von Glivenko-Cantelli findet man in (Shorack und Wellner 1986, p. 106):

Satz 1.6.7. *Die empirische Verteilungsfunktion konvergiert fast sicher gegen die mittlere Verteilungsfunktion mit einer Geschwindigkeit, die nicht von den Verteilungsfunktionen abhängt.*

Der folgende Satz ist ein Spezialfall von (Shorack und Wellner 1986, p. 109):

Satz 1.6.8. *Sei X_{n1}, \dots, X_{nm} eine Dreieckschema von zeilenweise unabhängigen Zufallsvariablen mit Verteilungsfunktionen F_{n1}, \dots, F_{nm} auf $[0, 1]$. Sei*

$$\limsup_{n \rightarrow \infty} \max_{1 \leq k \leq m} \bar{F}_n\left(\frac{k}{m}\right) - \bar{F}_n\left(\frac{k-1}{m}\right) \rightarrow 0 \quad \text{für } m \rightarrow \infty. \quad (1.6.4)$$

Dann ist \mathbb{E}_n schwach kompakt auf dem Skorohod-Raum mit der Metrik der gleichmäßigen Konvergenz.

$$\mathbb{E}_n \xrightarrow{\mathcal{D}} \mathbb{E} \quad \text{einen Prozeß } \mathbb{E} \quad (1.6.5)$$

im selben Raum genau dann, wenn eine Funktion $K(s, t)$ existiert, so daß

$$K_n(s, t) = \text{Cov}(\mathbb{E}_n(s), \mathbb{E}_n(t)) \rightarrow K(s, t) \quad (1.6.6)$$

für $n \rightarrow \infty$ und $0 \leq s, t \leq 1$. In diesem Fall sind die endlichdimensionalen Projektionen von \mathbb{E} normalverteilt mit Kovarianzfunktion K und die Pfade von \mathbb{E} sind fast sicher stetig.

Das Ergebnis von Satz 1.6.8 kann durch gewisse Transformationen auf andere Bereiche als $[0, 1]$ erweitert werden. Dazu definiert man die *rechtsseitig stetige inverse Verteilungsfunktion* durch

$$F_+^{-1}(t) = \sup_x \{F(x) \leq t\}. \quad (1.6.7)$$

Diese Funktion hat die Eigenschaft, daß

$$F_+^{-1} \circ F(x) = \inf_{y \in [-\infty, x]} \{y : F(y) = F(x)\} \quad (1.6.8)$$

(Shorack und Wellner 1986, p. 8) wieder eine Verteilungsfunktion ist.

Satz 1.6.9. *Sei \mathbb{E}_n der zu den zum Dreieckschema X_{n1}, \dots, X_{nn} von zeilenweise unabhängigen Zufallsvariablen mit Verteilungsfunktion F_{n1}, \dots, F_{nn} gehörige empirische Prozeß. Sei F eine Verteilungsfunktion, die nur in denjenigen Abschnitten nicht monoton ist, in denen auch alle F_{ni} nicht monoton sind. Seien Zufallsvariable $\xi_{n1}, \dots, \xi_{nn}$ mit Verteilungsfunktionen $F_{ni} \circ F_+^{-1}$ und empirischem Prozeß \mathbb{U}_n gegeben.*

Wenn der empirische Prozeß \mathbb{U}_n in Verteilung gegen den Prozeß \mathbb{U} konvergiert, dann konvergiert \mathbb{E}_n gegen $\mathbb{U} \circ F$.

Beweis. $H(x) = x \circ F$ ist für $x \in D[0, 1]$ eine meßbare Abbildung nach $D[-\infty, \infty]$ (Die σ -Algebren werden jeweils durch die Koordinatenprojektionen erzeugt.). Darüberhinaus ist diese Abbildung stetig bezüglich gleichmäßiger Konvergenz, denn $|H(x) - H(y)| = \sup_r |x(F(r)) - y(F(r))| \leq |x - y|$. Nach dem Satz von der stetigen Abbildung 1.6.3 konvergiert daher $H(\mathbb{U}_n)$ in Verteilung gegen $H(\mathbb{U})$.

Es gilt für jeden Index in :

$$\begin{aligned} & \mathbb{P}(\xi_n \circ F \leq x) \\ &= \mathbb{P}(\xi_n \leq F(x)) \\ &= F_n \circ F_+^{-1} \circ F(x) \\ &= F_n \left(\inf_{y \in [-\infty, x]} \{y : F(y) = F(x)\} \right) \\ &= F_n(x) \end{aligned}$$

Daher hat $\xi_{in} \circ F$ die selben Wahrscheinlichkeiten wie X_{in} .

Die Verteilung von $H(\mathbb{U}_n)$ stimmt daher mit der Verteilung von \mathbb{E}_n überein. Da die Verteilungen der endlichdimensionalen Projektionen von \mathbb{E}_n mit denen von $H(\mathbb{U}_n)$ übereinstimmen, konvergiert \mathbb{E}_n gegen $H(\mathbb{U}) = \mathbb{U} \circ F$. \square

In den Anwendungen treten oft Familien von Verteilungsfunktionen auf, durch zufällige Parameter erzeugt werden. In solchen Fällen liefert der folgende Satz von Daffer und Taylor (1979) ein Kriterium für (1.6.4):

Satz 1.6.10. *Sei (X_n) eine Folge von unabhängigen Zufallselementen von $D([0, 1])$ mit*

1. X_n ist fast sicher monoton wachsend für alle n ;
2. $\sum_{n=1}^{\infty} \frac{\mathbb{E}|X_n|_{\infty}^r}{n^r} < \infty$ für ein $1 \leq r \leq 2$;
3. $\mathbb{E} X_n = \mathbb{E} X_1$ für alle n .

Dann ist

$$\lim_{n \rightarrow \infty} \left| n^{-1} \sum_{i=1}^n X_i - \mathbb{E} X_1 \right|_{\infty} = 0 \quad \text{fast sicher.}$$

1.7 Grenzwertsätze und Ungleichungen

1.7.1 Der zentrale Grenzwertsatz

Satz 1.7.1. Sei (X_n) eine Folge von unabhängigen p -dimensionalen Zufallsvariablen mit Verteilungsfunktionen $F_n(x)$. Sei $\mathbb{E} X_n = 0$ und es existiere $\text{Var} X_n = \Sigma_n$. Für $n \rightarrow \infty$ konvergiere $1/n \sum_{i=1}^n \Sigma_i$ gegen $\Sigma \neq 0$ und für alle $\epsilon > 0$ gelte

$$\frac{1}{n} \sum_{i=1}^n \int_{|x| > \epsilon \sqrt{n}} |x|^2 dF_i \rightarrow 0. \quad (1.7.1)$$

wobei $|\cdot|$ die euklidische Norm bezeichnet.

Dann konvergiert $n^{-1/2} \sum_{i=1}^n X_i$ gegen die p -variate Normalverteilung mit Erwartungsvektor 0 und Kovarianzmatrix Σ .

1.7.2 Lenglart'sche Ungleichung

Sei \tilde{X} der nichtfallende Kompensator eines lokalen Submartingals X . Es gilt die nachstehende Ungleichung, die auch im Beweis von Rebolledo's Grenzwertsatz verwendet wird für $\eta > 0$ und $\delta > 0$ (Lenglart 1977):

$$\mathbb{P}(\sup_{\mathcal{T}} X > \eta) \leq \frac{\delta}{\eta} + \mathbb{P}(\tilde{X}(\tau) > \delta) \quad \text{für beliebige } \epsilon > 0, \delta > 0. \quad (1.7.2)$$

Im Spezialfall, daß X das Quadrat eines lokal quadratisch integrierbaren Martingals M ist, und somit $\tilde{X} = \langle M \rangle$ gilt, erhält man aus (1.7.2)

$$\mathbb{P}(\sup_{\mathcal{T}} |M| > \eta) \leq \frac{\delta}{\eta^2} + \mathbb{P}(\langle M \rangle(\tau) > \delta). \quad (1.7.3)$$

1.7.3 Funktionen, Reihen und Konvergenz in Wahrscheinlichkeit

Satz II.1 aus Andersen und Gill (1982, p1116) lautet:

Satz 1.7.2. *Sei E eine offene, konvexe Teilmenge des \mathbb{R}^p und sei (F_i) eine Folge von zufälligen fast sicher konkaven Funktionen auf E derart, daß für alle $x \in E$: $F_i(x) \xrightarrow{P} f(x)$, wobei f eine reelle Funktion auf E sein soll. Dann ist auch f konkav und für jede kompakte Menge $A \subset E$ gilt*

$$|F_n(\cdot) - f(\cdot)|_A \xrightarrow{P} 0. \quad (1.7.4)$$

Der folgende Satz wird in (Andersen und Gill 1982, p1116) ohne Beweis angegeben:

Satz 1.7.3. *Sei E eine offene Menge in \mathbb{R}^p und (F_i) eine Folge von zufälligen fast sicher konkaven Funktionen auf E , die auf jeder kompakten Teilmenge von E gleichmäßig in Wahrscheinlichkeit gegen eine (konkave) Funktion f konvergieren. Wenn f ein eindeutiges Maximum bei $\hat{x} \in E$ hat und \hat{X}_i die Funktion F_i maximiert, dann gilt $\hat{X}_i \xrightarrow{P} \hat{x}$.*

Beweis. Zunächst stellt man fest, daß die Wahrscheinlichkeit, daß \hat{X}_i in einer kompakten Menge liegt, gegen eins konvergiert, mit folgender Begründung:

Da f konkav ist, gibt es eine kompakte Menge K und ein $c > 0$ mit der Eigenschaft, daß für alle $x \notin K$: $f(x) < f(\hat{x}) - c$. Nun gilt wegen $F_i(\hat{X}_i) \geq F_i(\hat{x})$ die Abschätzung:

$$\mathbb{P}\left(F_i(\hat{X}_i) < f(\hat{x}) - c\right) \leq \mathbb{P}\left(F_i(\hat{x}) < f(\hat{x}) - c\right) \rightarrow 0 \quad \text{für } i \rightarrow \infty.$$

Sei $U_n = B(\hat{x}, n^{-1}) \subseteq E$ für n genügend groß. Sei $\epsilon_n = f(\hat{x}) - \max f(\partial U_n)$. Dann ist (wegen der Eindeutigkeit von \hat{x}) $\epsilon_n > 0$ und $\lim_{n \rightarrow \infty} \epsilon_n = 0$. Es gilt

$$\begin{aligned} & \mathbb{P}\left(|\hat{x} - \hat{X}_k| < 1/n\right) \\ & \geq \mathbb{P}(\hat{X}_k \in U_n) \\ & \geq \mathbb{P}\left(f(\hat{x}) - f(\hat{X}_k) < \epsilon_n\right) \end{aligned}$$

und

$$\begin{aligned} f(\hat{x}) - f(\hat{X}_k) &= f(\hat{x}) - F_k(\hat{x}) + \underbrace{F_k(\hat{x}) - F_k(\hat{X}_k)}_{< 0} + F_k(\hat{X}_k) - f(\hat{X}_k) \\ &\leq \left|f(\hat{x}) - F_k(\hat{x})\right| + \left|F_k(\hat{X}_k) - f(\hat{X}_k)\right|. \end{aligned}$$

Die letzten beiden Summanden konvergieren für $\hat{X}_k \in K$ wegen der gleichmäßigen Konvergenz auf kompakten Mengen von f_n in Wahrscheinlichkeit gegen 0, so daß $\mathbb{P}(f(\hat{x}) - f(\hat{X}_k) < \epsilon_n)$ gegen 1 konvergiert. Aber auch $\mathbb{P}(\hat{X}_k \notin K)$ konvergiert gegen 0. \square

Lemma 1.7.4. (*Varianzkriterium für Reihen von Khinchin und Kolmogorov*) Seien ξ_1, ξ_2, \dots unabhängige Zufallsvariable mit Mittelwert 0 und $\sum_{n=1}^{\infty} \mathbb{E} \xi_n^2 < \infty$. Dann konvergiert $\sum_{n=1}^{\infty} \xi_n$ fast sicher.

Beweis. Siehe (Kallenberg 1997, p. 47). \square

1.8 Das Produkt-Integral

Das Produkt-Integral \prod ist eine Verallgemeinerung des gewöhnlichen Produktes \prod , so wie das Integral \int die Summe \sum verallgemeinert. Mit Hilfe des Produkt-Integrals kann die beobachtete Likelihood-Funktion eines Prozesses als unendliches Produkt von bedingten Likelihoods für die Entwicklung des Prozesses in infinitesimalen Zeitintervallen bezüglich der vorhergehenden Geschichte dargestellt werden. Die gleiche Konstruktion kommt in der Beziehung zwischen Überlebensfunktion und Hasardrate vor. Bei der Analyse von Ereignis-Geschichten mit mehreren Zuständen wird daraus die Beziehung zwischen der Matrix der Übertragungsraten und der Matrix der Übergangswahrscheinlichkeiten eines Markov-Prozesses in kontinuierlicher Zeit.

Statistische Modelle werden oft in kontinuierlicher Zeit spezifiziert, die natürlichen Schätzer sind hingegen oft diskret. Daher ist es notwendig, daß der Kalkül diskrete und kontinuierliche Versionen einheitlich behandeln kann. Der Kalkül wird in (Gill und Johansen 1990) behandelt.

Sei $X(t)$, $t \in \mathcal{T}$, eine $p \times p$ -Matrix von cadlag Funktionen von lokal beschränkter Variation. Man definiert

$$Y = \prod (I + dX), \quad (1.8.1)$$

das *Produkt-Integral von X* über die Intervalle der Form $[0, t]$, $t \in \mathcal{T}$ als folgende $p \times p$ -matrixwertige Funktion

$$Y(t) = \prod_{s \in [0, t]} (I + dX(s)) = \lim_{\max |t_i - t_{i-1}| \rightarrow 0} \prod (I + X(t_i) - X(t_{i-1})), \quad (1.8.2)$$

wobei $0 = t_0 < t_1 < \dots < t_n = t$ eine Partition von $[0, t]$ ist und das Matrizenprodukt in der natürlichen Ordnung von links nach rechts ausgeführt wird. Im äußerst linken Ausdruck des Produktes muß $X(0)$ durch $X(0-) = 0$ ersetzt werden, da der linke Endpunkt 0 im Intervall $[0, t]$ enthalten ist.

Das Produkt-Integral über ein Teilintervall von \mathcal{T} ist unter natürlicher Behandlung der Endpunkte analog definiert. Produkt-Integrale über die disjunkte Vereinigung von Intervallen sind multiplikativ. Wenn X eine Treppenfunktion ist, dann ist das Produkt-Integral nur das endliche Produkt der Sprungzeiten von X der Identitätsmatrix plus der Sprunghöhen von X . Für *skalare* Funktionen X muß die Reihenfolge der Multiplikationen nicht berücksichtigt werden und man kann den stetigen Anteil X^c und die Sprünge

ΔX getrennt behandeln. Man erhält

$$\prod (1 + dX) = \exp(X^c) \prod (1 + \Delta X) \quad (1.8.3)$$

mit $\Delta X = X - X_-$.

Für stetige Funktionen ist das Produktintegral daher die Exponentialfunktion.

Das Produkt-Integral geht auf Volterra (1887) zurück, der es im Zusammenhang mit einer bestimmten Integralgleichung einführte.

Satz 1.8.1. $\prod (I + dX)$ existiert und ist komponentenweise eine cadlag Funktion von lokal beschränkter Variation. Es ist die eindeutige Lösung der Integralgleichung

$$Y(t) = I + \int_{s \in [0, t]} Y(s-) dX(s) \quad (1.8.4)$$

Die Beziehung zwischen der Überlebensfunktion und Hasard-Maßen kann durch Produkt-Integrale folgendermaßen hergestellt werden:

Satz 1.8.2. Sei S die Überlebensfunktion einer positiven Zufallsvariablen (der Überlebenszeit) T , d.h. $S(t) = P(T > t)$ für alle $t \geq 0$ und $S(\tau) = 0$. Die kumulative Hasard-Funktion sei definiert durch

$$A(t) = - \int_0^t \frac{dS(s)}{S(s-)} \quad (1.8.5)$$

Dann gilt für alle t mit $A(t) < \infty$

$$S(t) = \prod_{[0, t]} (1 - dA). \quad (1.8.6)$$

Wenn S absolut stetig ist, sodaß die Verteilungsfunktion $F = 1 - S$ die Dichte f hat, dann ist die Hasardrate als $\alpha = f/(1 - F)$ definiert. Die kumulative Hasard-Funktion ist dann $A(t) = \int_0^t \alpha(s) ds$ und die Gleichung (1.8.6) wird wegen (1.8.3) zu

$$S(t) = \prod_0^t (1 - \alpha(s) ds) = \exp \left(- \int_0^t \alpha(s) ds \right). \quad (1.8.7)$$

Wenn S diskret ist, kann eine zur Hasardrate analoge Funktion definiert werden und das Produktintegral (1.8.6) wird zu einem endlichen Produkt.

Die kumulative Hasardfunktion wurde schon auf Seite 9 beim Einsprung-Prozeß behandelt, wo sie Bestandteil des Kompensators war.

1.8.1 Markov-Prozesse

Zählprozesse und Produkt-Integrale können für die Modellierung von kontinuierlichen Markov-Prozessen mit endlichem Zustandsraum verwendet werden. Eine $p \times p$ -matrixwertige Funktion A ist ein lokal endliches *Intensitätsmaß eines Markov-Prozesses* auf dem Zeitintervall \mathcal{T} , wenn die Komponenten A_{hj} , $h \neq j$ nichtfallende cadlag-Funktionen sind, die zur Zeit 0 gleich 0 sind, $A_{hh} = -\sum_{j \neq h} A_{hj}$ und $\Delta A_{hh}(t) \geq -1$ für alle t . A_{hj} heißt *integrierte Intensitätsfunktion* des Übergangs von Zustand h in Zustand j . A_{hh} ist die *negative integrierte Intensitätsfunktion* der Übergänge aus Zustand h heraus. Der folgende Satz stellt dann den Zusammenhang zum Produktintegral her:

Satz 1.8.3. *Sei A ein Intensitätsmaß. Definiere*

$$P(s, t) = \prod_{(s, t]} (I + dA), \quad s \leq t; \quad s, t \in \mathcal{T}. \quad (1.8.8)$$

Dann ist P die Übergangsmatrix eines Markov-Prozesses mit Zustandsraum $(1, \dots, p)$ und Intensitätsmaß A . Der Prozeß kann folgendermaßen rekonstruiert werden: Wenn der Prozeß zur Zeit t_0 in Zustand h ist, dann bleibt er in diesem Zustand für eine Dauer mit integrierter Hasardfunktion

$$-(A_{hh}(t) - A_{hh}(t_0)), \quad t_0 \leq t \leq \inf(u \geq t_0 : \Delta A_{hh}(u) = -1).$$

Wenn der Prozeß zur Zeit t aus Zustand h heraus springt, dann springt er in Zustand $j \neq h$ mit Wahrscheinlichkeit $(dA_{hj}/(-dA_{hh}))(t)$.

Die Markov-Prozesse, deren Übergangs-Matrizen nicht durch ein Produktintegral (1.8.8) dargestellt werden können, haben für einen Zustand h und Zeiten u und v , wenn der Prozeß zur Zeit u in Zustand h ist, bis v eine unendliche kumulative Hasardrate.

Der folgende Satz stellt den Zusammenhang zu Zählprozessen her:

Satz 1.8.4. *Sei A das Intensitätsmaß eines Markov-Prozesses X . Sei $\mathcal{F}_t = \sigma(X(s) : s \leq t)$; definiere*

$$\begin{aligned} Y_h(t) &= I(X(t-) = h), \\ N_{hj}(t) &= \#(s \leq t : X(s-) = h, X(s) = j), \quad h \neq j. \end{aligned}$$

Dann ist $N = (N_{hj}, h \neq j)$ ein multivariater Zählprozeß mit Komponenten des Kompensators bezüglich $(\mathcal{F}_t) = (\sigma(X(0)) \vee \mathcal{N}_t)$

$$\Lambda_{hj}(t) = \int_0^t Y_h(s) dA_{hj}(s). \quad (1.8.9)$$

Die Intensität des Einsprung-Prozesses **1.5.1** ist ein Spezialfall dieses Satzes, wenn T der einzige Übergang von 0 nach 1 eines Markov-Prozesses mit den Stadien 0 und 1 ist, und kein Übergang von 1 nach 0 möglich ist.

1.9 Likelihoods für Zählprozesse

Gegeben sei eine multivariater Zählprozeß $N = (N_1, \dots, N_k)$. Er soll unter verschiedenen Wahrscheinlichkeitsmaßen und daher auch Intensitätsprozessen betrachtet werden. Folgende Annahmen sollen gelten:

$$\begin{aligned} \mathcal{F}_t &= \mathcal{F}_0 \vee \sigma(N(s) : s \leq t) = \mathcal{F}_0 \vee \mathcal{N}_t, \\ \mathcal{T} &= \bar{\mathcal{T}} = [0, \tau], \quad \mathcal{F} = \mathcal{F}_\tau. \end{aligned} \quad (1.9.1)$$

Schon beim Einsprung-Prozeß (siehe Seite 9) wurde ein Zusammenhang zwischen einem Wahrscheinlichkeitsmaß (festgelegt durch die Verteilungsfunktion F) und einem Intensitätsprozeß hergestellt. Diese Konstruktion könnte auf allgemeine Zählprozesse erweitert werden: Zuerst wird \mathbb{P} auf \mathcal{F}_0 angegeben. Dann gibt man die bedingte Verteilung des ersten Sprunges $T_1 \in (0, \tau]$ von N bezüglich \mathcal{F}_0 an. Sodann gibt man die bedingte Verteilung des Typs $J_1 \in \{0, 1, \dots, k\}$ des ersten Sprunges bezüglich \mathcal{F}_0 und T_1 an. Der Typ 0 ist dabei nur dann zulässig, wenn $T_1 = \tau$. Dann fährt man fort mit dem zweiten Sprung $T_2 \in (T_1, \tau]$ bezüglich $\mathcal{F}_0 \vee \sigma\{T_1, J_1\}$. Die Konstruktion schlägt dann fehl, wenn es mit positiver Wahrscheinlichkeit unendlich viele Sprünge in $(0, \tau]$ gibt. Die schwächsten Bedingungen, unter denen es nur endlich viele Sprünge gibt, werden im Satz von Girsanov (Andersen et al. 1993, p. 100) gegeben. Allerdings sind diese Bedingungen kaum nachprüfbar.

Die bedingten Verteilungen können auch durch bedingte kumulative Hasard-Funktionen ersetzt werden. Diese können zu typ-spezifischen Hasard-Funktionen für (T_n, J_n) gemacht werden, indem man sie mit der Wahrscheinlichkeit des Typs gegeben die vorhergehenden Sprünge, multipliziert. Typ-spezifische Hasard-Funktionen könnten aber den Intensitätsprozeß von N festlegen (siehe **1.5.1**). Diese Konstruktion ist möglich und wird im folgenden Satz zusammengefaßt:

Satz 1.9.1. (*Jacod'sche Formel für den Intensitätsprozeß*).

Sei (1.9.1) erfüllt. Sei \mathbb{P}_n eine reguläre Version der gemeinsamen bedingten Verteilung von (T_n, J_n) bezüglich

$$\mathcal{F}_{T_{n-1}} = \mathcal{F}_0 \vee \sigma\{T_1, J_1, \dots, T_{n-1}, J_{n-1}\}.$$

Dann gilt auf dem Intervall $(T_{n-1}, T_n]$

$${}_{T_{n-1}}\Lambda_h(t) = \int_{T_{n-1}}^t \frac{P_n(ds, \{h\})}{P_n([s, \tau], \bar{E})}, \quad h = 1, \dots, k,$$

mit $E = \{1, \dots, k\}$ und $\bar{E} = E \cup \{0\}$. Umgekehrt gilt auf (T_{n-1}, T_n)

$$P_n((t, \tau], \bar{E}) = \prod_{T_{n-1}}^t (1 - d\Lambda(s))$$

(\cdot bedeutet hier die Summe über $1, \dots, k$) und für $P_n(h|t)$, die P_n -bedingte Verteilung von J_n gegeben $T_n = t$, gilt

$$P_n(h|t) = \frac{d\Lambda_h(t)}{d\Lambda(t)}, \quad h = 1, \dots, k.$$

Durch Ersetzung der diskreten Marken h durch ein Element dx eines beliebigen Meßraums (E, \mathcal{E}) von Marken kann dieses Ergebnis verallgemeinert werden.

Satz 1.9.2. (Jacod-Formel für die Likelihood-Ratio). Es gelte (1.9.1). Seien P und \tilde{P} zwei Wahrscheinlichkeitsmaße, sodaß die Kompensatoren von N die Prozesse Λ beziehungsweise $\tilde{\Lambda}$ sind. Weiters sei \tilde{P} absolut stetig bezüglich P , was als $\tilde{P} \ll P$ geschrieben wird. Dann gilt

$$\tilde{\Lambda}_h \ll \Lambda_h \text{ für alle } h, \text{ } P\text{-fast-sicher,}$$

$$\Delta\Lambda(t) = 1 \text{ für ein } t \text{ impliziert } \Delta\tilde{\Lambda}(t) = 1, \text{ } P\text{-fast-sicher}$$

und

$$\frac{d\tilde{P}}{dP} = \frac{d\tilde{P}}{dP} \bigg|_{\mathcal{F}_0} \frac{\prod_{t \in [0, \tau]} \left(\prod_h d\tilde{\Lambda}_h(t)^{\Delta N_h(t)} \left(1 - d\tilde{\Lambda}(t) \right)^{1 - \Delta N(t)} \right)}{\prod_{t \in [0, \tau]} \left(\prod_h d\Lambda_h(t)^{\Delta N_h(t)} \left(1 - d\Lambda(t) \right)^{1 - \Delta N(t)} \right)} \quad (1.9.2)$$

$$= \frac{d\tilde{P}}{dP} \bigg|_{\mathcal{F}_0} \prod_t \prod_h \left(\frac{d\tilde{\Lambda}_h(t)}{d\Lambda_h(t)} \right)^{\Delta N_h(t)} \frac{\prod_{t \in [0, \tau]: \Delta N(t) \neq 1} \left(1 - d\tilde{\Lambda}(t) \right)}{\prod_{t \in [0, \tau]: \Delta N(t) \neq 1} \left(1 - d\Lambda(t) \right)} \quad (1.9.3)$$

Dabei ist $\Lambda = \sum_h \Lambda_h$. (1.9.2) bedeutet das endliche Produkt über den ersten Faktor mal das Produktintegral über den kontinuierlichen Anteil des zweiten Faktors. Abgesehen von der vereinfachten Schreibweise ist auch eine Ähnlichkeit mit einer multinomialen Likelihood von „infinitesimalen“ Experimenten beabsichtigt. Satz 1.9.2 gilt auch für markierte Punkt-Prozesse.

Folgerung 1.9.3. (*Stetiger und absolut stetiger Fall*). Wenn Λ und $\tilde{\Lambda}$ P -fast-sicher stetig sind, dann

$$\frac{d\tilde{P}}{dP} = \frac{d\tilde{P}}{dP} \Bigg|_{\mathcal{F}_0} \frac{\prod_{t,h} d\tilde{\Lambda}_h(t)^{\Delta N_h(t)} \exp(-\tilde{\Lambda}(\tau))}{\prod_{t,h} d\Lambda_h(t)^{\Delta N_h(t)} \exp(-\Lambda(\tau))} \quad (1.9.4)$$

und wenn diese sogar absolut stetig sind, dann

$$\frac{d\tilde{P}}{dP} = \frac{d\tilde{P}}{dP} \Bigg|_{\mathcal{F}_0} \frac{\prod_{t,h} \tilde{\lambda}_h(t)^{\Delta N_h(t)} \exp(-\tilde{\Lambda}(\tau))}{\prod_{t,h} \lambda_h(t)^{\Delta N_h(t)} \exp(-\Lambda(\tau))}. \quad (1.9.5)$$

In (1.9.5) wurde $d\Lambda_h(t) = \lambda_h(t) dt$ verwendet und schließlich dt weggelassen (gekürzt). Zum Zweck der statistischen Schätzung werden Likelihood Ratios als Radon-Nikodym-Ableitungen von Elementen einer Familie von Wahrscheinlichkeitsmaßen bezüglich einer festen Referenz-Verteilung gebildet. Likelihoods müssen nur bis auf einen Proportionalitätsfaktor angegeben werden. Daher wird in (1.9.2) bis (1.9.5) nur der Zähler benötigt. Die Ergebnisse von Folgerung 1.9.3 kann man daher ohne Tilden auch anschreiben als

$$dP = dP|_{\mathcal{F}_0} \prod_{t,h} d\Lambda_h(t)^{\Delta N_h(t)} \exp(-\Lambda(\tau)) \quad (1.9.6)$$

beziehungsweise (wobei dt weggelassen wird)

$$\begin{aligned} dP &= dP|_{\mathcal{F}_0} \prod_{t,h} \lambda_h(t)^{\Delta N_h(t)} \prod_0^\tau (1 - d\Lambda.) \\ dP &= dP|_{\mathcal{F}_0} \prod_{t,h} \lambda_h(t)^{\Delta N_h(t)} \exp(-\Lambda(\tau)) \end{aligned} \quad (1.9.7)$$

Die nachstehende Notation suggeriert noch stärker ein Produkt von infinitesimalen multinomialen Experimenten als (1.9.6)

$$dP = dP|_{\mathcal{F}_0} \prod_{t \in [0, \tau]} \left(\prod_h d\Lambda_h(t)^{\Delta N_h(t)} (1 - d\Lambda(t))^{1 - \Delta N(t)} \right). \quad (1.9.8)$$

Die Interpretation ist die gleiche, aber es soll suggeriert werden, daß man die Radon-Nikodym-Ableitung bezüglich infinitesimal kleiner Abschnitte auf $[0, \tau]$ faktorisieren kann, indem man das Verhältnis der entsprechenden (endlich vielen) Teilintervalle bildet und dann die Feinheit der Partition gegen 0 gehen läßt (Eine solche Konstruktion ist auch formal korrekt durchführbar).

In den statistischen Anwendungen werden Likelihood-Ratios nur für Radon-Nikodym-Ableitungen bezüglich der Mitglieder einer Familie von Wahrscheinlichkeitsmaßen bezüglich einer festen Referenz-Verteilung verwendet. Die Likelihoods werden dann nur bis auf einen Proportionalitätsfaktor benötigt, weshalb der Nenner der Likelihood-Ratios nicht angegeben werden muß.

Angenommen, es gilt (1.9.1) für Wahrscheinlichkeiten \tilde{P} und P , $\tilde{P} \ll P$, auf einer gemeinsamen Filtrierung (\mathcal{F}_t) . Dann ist der *Likelihood-Prozeß* gegeben durch

$$L(t) = \frac{d\tilde{P}}{dP} \Big|_{\mathcal{F}_t} \quad (1.9.9)$$

Man kann zeigen, daß der so definierte Prozeß ein Martingal ist. $L(t)$ besitzt auch die Darstellung (1.9.3). Eine derart definierter Prozeß ist unter bestimmten Umständen auch dann ein Martingal, wenn die Jacodformel (1.9.2) für die Likelihoodratio nicht erfüllt ist. Ein derartiger Prozeß ist beispielsweise der weiter unten definierte partielle Likelihood-Prozeß. Seine Eigenschaften hängen nur von den Kompensatoren Λ und $\tilde{\Lambda}$ von N unter P beziehungsweise \tilde{P} ab.

Angenommen, es gilt wiederum (1.9.1) für alle $\{P_\theta : \theta \in \Theta\}$, wobei $\Theta \subseteq \mathbb{R}$, offen. Alle P_θ werden von einem Maß Q dominiert. Sie (zur Vereinfachung) sollen auf \mathcal{F}_0 übereinstimmen und die Kompensatoren sollen absolut stetig sein, das heißt unter P_θ soll N den Kompensator $\Lambda^\theta = \int \lambda^\theta$ haben. Der Likelihood-Prozeß hängt nun von θ ab. Da der Nenner nicht von θ abhängt, kann er weggelassen werden und die Likelihood ist proportional zu

$$L(\theta; t) = \exp(-\Lambda^\theta(t)) \prod_{T_n \leq t} \lambda_{J_n}^\theta(T_n) \quad (1.9.10)$$

und daher

$$\begin{aligned} \log L(\theta; t) &= \sum_h \int_0^t \log \lambda_h^\theta(s) dN_h(s) - \Lambda^\theta(t) \\ &= \sum_h \int_0^t (\log \lambda_h^\theta(s) dN_h(s) - \lambda_h^\theta(s) ds). \end{aligned}$$

Der *Score-Prozeß* ist nun die Ableitung bezüglich θ . Unter der Annahme, daß die Ableitung- und Integral-Zeichen vertauscht werden können, erhält man:

$$\frac{\partial}{\partial \theta} \log L(\theta; t) = \sum_h \int_0^t \frac{\partial}{\partial \theta} \log \lambda_h^\theta(s) (dN_h(s) - \lambda_h^\theta(s) ds). \quad (1.9.11)$$

Wenn vorausgesetzt werden kann, daß $\frac{\partial}{\partial \theta} \log \lambda_h^\theta(s)$ integrierbar ist, dann ist der Score-Prozeß ein $(P_\theta, (\mathcal{F}_t))$ -lokales Martingal. Das Ergebnis gilt jedoch auch, ohne daß ein Likelihoodprozeß im Sinn von (1.9.9) vorliegt.

Wenn Ableitung und Integral vertauscht werden darf, dann ergibt nochmaliges Differenzieren bezüglich θ die *beobachtete Information* bei θ :

$$\frac{\partial^2}{\partial \theta^2} \log L(\theta; t) = \sum_h \int_0^t \frac{\partial^2}{\partial \theta^2} \log \lambda_h^\theta(s) dM_h(s) - \int_0^t \left(\frac{\partial}{\partial \theta} \log \lambda_h^\theta(s) \right)^2 \lambda_h^\theta(s) ds \quad (1.9.12)$$

Es gilt

$$\langle \partial / \partial \theta \log L(\theta; \cdot) \rangle = - \sum_h \int \left(\frac{\partial}{\partial \theta} \log \lambda_h^\theta \right)^2 \lambda_h^\theta. \quad (1.9.13)$$

$-\langle \partial / \partial \theta \log L(\theta; \cdot) \rangle$ ist der Kompensator sowohl von $-(\frac{\partial}{\partial \theta} \log L(\theta; \cdot))^2$ als auch von $(\frac{\partial^2}{\partial \theta^2} \log L(\theta; \cdot))$. Das entspricht dem Ergebnis der Likelihood-Theorie für reelle Zufallsvariable, daß die Varianz der Score-Funktion mit der erwarteten Information übereinstimmt.

1.10 Partielle Likelihoods für Zählprozesse

Im vorangegangenen Abschnitt wurde die Likelihood eines Zählprozesses bezüglich der selbsterzeugten Filtrierung angegeben (1.9.8). Es wurde aber bemerkt, daß auch bezüglich größerer Filtrierungen als der selbsterzeugten die rechte Seite von (1.9.8) die Martingaleigenschaft beibehält, die für die statistische Inferenz wichtig ist. Eine weitere Möglichkeit zur Herleitung von Likelihood-ähnlichen Ausdrücken sind gefilterte Zählprozesse (siehe 2.3).

Hier wird zunächst eine Methode beschrieben, mit der man Hierarchien von Zählprozessen definieren kann. Die Likelihood des Zählprozesses kann dann mit Hilfe dieser Reduktionen in einer Weise faktorisiert werden, woraus sich die für die Überlebensanalyse typischen partiellen Likelihoods ergeben.

In Verallgemeinerung von (1.9.8) wird für einen markierten Punkt-Prozeß folgende Likelihood angeschrieben

$$dP = \prod_t \left(\prod_{x \in E} \Lambda(dt, dx)^{N(dt, dx)} (1 - \Lambda(dt, E))^{1 - N(dt, E)} \right). \quad (1.10.1)$$

Dabei ist N ein multivariater Zählprozeß auf dem Raum der Marken (E, \mathcal{E}) mit Kompensator Λ bezüglich der selbsterzeugten Filtrierung (\mathcal{F}_t) . N und

Λ werden als zufällige Maße auf $(\mathcal{T} \times E, \mathcal{B}(\mathcal{T}) \otimes \mathcal{E})$ betrachtet. Die Interpretation von (1.10.1) ist, daß bezüglich \mathcal{F}_{t-} auf $[t, t + dt)$ (oder kürzer: auf dt), ein Ereignis mit Marke in dx (also $N(dt, dx) = 1$) mit Wahrscheinlichkeit $\Lambda(dt, dx)$ beobachtet wird. Kein Ereignis, also $N(dt, E) = 0$, wird mit Wahrscheinlichkeit $1 - \Lambda(dt, E)$ beobachtet. Diese Schreibweise soll die nachfolgende Konstruktion motivieren.

Nun wird eine Hierarchie der Marken definiert. Sei also N ein markierter Punkt-Prozeß mit Raum von Marken (E, \mathcal{E}) und mit Kompensator Λ bezüglich der selbsterzeugten Filtrierung (\mathcal{F}_t) . Sei \emptyset ein Punkt, der nicht in E liegt und *leere Marke* genannt wird. \emptyset soll eine Marke für den Fall sein, daß kein Ereignis eintritt. Sei $\bar{E} = E \cup \{\emptyset\}$. Sei (G, \mathcal{G}) ein anderer Raum von Marken, der \emptyset nicht enthält und $\bar{G} = G \cup \{\emptyset\}$. Die σ -Algebren auf \bar{E} und \bar{G} seien jeweils von \mathcal{E} und \mathcal{G} erzeugt.

Sei g eine meßbare Abbildung von \bar{E} nach \bar{G} mit $g(\emptyset) = \emptyset$. N^g sei ein markierter Punkt-Prozeß mit Raum von Marken (G, \mathcal{G}) , der durch $N^g((0, t] \times A) = N((0, t] \times g^{-1}(A))$ definiert ist. Der Kompensator Λ^g von N^g ist definiert durch

$$\Lambda^g((0, t] \times A) = \Lambda((0, t] \times g^{-1}(A)).$$

Dann kann man aufgrund der Schreibweise (1.10.1) folgenden Ansatz für die bedingte Verteilung von N^g gegeben \mathcal{F}_{t-} machen:

N^g hat ein Ereignis in $dt \times dy$ (das heißt, $N(dt, dy) = 1$) mit Wahrscheinlichkeit $\Lambda^g(dt, dy)$. Kein Ereignis, das heißt $N^g(dt, G) = 0$, tritt mit Wahrscheinlichkeit $1 - \Lambda^g(dt, G)$ ein.

Unter der Bedingung, daß $dt \times dy$ ein Ereignis hat, hat auch der ursprüngliche Prozeß ein Ereignis in $dt \times dx$ (wenn $g(x) = y$ erfüllt ist) mit bedingter Wahrscheinlichkeit $\Lambda(dt, dx)/\Lambda^g(dt, dy)$.

Unter der Bedingung, daß N^g kein Ereignis in $dt \times dy$ hat, hat der ursprüngliche Prozeß ein Ereignis in $dt \times dx$ (wenn $g(x) = \emptyset$ erfüllt ist) mit bedingter Wahrscheinlichkeit $\Lambda(dt, dx)/(1 - \Lambda^g(dt, G))$. Der ursprüngliche Prozeß hat kein Ereignis mit Wahrscheinlichkeit $(1 - \Lambda(dt, g^{-1}(\emptyset)))/(1 - \Lambda^g(dt, G))$.

Das motiviert folgende Umformung für (1.10.1):

$$\begin{aligned}
dP = & \prod_t \left\{ \left(\prod_{y \in G} \Lambda^g(dt, dy)^{N^g(dt, dy)} (1 - \Lambda^g(dt, G))^{1 - N^g(dt, G)} \right) \right. \\
& \cdot \prod_{y \in G} \left(\prod_{x: g(x)=y} \left(\frac{\Lambda(dt, dx)}{\Lambda^g(dt, dy)} \right)^{N(dt, dx)} \right)^{N^g(dt, dy)} \\
& \cdot \left(\prod_{x: g(x)=\emptyset} \left(\frac{\Lambda(dt, dx)}{1 - \Lambda^g(dt, G)} \right)^{N(dt, dx)} \right. \\
& \left. \left. \cdot \left(1 - \frac{\Lambda(dt, g^{-1}(\emptyset))}{1 - \Lambda^g(dt, G)} \right)^{1 - N(dt, E)} \right)^{1 - N^g(dt, G)} \right\}. \tag{1.10.2}
\end{aligned}$$

Die Faktorisierung (1.10.2) benötigt noch eine mathematische Interpretation. Die erste Zeile verwendet die in (1.9.2) eingeführte Schreibweise. Dieser Ausdruck wird zumeist auch die *partielle Likelihood* bezüglich N^g sein. Sie kommt dadurch zustande, daß man den Rest der Information aus der vollständigen Likelihood wegläßt. Es ist die Likelihood, die man bezüglich der selbsterzeugten Filtrierung von N^g erhalten würde. Allerdings kann Λ^g von der gesamten Vergangenheit von N abhängen, nicht nur von N^g .

Für eine Erklärung der übrigen Ausdrücke wird auf (Andersen et al. 1993, p107ff), beziehungsweise auf die dort zitierte Literatur verwiesen.

Kapitel 2

Modelle für Zählprozesse

In diesem Abschnitt werden Zählprozeßmodelle behandelt (siehe Andersen et al. (1993), pp 121). In den Anwendungen werden die Prozesse nicht vollständig beobachtet, sondern sie sind zensiert, gestutzt oder gefiltert. Die resultierenden Likelihoods und partiellen Likelihoods werden hergeleitet. Unvollständige Beobachtungen führen auch zu marginalen Modellen und Gebrechlichkeits-Modellen, die in den letzten beiden Abschnitten vorgestellt werden.

Der allgemeine Rahmen für Zählprozeßmodelle ist ein multivariater Zählprozeß $N = (N(t), t \in \mathcal{T})$, definiert auf einem Maßraum (Ω, \mathcal{F}) mit Filtrierung (\mathcal{F}_t) , an die N angepaßt ist.

Darüberhinaus ist eine Familie von Wahrscheinlichkeitsmaßen

$$\mathcal{P} = \{P_{\theta\phi} : (\theta, \phi) \in \Theta \times \Phi\}$$

auf (Ω, \mathcal{F}) definiert. $P_{\theta\phi}$ soll zusammen mit der Filtrierung (\mathcal{F}_t) die üblichen Bedingungen (außer Vollständigkeit) erfüllen. θ ist ein Parameter, der geschätzt werden soll, während der Neben-Parameter ϕ weitere Parameter enthält, die beispielsweise den Zensierungsprozeß oder die Verteilung von Kovariablen charakterisieren.

Der Kompensator von N bezüglich $P_{\theta\phi}$ wird mit $\Lambda^\theta(\cdot)$ oder $\Lambda(\cdot, \theta)$ bezeichnet. Es wird also angenommen, daß der Kompensator nicht von ϕ abhängt. Θ und Φ sind jeweils Teilmengen von nicht notwendigerweise endlichdimensionalen Vektorräumen.

2.1 Vollständig beobachtete Prozesse

Sei die selbsterzeugte Filtrierung gegeben durch $\mathcal{N}_t = \sigma\{N(u), 0 \leq u \leq t\}$ und $\mathcal{F}_t = \mathcal{F}_0 \vee \mathcal{N}_t$. \mathcal{F}_0 soll durch die Zufallsvariable X_0 erzeugt werden, die zum Zeitpunkt 0 realisiert wurde.

Beispiel 2.1.1. Eine einzelne nicht-negative Zufallsvariable

Der schon auf Seite 9 definierte Einsprung-Prozeß entspricht der Realisierung einer einzelnen nicht-negativen reellen Zufallsvariablen X . Dort war F deren Verteilungsfunktion, $S = 1 - F$ deren Überlebensfunktion, f deren Dichte, $\alpha = f/S$ deren Hasard-Rate. Nun wird diese Definition erweitert um einen (nicht notwendigerweise endlich-dimensionalen) Parameter θ , von dem die Verteilungsfunktion F^θ und die Hasardrate $\alpha^\theta = \alpha(\cdot, \theta)$ abhängen. Sei $\tau = \infty$, $\mathcal{T} = [0, \tau]$. Dann ist ein Zählprozeß definiert durch:

$$N(t) = I(X \leq t). \quad (2.1.1)$$

Der Kompensator bezüglich der selbsterzeugten Filtrierung $(\mathcal{N}_t) = (\mathcal{F}_t)$ und der Wahrscheinlichkeit P_θ , die der Verteilungsfunktion F^θ von X entspricht, ist gegeben durch

$$\Lambda(t, \theta) = \Lambda^\theta(t) = \int_0^t \alpha^\theta(u) Y(u) du. \quad (2.1.2)$$

Darin ist Y der vorhersagbare Prozeß

$$Y(t) = I(X \geq t) = 1 - N(t-). \quad (2.1.3)$$

Die Likelihood für θ aufgrund der Beobachtung von N auf \mathcal{T} ist gerade die Dichte f^θ , ausgewertet bei X . Mit Hilfe von (1.9.8) erhält man

$$\begin{aligned} L(\theta) &= \prod_t (\alpha^\theta(t) Y(t))^{\Delta N(t)} \prod_{\mathcal{T}} (1 - d\Lambda^\theta) \\ &= \prod_t (\alpha^\theta(t) Y(t))^{\Delta N(t)} \exp\left(-\int_0^t \alpha^\theta(u) Y(u) du\right) \\ &= \alpha^\theta(X) \exp\left(-\int_0^X \alpha^\theta(t) dt\right) = \alpha^\theta(X) S^\theta(X) = f^\theta(X). \end{aligned} \quad (2.1.4)$$

Beispiel 2.1.2. Unzensierte Überlebensdaten Sei nun eine Stichprobe X_1, \dots, X_n von unabhängigen nicht-negativen Zufallsvariablen mit Hasard-Funktionen α_i^θ gegeben. Sei wiederum $\tau = \infty$. Dadurch werden wie in (2.1.1)

und (2.1.3) Prozesse $N_i(t)$ und $Y_i(t)$ definiert. In den Anwendungen ist X_i die Zeit bis zum Eintreten eines Ereignisses, beispielsweise die Beendigung der Liegezeit einer Teleskopkrone.

Der multivariate Zählprozeß $N = (N_1, \dots, N_n)$ bezüglich der selbsterzeugten Filtrierung (\mathcal{N}_t) hat aufgrund der Unabhängigkeit der Komponenten bezüglich P_θ , der gemeinsamen Verteilung der X_i , den komponentenweisen Intensitätsprozeß $\alpha_i^\theta Y_i$ (Das Ergebnis könnte auch mit der Jacodformel (1.9.1) erhalten werden). Die Likelihood ist wegen (1.9.7)

$$\begin{aligned} L(\theta) &= \prod_{i=1}^n \prod_t (\alpha_i^\theta(t) Y_i(t))^{\Delta N_i(t)} \prod_{\mathcal{T}} \left(1 - \sum_{i=1}^n \alpha_i^\theta Y_i(t) dt \right) \\ &= \prod_{i=1}^n \alpha_i^\theta(X_i) \exp \left(- \sum_{i=1}^n \int_0^{X_i} \alpha_i^\theta(t) dt \right) \\ &= \prod_{i=1}^n \alpha_i^\theta(X_i) S_i^\theta(X_i) = \prod_{i=1}^n f_i^\theta(X_i). \end{aligned} \quad (2.1.5)$$

Das Modell der *multiplikativen Intensitäten* (Aalen 1978) ist eine Verallgemeinerung des Modells mit Intensitätsprozeß $\alpha_i^\theta Y_i$. Die Indikatorfunktion Y wird durch einen vorhersagbaren Prozeß ersetzt, der *beobachtbar* ist in dem Sinn, daß er nicht von θ abhängt. α sei deterministisch:

$$\lambda_i^\theta(t) = \alpha_i^\theta(t) Y_i(t). \quad (2.1.6)$$

$Y_i(t)$ ist oft der Indikator dafür, ob bei Individuum i zum Zeitpunkt t überhaupt ein Ereignis eintreten kann; man sagt dann, das Individuum gehört zum Zeitpunkt t der *Risikomenge* an. $\alpha_i^\theta(t)$ ist die individuelle Hasardrate für ein Ereignis, die Übergangintensität eines Markov-Prozesses (siehe unten) oder eine typ-spezifische Hasardrate. Wenn Hasardraten für alle Individuen gleich sind, also $\alpha_i^\theta = \alpha^\theta$, dann erfüllt auch der aggregierte Prozeß $N = \sum_{i=1}^n N_i$ das Modell der multiplikativen Intensitäten durch

$$\lambda^\theta(t) = \alpha^\theta(t) Y(t). \quad (2.1.7)$$

Dabei ist $Y(t)$ der Umfang der Risikomenge zur Zeit t .

2.1.1 Modelle von Markov-Prozessen

Sei $(X(t), t \in \mathcal{T})$ ein Markov-Prozeß mit endlichen Zustandsraum S und rechtsseitig stetigen Pfaden. Der Anfangszustand sei $X(0) = X_0$ und hänge von Parametern ϕ und (nicht notwendigerweise) θ ab.

Es sollen lokal integrierbare Übergangintensitäten $\alpha_{hj}^\theta(t) = \alpha_{hj}(t, \theta)$ vom Zustand h in den Zustand j , $h \neq j$, gegeben sein (siehe 1.8.1). Die Übergangskraft aus Zustand h heraus ist

$$\mu_h^\theta = \sum_{j \in S \setminus \{h\}} \alpha_{hj}^\theta$$

h wird als *absorbierend bezeichnet*, wenn $\mu_h^\theta(t) = 0$ für alle $t \in \mathcal{T}$.

Sei $N_{hj}(t)$ die Anzahl der direkten Übergänge für X von h nach j , $h \neq j$, in $[0, t]$. Dann sind $N = (N_{hj}(\cdot), h \neq j)$ zusammen mit X_0 äquivalent zu X in dem Sinn, daß die Beobachtung von $X(u)$, $0 \leq u \leq t$, die selben Daten liefert, wie die Beobachtung von X_0 und N auf $[0, t]$. Sei (\mathcal{N}_t) die selbsterzeugte Filtrierung von N und sei $\mathcal{F}_t = \mathcal{N}_t \vee \mathcal{F}_0$, wobei \mathcal{F}_0 von X_0 erzeugt wird. Der $P_{\theta\phi}$ -Intensitätsprozeß von N bezüglich (\mathcal{F}_t) ist dann aufgrund von Satz 1.8.4

$$\lambda_{hj}^\theta(t) = \alpha_{hj}^\theta(t) Y_h(t), \quad (2.1.8)$$

wobei $Y_h(t) = I(X(t-) = h)$ die Indikatorfunktion dafür ist, daß X vor t im Zustand h ist. Der Prozeß N hat also multiplikative Intensitäten.

Beispiel 2.1.3. (Konkurrierende Risiken)

Ein Spezialfall eines Markov-Prozesses wäre ein Prozeß mit 2 Zuständen, 0 (lebend) und 1 (tot). Sei $\alpha_{10}(t, \theta) \equiv 0$ (Zustand 1 ist absorbierend). Das ergibt eine unzensurierte Überlebenszeit mit der Hasardfunktion als Übergangintensität $\alpha_{01}(t, \theta)$ (Beispiel 2.1.1).

Konkurrierende Risiken sind ein weiteres Beispiel. Die Zustände sind der transiente Zustand 0 (lebend) und die absorbierenden Zustände $h = 1, \dots, k$, die jeweils einer Todesursache entsprechen. Der Anfangszustand ist 0 und die Übergangintensitäten $\alpha_{0h}(t, \theta)$, $h = 1, \dots, k$ werden als *für die Todesursache spezifische Hasardfunktionen* bezeichnet.

Dieses Modell der konkurrierenden Risiken ist äquivalent zu einem Modell mit unabhängigen Zufallsvariablen X_{i1}, \dots, X_{ik} , $i = 1, \dots, n$ mit Hasardfunktionen $\alpha_{01}(t, \theta), \dots, \alpha_{0k}(t, \theta)$ und dem multivariaten Zählprozeß $N = N_1, \dots, N_k$ mit

$$N_h(t) = \sum_{i=1}^n I\left(\min_l X_{il} = X_{ih} \leq t\right).$$

In der Zuverlässigkeitstheorie wird $\min_l X_{il}$ als Lebensdauer eines Systems interpretiert, das aus k unabhängigen Komponenten besteht mit Lebensdauern X_{i1}, \dots, X_{ik} . Der Nutzen dieser Interpretation ist in der Medizin umstritten. Denn oft sind die Lebensdauern der einzelnen Komponenten hypothetische (oder „latente“) Lebensdauern und es gibt keine physischen Entsprechungen.

Beispiel 2.1.4. Der Erkrankungs-Tod Prozeß mit Dauer als Kovariable

Seien 0, 1 und 2 die Zustände gesund, erkrankt und tot. Der Zählprozeß der Übergänge sei definiert durch $N(t) = (N_{01}(t), N_{02}(t), N_{12}(t))$. Die Komponenten mit $h = 1, 2$ haben jeweils Intensitätsprozeß $\alpha_{0h}(t)Y_0(t)$ mit $Y_0(t) = 1 - N_{01}(t-) - N_{02}(t-)$, während $N_{12}(t)$ den Intensitätsprozeß $\alpha_{12}(t, t-T)Y_1(t)$ mit $Y_1(t) = N_{01}(t-) - N_{12}(t-)$ hat, mit $T = \inf(t : N_{01}(t) = 1)$. Y_h ist der Indikator dafür, daß das Individuum zur Zeit $t-$ in Zustand h ist, während T der Zeitpunkt des Übergangs von 0 auf 1 ist, falls dieser überhaupt eintritt. Die Intensität $\alpha_{12}(t, d)$ hängt also von der Zeit t und der Dauer d bis zur Erkrankung ab. Da sie im allgemeinen nicht als Produkt einer deterministischen Funktion und eines stochastischen Prozesses, der nicht vom Parameter abhängt, angeschrieben werden kann, liegt kein Modell der multiplikativen Intensitäten vor. Das Modell ist ein Semi-Markov- oder Erneuerungs-Prozeß. Ein Markov-Prozeß liegt nur dann vor, wenn $\alpha_{12}(t, d)$ nicht von d abhängt.

2.1.2 Das Regressionsmodell für das relative Risiko

Die Daten sollen aus Paaren (X_i, Z_i) , $i = 1, \dots, n$, bestehen. Dabei ist X_i eine nicht-negative Zufallsvariable und Z_i ein p -dimensionaler Zufallsvektor von Regressoren. Die X_1, \dots, X_n seien bedingt unabhängig bezüglich $Z = (Z_1, \dots, Z_n)$. Die bedingte Verteilung von X_i gegeben $Z = z = (z_1, \dots, z_n)$ (Z hat festen Wert z) soll durch eine Hasardfunktion α_i^θ (siehe Einsprung-Prozeß, Seite 36) von der Form

$$\alpha_i^\theta(t) = \alpha_0^\gamma(t) r(\beta^\top z_i) \quad (2.1.9)$$

gegeben sein, mit $\theta = (\gamma, \beta)$, $\beta \in \mathbb{R}^p$, und *relativer Risiko-Funktion* $r(\cdot)$, die nicht-negativ ist und $r(0) = 1$ erfüllt. β ist der Vektor der *Regressionsparameter*. Das wichtigste Beispiel dieses Modells ist das *Cox-Modell* (Cox 1972) mit $r = \exp$ und γ unendlich-dimensional (auch Modelle mit endlichdimensionalem γ können betrachtet werden). Z_1, \dots, Z_n sind die *Kovariablen*, deren marginale Verteilung von den Parametern θ und ϕ abhängen kann. Die bedingten Verteilungen der X_i bezüglich Z sind unabhängig.

Seien nun die Zählprozesse N_i und Prozesse Y_i definiert wie in (2.1.1) und (2.1.3). Die Filtrierung (\mathcal{N}_t) sei durch den multivariaten Zählprozeß $N = (N_1, \dots, N_n)$ erzeugt. Sei \mathcal{F}_0 durch Z erzeugt. Sei $\mathcal{F}_t = \mathcal{F}_0 \vee \mathcal{N}_t$. Sei $P_{\theta\phi}$ die gemeinsame Wahrscheinlichkeitsverteilung der Daten (X_i, Z_i) , $i = 1, \dots, n$. Der Kompensator von N bezüglich $P_{\theta\phi}$ und (\mathcal{F}_t) kann nun aus der Produktkonstruktion für bedingt unabhängige Zählprozesse (siehe Seite 12) bestimmt werden. Der Intensitätsprozeß von N_i hängt nur von θ ab und ist gegeben durch $Y_i(t)\alpha_i^\theta(t)$.

Die Likelihood (1.9.7) ist dann von der Form

$$L(\theta, \phi) = L_0(\theta, \phi)L_\tau(\theta)$$

mit

$$L_0(\theta, \phi) = P_{\theta, \phi}(Z)$$

und

$$\begin{aligned} L_\tau(\theta) &= \prod_{i=1}^n \prod_{t \in \mathcal{T}} (\alpha_i^\theta(t) Y_i(t))^{\Delta N_i(t)} \prod_{t \in \mathcal{T}} \left(1 - \sum_{i=1}^n \alpha_i^\theta(t) Y_i(t) dt \right)^{1 - \Delta N_i(t)} \\ &= \prod_{i=1}^n \alpha_0(X_i, \gamma) r(\beta^\top z_i) \exp \left(-r(\beta^\top z_i) \int_0^{X_i} \alpha_0(u, \gamma) du \right). \end{aligned}$$

Der Exponent im Produktintegral ist als Schreibweise zu verstehen: Das Integral soll nur über Bereiche berechnet werden, auf denen $\Delta N_i(t) = 0$ (vgl. (1.9.3)). Er könnte bei absolut stetigen Kompensatoren entfallen.

Wenn die Verteilung von Z nicht von θ abhängt, dann ist für festes ϕ durch $L_\tau(\theta)$ die volle Likelihood für θ gegeben. Sonst ist $L_\tau(\theta)$ eine partielle Likelihood beziehungsweise die volle bedingte Likelihood bezüglich Z .

Der obenstehende Ausdruck für $L_\tau(\theta)$ ist die Likelihood für die bedingte Verteilung von X_1, \dots, X_n , bezüglich Z , ausgewertet bei X_1, \dots, X_n .

2.2 Rechts-Zensierung

2.2.1 Einleitung

Der vorangegangene Abschnitt setzte die Beobachtung von Zählprozessen auf ganz \mathbb{R}^+ voraus, da diese durch nicht-negative Zufallsvariable erzeugt wurden, die die Zeit bis zum Eintreten eines Ereignisses modellieren. In den Anwendungen wird das Eintreten dieses Ereignisses oft nicht beobachtet, sei es, weil man nicht so lange beobachten will, oder sei es, weil durch geänderte Umstände der Eintritt des Ereignisses nicht mehr möglich ist.

In diesem Unterabschnitt sollen multivariate Zählprozesse auf \mathcal{T} betrachtet werden, denen ein Prozeß der *rechtsseitigen Zensierung* „überlagert“ wurde. Es werden diejenigen Bedingungen behandelt, unter welchen im unvollständig beobachteten Zählprozeß die Modelleigenschaften der vollständig beobachteten Zählprozesse im wesentlichen erhalten bleiben.

Gegeben sei der multivariate Zählprozeß

$$N = (N_1, \dots, N_n).$$

Durch Rechts-Zensierung kann der Prozeß N_i nur auf einer Teilmenge $E_i = \{t \leq U_i\} \subseteq \mathcal{T}$ beobachtet werden, wobei U_i eine nicht-negative Zufallsvariable ist. Es ist der Bereich, auf dem der folgendermaßen definierte *individuelle Zensierungsprozeß* C_i gleich 1 ist:

$$C_i(t) = I(t \in E_i) = I(t \leq U_i). \quad (2.2.1)$$

Der Zensierungsprozeß

$$C(t) = (C_i(t), i = 1, \dots, n).$$

muß an die Filtrierung (\mathcal{F}_t) des vorhergehenden Unterabschnitts nicht angepaßt sein, sondern kann zusätzliche Variabilität enthalten. Sei also $\mathcal{F}_t = \mathcal{F}_0 \vee \mathcal{N}_t$ mit $\mathcal{F}_0 = \sigma(X_0)$. Nun wird eine größere Filtrierung betrachtet:

$$(\mathcal{G}_t) \supseteq (\mathcal{F}_t).$$

Der Zensierungsprozeß $C(\cdot)$ soll (\mathcal{G}_t) -vorhersagbar sein. Da $C_i(\cdot)$ linksseitig stetig ist, ist das dann der Fall, wenn $C(\cdot)$ angepaßt ist an (\mathcal{G}_t) . Das heißt, die Zensierungszeiten U_i sind Stoppzeiten bezüglich (\mathcal{G}_t) . Ein Beispiel für eine derartige Filtrierung ist

$$\mathcal{G}_t = \mathcal{F}_t \vee \sigma\{C(u), u \leq t\}.$$

Der Kompensator von N bezüglich (\mathcal{G}_t) kann vom Kompensator bezüglich (\mathcal{F}_t) verschieden sein.

Es tritt nun auch das Problem auf, daß die beobachteten Daten die Filtrierung (\mathcal{G}_t) nicht erzeugen können, da weder N noch C vollständig beobachtet werden können. Der beobachtbare Teil von N ist der *rechtsseitig zensierte Zählprozeß* $N^c = (N_1^c, \dots, N_n^c)$, gegeben durch

$$N_i^c(t) = \int_0^t C_i(s) dN_i(s). \quad (2.2.2)$$

Weiters soll \mathcal{F}_0 von der beobachtbaren Zufallsvariablen X_0 erzeugt werden. Auch der Zensierungsprozeß ist nur teilweise beobachtbar. Man kann sich das Konzept der Zensierung mit Hilfe einer Absorptionszeit τ_i vorstellen, nach der die Intensität von N_i Null ist. τ_i ist eine Zufallsvariable auf $\bar{\mathcal{T}}$. Beim Einsprung-Prozeß beziehungsweise unzensierten Überlebensdaten ist das die

Zeit des Eintretens des Ereignisses, also $\tau_i = X_i$. Wenn die Absorptionszeit $\tau_i = X_i$ vor U_i ist, dann wird $N_i(t)$ beobachtet, und $N_i = N_i^c$, U_i wird jedoch (im allgemeinen) nicht beobachtet. Wenn der Prozeß nicht vor U_i absorbiert wird, dann wird U_i und N_i^c beobachtet, während N_i hingegen nicht beobachtet wird. Die Filtrierung, die durch die beobachteten Daten erzeugt wird, ist (\mathcal{F}_t^c) .

Zensierte Überlebensdaten werden üblicherweise angegeben als Beobachtungen der Form

$$\left(\tilde{X}_i, D_i; i = 1, \dots, n \right),$$

dabei ist

$$\tilde{X}_i = X_i \wedge U_i$$

und

$$D_i = I(\tilde{X}_i = X_i)$$

definiert. Es ist also für jedes Individuum eine Beobachtungszeit \tilde{X}_i und die Information, ob das Ende der Beobachtungszeit durch ein Ereignis oder durch eine Zensierung verursacht war. Wenn $D_i = 1$, dann wird die Zensierungszeit üblicherweise nicht beobachtet. Zusammen mit der Zufallsvariablen X_0 sind die Daten zur Zeit t gegeben durch

$$(N^c(u), Y^c(u); 0 \leq u \leq t),$$

mit $N^c = (N_i^c; i = 1, \dots, n)$ und $Y^c = (Y_i^c; i = 1, \dots, n)$. Dabei ist

$$N_i^c(t) = I(\tilde{X}_i \leq t, D_i = 1) \tag{2.2.3}$$

der rechts-zensierte Zählprozeß und

$$Y_i^c(t) = I(\tilde{X}_i \geq t)$$

ist ein Indikator dafür, ob das Individuum i kurz vor t unter Risiko war.

In den folgenden Beispielen ist der Zensierungsprozeß vorhersagbar bezüglich der Filtrierung \mathcal{N}_t , sodaß die ursprüngliche Filtrierung zur Aufnahme des Zensierungsprozesses nicht vergrößert werden muß.

Beispiel 2.2.1. Überlebensdaten mit einfacher Zensierung vom Typ I.

Die Beobachtung jedes Individuums wird zu einem gemeinsamen, *deterministischen* Zeitpunkt u_0 beendet. Da der Prozeß nicht zufällig ist, ist er trivialerweise vorhersagbar bezüglich jeder Filtrierung. Diese Art der Zensierung ist in der Industrie üblich, wo für alle Versuchseinheiten ein Lebensdauer-Test zur gleichen Zeit begonnen und beendet wird.

Beispiel 2.2.2. Überlebensdaten mit einfacher Zensierung vom Typ II.

Das Experiment wird nach dem r -ten Ausfall, $r \leq n$ beendet. Daher ist $U_i = X_{(r)}$, $i = 1, \dots, n$ ($X_{(r)}$ bezeichnet die r -t größte Beobachtung). $C_i = I(t \leq X_{(r)})$ ist dann vorhersagbar bezüglich (\mathcal{N}_t) , da $X_{(r)}$ bezüglich dieser Filtrierung ein Stoppzeit ist. Die zensierten Beobachtungen $\tilde{X}_1, \dots, \tilde{X}_n$ sind abhängige Zufallsvariable. Auch dieser Zensierungstyp wird meist in der Industrie angewendet.

Nun folgen Beispiele, in denen (\mathcal{G}_t) echt größer ist als (\mathcal{F}_t) .

Beispiel 2.2.3. Überlebensdaten und progressive Zensierung vom Typ I

In klinischen Studien treten die Patienten in die Studie nacheinander ein (engl.: staggered entry). Zu einem bestimmten Zeitpunkt wird die Studie beendet. Wenn man an einem Modell für die Überlebenszeit vom Eintritt des Patienten in die Studie interessiert ist, ist die maximale Beobachtungszeit vom Eintritt in die Studie bis zum Abschluß - diese Zeitspanne soll U_i sein. Wenn die Überlebenszeiten der Patienten X_i ab Eintritt in die Studie unabhängig sind von den Eintrittszeiten X_0 , und \mathcal{G}_0 von den Eintrittszeiten erzeugt ist, dann ist der Zensierungsprozeß C vorhersagbar bezüglich (\mathcal{G}_t) mit $\mathcal{G}_t = \mathcal{G}_0 \vee \mathcal{F}_t$. Eine andere Sichtweise ist, daß die Zensierungszeiten, gegeben \mathcal{G}_0 , deterministisch sind, und somit eine Verallgemeinerung von Beispiel 2.2.1 vorliegt. In diesem Beispiel ist die Zensierung vollständig beobachtbar, zumindest, wenn man die (unrealistische) Annahme macht, daß die Beendigung der Studie die einzige Ursache für Zensierung ist.

Beispiel 2.2.4. Überlebensdaten mit zufälliger Zensierung

Die Zensierungszeiten $U = (U_1, \dots, U_n)$ haben in diesem Modell beliebige Verteilung und sind unabhängig von den Überlebenszeiten $X = (X_1, \dots, X_n)$. Das einfachste und mathematisch am besten zu bewältigende Modell nimmt unabhängig und identisch verteilte Zensierungszeiten an. Aber auch im allgemeinen Fall kann (\mathcal{G}_t) durch $\mathcal{G}_t = \mathcal{F}_t \vee \sigma(C(u); 0 \leq u \leq t)$ angegeben werden.

Man kann aber auch $\mathcal{G}_0 = \sigma(U)$ und $\mathcal{G}_t = \mathcal{G}_0 \vee \mathcal{N}_t$ setzen. Das hat den Nachteil, daß die Annahme, daß Zensierungszeiten schon zur Zeit 0 bekannt sind, gegen die Intuition ist.

2.2.2 Unabhängige Rechts-Zensierung

Es wird angenommen, daß das Ausgangsmodell der Zählprozeß N ist, mit (\mathcal{F}_t) -Kompensator Λ^θ und Intensitätsprozeß λ^θ bezüglich eines Wahrschein-

lichkeitsmaßes $P_{\theta\phi}$. ϕ ist ein Nebenparameter; typischerweise ist die Verteilung der Zensierungen U_1, \dots, U_n von ϕ abhängig.

Die Zensierungen können den Kompensator und die Aussagen über das zu modellierende Ereignis verändern. Wenn beispielsweise in einer klinischen Studie gerade die Patienten mit einer guten Prognose zensiert werden, dann ist die verbleibende Population nicht mehr „repräsentativ“ für die Grundgesamtheit. Zensierungsmechanismen, die keine derartigen Verzerrungen bewirken, werden als *unabhängig* bezeichnet. Das ist der Inhalt der nachstehenden Definition:

Definition 2.2.1. *Sei N ein multivariater Zählprozeß mit Kompensator Λ^θ bezüglich der Filtrierung (\mathcal{F}_t) und dem Wahrscheinlichkeitsmaß $P_{\theta\phi}$. Sei C ein Rechts-Zensierungsprozeß, der bezüglich $(\mathcal{G}_t) \supseteq (\mathcal{F}_t)$ vorhersagbar ist. Man sagt, die Rechts-Zensierung von N durch C ist unabhängig, wenn Λ^θ auch der Kompensator von N bezüglich (\mathcal{G}_t) und $P_{\theta\phi}$ ist.*

Einfache Zensierung vom Typ I und vom Typ II (siehe Beispiel 2.2.1 und Beispiel 2.2.2) ist offensichtlich unabhängig in diesem Sinn wegen $(\mathcal{G}_t) = (\mathcal{F}_t)$.

Aufgrund der Definition von unabhängiger Zensierung gibt es die Zerlegung

$$N_i(t) = \Lambda_i^\theta(t) + M_i(t). \quad (2.2.4)$$

Dabei ist $M_i(t)$ ein lokal quadratisch integrierbares Martingal bezüglich (\mathcal{G}_t) und wegen (2.2.2) gilt

$$\begin{aligned} N_i^c(t) &= \int_0^t C_i(s) dN_i(s) \\ &= \int_0^t C_i(s) d\Lambda_i^\theta(s) + \int_0^t C_i(s) dM_i(s) \\ &= \Lambda_i^c(t, \theta) + M_i^c(t) \end{aligned}$$

Da $C_i(\cdot)$ vorhersagbar und beschränkt ist, ist M_i^c wiederum ein lokal quadratisch integrierbares Martingal bezüglich (\mathcal{G}_t) (wegen Satz 1.4.1). Folglich ist der Kompensator von N_i^c bezüglich $P_{\theta\phi}$ und (\mathcal{G}_t) für alle $\phi \in \Phi$ gegeben durch:

$$\Lambda_i^c(t, \theta) = \int_0^t C_i(s) d\Lambda_i^\theta(s). \quad (2.2.5)$$

Wenn N das Modell der multiplikativen Intensitäten von Aalen mit $\lambda_i^\theta(t) = \alpha_i^\theta(t) Y_i(t)$ bezüglich (\mathcal{G}_t) erfüllt (siehe Seite 37), dann erfüllt auch N^c ein Modell mit multiplikativen Intensitäten, gegeben durch:

$$\lambda_i^c(t, \theta) = \alpha_i^\theta(t) Y_i^c(t), \quad (2.2.6)$$

wobei

$$Y_i^c(t) = C_i(t)Y_i(t).$$

Der beobachtbare Prozeß N_i^c hat also den selben deterministischen Teil wie N_i , der zufällige Anteil Y_i , der Indikatorprozeß dafür, daß das Individuum unter Risiko ist, muß ersetzt werden durch Y_i^c , also den Indikatorprozeß dafür, daß der Patient *beobachtet wird und unter Risiko* ist. Die Beobachtungen, die zur Zeit t vorliegen, können also mittels $N^c = (N_1^c, \dots, N_n^c)$ und $Y^c = (Y_1^c, \dots, Y_n^c)$ beschrieben werden durch

$$(X_0, N^c(u), Y^c(u)); 0 \leq u \leq t)$$

Eine äquivalente Charakterisierung der Daten, die zur Zeit t beobachtet werden, ist durch die σ -Algebra (\mathcal{F}_t^c) gegeben:

$$\mathcal{F}_t^c = \sigma(X_0, (N^c(u), Y^c(u)); 0 \leq u \leq t). \quad (2.2.7)$$

Der beobachtbare Prozeß N^c erfüllt dann aufgrund des Erneuerungssatzes 1.5.3 ebenfalls ein Modell mit multiplikativen Intensitäten bezüglich der Filtrierung (\mathcal{F}_t^c) , die von den beobachtbaren Daten erzeugt wurde. Da aufgrund der Definition Y_i^c an (\mathcal{F}_t^c) angepaßt ist, gilt

$$\mathbb{E}_{\theta\phi}(\lambda_i^c(t, \theta) | \mathcal{F}_t^c) = \alpha_i^\theta(t) Y_i^c(t), \quad (2.2.8)$$

was wiederum ein Modell mit multiplikativen Intensitäten ergibt.

Im allgemeinen Fall ist der Kompensator $\Lambda_i^c(\cdot, \theta)$ abhängig von θ und den vergangenen Beobachtungen. Zum Zweck der Parameterschätzung wird angenommen, daß man $\Lambda_i^c(t, \theta)$ für jeden Wert des Parameters θ mittels der beobachtbaren Daten berechnen kann. Daher kann es zweckmäßig sein, die Familie der $P_{\theta\phi}$ -Kompensatoren von N^c bezüglich (\mathcal{G}_t) zu definieren durch

$$\mathcal{L}^c(t) = (\Lambda^c(u, \theta); \theta \in \Theta, 0 \leq u \leq t). \quad (2.2.9)$$

Die beobachtbaren Daten zur Zeit t sollen dann gegeben sein durch

$$(X_0, (N^c(u), 0 \leq u \leq t); \mathcal{L}^c(t)).$$

Die beobachtbaren Daten können auch spezifiziert werden durch die σ -Algebra, die sie erzeugen:

$$\mathcal{F}_t^c = \sigma(X_0, (N^c(u), 0 \leq u \leq t); \mathcal{L}^c(t)) \quad (2.2.10)$$

Aufgrund des Erneuerungssatzes 1.5.3 ist 2.2.9 gleichzeitig auch die Familie der $(P_{\theta\phi}, (\mathcal{F}_t^c))$ -Kompensatoren von N^c .

Wenn das Modell nicht entartet ist (beispielsweise durch nicht-positives α_i^θ), dann stimmen die beiden Definitionen von (\mathcal{F}_t^c) überein.

2.2.3 Die Konstruktion von Likelihoods

Sei N ein vollständig beobachteter Zählprozeß. Die (\mathcal{F}_t) -Likelihood soll auf N und der Zufallsvariablen X_0 , deren Verteilung von den Parametern ϕ und θ abhängen kann, beruhen. Dann ist die Likelihood aufgrund von (1.9.6) gegeben durch

$$\begin{aligned} L(\theta, \phi) &= L_0(\theta, \phi) \prod_{t \in \mathcal{T}} \left\{ \prod_{i=1}^n d\Lambda_i^\theta(t)^{\Delta N_i(t)} (1 - d\Lambda^\theta(t))^{1 - \Delta N(t)} \right\} \\ &= L_0(\theta, \phi) L_\tau(\theta). \end{aligned} \quad (2.2.11)$$

Dabei ist $N = \sum_{i=1}^n N_i$ und Λ^θ ist der Kompensator von N bezüglich (\mathcal{F}_t) .

Die verfügbaren Beobachtungen zur Zeit t sind X_0 , $(N^c(u), 0 \leq u \leq t)$ und die für Individuen, bei denen die Zeit der Zensierung der Absorptionszeit vorhergeht, auch U_i . Sie sind formal durch $(X_0, N^c, \mathcal{L}^c(t))$ gegeben, wobei $\mathcal{L}^c(t)$ die Familie der $P_{\theta\phi}$ -Kompensatoren bezüglich (\mathcal{G}_t) ist.

Nun sollen die verfügbaren Beobachtungen (X_i oder U_i) als markierter Punktprozeß N^* dargestellt werden. N^* hat Marken von der Form $x = (y, u)$. y enthält die Information, daß Individuum i ein Ereignis erfahren hat. Sonst ist $y = \emptyset$, was „kein Ereignis“ bedeutet. Die zweite Komponente u bezeichnet, welches Individuum zensiert wurde. $u = \emptyset$ bedeutet keine Zensierung. N^* hat fast sicher keine gänzlich leere Marke $x = (\emptyset, \emptyset)$. N^c ist dann der reduzierte Prozeß aufgrund einer Aggregation von N^* , definiert durch

$$N_y^c = \sum_{x: g(x)=y} N_x^*,$$

wobei $g(y, u) = y$ und $N^* = (N_x^*)$ (der Vektor mit Komponenten N_x^*) und $N^c = (N_y^c)$ (der Vektor mit Komponenten N_y^c).

Die beiden Arten der Beschreibung der verfügbaren Beobachtungen sind äquivalent in dem Sinn, daß (für jedes $t \in \mathcal{T}$) $\mathcal{F}_t^c = \sigma(X_0, N^*(u), 0 \leq u \leq t)$ äquivalent ist \mathcal{F}_t^c aus (2.2.10). Die Likelihood von N^* bezüglich (\mathcal{F}_t^c) ist gemäß (2.2.11)

$$L_\tau^*(\theta, \phi) = L_0(\theta, \phi) L_\tau^*(\theta, \phi)$$

Aufgrund der Konstruktion, die zu (1.10.2) führt, ist folgende Faktorisierung der Likelihood möglich:

$$L_\tau^*(\theta, \phi) = L_\tau^c(\theta) L_\tau''(\theta, \phi) \quad (2.2.12)$$

Dabei wurde $L_0(\theta, \phi)$ in den zweiten Faktor $L_\tau''(\theta, \phi)$ einbezogen. Der erste Faktor in (2.2.12), der dem ersten Faktor in (1.10.2) entspricht, ist

$$L_\tau^c(\theta) = \prod_t \left\{ \prod_i d\Lambda_i^c(t, \theta)^{\Delta N_i^c(t)} (1 - d\Lambda_i^c(t, \theta))^{1 - \Delta N_i^c(t)} \right\}, \quad (2.2.13)$$

mit $N^c = \sum_{i=1}^n N_i^c$ und $\Lambda^c(\cdot, \theta)$ ist der $(P_{\theta\phi}, (\mathcal{F}_t^c))$ -Kompensator von N^c . Aufgrund der Annahmen hängt Λ^c nicht von ϕ ab und die *partielle Likelihoodfunktion* (2.2.13) der *innovativen Ereignisse* hat die gleiche Form, wie die (partielle) Likelihood $L_\tau(\theta)$ des unzensierten Prozesses N in (2.2.11). Daher kann man feststellen, daß unabhängige Rechts-Zensierung die Gestalt der (partiellen) Likelihood beibehält.

Eine Konsequenz davon ist, daß die partielle Likelihood die Martingal-Eigenschaften der Likelihood des unzensierten Prozesses beibehält. Auch die asymptotischen Tests sind bei unabhängig zensierten Prozessen sehr ähnlich unzensierten Beobachtungen mit vollständiger Likelihood, da die Martingal-Theorie in der asymptotischen Theorie eine zentrale Stelle einnimmt.

Beispiel 2.2.5. Überlebensdaten mit zufälliger Zensierung

Zunächst wird nur ein Individuum betrachtet, mit Zeit des Ereignisses X und Zensierungszeit U . Die nicht-negative Zufallsvariable X hat Hasardrate $\alpha^\theta(t)$ und die Verteilung von U muß nicht notwendigerweise absolut stetig sein. Man kann die Daten durch einen markierten Punkt-Prozeß auf \mathcal{T} beschreiben mit Marke $(1, \emptyset)$ bei X und $(\emptyset, 1)$ bei U (Die Marke (\emptyset, \emptyset) kommt mit Wahrscheinlichkeit 0 vor). Sei (\mathcal{G}_t) die Filtrierung, die durch diesen markierten Zählprozeß erzeugt wird. Dann hat der Prozeß $N(t) = I(X \leq t)$ bezüglich (\mathcal{G}_t) den selben Kompensator

$$\Lambda^\theta(t) = \int_0^t \alpha^\theta(u) I(X \geq u) du$$

wie das Modell ohne Zensierung, also bezüglich $(\mathcal{F}_t) = (\mathcal{N}_t)$. Das folgt wegen der Unabhängigkeit von X und U aus Satz 1.9.1.

Die Verhinderung der Beobachtung des Ereignisses X durch die Rechts-Zensierung U entspricht nun der Situation, in der die Beobachtung durch eine (\mathcal{G}_t) -Stopzeit $X \wedge U$ beendet wird und die Marke, die kennzeichnet, ob ein Ereignis oder Zensierung vorliegt, und die zu dieser Sprung-Zeit gehört, beobachtet wird. Der so definierte markierte Punkt-Prozeß ist N^* . Der Zensierungsprozeß $C(t) = I(U \geq t)$ ist daher (\mathcal{G}_t) -vorhersagbar und wegen (2.2.6) hat der zensierte Prozeß N^c Kompensator

$$\Lambda^c(t, \theta) = \int_0^t \alpha^\theta(u) I(X \wedge U \geq u) du$$

bezüglich (\mathcal{G}_t) und wegen (2.2.8) auch bezüglich (\mathcal{F}_t^c) , der Filtrierung, die durch den zensierten markierten Punkt-Prozeß N^* erzeugt wird (siehe (2.2.10)).

Wenn eine Stichprobe X_1, \dots, X_n von unabhängig, identisch verteilten Zufallsvariablen und eine davon unabhängige Zensierung $U = (U_1, \dots, U_n)$ vorliegt, dann kann man genauso vorgehen. Vom „großen“ markierten Punkt-Prozeß, der aus den N_i und U_i besteht, und der (\mathcal{G}_t) erzeugt, leitet man zuerst den multivariaten Zählprozeß $N = (N_1, \dots, N_n)$ ab, mit $N_i(t) = I(X_i \leq t)$ und Marken der Form (i, \emptyset) bei Ereignissen. Der $P_{\theta\phi}$ -Kompensator $\Lambda(\cdot, \theta)$ für N bezüglich (\mathcal{G}_t) stimmt wegen der Unabhängigkeit von X und U mit dem ursprünglichen (\mathcal{F}_t) -Kompensator des unzensierten Modells überein. Der Kompensator des zensierten Modells kann analog zu oben angegeben werden.

Sei N^* wiederum der markierte Punkt-Prozeß, der aus den $X_i \wedge U_i$ gebildet wird, mit Marken aus dem „großen“ Prozeß. $C(t)$ ist vorhersagbar bezüglich (\mathcal{G}_t) . Dann hat der zensierte Prozeß N^c die Darstellung $N^c(t) = \int_0^t C(u) dN(u)$.

Beispiel 2.2.6. Zensierung durch konkurrierende Risiken

Siehe Beispiel 2.1.3. Man will eine von mehreren Todesursachen untersuchen. Die anderen Todesursachen sind in diesem Fall Rechts-Zensierungen. Aufgrund des Konzepts der „latenten“ Zeitpunkte des Todes ist klar, daß unabhängige Zensierung vorliegt. Da aber hypothetische Zeitpunkte des Todes umstritten sind, ist es nützlich, zu zeigen, daß dieses Ergebnis auch mittels eines markierten Punkt-Prozesses erhalten werden kann.

Der Einfachheit halber wird nur eine konkurrierende Todesursache betrachtet. Sei $\alpha_{01}(t, \theta)$ die Hasardfunktion der untersuchten Todesursache und sei $\alpha_{02}(t, \theta, \phi)$ die Hasardfunktion einer weiteren Todesursache. Für das Individuum i definiert man $N_i = (N_{0ji}, j = 1, 2)$ wie in Beispiel 2.1.3, sowie die selbsterzeugte Filtrierung (\mathcal{N}_t) . Nun kann N_i als markierter Punkt-Prozeß definiert werden mit Raum von Marken $E = \{(1, \emptyset), (\emptyset, 2)\}$. Die Marke $(1, \emptyset)$ entspricht einem Tod durch Ursache 1, die Gegenstand der Untersuchung ist, und die Marke $(\emptyset, 2)$ entspricht dem Tod durch die konkurrierende Ursache. Der $(P_{\theta\phi}, (\mathcal{N}_t))$ -Kompensator der Komponente $N_{01i}(t)$ zählt die Anzahl der Marken, die Gegenstand der Untersuchung sind, im Zeitraum $[0, t]$. Er ist

$$\Lambda_{1,i}(t, \theta) = \int_0^t \alpha_{01}(u, \theta) Y_{0i}(u) du$$

Dabei gibt $Y_{0i}(t) = 1 - N_{0\cdot i}(t-)$ an, ob das Individuum i zur Zeit $t-$ lebt (Zustand 0). Nun kann der Rechts-Zensierungsprozeß definiert werden durch

$C_i(t) = I(t \leq U_i)$, wobei U_i die folgende (\mathcal{N}_t) -Stoppzeit ist:

$$U_i = \inf_{t \in \mathcal{T}} \{t : N_{02i}(t) = 1\}.$$

Der *zensierte markierte Punkt-Prozeß*, das ist der beobachtbare Anteil von N_i , entspricht der Zufallsvariablen $\tilde{X}_i = X_i \wedge U_i$ mit

$$X_i = \inf_{t \in \mathcal{T}} \{t : N_{01i}(t) = 1\}.$$

Er trägt die Marke $(1, \emptyset)$, wenn $\tilde{X}_i = X_i$, und $(\emptyset, 2)$, wenn $\tilde{X}_i = U_i$. Die Komponente $N_{01i}^c(t) = N_{01i}(t)$ zählt die Marken der Form $(1, \emptyset)$ in $[0, t]$ im zensierten markierten Punkt-Prozeß, und hat $(P_{\theta\phi}, (\mathcal{N}_t))$ -Kompensator $\Lambda_{1i}(\cdot, \theta)$, der wiederum an (\mathcal{F}_t^c) , die vom zensierten markierten Punkt-Prozeß erzeugte Filtrierung, angepaßt ist, da für diesen Zählprozeß $Y_{0i}(t) = I(\tilde{X}_i \geq t)$. Daher ist $\Lambda_{1i}(\cdot, \theta)$ auch der $(P_{\theta\phi}, (\mathcal{F}_t^c))$ -Kompensator von N_{01i}^c . Das bedeutet wiederum, daß man statistische Schätzung für θ (und damit auch die Hazardfunktion α_{01}^θ der Todesursache, die Gegenstand der Untersuchung ist) bei konkurrierenden Todesursachen auch dann durchführen kann, wenn man die anderen Todesursachen als unabhängigen Zensierungsprozeß betrachtet.

Beispiel 2.2.7. Zensierung in Abhängigkeit von Kovariablen

Im Regressionsmodell der relativen Risiken (siehe **2.1.2**) wurde eine Filtrierung der Form $\mathcal{F}_t = \mathcal{F}_0 \vee \mathcal{N}_t$ betrachtet. \mathcal{F}_0 wurde von den zeitunabhängigen Kovariablen Z_1, \dots, Z_n erzeugt. Die vorhergehenden Modelle für die Zensierung können also auch von den Kovariablen abhängen, die \mathcal{F}_0 erzeugen. In einer Studie mit zwei Stichproben können beispielsweise die Verteilungen der Zensierungen in den beiden Stichproben unterschiedlich sein. Eine weitere Art dieser Zensierung in Abhängigkeit von Kovariablen, die mit diesem Modell verträglich ist, wäre eine Studie über die Überlebenszeit mit Zeit seit Eintritt in die Studie als Zeitskala und Alter und Geschlecht als Kovariable. Wenn jedes Jahr die älteste Frau, die noch lebt, zensiert wird, dann ist eine derartige Zensierung unabhängig.

Da der Zensierungsprozeß an eine erweiterte Filtrierung (\mathcal{G}_t) angepaßt sein muß, ist der entscheidende Punkt, daß die Erweiterung, die durch die Kovariablen erzeugt wird, den Kompensator von N nicht ändert. *Abhängige Rechts-Zensierung* läge beispielsweise vor, wenn der Zensierungsprozeß von Kovariablen abhänge, die nicht im Modell enthalten sind.

Beispiel 2.2.8. Eine randomisierte klinische Studie über Leberzirrhose

In einer randomisierten Studie über den Effekt der Behandlung mit Prednison gegen Placebo auf das Überleben der Patienten mit Leberzirrhose, wurden die

meisten Patienten dadurch zensiert, daß sie am Ende der Studie noch lebten. Allerdings schieden einige Patienten wegen Nebenwirkungen von Prednison frühzeitig aus der Studie aus. Es ist schwer zu entscheiden, ob diese Zensierung unabhängig ist. Formell ist der Zensierungsprozeß angepaßt, wenn alle Kovariablen, von denen das Ausscheiden aus der Studie abhängt, im Modell für den Intensitätsprozeß des Todes enthalten sind, das heißt zu \mathcal{F}_0 . Weiters müssen die Zeiten des Ausscheidens U_i in (\mathcal{G}_t) enthalten sein. Das entscheidende Problem ist daher, ob der Einschluß der Zeiten des Ausscheidens den Intensitätsprozeß des Todes ändert. Das kann im Prinzip dadurch entschieden werden, daß X_i und U_i unabhängig bezüglich \mathcal{F}_0 sind. Dies ist insbesondere dann nicht der Fall, wenn Tod und das Ausscheiden aus der Studie von einer nicht beobachteten Kovariablen abhängen, die in \mathcal{F}_0 nicht berücksichtigt werden konnte.

Aus diesem Grund wird empfohlen, in klinischen Studien Patienten weiterhin zu beobachten, auch wenn sie nicht mehr behandelt werden. Die Analyse der Daten sollte dann auf der Gruppenzugehörigkeit gemäß Randomisierung beruhen (intention to treat principle).

In der Definition von unabhängiger Rechts-Zensierung wurde verlangt, daß der Zensierungsprozeß C bezüglich der erweiterten Filtrierung (\mathcal{G}_t) vorhersehbar ist, und der Kompensator von N unverändert bleibt. Ein Beispiel C für *abhängige* Zensierung darf daher nicht an (\mathcal{G}_t) angepaßt sein (siehe die beiden vorhergehenden Beispiele).

2.2.4 Nicht-informative Rechts-Zensierung

Sei ein multivariater Zählprozeß $N = (N_1, \dots, N_n)$ gegeben, der bezüglich eines unabhängigen Rechts-Zensierungs-Prozesses C zensiert wird. Die Beobachtungen werden wie oben als markierter Punkt-Prozeß N^* betrachtet. Die $P_{\theta\phi}$ -Likelihood für N^* bezüglich (\mathcal{F}_t^c) ist nach (2.2.12) gegeben durch

$$L_\tau^*(\theta, \phi) = L_\tau^c(\theta) L_\tau''(\theta, \phi).$$

Die partielle Likelihood $L_\tau^c(\theta)$ für θ aufgrund des beobachteten Zählprozesses N^c (definiert in (2.2.4)) ist wegen (2.2.13)

$$L_\tau^c(\theta) = \prod_t \left\{ \prod_i d\Lambda_i^c(t, \theta)^{\Delta N_i^c(t)} (1 - d\Lambda^c(t, \theta))^{1 - \Delta N^c(t)} \right\}.$$

Die Frage, ob durch $L_\tau^c(\theta)$ bei festem $\phi \in \Phi$ eine volle Likelihood für θ aufgrund der Beobachtung von N^* gegeben ist, ist Anlaß für die nachstehende Definition:

Definition 2.2.2. Wenn für jedes feste $\phi \in \Phi$ die Likelihood (2.2.13) durch

$$L_{\tau}^*(\theta, \phi) = L_{\tau}^c(\theta)L_{\tau}''(\phi)$$

faktoriell werden kann, und daher L_{τ}'' nicht von θ abhängt, dann wird die Rechts-Zensierung C als nicht-informativ für θ bezeichnet.

Eine Interpretation von nicht-informativer Zensierung durch C wäre, daß für festes $\phi \in \Phi$ die bedingte Verteilung von $dN^*(t)$ bezüglich \mathcal{F}_{t-}^c und bezüglich $dN^c(t)$ nicht von θ abhängt. Oder in anderen Worten, die Zensierung ist nicht-informativ, wenn für festes $\phi \in \Phi$ und $t \in \mathcal{T}$ die bedingte Intensität für die Zensierung eines bestimmten Individuums bei t bezüglich der Vergangenheit bis vor t und bezüglich einem Ereignis bei t nicht von θ abhängt.

In Beispiel 2.2.3 liegt nicht-informative Zensierung vor, wenn die Eintrittszeiten nicht von θ abhängen. In Beispiel 2.2.4 und Beispiel 2.2.5 ist die Zensierung nicht-informativ, wenn deren Verteilung nicht von θ abhängt. In Beispiel 2.2.7 ist die Zensierung nicht-informativ, wenn die Zensierung zufällig ist und die Verteilung der Zensierung in Abhängigkeit von den Kovariablen nicht von θ abhängt.

Informative Rechts-Zensierung liegt beispielsweise vor, wenn konkurrierende Todesursachen mit jeweils spezifischen Intensitäten von θ abhängen (siehe Beispiel 2.2.6). Auch das Koziol-Green-Modell (Koziol und Green 1976), bei dem die Intensitäten von Ereignissen und Zensierungen proportional sind, hat informative Zensierung. Statistiken, die nur auf L_{τ}^c beruhen, werden nur bei informativer Zensierung Information verlieren. Allerdings sind die sinnvollsten Rechts-Zensierungs-Mechanismen nicht-informativ.

2.2.5 Links-Stutzen

Diese Art von Daten sind in Demographie und Epidemiologie sehr häufig. Sie sollen an folgendem Beispiel illustriert werden:

Beispiel 2.2.9. Überleben von insulin-abhängigen Diabetikern im Bezirk Fyn

Im dänischen Bezirk Fyn mit 450000 Einwohnern gab es am 1. Juli 1973 1499 Diabetiker. Am 1. Jänner 1982 wurde mit Hilfe eines Sterberegisters der Überlebensstatus dieser Patienten bestimmt. Man wollte den Einfluß von Alter und Dauer der Erkrankung auf das Überleben untersuchen. Die Zeitskala beginnt mit der Erkrankung an Diabetes. Um eine Verfälschung durch die Länge des Überlebens (In Querschnittstudien haben längere Überlebenszeiten eine größere Wahrscheinlichkeit, in die Studie aufgenommen zu werden), sollen die Patienten nur vom 1. Juli 1973 bis zum

1. Jänner 1982 beziehungsweise bis zur Zeit der Emigration Informationen beitragen können.

Da ein Diabetiker nur dann in die Studie aufgenommen wurde, wenn er am 1. Juli 1973 lebte, ist für die Überlebenszeiten X_i , $i = 1, \dots, n$, die relevante Verteilung die bedingte Verteilung von X gegeben $X > V$. Dabei ist die Eintrittszeit V die Dauer der Krankheit bis 1. Juli 1973. Dies ist ein Beispiel für links-gestutzte Daten.

Man betrachtet wiederum einen multivariaten Zählprozeß, der zu einem Individuum gehört. Sei $N = (N_h : h = 1, \dots, k)$ definiert auf (Ω, \mathcal{F}) mit $P_{\theta\phi}$ -Kompensator Λ^θ und Intensitätsprozeß λ^θ bezüglich der Filtrierung (\mathcal{F}_t) vom Typ $\mathcal{F}_t = \mathcal{F}_0 \vee \mathcal{N}_t$. Es wird angenommen, daß es eine größere Filtrierung $(\mathcal{G}_t) \supseteq (\mathcal{F}_t)$ gibt, mit der Eigenschaft, daß der $(P_{\theta\phi}, (\mathcal{G}_t))$ -Kompensator von N wiederum Λ^θ ist. Man betrachtet also nur *unabhängiges Links-Stutzen* (analog zur unabhängigen Rechts-Zensierung). Die größere Filtrierung (\mathcal{G}_t) soll die zusätzliche zufällige Variabilität enthalten, die durch die Stutzzeit entstanden ist, so daß V insbesondere eine (\mathcal{G}_t) -Stoppzeit ist. Sei $A \in \mathcal{G}_V$ mit $P_{\theta\phi}(A) > 0$. Der Prozeß N , der bei V gestartet wurde, ist definiert durch

$${}_V N(t) = N(t) - N(t \wedge V). \quad (2.2.14)$$

Es soll der Prozeß N , gestartet zur Zeit V , bezüglich des Ereignisses A , das vor V stattfindet, untersucht werden. Unter dieser bedingten Verteilung heißt ${}_V N$ *links-gestutzter Prozeß*. ${}_V N$ hat wegen Proposition 1.5.2 den Intensitätsprozeß

$${}_V \lambda^\theta(t) = \lambda^\theta I(t > V)$$

bezüglich der Filtrierung $({}_V \mathcal{G}_t)$, gegeben durch

$${}_V \mathcal{G}_t = \mathcal{G}_t \vee \mathcal{G}_V$$

und eine bedingte Verteilung $P_{\theta\phi}^A$, gegeben durch

$$P_{\theta\phi}^A(F) = P_{\theta\phi}(F \cap A) / P_{\theta\phi}(A), \quad F \in \mathcal{F}.$$

Wenn \mathcal{G}_t von der Form $\mathcal{G}_t = \mathcal{G}_0 \vee \mathcal{N}_t$ ist, dann ist

$${}_V \mathcal{G}_t = \mathcal{G}_V \vee \sigma \{ {}_V N(u); 0 \leq u \leq t \}.$$

Die $(P_{\theta\phi}^A, ({}_V \mathcal{G}_t))$ -bedingte Likelihood für ${}_V N$ bezüglich \mathcal{G}_V (die auch eine partielle Likelihood ist, siehe unten), ist gegeben durch

$${}_V L(\theta) = \prod_{t > V} \left\{ \prod_h \lambda_h^\theta(t)^{\Delta_V N_h(t)} (1 - \lambda^\theta(t) dt)^{1 - \Delta_V N \cdot (t)} \right\}. \quad (2.2.15)$$

Analog zu rechts-zensierten Daten (siehe (2.2.9) und (2.2.10)) soll nun untersucht werden, wie die Daten mit dem Fortschritt der Zeit verfügbar werden. Unter der Bedingung, daß A realisiert wurde, soll ${}_V N$ beobachtet werden. Außerdem soll es möglich sein, mit den verfügbaren Daten den Intensitätsprozeß ${}_V \lambda^\theta$ für jedes $\theta \in \Theta$ anzugeben. Die beobachtete σ -Algebra zur Zeit t , $t > V$ ist dann gegeben durch

$${}_V \mathcal{F}_t^c = \sigma \{ {}_V N(u), V \leq u \leq t, {}_V \mathcal{L}(t) \}$$

mit

$${}_V \mathcal{L}(t) = \{ {}_V \lambda^\theta(u); \theta \in \Theta; V \leq u \leq t \}.$$

Aufgrund des Erneuerungssatzes Satz 1.5.3 ist ${}_V \lambda^\theta$ auch der $({}_V \mathcal{F}_t^c, P_{\theta\phi}^A)$ -Intensitätsprozeß von ${}_V N$ und (2.2.15) ist eine partielle Likelihood bezüglich $({}_V \mathcal{F}_t^c)$.

Beispiel 2.2.10. Zufälliges Links-Stutzen einer Überlebenszeit

Gegeben sei eine Zufallsvariable $X > 0$ mit Hasardfunktion α_X^θ . Sei $N(t) = N_X(t) = I(X \leq t)$ und sei (\mathcal{N}_t) die von N_X erzeugte Filtrierung. Sei $V > 0$ unabhängig von X mit einer Verteilung, die von ϕ abhängt. Sei ein bivariater Zählprozeß mit Komponenten (N_X, N_V) , $N_V(t) = I(V \leq t)$ gegeben. Sei (\mathcal{G}_t) dessen selbsterzeugte Filtrierung. Der (\mathcal{G}_t) -Intensitätsprozeß $\lambda_X^{\theta\phi}$ von N_X bezüglich der gemeinsamen Verteilung $P_{\theta\phi}$ von X und V ist (vergleiche Beispiel 2.2.5)

$$\lambda_X^{\theta\phi}(t) = \lambda_X^\theta(t) = \alpha_X^\theta(t) I(t \leq X).$$

Das ist der selbe Intensitätsprozeß, wie bezüglich (\mathcal{N}_t) . Wenn das Ereignis $A = \{X > V\}$ positive Wahrscheinlichkeit hat, dann ist der Intensitätsprozeß des links-gestutzten Prozesses ${}_V N_X(t) = N_X(t) - N_X(t \wedge V)$ bezüglich der bedingten gemeinsamen Verteilung $P_{\theta\phi}^A$ von X und V gegeben A

$${}_V \lambda_X^{\theta\phi}(t) = {}_V \lambda_X^\theta(t) = \alpha_X^\theta(t) I(V < t \leq X).$$

Daher ist zufälliges Links-Stutzenunabhängig und erhält die multiplikative Struktur des Intensitätsprozesses. Insbesondere hängt der Intensitätsprozeß nur von θ ab. Die bedingte (bezüglich V) beziehungsweise partielle Likelihood für ${}_V N_X$ bezüglich der bedingten Verteilung gegeben $X > V$ ist wegen (2.2.15)

$$\begin{aligned} {}_V L(\theta) &= \prod_{t>V} \left({}_V \lambda_X^\theta(t)^{\Delta_V N_X(t)} (1 - {}_V \lambda_X^\theta(t) dt)^{1 - \Delta_V N_X(t)} \right) \\ &= \alpha_X^\theta(X) \exp \left(- \int_V^X \alpha_X^\theta(t) dt \right) = S_X^\theta(X) \alpha_X^\theta(X) / S_X^\theta(V). \end{aligned}$$

2.3 Verallgemeinerte Zensierung, Filterung und Stutzen

Die *Rechts-Zensierung* ist die Beschränkung der Beobachtung von N_i auf ein Intervall $E_i = [0, U_i]$. Beliebige Teilmengen E_i von \mathcal{T} als Beobachtungspläne sind denkbar. *Links-Zensierung* entspricht einer Menge $E_i = (V_i, \tau]$, $V_i \geq 0$. Im Unterschied dazu wird beim Links-Stutzen der Zählprozeß N_i bei V_i nur unter der Bedingung gestartet, daß vor V ein Ereignis stattgefunden hat.

Ganz allgemein kann man Zensierung auf zufälligen Intervallen definieren durch

$$E_i = \bigcup_{j=1}^r (V_{ji}, U_{ji}], \quad (2.3.1)$$

wobei $0 \leq V_{1i} \leq U_{1i} \leq \dots \leq V_{ri} \leq U_{ri} \leq \tau$.

Bei der *Filterung* des Zählprozesses werden keine Ereignisse registriert. Ein Beispiel dafür ist ein Markov-Prozeß mit den beiden Zuständen krank und gesund. Der Patient wechselt zwischen diesen beiden Zuständen. Wenn der Patient im Intervall $[0, U_{1i}]$ und $(V_{2i}, U_{2i}]$, $U_{1i} < V_{2i}$ beobachtet wird, dann ist zwar am Ende der Zeit ohne Beobachtungen der Zustand des Patienten bekannt, aber nicht, wie oft der Zustand gewechselt hat. Wenn der Patient zensiert worden wäre, dann wäre am Ende der Zensierungszeit die Anzahl der Ereignisse bekannt, die in der Zwischenzeit stattgefunden haben.

Dies motiviert die Einführung eines *Zensierungs- und Filterungs-Prozesses* C durch $C_i = I(t \in E_i)$ und eines *gefilterten* Zählprozesses durch

$$N_i^c(t) = \int_0^t C_i(u) dN_i(u).$$

Wenn man sich auf *unabhängige* Filterung beschränkt, dann kann man die Existenz einer Filtrierung $(\mathcal{G}_t) \supseteq (\mathcal{F}_t)$ annehmen, sodaß $N = (N_1, \dots, N_n)$ bezüglich der beiden Filtrierungen den gleichen $P_{\theta\phi}$ -Kompensator Λ^θ hat. Die U_{ij} und V_{ij} sollen (\mathcal{G}_t) -Stoppzeiten sein, so daß die C_i (\mathcal{G}_t) -vorhersagbar sind. Dann ist der $P_{\theta\phi}$ -Kompensator von N_i^c bezüglich (\mathcal{G}_t)

$$\Lambda_i^c(t, \theta) = \int_0^t C_i(u) d\Lambda_i(u, \theta).$$

Die Daten, die zur Verfügung stehen, seien X_0 und N^c . Für eine Menge E_i der Form (2.3.1) werden auch die V_{ji} und U_{ji} beobachtet, wenn sie vor der Absorptionszeit τ_i eintreten. Wenn die Absorptionszeit nicht beobachtet wird,

das heißt $\tau_i \in (U_{j_i}, V_{j+1,i}]$, dann beobachtet man auch das kleinste V_{j_i} mit $V_{j_i} \geq \tau_i$. Bei Zensierung werden auch die Werte von N_i bei der Eintrittszeit V_{j_i} registriert, bei Filterung werden sie hingegen nicht beobachtet.

Es soll jedoch angenommen werden, daß $\Lambda^c(t, \theta)$ für jeden vorgegebenen Wert von θ berechnet werden kann. Betrachtet man diese Daten als markierten Punkt-Prozeß N^* , dann kann man die Likelihood $L_\tau^*(\theta, \phi)$ bezüglich der Filterung (\mathcal{F}_t^c) , die durch N^* erzeugt wird, angeben. Die partielle Likelihood für N^c ist (vergleiche (2.2.13))

$$L_\tau^c(\theta) = \prod_{t \in \mathcal{T}} \prod_i d\Lambda_i^c(t, \theta)^{\Delta N_i^c(t)} (1 - d\Lambda^c(t, \theta))^{1 - \Delta N^c(t)}. \quad (2.3.2)$$

Wenn nur die partielle Likelihood L_τ^c für die Schätzung herangezogen wird, dann wird Information über Marken von N^* und über vorhergehende Ereignisse, deren Zeit nicht bekannt ist, verloren gehen. Üblicherweise werden diese Marken Information über θ enthalten, sodaß der Filter-Prozeß als *informativ für θ* bezeichnet wird. Ansonsten ist der Filter-Prozeß nicht-informativ für θ und $L_\tau^c(\theta)$ ist die Likelihood für θ .

In einem Extrembeispiel der Intervallzensierung wird nur ein diskretes Skelett des Prozesses N beobachtet zu Zeiten $0 = \tau_0 < \tau_1 < \dots < \tau_r \leq \tau$. Da man mit Wahrscheinlichkeit 1 niemals einen Sprung beobachten wird, wird der beobachtete Prozeß auf nichts reduziert, die partielle Likelihood $L_\tau^c(\theta) = 1$ ist konstant.

2.4 Teilweise spezifizierte Modelle

Bei unabhängiger Zensierung oder Filterung kann man eine partielle Likelihood für θ , den zu schätzenden Parameter, anschreiben, die die selbe Form hat, wie die Likelihood für die vollständigen Daten und nicht vom Nebenparameter ϕ abhängt. Deshalb kann die Likelihood berechnet werden, ohne daß man ein Modell für den Zensierungsmechanismus angeben muß. Man kann die Zensierungen so behandeln, als hätten sie zu fixen Zeiten stattgefunden.

Auch Gebrechlichkeits-Modelle (siehe 2.5) und marginale Modelle (siehe 2.6) sind teilweise spezifizierte Modelle.

Das Cox-Modell 2.1.2 ist deswegen nur teilweise spezifiziert, weil es für gegebene Kovariable geschätzt werden kann, ohne daß man deren Verteilung angeben muß. Allerdings sind die Kovariablen im allgemeinen Fall zeitabhängig und nicht fest bezüglich \mathcal{F}_0 , aber auch zeitabhängige Kovariable können bezüglich \mathcal{F}_0 deterministisch oder fix sein.

Beispiel 2.4.1. Überleben von insulin-abhängigen Diabetikern

Das Problem wurde schon in Beispiel 2.2.9 beschrieben. Die Zeitskala ist darin das Alter des Patienten, während das Alter bei der Diagnose eine zeitunabhängige Kovariable ist. Der Intensitätsprozeß des Ereignisses „Tod“ kann auch von der Kovariablen „Dauer der Erkrankung“ abhängen, die aus dem Alter bei der Diagnose berechnet werden kann. Daher ist der stochastische Prozeß $Z_i(t)$ = „Dauer der Erkrankung bei Patient i zur Zeit t “ an die Filtrierung (\mathcal{F}_t^c) , die durch die Daten erzeugt wird, angepaßt.

Eine zeitabhängige Kovariable wie in diesem Beispiel wird als *definierte zeitabhängige Kovariable* bezeichnet. Die (partielle) Likelihood bleibt unverändert und die Analyse kann so durchgeführt werden, als wären die Pfade im voraus fixiert worden.

Darüberhinaus gibt es *wirklich zufällige* Kovariable in dem Sinn, daß die Prozesse $Z_i(\cdot)$, $i = 1, \dots, n$, nicht automatisch an die betrachtete Filtrierung des Zählprozesses angepaßt sind. Dabei werden Hilfskovariable und interne Kovariable unterschieden. Eine Hilfskovariable ist beispielsweise die Luftverschmutzung bei einer Studie über Asthma-Anfälle. Interne Kovariable sind beispielsweise Laborbefunde, die in der Beobachtungszeit von den Patienten erhoben wurden und prognostisch relevante Information über Patienten enthalten.

Derartige Kovariable werden durch Vergrößerung der Filtrierung in das Modell des Zählprozesses einbezogen. Eine Möglichkeit ist es, das ganze System der unzensierten Beobachtungen als markierten Punkt-Prozeß N_Z zu betrachten. Darin gehören *innovative* Marken zu Ereignissen, die Gegenstand der Untersuchung sind, während *nicht-innovative* Marken insbesondere Änderungen der Kovariablen betreffen. Allerdings können durch Marken eines Punkt-Prozesses nur treppenförmige Pfade von Kovariablen modelliert werden. Für deterministische Kovariable sind kontinuierliche Änderungen weiterhin zulässig.

Über den vollständig beobachteten Prozeß N_Z mit Filtrierung (\mathcal{F}_t) soll nun ein Zensierungs- und Filterungsprozeß C gelegt werden, der an eine größere Filtrierung (\mathcal{G}_t) angepaßt ist. Auf diese Weise wird ein zensierter beziehungsweise gefilterter markierter Punkt-Prozeß N_Z^* erhalten. Nur ein Teil N^c davon zählt die Ereignisse, die Gegenstand der Untersuchung sind. Darüberhinaus erhält N^c die innovativen Marken. Es wird angenommen, daß man mittels N_Z^* für jeden Wert von θ den $(P_{\theta\phi}, (\mathcal{G}_t))$ -Kompensator von N^c angeben kann.

Sei (\mathcal{F}_t^c) die Filtrierung, die durch die Daten N_Z^* erzeugt wird. Die volle Likelihood $L_\tau^*(\theta, \phi)$ von N_Z^* bezüglich (\mathcal{F}_t^c) kann in eine partielle Likelihood

$L_\tau^c(\theta)$, die auf N^c beruht und nicht vom Neben-Parameter ϕ abhängt, und einen zweiten Faktor $L_\tau''(\theta, \phi)$ zerlegt werden, der möglicherweise nicht mehr von θ abhängt. Daher kann die statistische Schätzung auf L_τ^c allein beruhen, ohne den Zensierungsmechanismus oder den Kovariablen-Prozeß zu modellieren. Wenn jedoch der zweite Faktor tatsächlich von θ abhängt (die Zensierung oder der Kovariablen-Prozeß sind *informativ*), dann ist die statistische Schätzung mittels voller Likelihood $L_\tau^*(\theta, \phi)$ effizienter.

Ein partiell spezifiziertes Modell, bei dem nur der (\mathcal{F}_t^c) -Kompensator Λ_θ^c von N^c angegeben wird, hat nur beschränkte Vorhersagefähigkeit. Insbesondere sind zeitabhängige Kovariablen nicht erfaßt. Wenn man sie berücksichtigen will, muß man die entsprechenden Marken als innovativ betrachten und die entsprechenden Parameter der Verteilung θ statt ϕ zurechnen. Auch das Modell der konkurrierenden Risiken hat eingeschränkte Vorhersagefähigkeit.

Oft werden zeitabhängige Kovariablen nicht beobachtet, sodaß nicht einmal die partielle Likelihood L_τ^c berechnet werden kann. Das ist oft der Fall bei Modellen für Erkrankung und Tod (Beispiel 2.1.4). Wenn der Übergang von Zustand 1 in Zustand 2 von der zeitabhängigen Kovariablen $Z(t) = t - T$, das ist die Aufenthaltsdauer in Zustand 1 zur Zeit t , abhängt, dann ist bei jeder Eintrittszeit $V > T$ der Wert von T unbekannt und $Z(t)$ nicht beobachtbar.

2.5 Gebrechlichkeits-Modelle

Es sollen Modelle betrachtet werden, in denen der Intensitätsprozeß von einer nicht beobachtbaren Zufallsvariablen abhängt, die multiplikativ auf die Intensität wirkt. Bei der Analyse von Ausfällen oder Tod kann diese Zufallsvariable als Gebrechlichkeit (englisch: frailty) oder Anfälligkeit für Unfälle und ähnliches interpretiert werden. Mehrere Komponenten des Zählprozesses können den selben Wert der Gebrechlichkeits-Variablen teilen, so daß eine positive statistische Abhängigkeit zwischen den einzelnen Komponenten induziert wird. Der Begriff wurde von Vaupel, Manton und Stallard (1979) eingeführt.

Die Darstellung beruht auf (Andersen et al. 1993, pp 660). Gebrechlichkeits-Modelle werden kurz vorgestellt, da sie in gewissen Anwendungen eine Alternative zu marginalen Modellen sind. Gamma- und positiv stabile Gebrechlichkeits-Modelle werden aber als Variante bei der Simulation von abhängigen Überlebensdaten behandelt (siehe 6.1). Es wird hier nicht auf die Schätzung von Gebrechlichkeits-Modellen eingegangen.

Man betrachtet den multivariaten Zählprozeß N mit Komponenten N_{ihl} .

Komponenten mit dem selben ersten Index i haben den selben Wert der Gebrechlichkeits-Variablen Z_i . Man denke bei i an einen Wurf in einem Tierexperiment, h ist ein Stratum (zum Beispiel eine Behandlungsgruppe) und l sind Individuen innerhalb einer Familie beziehungsweise eines Stratums. Es wird keine Balanzierung des Designs verlangt.

Beispiel 2.5.1. Gebrechlichkeits-Modell mit k Stichproben

Der Zählprozeß

$$N = (N_{ihl} : i = 1, \dots, n; h = 1, \dots, k; l = 1, \dots, n_{ih})$$

habe Intensitätsprozeß

$$\lambda_{ihl}(t) = Z_i Y_{ihl}(t) \alpha_h(t).$$

Konkret könnte der Zusammenhang zwischen gewissen Individuen dieses k -Stichproben-Modells beispielsweise in einer Überlebensstudie einer Anzahl von n Familien mit $k = 4$ bestehen, wobei für die Mütter $h = 1$, $n_{i1} = 1$, für die Väter $h = 2$, $n_{i2} = 1$, für die Töchter $h = 3$, $n_{i3} = 0, 1, 2, \dots$ und für die Söhne $h = 4$, $n_{i4} = 0, 1, 2, \dots$ gilt.

Eine andere Anwendung wäre gegeben durch eine Wurf-geschichtete Studie, bei der ein Individuum (h, l) in Wurf i Behandlung h , $h = 1, \dots, k$ erhält und die Individuen $l = 1, \dots, n_{ih}$ vom gleichen Wurf abhängig sind.

Beispiel 2.5.2. Gebrechlichkeit im Regressionsmodell für das relative Risiko

Das Regressionsmodell in **2.1.2** kann für beobachtbare zeitabhängige Kovariable $X_{ihl}(t)$ und einen Wert der Gebrechlichkeits-Variablen Z_i gegeben werden durch

$$\lambda_{ihl}(t) = Z_i Y_{ihl}(t) \exp \{ \beta^\top X_{ihl}(t) \} \alpha_{h0}(t). \quad (2.5.1)$$

Wie im vorhergehenden Beispiel, kann damit die Abhängigkeit von Individuen berücksichtigt werden.

Eine univariate Version (also $k = 1$ und $n_1 = 1$) dieses Modells ist ein Cox'sches Regressionsmodell, in dem einige zeitunabhängigen Kovariablen X_{1i}, \dots, X_{mi} mit Regressionskoeffizienten $\gamma_1, \dots, \gamma_m$ entfallen oder nicht beobachtet wurden, so daß

$$Z_i = \exp(\gamma_1 X_{1i} + \dots + \gamma_m X_{mi}).$$

Diese Ansätze sind insofern einschränkend, als der Gebrechlichkeits-Parameter multiplikativ auf den Intensitätsprozeß wirkt und pro Familie i nur ein Gebrechlichkeits-Parameter Z_i vorgesehen ist, obwohl denkbar wäre, daß es für

jeden Typ von Ereignis einen eigenen Wert des Gebrechlichkeits-Parameters gibt. Zumeist ist das Modell nur unter bestimmten Annahmen über die Verteilung von Z schätzbar.

Zur weiteren Entwicklung des Modells wird auf den Fall $k = 1$ und $n_{i1} = 1$ eingeschränkt, da er von der Schreibweise her leicht zu behandeln ist und die Verallgemeinerbarkeit offensichtlich ist. Das Modell umfaßt daher den multivariaten Zählprozeß $N = (N_1, \dots, N_n)$ mit Intensitätsprozeß λ mit Komponenten von der Form

$$\lambda_i(t) = Z_i Y_i(t) \alpha(t). \quad (2.5.2)$$

Darin ist Y_i ein beobachtbarer, vorhersagbarer Prozeß, α eine Hasardfunktion und Z_i eine nicht beobachtbare Zufallsvariable. Die Komponenten werden zu Vektoren Y und Z zusammengefaßt.

Das Schätzproblem kann als Schätzung von fehlenden Werten aufgefaßt werden. Die vollständigen Daten bestehen aus dem Tripel N, Y und Z , während die unvollständigen Daten nur aus N und Y bestehen. Bei vollständigen Daten kann Z in Y aufgehen, sodaß ein Standardproblem mit multiplikativen Intensitäten vorliegt (vergleiche (2.1.7)). Die Schätzung der Gebrechlichkeits-Verteilung ist unabhängig von der Schätzung der Hasardraten möglich. Bei Skalenparametern der Gebrechlichkeits-Verteilung kann ein Identifikationsproblem auftreten (siehe 6.1.2).

Die Intensität (2.5.2) impliziert, daß der nicht beobachtbare Zufallsvektor Z \mathcal{G}_0 -meßbar in der Filtrierung (\mathcal{G}_t) , bezüglich der die Intensität von N definiert wurde. Bei der Berechnung von Likelihoods muß man auf Intensitäten zurückgreifen, die bezüglich kleinerer Filtrierungen definiert wurden, die von den Daten erzeugt werden, wie zum Beispiel die Filtrierung $(\mathcal{F}_t) = (\sigma\{N(s), Y(s), s \leq t\})$. Dieser Schritt ist auf einfache Weise dann möglich, wenn eine Produkt-Struktur angenommen wird. Diese Bedingung ergänzt das Modell (2.5.2). Da Z aufgrund von (2.5.2) implizit \mathcal{G}_0 -meßbar ist, würde man gerne haben, daß bezüglich $Z = z$, der Intensitätsprozeß λ von N gegeben ist durch

$$\lambda_i(t) = z_i Y_i(t) \alpha(t) \quad (2.5.3)$$

Wenn man die benötigte Struktur gleich in das Modell einbaut, dann kann der mathematische Aufwand klein gehalten werden. (Sonst wird die Blackwell-Theorie benötigt (Jacobsen 1982).) Sei daher $(\Omega', \mathcal{G}', (P'_\eta))$ der natürliche Wahrscheinlichkeitsraum, auf dem Z definiert ist: Es ist dies der n -dimensionale euklidische Raum mit der Borel'schen σ -Algebra. In der

nachstehenden Produktkonstruktion wird die Schreibweise $Z = \omega'$ verwendet. Sei $(\Omega'', (\mathcal{G}_t''), (P''_{\alpha z}))$ ein gefilterter Wahrscheinlichkeitsraum mit einer Familie von Wahrscheinlichkeitsmaßen und einem multivariaten Zählprozeß N , der so definiert ist, daß er unter dem Wahrscheinlichkeitsmaß $P''_{\alpha z}$ Intensitätsprozeß λ von der Form (2.5.3) hat. Sei $\omega'' = (N, Y)$. Der Produktraum $(\Omega, (\mathcal{G}_t), (P_{\eta\alpha}))$ trägt das Wahrscheinlichkeitsmaß $P_{\eta\alpha}$, das durch Mischung von $P''_{\alpha z}$ über z mit Wahrscheinlichkeit P'_η entstanden ist, also $P_{\eta\alpha}(A \times B) = \int_A P_{z\theta}(B) dP'_\eta(z)$. Auf dem neuen Wahrscheinlichkeitsraum sind sowohl Z als auch (N, Y) durch die selben Funktionen definiert, wie die Koordinaten $\omega = (\omega', \omega'')$ in den ursprünglichen Räumen. Bezüglich der Filtrierung (\mathcal{G}_t) mit $\mathcal{G}_t = \mathcal{G}' \otimes \mathcal{G}_t''$ hat N die Intensität (2.5.2), während sie bezüglich Z (z ist fix), oder bezüglich $\mathcal{G}' \otimes \{\Omega'', \emptyset\}$ die Intensität (2.5.3) hat.

Eine kleinere σ -Algebra als (\mathcal{G}_t) ist gegeben durch (\mathcal{F}_t) , wobei $\mathcal{F}_t = \{\Omega', \emptyset\} \otimes \mathcal{G}_t''$. Daher kann \mathcal{F}_t mit \mathcal{G}_t'' identifiziert werden, die somit die Interpretation der vom beobachtbaren Teil des Modells erzeugten Filtrierung hat. Die Filtrierung (\mathcal{G}_t) entsteht daraus dadurch, daß man zur Zeit 0 die hypothetische Beobachtung Z hinzufügt.

Zum Zweck der Schätzung wird nun angenommen, daß α von einem Parameter θ abhängt. θ kann euklidisch oder auch non-parametrisch, also unendlichdimensional, sein. Während bisher informative Zensierung nur möglicherweise die Effizienz der statistischen Schätzung beeinträchtigte, ist nun die Schätzung der Verteilung von Z überhaupt nur bei *nicht-informativer* Zensierung möglich. Der Grund dafür ist, daß über z integriert werden muß, damit man den beobachtbaren Teil des Modells erhält.

Sowohl (2.5.2) als auch (2.5.3) können nicht direkt für die Schätzung verwendet werden. Da die selbsterzeugte Filtrierung nur eine partielle Likelihood erzeugt, liegt ein partiell spezifiziertes Modell vor (siehe 2.4). Wenn z in den nicht spezifizierten Teilen des Modells z enthalten ist, dann kann nicht bestimmt werden, was die volle Likelihood ist, nachdem über z integriert wurde (siehe auch Beispiel 2.2.7).

Sei nun \mathcal{G}_t'' vom Typ, der in 2.2, 2.3 und 2.4 behandelt wurde: Ein „großer“ markierter Punkt-Prozeß, dessen Marken in innovative, die Sprünge von N entsprechen, und nicht-innovative Marken, die unter anderem Zensierungen (die Sprünge von Y) enthalten, unterteilt wurden. Dabei wurde schon auf den beobachtbaren Teil eines noch größeren Prozesses eingeschränkt. Wenn nun die bedingte Intensität von N tatsächlich durch (2.5.3) angegeben wird, dann wurde darin schon die versteckte Annahme einer unabhängigen Zensierung gemacht:

Bedingung 2.5.1. *Bezüglich $Z = z$ ist die Zensierung unabhängig.*

Für die weitere Analyse wird auch noch die folgende Bedingung benötigt:

Bedingung 2.5.2. Für $Z = z$ ist die Zensierung für z nicht-informativ.

Die partielle bedingte Likelihood von N ist dann

$$dP''_{\alpha z} = \prod_t \left\{ \prod_i (\lambda_{i\alpha z}(t))^{\Delta N_i(t)} (1 - \lambda_{\alpha z}(t) dt)^{1 - \Delta N_i(t)} \right\}. \quad (2.5.4)$$

Dabei ist $\lambda_{i\alpha z}(t) = z_i Y_i(t) \alpha(t)$ und die Punkte kennzeichnen die Summe über $i = 1, \dots, n$. Als Funktion von z ist (2.5.4) proportional zur bedingten Dichte der Daten (N, Y) bezüglich $Z = z$. Daher ist die bedingte Dichte von Z gegeben die Daten aufgrund der Bayes-Regel proportional zur (partiellen) Likelihood der vollständigen Daten $L^G(\alpha, \eta)$, das ist das Produkt von (2.5.4) und der marginalen Dichte von Z :

$$L^G(\alpha, \eta) = \prod_i \left\{ p(z_i; \eta) \prod_t (z_i Y_i(t) dA(t))^{\Delta N_i(t)} \exp \left(-z_i \int_0^\tau Y_i(s) dA(s) \right) \right\}. \quad (2.5.5)$$

Dabei ist $p(z_i; \eta)$ die Dichte der Gebrechlichkeits-Verteilung. (2.5.5) ist bei gegebenem η die volle beziehungsweise partielle Likelihood für α aufgrund der Beobachtung von Z , N und Y . Darüberhinaus ist es die volle marginale Likelihood bezüglich $Z = z$, wenn die Zensierung für α nicht-informativ ist. Wenn man aus diesem Ausdruck z herausintegriert, dann erhält man entweder eine volle Likelihood für α aufgrund der Beobachtung von N und Y oder die *marginale partielle Likelihood*. Aufgrund von Bedingung 2.5.2 ist die marginale partielle Likelihood gleich der *partiellen marginalen Likelihood*, das ist die partielle Likelihood für α aufgrund von N aus dem „marginalen“ oder „Experiment mit unvollständigen Daten“ (ohne Beobachtung von Z)(Gill 1992). Dies gilt im allgemeinen, wenn der Zensierungsprozeß keine Information über die Zufallsvariable enthält, über die integriert wird. Dasselbe gilt für η : Wenn für $Z = z$ die Zensierung keine Information über η enthält, dann ist (2.5.5) die volle Likelihood für die „vollständigen Daten“ Z , N und Y . Durch Herausintegrieren von z erhält man die volle Likelihood aufgrund der „unvollständigen Daten“ N und Y . Wenn (2.5.5) eine partielle Likelihood ist, dann liegt aufgrund von (Gill 1992) nach Integration über z eine partielle (marginale) Likelihood für η vor.

Die marginale partielle Likelihood ist

$$L^{\mathcal{F}}(\alpha, \eta) = \prod_i \int_0^\tau \prod_t (z_i Y_i(t) dA(t))^{\Delta N_i(t)} \exp\left(-z_i \int_0^\tau Y_i(s) dA(s)\right) p(z_i; \eta) dz_i. \quad (2.5.6)$$

Die ganze Konstruktion für die beobachtbaren Daten des Gebrechlichkeits-Modells ist im Ergebnis identisch mit der Spezifikation des Intensitätsprozesses $\lambda_{i\eta\alpha}^{\mathcal{F}}$ von N bezüglich der von den beobachteten Daten erzeugten Filtrierung als eigenständiges Modell. Im nachhinein könnte die Darstellung der Likelihood (2.5.6) als Resultat der Integration über eine Gebrechlichkeits-Verteilung als Motivation für die spezielle Form der Likelihood aufgefaßt werden.

Diese Art der Betrachtung ermöglicht es, unter Annahme von Γ -verteilter Gebrechlichkeit, die Nullhypothese eines Skalierungsparameters 0 zu testen, der am Rand des Parameterbereichs liegt, da im marginalen Intensitätsprozeß auch negative Werte des Skalierungsparameters zulässig sind (Andersen et al. 1993, p. 667).

2.6 Marginale Modelle

Der marginale Ansatz ist eine weitere Möglichkeit zur Modellierung von Daten, die durch Gruppenzugehörigkeit voneinander abhängig sind. Das Konzept geht auf (Liang und Zeger 1986) zurück, wo eine Methode angegeben wurde, wie im verallgemeinerten linearen Modell bei abhängigen Daten mittels (konsistenter) Schätzugleichungen und Gewichtsfunktionen konsistente Parameterschätzung möglich ist. Dabei muß nur die marginale Verteilung der Daten eines Gruppenmitglieds spezifiziert werden. Diese Schätzmethode kann auch auf Ereignisdaten angewendet werden. Allerdings will man die Methode mit multivariaten Zählprozessen formulieren, da dadurch das Problem der unvollständigen Daten, also Zensierung und Filterung, besser behandelt werden kann.

Seien Zählprozesse N_{ik} , $i = 1, \dots, n$, $k = 1, \dots, K$ gegeben. Zählprozesse mit dem selben Index i sind voneinander abhängig, während Prozesse mit unterschiedlichem erstem Index unabhängig sind. Die Zählprozesse N_{i1}, \dots, N_{iK} mit dem selben Index k werden als zum selben *Haufen* gehörig betrachtet. Diese Bezeichnung ist jedoch nicht vorteilhaft, wenn verschiedene

Arten von Ereignissen beim selben Individuum i indiziert werden und wird in diesem Fall daher nicht verwendet.

In vielen Anwendungen ist diese Abhängigkeit innerhalb des Haufens nicht von Interesse, aber aufwendig zu modellieren. Sie soll daher nicht in den Schätzer eingehen. Ein solcher Schätzer ist dann jedenfalls kein Maximum-Likelihood-Schätzer.

Betrachtet man jeweils nur das k -te Individuum für $i = 1, \dots, n$, dann sind die Daten unabhängig und diese „marginale“ Sichtweise stimmt mit einem Modell mit unabhängigen Beobachtungen überein. Das ist der Ausgangspunkt für die folgenden Definitionen:

Für einen festen Index k seien die Zählprozesse N_{ik} unabhängig. Der multivariate Zählprozeß $N_k = (N_{1k}, \dots, N_{nk})$ erzeugt die Filtrierung $(\mathcal{N}_t^{(k)})$. Eine Zufallsvariable $X_0^{(k)}$ erzeuge die σ -Algebra $\mathcal{F}_0^{(k)}$. Sei $\mathcal{F}_t^{(k)} = \mathcal{F}_0^{(k)} \vee \mathcal{N}_t^{(k)}$. Sei $(\mathcal{G}_t^{(k)}) \supseteq (\mathcal{F}_t^{(k)})$ eine Filtrierung, bezüglich der ein Filterungsprozeß $C_i^{(k)}$ vorhersagbar ist (siehe (2.3)). Wenn man sich auf unabhängige Filtrierung beschränkt, dann ist der $P_{\theta\phi}$ -Kompensator bezüglich $(\mathcal{F}_t^{(k)})$ und $(\mathcal{G}_t^{(k)})$ jeweils $\Lambda_i^{(k)}(\cdot, \theta)$. Der Kompensator des gefilterten Prozesses $N_{ik}^c(t) = \int_0^t C_{ik}(u) dN_{ik}(u)$ ist dann

$$\Lambda_{ik}^c(t, \theta) = \int_0^t C_{ik}(u) d\Lambda_{ik}(u, \theta). \quad (2.6.1)$$

Die Daten umfassen $X_0^{(k)}$, N_k^c und weitere Angaben über Zensierung, Stutzen und Filterung. Darüberhinaus können Angaben als Marken vorliegen, der entsprechende markierte Punkt-Prozeß wird mit N_k^* bezeichnet. Sei $L_\tau^{(k)*}$ die marginale Likelihood bezüglich $(\mathcal{F}_t^{(k)c})$. Dann ist die partielle marginale Likelihood gegeben durch (vergleiche mit (2.3.2)):

$$L_\tau^{(k)c}(\theta) = \prod_{t \in \mathcal{T}} \prod_i d\Lambda_i^{(k)c}(t, \theta)^{\Delta N_i^{(k)c}(t)} (1 - d\Lambda_i^{(k)c}(t, \theta))^{1 - \Delta N_i^{(k)c}(t)}. \quad (2.6.2)$$

Das marginale Modell kann auch dann angewendet werden, wenn ein Individuum mehrere Ereignisse erfahren kann und zuwenig Information zur Verfügung steht, um den multivariaten Zählprozeß adäquat zu modellieren (siehe Beispiel 2.6.2).

Beispiel 2.6.1. Eine Ribavirin-Studie

In (Wei, Lin und Weissfeld 1989) wird folgendes Beispiel gegeben: In einer randomisierten klinischen Studie wurden Patienten mit AIDS (englische Abkürzung für „acquired immune deficiency syndrome“) zufallsgesteuert den 3 Behandlungen Placebo, niedrig dosiertes Ribavirin und hoch dosiertes Ribavirin zugewiesen. Die Wirksamkeit von Ribavirin gegen Retroviren sollte gezeigt werden. 4, 8 und 12 Wochen nach Behandlungsbeginn wurden jeweils Blutproben genommen, und in regelmäßigen Abständen wurde aus dem Serum der Gehalt an p24-Antigen bestimmt. Aus der Dauer bis zur Nachweisbarkeit des Virus in der Probe wurde auf den Virusgehalt geschlossen. Fehlende Werte entstanden dadurch, daß die Patienten für die Probenahme nicht verfügbar waren, oder weil es keine Messung gab, da die Probe verdorben war, oder weil das Virus in der Probe länger als 4 Wochen lang nicht nachweisbar war.

Da aus der Dauer bis zum Virusnachweis auf die Stärke der Erkrankung des Patienten geschlossen werden soll, soll ein Modell dafür angegeben werden. Mittels eines Gebrechlichkeits-Modells könnte die Abhängigkeit der Werte eines Patienten modelliert werden, da man annehmen könnte, daß es innerhalb der Patienten einer gewisse Verteilung der Stärke des Virus-Befalls gibt, der sich multiplikativ auf die Hasardfunktion der Dauer bis zum Virusnachweis auswirkt. Allerdings sind auch viele andere multivariate Verteilungen möglich.

Betrachtet man jeden Zeitpunkt der Probenahme als eigenes Modell, so kann man den Vergleich der Behandlungsgruppen in einem Cox-Modell mit der Gruppenzugehörigkeit als Kovariable modellieren. Mit Hilfe von marginalen Schätz- und Testmethoden kann ein statistischer Test angegeben werden für die Nullhypothese, daß sich die 3 Gruppen in keinem der 3 Zeitpunkte der Probenahme unterscheiden.

Das Cox-Modell für die Zeitpunkte $k = 1, 2, 3$ enthält 2 Indikator-Variable Z_1 und Z_2 für die Zugehörigkeit zur Behandlungsgruppe. Sei $X_0^{(k)} = Z = (Z_1, Z_2)$. Der erste Index i bezeichnet das Individuum, der zweite Index k bezeichnet den Zeitpunkt der Probenahme. Der marginale Intensitätsprozeß ist dann $\lambda_{ik}(t, \beta_k) = I(X_{ik} \leq t) \alpha_0^{(k)}(t) \exp \{ \beta_k^\top Z_{ik} \}$. Fehlende Werte wegen nicht vorhandener oder verdorbener Blutprobe werden als unabhängige zufällige Rechts-Zensierung betrachtet (Beispiel 2.2.4), während die Rechts-Zensierung nach 28 Tagen eine Zensierung vom Typ I ist (Beispiel 2.2.1). Bezüglich der von den beobachtbaren Daten erzeugten Filtrierung $(\mathcal{F}_t^{(k)c})$ ist der Intensitätsprozeß für die Nachweisbarkeit vor Zensierung (das ist der „zensierte Zählprozeß“) $\lambda_{ik}^c(t, \beta_k) = Y_{ik}(t) \alpha_0^{(k)}(t) \exp \{ \beta_k^\top Z_{ik} \}$. Der Indikator-Prozeß Y_{ik} steht für die Beobachtbarkeit des Beginns der Nach-

weisbarkeit von p24-Antigenen. Die partielle marginale Likelihood berechnet sich dann aufgrund von (2.6.2). Der Effekt von Ribavirin könnte dann durch einen Test für die Nullhypothese $\beta_k = 0$ für $k = 1, 2, 3$ geprüft werden. Als Alternative dazu kommt auch ein in (Wei et al. 1989) gegebener marginaler multipler Test in Frage.

Beispiel 2.6.2. Studie über kolorektales Karzinom

Dieses Beispiel wird in (Lin 1994) angegeben. Der Effekt von adjuvanter Therapie mit Levamisol und Fluorouracil auf Patienten mit reseziertem kolorektalem Karzinom soll untersucht werden. Sowohl Rezidiv¹ als auch die tumorbezogene Überlebenszeit sind als Zielgrößen von Interesse.

Es kann nicht ausgeschlossen werden, daß Patienten, die ohne Rezidiv versterben, Information über Rezidiv tragen. Sie werden daher in der Analyse der Rezidive nicht als Zensierungen, sondern als Ereignisse betrachtet. Man erhält die *tumorfremie Überlebenszeit*.

Tumorfremie Überlebenszeit und Tod am Tumor werden getrennt modelliert. Die Zugehörigkeit zu einer Behandlungsgruppe wird durch zeitunabhängige Kovariable im Cox-Modell erfaßt. Auch prognostische Faktoren können in das Modell einbezogen werden. Wie im vorhergehenden Beispiel wird für jede Art von Ereignis eine eigene Hasardrate angenommen.

$j = 1$ beziehe sich auf tumorfreies Überleben, während $j = 2$ sich auf Tod durch Tumor bezieht. $Y_{ij}(t)$ ist ein Indikator dafür, daß Patient i der Risikomenge für das Ereignis j angehört. Bei $j = 1$ ist dies das Minimum der Zeit von der Operation bis zum Rezidiv, bis zum Tod, und bis zum Ende der Beobachtungszeit von Patient i . Analog ist die Zugehörigkeit zur Risikomenge für $j = 2$ definiert durch die Zeit von der Operation bis zum Tod mit Tumor beziehungsweise dem Ende der Beobachtungszeit. N_{ij}^c springt bei der Beobachtung eines Rezidivs oder des Todes vor Rezidiv beziehungsweise beim Tod des Patienten nach Rezidiv, und ist danach jeweils absorbiert. Das Modell mit proportionalen Hasardraten ist gegeben durch

$$\lambda_{ij}^c(t, \beta_j) = Y_{ij}(t) \lambda_j(t) \exp \{ \beta_j^\top Z_{ij} \}.$$

Z_{ij} enthält Angaben über die Gruppenzugehörigkeit des Patienten und Werte von prognostischen Faktoren die zur Zeit der Operation bekannt sind. Zumeist wird in einer derartigen Anwendung $Z_{i1} = Z_{i2}$ gelten.

¹Ein Tumorrezidiv ist das Wiederauftreten eines histologisch gleichartigen Tumors am gleichen Ort oder im gleichen Organ nach vorausgegangener radikaler Behandlung (Pschyrembel 1994)

Beispiel 2.6.3. GCD-Studie

Auch dieses Beispiel wird in (Lin 1994) angegeben. Bei der septischen Granulomatose (engl.: chronic granulomatous disease, abgekürzt GCD), einer vererbten, seltenen Störung des oxidativen Metabolismus der Granulozyten, kommt es zu wiederholten pyogenen Infektionen, die tödlich sein können. Zur Prüfung der Wirksamkeit von Gamma-Interferon bei der Behandlung dieser Störung wurde eine placebo-kontrollierte klinische Studie durchgeführt.

In der für die Beurteilung des Behandlungseffekts maßgebenden Analyse wurde nur die Zeit bis zur ersten Infektion ausgewertet. Da die Ärzte jedoch eigentlich die Wirkung auf die *Häufigkeit* von Infektionen untersuchen wollten, wurde nach einer dafür adäquaten Analyse gesucht. Ein Modell mit multiplikativen Intensitäten (siehe Seite 37) wie in (Andersen und Gill 1982) berücksichtigt keine Änderungen der Hazardrate durch vorhergegangene Infektionen. Eine zeitabhängige Kovariable, die bis 60 Tage nach einer Infektion 1 ist und sonst 0, könnte diesen Effekt berücksichtigen. Allerdings erscheint die Annahme eines derartigen Effekts willkürlich.

In (Lin 1994) wird vorgeschlagen, analog zum vorhergehenden Beispiel, die erste, zweite und dritte Infektion jeweils als verschiedene Typen von Ereignissen zu betrachten. Die Zeit bis zum Auftreten der k -ten Infektion ist dabei die zu modellierende Zufallsvariable. Der Regressionsparameter des Cox-Modells für die Indikatorvariable der Behandlungsgruppe wird für jede Infektion getrennt geschätzt. In einer Modellvariante wird für alle Infektionen ein gemeinsamer Parameter für den Behandlungseffekt angenommen.

Auch ein Gebrechlichkeits-Modell könnte höhere individuelle Anfälligkeit für Infektionen berücksichtigen.

Beispiel 2.6.4. Retinopathia diabetica

In einer klinischen Studie wurde bei Patienten mit Retinopathia diabetica zufällig ein Auge ausgewählt und durch Photokoagulation mittels Laser behandelt. Es sollte untersucht werden, ob mit dieser Therapie die Erblindung des Auges hinausgezögert werden kann. Dazu wurden die Patienten nach der Operation durch mehrere Jahre beobachtet, um den Zeitpunkt des Erblindens eines Auges festzustellen. Es wird angenommen, daß es einen Zusammenhang bei der Erblindung der beiden Augen eines Patienten gibt (nach Lin (1994)).

Beispiel 2.6.5. Schizophrenie-Studie

In einer genetisch-epidemiologischen Studie wurden 487 verwandte ersten Grades von 93 weiblichen Probanden einbezogen (1 bis 12 Verwandte pro Proband) (Lin 1994). Unter anderem wurde untersucht, ob das Risiko einer

affektiven Erkrankung (Depression, Manie oder beides) eines Verwandten vom Alter des Patienten beim Ausbrechen von Schizophrenie beim Probanden abhängt.

Das Alter des Probanden geht durch eine Indikatorfunktion ein (1 für Alter unter 16 und 0 für Alter ab 16 Jahren). Das Geschlecht des Verwandten ist eine weitere Kovariable. Die Verwandten eines Probanden werden als Haufen betrachtet und durchnummeriert. Wie im vorhergehenden Beispiel werden die Verwandten als austauschbar angesehen, so daß für alle eine gemeinsame marginale Hasardfunktion angenommen werden kann. Eine naheliegende Zeitskala ist die Zeit ab Diagnose der Erkrankung beim Probanden. Die Kovariablen wirken multiplikativ auf die zugrundeliegende Hasardfunktion (Cox-Modell).

Beispiel 2.6.6. Reverschlüsse nach PTLA

Okkludierte Beinarterien können durch perkutane transluminale Laser-Angioplastie (PTLA) wieder durchgängig gemacht werden. In einer Studie an der Medizinischen Universitätsklinik der Karl-Franzens-Universität wurden 74 Patienten nach gelungenem Eingriff regelmäßig klinisch und angiographisch untersucht. Ein *klinischer Reverschluß* ist dadurch definiert, daß der Patient befindet, daß der Zustand wieder so schlecht ist, wie vor dem Eingriff. Ein *angiographischer Reverschluß* liegt vor, wenn der behandelte Abschnitt der Beinarterie in der Angiographie zu 100% verschlossen ist. Wenn die Arterie sowohl klinisch als auch angiographisch verschlossen ist, dann wurde eine *Reintervention* (nochmalige PTLA) durchgeführt, oder der Patient wurde Bypass-operiert. Beim dritten Verschluß wurde immer Bypass-operiert.

Die Fragestellung war, wie der Verlauf der Offenheitsrate bei den verschiedenen Kriterien für Reverschlüsse ist, und ob es Unterschiede zwischen den Kriterien gibt. Die Kriterien sind der erste angiographische und klinische Reverschluß, sowie der zweite angiographische Reverschluß nach Reintervention. Für die Beobachtung eines dieser Ereignisse liegt zunächst einmal progressive Zensierung vom Typ I mit „staggered entry“ vor. Für erstmalige klinische und angiographische Reverschlüsse ist das die einzige Möglichkeit der Zensierung. Bei den klinischen und angiographischen Reverschlüssen nach Reintervention kommt noch die Zensierung durch eine Bypass-Operation nach erstmaligem Reverschluß dazu. Es wird angenommen, daß diese Zensierung unabhängig ist.

Der Verlauf der Offenheitsraten kann nun durch einen Schätzer der Überlebensfunktion bestimmt werden. Beim Vergleich von solchen Verläufen muß die Abhängigkeit bei Daten vom selben Individuum berücksichtigt werden. Die Daten für die 4 Verläufe der Offenheitsraten werden so betrachtet, als

wären sie durch jeweils 4 Individuen, die zu einer Gruppe (Haufen) gehören, entstanden. Der Index $j = 1$ gehört dann zum Kriterium „erster klinischer Reverschluß“ und so weiter.

Beispiel 2.6.7. Liegedauer von Teleskopkronen

In einer retrospektiven Kohortenstudie wurden an der Universitätsklinik für Zahn-, Mund- und Kieferheilkunde der Karl-Franzens-Universität alle Patienten, die Teleskopkronen erhalten hatten, zu einer Nachuntersuchung gebeten (Jerebic 1996). Im Zuge dieser Nachuntersuchung wurde der Zahnstatus und die Liegedauer der Teleskopkronen erhoben. Abgesehen davon, daß auch das Erscheinen zur Nachuntersuchung als unabhängige Zensierung betrachtet werden kann, liegt progressive Zensierung vom Typ I vor (Beispiel 2.2.3). Bezüglich der Liegedauer war der Einfluß von Kovariablen, die den Typ der Prothese, die Zahnhygiene, die Art des Zahnes und den Kiefer repräsentieren, in ein Cox-Modell aufgenommen worden. Für alle Teleskopkronen wurde eine gemeinsame zugrundeliegende Hasard-Funktion angenommen. Das heißt, die Teleskopkronen eines Patienten wurden diesbezüglich als austauschbar angesehen, alle Unterschiede zwischen den Zähnen werden durch die Kovariablen erfaßt. Eine entsprechende statische Modellierung und Analyse wird in Kapitel 7 vorgestellt.

Beispiel 2.6.8. Mortalität und diabetische Nephropathie (DN) bei insulinabhängigen Diabetikern

Dieses Modell wurde schon in Beispiel 2.1.4 vorgestellt. Es wird in (Andersen et al. 1993, pp478ff) als Beispiel für das Cox-Modell bei multivariaten Zählprozessen gegeben. Die Übergänge sind $h = 02$, Tod ohne DN, $h = 01$, Auftreten von DN und $h = 12$, Tod nach Auftreten von DN. Ein mögliches Modell enthält eine Indikatorvariable für das Geschlecht als Kovariable in einem Cox-Modell für einen multivariaten Zählprozeß (Die Komponenten des Zählprozesses erfüllen jeweils ein Cox-Modell, $\beta = (\beta_{01}, \beta_{02}, \beta_{12})$)

$$\alpha_{h,i}(t)Y_j(t) = \alpha_{h,0}Y_{ji}(t) \exp \{ \beta_{jk}^\top Z_{hi} \}. \quad (2.6.3)$$

Das Modell ist als Markov-Prozeß spezifiziert. Man kann jedoch zusätzlich annehmen, daß auch das Alter des Patienten, die Dauer der DN und auch die Zeit, die bis zur Erkrankung an DN vergeht, Einfluß auf die Intensität des Übergangs von 1 nach 2 hat. In einem marginalen Modell wird für jedes der 3 Ereignisse (Auftreten von DN, Tod mit DN und Tod ohne DN) jeweils ein Cox-Modell geschätzt. Der Unterschied zum multivariaten Modell wirkt sich erst dann aus, wenn die gemeinsame Verteilung der Schätzungen der Regressionsparameter benötigt wird.

Kapitel 3

U -Statistiken

3.1 Definitionen

Üblicherweise werden U -Statistiken auf unabhängig identisch verteilte Zufallsvariable angewendet. Dennoch können die wichtigsten Ergebnisse, nämlich der zentrale Grenzwertsatz für U -Statistiken und der Jackknife-Schätzer für die Varianz der U -Statistik auch für nicht identisch verteilt Zufallsvariable gegeben werden (Puri und Sen 1971, Lee 1990).

In dieser Abhandlung werden p -dimensionale U -Statistiken benötigt. Es erscheint zweckmäßig, die Kovarianz zweier vektor-wertiger Zufallsvariablen X und Y durch $\text{Cov}(X, Y) = \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y)^\top$ zu definieren und unter der Varianz Var einer vektor-wertigen Zufallsvariablen $\text{Cov}(X, X)$, die üblicherweise als Kovarianzmatrix bezeichnet wird, zu verstehen.

Ab nun sei X_1, \dots, X_n eine Stichprobe von unabhängigen Zufallselementen eines metrischen Raumes. Die Kernfunktion ψ vom Grad k sei eine symmetrische Funktion mit k Argumente und mit Werten in \mathbb{R}^p . Dann ist die U -Statistik vom Grad k zum Kern ψ definiert durch

$$U_n = \binom{n}{k}^{-1} \sum_{(n,k)} \psi(X_{i_1}, \dots, X_{i_k}). \quad (3.1.1)$$

Mit der Indexmenge (n, k) wird die Summe über alle $\binom{n}{k}$ (ungeordneten) Teilmengen von $\{1, \dots, n\}$ vom Umfang k bezeichnet. Sei $\theta(i_1, \dots, i_k) = \mathbb{E} \psi(X_{i_1}, \dots, X_{i_k})$. Der Erwartungswert ist

$$\mathbb{E} U_n = \binom{n}{k}^{-1} \sum_{(n,k)} \theta(i_1, \dots, i_k). \quad (3.1.2)$$

Seien nun S und T Teilmengen von $\{1, \dots, n\}$ vom Umfang k und $\psi(S)$ sei der Kern ψ angewandt auf die Zufallselemente mit den Indizes aus S als Argumenten. Sei $\sigma^2(S, T) = \text{Cov}(\psi(S), \psi(T))$. Dann erhält man

$$\begin{aligned} \text{Var } U_n &= \binom{n}{k}^{-2} \sum_{(n,k)} \sum_{(n,k)} \sigma^2(S, T) \\ &= \binom{n}{k}^{-1} \sum_{c=1}^k \binom{k}{c} \binom{n-k}{k-c} \bar{\sigma}_{c,n}^2. \end{aligned} \quad (3.1.3)$$

Darin ist $\bar{\sigma}_{c,n}^2$ die mittlere Kovarianz aller Ausdrücke mit c gemeinsamen Indizes:

$$\bar{\sigma}_{c,n}^2 = \left\{ \binom{n}{k} \binom{k}{c} \binom{n-k}{k-c} \right\}^{-1} \sum_{|T \cap S|=c} \sigma^2(S, T). \quad (3.1.4)$$

Für den zentralen Grenzwertsatz für U -Statistiken wird die gleichmäßige Beschränktheit von $|\text{Var } \psi(T)|$ durch eine Konstante A über alle k -Kombinationen T von Indizes aus $\{1, \dots, n\}$ benötigt¹. Dann gilt auch $|\sigma^2(S, T)| < A$ und $|\bar{\sigma}_{c,n}^2| < A$ für $c = 1, \dots, k$. Daher ist (Lee 1990, p. 42)

$$\text{Var } U_n = k^2 n^{-1} \bar{\sigma}_{1,n}^2 + o(n^{-1}). \quad (3.1.5)$$

Folgerung 3.1.1. *Sei ψ_n ein Kern vom Grad k , der von n abhängt. Es gilt für den damit gemäß (3.1.1) gebildeten Ausdruck U_n*

$$\text{Var } U_n = o(n^{-1}), \quad (3.1.6)$$

wenn es Konstanten $A_n = o(1)$ gibt mit $|\text{Var } \psi_n(T)| < A_n$ für alle Kombinationen T von Indizes aus $\{1, \dots, n\}$.

Beweis. Die obenstehenden Überlegungen gelten auch für einen Kern, der von n abhängt. Wegen (3.1.5) und $|\sigma^2(S, T)| < A$ ist $A_n = o(1)$ hinreichend für (3.1.6). \square

Der folgende Satz wird in den Anwendungen den zentralen Grenzwertsatz für U -Statistiken ersetzen, indem gezeigt wird, daß die U -Statistik wie eine Summe von unabhängigen Zufallsvariablen verteilt ist.

¹ $|\cdot|$ ist für Vektoren und Matrizen als Maximum der Beträge über die Komponenten definiert

Satz 3.1.2. Sei (X_i) eine Folge von unabhängigen Zufallselementen eines metrischen Raumes. Sei

$$\psi_{1,i}(x) = \binom{n-1}{k-1}^{-1} \sum_{(n-1,k-1)}^{(-i)} \psi_{1,i;i_2,\dots,i_k}(x), \quad (3.1.7)$$

worin die Summe über die $\binom{n-1}{k-1}$ -elementigen $(k-1)$ -Teilmengen $\{i_2, \dots, i_k\}$ von $\{1, \dots, n\}$ gemeint ist, die i nicht enthalten und

$$\psi_{1,i;i_2,\dots,i_k}(x) = \mathbb{E} \psi(x, X_{i_2}, \dots, X_{i_k}) - \mathbb{E} \psi(X_i, X_{i_2}, \dots, X_{i_k}) \quad (3.1.8)$$

definiert ist. Weiters sei $\text{Var} \psi(X_{i_1}, \dots, X_{i_k}) < A$ für eine Konstante A . Sei $S_n = \frac{k}{n} \sum_{i=1}^n \psi_{1,i}(X_i)$.

Dann gilt

$$n^{1/2} ((U_n - \mathbb{E} U_n) - S_n) \xrightarrow{P} 0. \quad (3.1.9)$$

Beweis. Die Vorgangsweise ist ähnlich wie in (Lee 1990, p144f). Es genügt zu zeigen, daß die Varianz von $n^{1/2} ((U_n - \mathbb{E} U_n) - S_n)$ in Wahrscheinlichkeit gegen 0 konvergiert. Da der Erwartungswert von S_n gleich 0 ist, genügt es,

$$n \mathbb{E} (U_n - \mathbb{E} U_n - S_n)^{\otimes 2}$$

zu betrachten.² Nun ist

$$\begin{aligned} & \mathbb{E} (U_n - \mathbb{E} U_n) S_n^\top \\ &= \frac{k}{n} \binom{n}{k}^{-1} \sum_{(n,k)} \sum_{i=1}^n \mathbb{E} (\psi(X_{i_1}, \dots, X_{i_k}) - \theta(i_1, \dots, i_k)) \psi_{1,i}(X_i)^\top \\ &= \frac{k}{n} \binom{n}{k}^{-1} \sum_{i=1}^n \sum_{(n-1,k-1)}^{(-i)} \mathbb{E} (\psi(X_i, X_{i_2}, \dots, X_{i_k}) - \theta(i, i_2, \dots, i_k)) \psi_{1,i}(X_i)^\top \\ &= \frac{k}{n} \binom{n}{k}^{-1} \sum_{i=1}^n \binom{n-1}{k-1} \text{Var} \psi_{1,i}(X_i) \\ &= \text{Var} S_n \end{aligned}$$

und

$$\begin{aligned} \text{Var} S_n &= \frac{k^2}{n^2} \sum_{i=1}^n \text{Var} \psi_{1,i}(X_i) = \frac{k^2}{n^2} \sum_{i=1}^n \binom{n-1}{k-1}^{-2} \times \\ & \sum_{(n-1,k-1)}^{(-i)} \sum_{(n-1,k-1)}^{(-j)} \text{Cov} (\psi(X_i, X_{i_2}, \dots, X_{i_k}), \psi(X_j, X_{j_2}, \dots, X_{j_k})). \quad (3.1.10) \end{aligned}$$

²Schreibweise: $x^{\otimes 2} = xx^\top$.

Daher gilt unter Verwendung von (3.1.5)

$$\begin{aligned}
 \text{Var } S_n &= \frac{k^2}{n^2} \binom{n-1}{k-1}^{-2} \sum_{|S \cap T|=1} \sigma^2(S, T) + o(n^{-1}) \\
 &= \binom{n}{k}^{-2} \binom{n}{k} \binom{k}{1} \binom{n-k}{k-1} \bar{\sigma}_{1,n}^2 + o(n^{-1}) \\
 &= k^2 n^{-1} \bar{\sigma}_{1,n}^2 + o(n^{-1}) \\
 &= \text{Var } U_n + o(n^{-1}).
 \end{aligned}$$

Daher konvergiert (3.1.9) in Wahrscheinlichkeit gegen 0. \square

3.2 Der Jackknife-Schätzer für die Varianz

Ein Jackknife-Schätzer zu einer Statistik $T_n = T_n(X_1, \dots, X_n)$ beruht auf den *Pseudowerten*

$$\hat{T}_i = nT_n - (n-1)T_{n-1}(-i).$$

Die Notation $T_{n-1}(-i)$ bedeutet

$$T_{n-1}(-i) = T_{n-1}(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n),$$

das heißt, daß die Statistik T_{n-1} auf die Stichprobe angewendet wird, nachdem die i -te Beobachtung entfernt wurde. Das könnte physisch mittels eines Taschenmessers (engl: Jackknife) vorgenommen werden, was die Bezeichnung des Schätzers erklärt.

Wiederum ist die Theorie zumeist für identisch verteilte Stichproben formuliert. Viele Ergebnisse gelten jedoch auch in allgemeineren Situationen. Der Jackknife-Schätzer \tilde{T}_n zu T_n ist der Mittelwert der Pseudowerte:

$$\tilde{T}_n = n^{-1} \sum_{i=1}^n \hat{T}_i. \quad (3.2.1)$$

Der Jackknife-Varianz-Schätzer von T_n beziehungsweise \tilde{T}_n ist definiert durch

$$\hat{\mathbf{V}}_{\text{Jack}}(T_n) = \frac{1}{n(n-1)} \sum_{i=1}^n \left(\hat{T}_i - \tilde{T}_n \right)^{\otimes 2} \quad (3.2.2)$$

Mit Hilfe der $T_{n-1}(-i)$ schreibt man

$$\hat{\mathbf{V}}_{\text{Jack}}(T_n) = \frac{n-1}{n} \sum_{i=1}^n (T_{n-1}(-i) - \bar{T}_{n-1})^{\otimes 2}, \quad (3.2.3)$$

wobei $\bar{T}_{n-1} = n^{-1} \sum_{i=1}^n T_{n-1}(-i)$.

Aufgrund einer Invarianz-Eigenschaft von U -Statistiken stimmt U_n mit \tilde{U}_n überein (Lee 1990, Satz 3 aus Abschnitt 4.1). Daher ist für U -Statistiken nur der Jackknife-Varianzschätzer interessant. Der folgende Satz aus (Lee 1990, p218) kann auf den nicht identisch verteilten Fall von vektorwertigen U -Statistiken unmittelbar übertragen werden:

Satz 3.2.1. *Sei U_n eine U -Statistik vom Grad k mit Kern ψ . Sei*

$$S_c = \sum_{|S \cap T|=c} \psi(S)\psi(T)^\top \quad c = 0, 1, \dots, k. \quad (3.2.4)$$

Dann gilt

$$\hat{\mathbf{V}}_{\text{Jack}}(U_n) = n^{-2}(n-1) \binom{n-1}{k}^{-2} \sum_{c=0}^k (cn - k^2) S_c. \quad (3.2.5)$$

Falls $\mathbb{E} \psi(S) = \theta$ und $|\text{Var} \psi(S)| < A$ für alle S aus (n, k) gilt auch

$$\mathbb{E} \hat{\mathbf{V}}_{\text{Jack}}(U_n) = \text{Var} U_n + o(n^{-1}). \quad (3.2.6)$$

Beweis. Der Beweis von (3.2.5) kann unmittelbar aus (Lee 1990) übernommen werden.

Satz 3.2.3 weiter unten kann zum Beweis von (3.2.6) nicht angewendet werden, da im nicht identisch verteilten Fall die sogenannte Hoeffding-Zerlegung nicht zur Verfügung steht. Daher muß ein direkter Weg beschritten werden.

Nach (3.1.4) ist $\mathbb{E} S_1 = \binom{n}{k} \binom{k}{1} \binom{n-k}{k-1} \bar{\sigma}_{1,n}^2 + \sum_{|S \cap T|=1} \theta^{\otimes 2}$. Man erhält daher

$$\begin{aligned}
& \mathbb{E} \hat{\mathbf{V}}_{\text{Jack}}(U_n) \\
&= \mathbb{E} n^{-2} (n-1) \binom{n-1}{k}^{-2} \sum_{c=0}^k (cn - k^2) S_c \\
&= n^{-2} (n-1) \binom{n-1}{k}^{-2} \left[(n-k^2) \binom{n}{k} \binom{k}{1} \binom{n-k}{k-1} \bar{\sigma}_{1,n}^2 \right. \\
&\quad \left. + \underbrace{(n-k^2) \binom{n}{k} \binom{k}{1} \binom{n-k}{k-1} \theta^{\otimes 2} - k^2 \binom{n}{k} \binom{n-k}{k} \theta^{\otimes 2}}_{O(n^{2k-1})} \right. \\
&\quad \left. + \underbrace{\sum_{c=2}^k (cn - k^2) S_c}_{=o(S_1)} \right] \\
&= \frac{(n-1) k!^2 (n-k-1)!^2}{n^2 (n-1)!^2} \frac{n!}{k!(n-k)!} \frac{k(n-k)!(n-k^2)}{(k-1)!(n-2k+1)!} \bar{\sigma}_{1,n}^2 + o(n^{-1}) \\
&= \frac{n-k^2}{n} \frac{(n-k-1)!^2}{(n-2)!(n-2k+1)!} k^2 \bar{\sigma}_{1,n}^2 + o(n^{-1}) \\
&= \frac{k^2}{n} \bar{\sigma}_{1,n}^2 + o(n^{-1}) \\
&= \text{Var } U_n + o(n^{-1}).
\end{aligned}$$

□

Die Darstellung (3.2.5) aus Satz 3.2.1 liefert auch die Konsistenz von $n \hat{\mathbf{V}}_{\text{Jack}}(U_n)$ für $n \text{Var } U_n$ (Lee 1990, p. 221). Die Ausdrücke S_c sind nämlich als U -Statistiken darstellbar: Sei $\Psi^{(c)}$ der Kern vom Grad $2k-c$, definiert durch

$$\begin{aligned}
& \Psi^{(c)}(x_1, \dots, x_{2k-c}) \\
&= \frac{1}{(2k-c)!} \sum_{(2k-c)} \psi(x_{i_1}, \dots, x_{i_k}) \psi(x_{i_{k-c+1}}, \dots, x_{i_{2k-c}})^\top \quad (3.2.7)
\end{aligned}$$

Dabei bedeutet $\sum_{(2k-c)}$ die Summe über alle Permutationen von Argumenten $x_{i_1}, \dots, x_{i_{2k-c}}$. $\Psi^{(c)}$ ist also eine symmetrisierte Version von $\psi(x_{i_1}, \dots, x_{i_k}) \psi(x_{i_{k-c+1}}, \dots, x_{i_{2k-c}})^\top$. Sei $U_n^{(c)}$ die auf $\Psi^{(c)}$ basierende U -Statistik mit

$$U_n^{(c)} = N_c^{-1} S_c \quad (3.2.8)$$

mit $N_c = \binom{n}{k} \binom{k}{c} \binom{n-k}{k-c}$, der Anzahl der Möglichkeiten für die Wahl von Mengen T und S vom Umfang k mit $|S \cap T| = c$. Aufgrund von Satz 3.1.2 konvergiert $U_n^{(c)}$ in Wahrscheinlichkeit gegen $\mathbb{E} U_n^{(c)}$. Der Faktor N_c^{-1} in (3.2.8) ist $O(n^{-2k+c})$, während S_c in (3.2.5) mit einem Faktor $O(n^{-2k-1+\min(c,1)})$ versehen ist. Folglich konvergieren die entsprechenden Summanden von $n\hat{V}_{\text{Jack}}(U_n)$ für $c = 0, 1$ in Wahrscheinlichkeit gegen die entsprechenden Erwartungswerte. Die Beiträge mit $c > 1$ konvergieren gegen 0, diese sind jedoch asymptotisch ohnehin zu vernachlässigen.

Eine Darstellung des Jackknife-Varianzschätzers für U -Statistiken kann nun mit Hilfe der folgenden Ausdrücke angegeben werden:

$$V_n^{(1)}(X_i) = \binom{n-1}{k-1}^{-1} \sum_{(i-1, k-1)}^{(-i)} \psi(X_i, X_{i_2}, \dots, X_{i_k}). \quad (3.2.9)$$

Es wird also über die Indexmenge aller Teilmengen S vom Umfang k von $\{1, \dots, n\}$, die i enthalten, summiert.

Satz 3.2.2. *Für $k < n$ gilt:*

$$\hat{V}_{\text{Jack}}(U_n) = \frac{n-1}{n} \frac{k^2}{(n-k)^2} \sum_{i=1}^n (V_n^{(1)}(X_i) - U_n)^{\otimes 2} \quad (3.2.10)$$

Beweis. Es gilt

$$V_n^{(1)}(X_i) = \binom{n-1}{k-1}^{-1} \left[\binom{n}{k} U_n - \binom{n-1}{k} U_{n-1}(-i) \right]$$

und somit

$$\begin{aligned} V_n^{(1)}(X_i) - U_n &= \binom{n-1}{k-1}^{-1} \binom{n}{k} U_n - \binom{n-1}{k-1}^{-1} \binom{n-1}{k} U_{n-1}(-i) - U_n \\ &= \left(\frac{n}{k} - 1 \right) U_n - \frac{n-k}{k} U_{n-1}(-i) \\ &= \frac{n-k}{k} (U_n - U_{n-1}(-i)). \end{aligned}$$

Unter Verwendung von $n^{-1} \sum_{i=1}^n U_{n-1}(-i) = U_n$ (Lee 1990, Satz 2 aus Abschnitt 4.1) erhält man

$$\begin{aligned} \sum_{i=1}^n (V_n^{(1)}(X_i) - U_n)^{\otimes 2} &= \left(\frac{n-k}{k} \right)^2 \sum_{i=1}^n (U_n - U_{n-1}(-i))^{\otimes 2} \\ &= \left(\frac{n-k}{k} \right)^2 \left(\sum_{i=1}^n U_{n-1}^{\otimes 2}(-i) - n U_n^{\otimes 2} \right) \end{aligned}$$

und

$$\begin{aligned}\hat{V}_{\text{Jack}}(U_n) &= \frac{n-1}{n} \left(\sum_{i=1}^n U_{n-1}^{\otimes 2}(-i) - nU_n^{\otimes 2} \right) \\ &= \frac{n-1}{n} \frac{k^2}{(n-k)^2} \sum_{i=1}^n (V_n^{(1)}(X_i) - U_n)^{\otimes 2}.\end{aligned}$$

□

Sei ab nun die Stichprobe X_1, \dots, X_n identisch verteilt. In derartigen Fällen wird die Notation $\sigma_{c,n}^2$ statt $\bar{\sigma}_{c,n}^2$ für $c = 1, \dots, k$ verwendet.

In (Puri und Sen 1971, p59) wird (für reellwertige Kerne)

$$S_1^2 = \frac{1}{n-1} \sum_{i=1}^n (V_n^{(1)}(X_i) - U_n)^2 \quad (3.2.11)$$

definiert. In (Puri und Sen 1971, p60) wird definiert

$$S_c^2 = \binom{n}{c}^{-1} \sum_{i_1 < \dots < i_c \leq n} [V_n^{(c)}(X_{i_1}, \dots, X_{i_c}) - U_n]^2 \quad (3.2.12)$$

für $c = 1, \dots, k$. Die Definition von $V_n^{(c)}$ für $c > 1$ wird nicht gegeben, da sie hier nicht benötigt wird. Man stellt fest, daß die Definitionen voneinander abweichen. Es ist vermutlich die Definition in (3.2.11) zusammen mit der Gleichung $\text{Var } U_n = \frac{k^2}{n} \text{Var } \sigma_{1,n}^2 + o(n^{-1})$, auf die sich Liang, Self und Chang (1993) beziehen, wenn sie unter Berufung auf (Puri und Sen 1971) folgenden Schätzer für $\text{Var } U_n$ vorschlagen:

$$\hat{V}(U_n) = \frac{k^2}{n(n-1)} \sum_{i=1}^n (V_n^{(1)}(X_i) - U_n)^{\otimes 2}. \quad (3.2.13)$$

(Dort wurde nur der Fall $k = 3$ behandelt). Weiter unten schlagen Liang et al. (1993) folgendes vor:

Es gilt für $1 \leq c < d \leq k$ (Lee 1990, p15)

$$0 \leq \frac{\sigma_{c,n}^2}{c} \leq \frac{\sigma_{d,n}^2}{d}. \quad (3.2.14)$$

Das impliziert $0 \leq d\sigma_{c,n}^2 \leq c\sigma_{d,n}^2$ für $1 \leq c < d \leq k$ und insbesondere $d\sigma_{1,n}^2 \leq \sigma_{d,n}^2$ für $1 \leq d \leq k$. Unter Verwendung von (3.1.3) erhält man

$$\frac{k^2}{n} \sigma_{1,n}^2 \leq \binom{n}{k}^{-1} \sum_{c=1}^k \binom{k}{c} \binom{n-k}{k-c} c \sigma_{1,n}^2 \leq \text{Var } U_n \quad (3.2.15)$$

Es gilt

$$\begin{aligned} & \frac{n}{9} \binom{n}{3}^{-1} \sum_{c=1}^3 \binom{3}{c} \binom{n-3}{3-c} c \\ &= \frac{2}{(n-1)(n-2)} \left(\frac{1}{2}(n-3)(n-4) + 2(n-3) + 1 \right) \\ &= \frac{2}{(n-1)(n-2)} \cdot (n^2 - 3n + 2) = 1. \end{aligned}$$

Liang et al. (1993) geben aufgrund von (3.2.15) für $k = 3$ folgenden Korrekturfaktor für den Schätzer (3.2.13) an:

$$\rho_3 = 1 + \frac{4}{n-4} + \frac{2}{(n-3)(n-4)}.$$

ρ_3 tritt auf, wenn folgende Faktorisierung durchgeführt wird:

$$\begin{aligned} \sum_{c=1}^3 \binom{3}{c} \binom{n-3}{3-c} c &= \frac{3}{2}(n-3)(n-4) + 6(n-3) + 3 \\ &= \frac{3}{2}(n-3)(n-4) \left(1 + \frac{4}{n-4} + \frac{2}{(n-3)(n-4)} \right). \end{aligned}$$

Als Schätzer für die Varianz einer U -Statistik sollte der Jackknife-Schätzer (3.2.2) verwendet werden. Bei identisch verteilter Stichprobe kann die Verfälschung angegeben werden:

Satz 3.2.3. *Seien δ_j^2 , $j = 1, \dots, k$, die bedingten Varianzen der Hoeffding-Zerlegung (Lee 1990, p29), dann gilt*

$$\mathbb{E} \hat{\mathbf{V}}_{Jack}(U_n) - \text{Var } U_n = n \sum_{j=2}^k \frac{j-1}{n-j} \binom{k}{j}^2 \binom{n}{j}^{-1} \delta_j^2. \quad (3.2.16)$$

3.3 U -Statistiken mit geschätztem Kern

Es soll ein Ergebnis für U -Statistiken hergeleitet werden, deren Kern von einem Parameter abhängt, der aus der Stichprobe geschätzt werden muß. Derartige Schätzer werden in (Randles 1982) behandelt. In diesem Abschnitt wird Satz 2.8 aus dieser Arbeit auf U -Statistiken auf nicht identisch verteilten Stichproben, die von einem Parameter aus einem normierten

Raum abhängen, verallgemeinert. Dafür werden auch Beweismethoden aus (Sukhatme 1958) verwendet.

Sei X_1, \dots, X_n eine Stichprobe von unabhängigen, nicht notwendigerweise identisch verteilten (und nicht notwendigerweise reellen) Zufallselementen vom Umfang n . Sei $h(x_1, \dots, x_k; \lambda)$ ein symmetrischer Kern mit Werten in \mathbb{R}^p vom Grad k . λ sei aus einem normierten Raum E mit Norm $|\cdot|$. Mit $|\cdot|$ für Werte aus \mathbb{R}^p wird das Maximum der Komponenten bezeichnet. Der Erwartungswert des Kerns bezüglich der Wahrscheinlichkeit P_ξ mit Parameter $\xi \in X$ sei

$$\theta_{\alpha_1, \dots, \alpha_k}(\lambda) = \mathbb{E}_\xi [h(X_{\alpha_1}, \dots, X_{\alpha_k}; \lambda)] \quad (3.3.1)$$

für Indizes $\alpha_1, \dots, \alpha_k$ aus (n, k) , der Menge der Kombinationen von k Indizes aus einer Indexmenge vom Umfang n . Sowohl der Kern als auch dessen Erwartungswert kann von λ abhängen. Die entsprechende U -Statistik ist

$$U_n(\lambda) = \binom{n}{k}^{-1} \sum_{\alpha \in (n, k)} h(X_{\alpha_1}, \dots, X_{\alpha_k}; \lambda). \quad (3.3.2)$$

Der Erwartungswert ist

$$\theta(\lambda) = \mathbb{E} U_n = \binom{n}{k}^{-1} \sum_{\alpha \in (n, k)} \theta_{\alpha_1, \dots, \alpha_k}(\lambda). \quad (3.3.3)$$

Sei $\hat{\lambda}$ ein konsistenter Schätzer von $\lambda \in E$. Es soll möglich sein, λ aus dem Parameter ξ zu berechnen. Ab nun wird der Index ξ von P_ξ weggelassen.

Man will nun hinreichende Bedingungen für die asymptotische Normalverteilung von

$$n^{1/2} [U_n(\hat{\lambda}) - \theta(\lambda)] \quad (3.3.4)$$

angeben.

Bedingung 3.3.1.

$$n^{1/2} (\hat{\lambda} - \lambda) \quad (3.3.5)$$

ist gleichmäßig straff in E . Das heißt, für jedes $\epsilon > 0$ gibt es einen kompakten Bereich $K_\epsilon \subset E$, sodaß $\limsup \mathbb{P} [n^{1/2} (\hat{\lambda} - \lambda) \notin G] < \epsilon$ für jede offene Menge $G \supseteq K_\epsilon$.

Insbesondere ist dann

$$n^{1/2} |\hat{\lambda} - \lambda| = O_p(1). \quad (3.3.6)$$

Bedingung 3.3.2. *Es gibt eine Umgebung $K(\lambda)$ von λ und Konstante $K_1 > 0$, $K_2 > 0$, $c > 0$, so daß für jede Kugel $B(\gamma, r) \subset K(\lambda)$, sowie für jede Kombination $\alpha_1, \dots, \alpha_k$ aus (n, k) und $0 < r \leq c$ gilt*

$$\mathbb{E} \sup_{\gamma' \in B(\gamma, r)} |h(X_{\alpha_1}, \dots, X_{\alpha_k}; \gamma') - h(X_{\alpha_1}, \dots, X_{\alpha_k}; \gamma)| \leq K_1 r; \quad (3.3.7)$$

und

$$\mathbb{E} \sup_{\gamma' \in B(\gamma, r)} |h(X_{\alpha_1}, \dots, X_{\alpha_k}; \gamma') - h(X_{\alpha_1}, \dots, X_{\alpha_k}; \gamma)|^2 \leq K_2 r; \quad (3.3.8)$$

Satz 3.3.3. *Seien die Bedingungen 3.3.1 und 3.3.2 erfüllt. Dann*

$$n^{1/2} \left(U_n(\hat{\lambda}) - \theta(\hat{\lambda}) - U_n(\lambda) + \theta(\lambda) \right) \xrightarrow{P} 0. \quad (3.3.9)$$

Beweis. Sei $\tilde{h}(X_{\alpha_1}, \dots, X_{\alpha_k}; \gamma) = h(X_{\alpha_1}, \dots, X_{\alpha_k}; \gamma) - \theta_{\alpha_1, \dots, \alpha_k}(\gamma)$. Weiters wird $W(X_{\alpha_1}, \dots, X_{\alpha_k}; r) = \tilde{h}(X_{\alpha_1}, \dots, X_{\alpha_k}; \lambda + r)$ gesetzt. Sei

$$Q_n(s) = \sqrt{n} \binom{n}{k}^{-1} \sum_{\alpha \in (k, n)} \tilde{h}(X_{\alpha_1}, \dots, X_{\alpha_k}; \lambda + s/\sqrt{n}) - \tilde{h}(X_{\alpha_1}, \dots, X_{\alpha_k}; \lambda). \quad (3.3.10)$$

Damit kann (3.3.9) geschrieben werden als

$$Q_n \left(n^{1/2}(\hat{\lambda} - \lambda) \right) \xrightarrow{P} 0. \quad (3.3.11)$$

Seien $\epsilon' > 0$ und $\delta' > 0$ beliebig gewählt.

Wegen Bedingung 3.3.1 gibt es eine kompakte Menge $C \subset E$, so daß für alle n

$$\mathbb{P} \left[n^{1/2}(\hat{\lambda} - \lambda) \notin C \right] \leq \delta'/2. \quad (3.3.12)$$

Es gilt

$$\mathbb{P} \left[\left| Q_n(n^{1/2}(\hat{\lambda} - \lambda)) \right| > \epsilon' \right] \leq \mathbb{P} \left[\sup_{s \in C} |Q_n(s)| > \epsilon' \right] + \frac{\delta'}{2}. \quad (3.3.13)$$

Daher ist zu zeigen, daß

$$\sup_{s \in C} |Q_n(s)| \xrightarrow{P} 0, \quad (3.3.14)$$

beziehungsweise

$$\mathbb{P} \left(\sup_{s \in C} |Q_n(s)| > \epsilon' \right) < \frac{\delta'}{2}$$

für alle n , die genügend groß sind.

Wie im folgenden gezeigt wird, kann man zu $\epsilon' > 0$ ein $\delta > 0$ angeben, so daß für beliebiges $\gamma \in E$

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\sup_{s \in B(\gamma\delta, \delta)} |Q_n(s)| > \epsilon' \right) = 0. \quad (3.3.15)$$

Endlich viele Kugeln $B(\gamma\delta, \delta)$ überdecken C , so daß (3.3.14) erfüllt ist.

Zunächst wird folgendes Lemma angegeben:

Lemma 3.3.4. *Sei $\delta > 0$. Sei $r \in E$. Seien*

$$H_{r,n} \left(X_1, \dots, X_k, \frac{\delta}{\sqrt{n}} \right) = \sup_{z \in B(r\frac{\delta}{\sqrt{n}}, \frac{\delta}{\sqrt{n}})} \left| W(X_{\alpha_1}, \dots, X_{\alpha_k}; z) - W \left(X_{\alpha_1}, \dots, X_{\alpha_k}; r\frac{\delta}{\sqrt{n}} \right) \right| \quad (3.3.16)$$

und

$$Q_{r,n}(z) = \sqrt{n} \binom{n}{k}^{-1} \sum_{\alpha \in (k,n)} \left[W \left(X_{\alpha_1}, \dots, X_{\alpha_k}; \frac{z}{\sqrt{n}} \right) - W \left(X_{\alpha_1}, \dots, X_{\alpha_k}; r\frac{\delta}{\sqrt{n}} \right) \right] \quad (3.3.17)$$

definiert. Dann gibt es von der Wahl von δ unabhängige Konstanten K_3 und

A_0 , so daß für ein genügend großes n gilt:

$$(i) \quad \mathbb{E} H_{r,n} \left(X_{\alpha_1}, \dots, X_{\alpha_k}, \frac{\delta}{\sqrt{n}} \right) \leq 2K_1 \frac{\delta}{\sqrt{n}}. \quad (3.3.18)$$

$$(ii) \quad \mathbb{E} \left| \left[H_{r,n} \left(X_{\alpha_1}, \dots, X_{\alpha_k}, \frac{\delta}{\sqrt{n}} \right) - \mathbb{E} H_{r,n} \left(X_{\alpha_1}, \dots, X_{\alpha_k}, \frac{\delta}{\sqrt{n}} \right) \right] \right. \\ \left. \cdot \left[H_{r,n} \left(X_{\beta_1}, \dots, X_{\beta_k}, \frac{\delta}{\sqrt{n}} \right) - \mathbb{E} H_{r,n} \left(X_{\beta_1}, \dots, X_{\beta_k}, \frac{\delta}{\sqrt{n}} \right) \right] \right| \\ < K_3 \frac{\delta}{\sqrt{n}}. \quad (3.3.19)$$

$$(iii) \quad \mathbb{E} |Q_{r,n}(t)|^2 \leq \frac{A_0 \delta}{\sqrt{n}} \quad \text{für } t \in B(r\delta, \delta). \quad (3.3.20)$$

Beweis. (i) Für hinreichend großes n liegt $B\left(\lambda + r\frac{\delta}{\sqrt{n}}, \frac{\delta}{\sqrt{n}}\right)$ in $K(\lambda)$. Dann gilt wegen (3.3.7) aus Bedingung 3.3.2:

$$\begin{aligned} & \mathbb{E} H_{r,n} \left(X_{\alpha_1}, \dots, X_{\alpha_k}, \frac{\delta}{\sqrt{n}} \right) \\ &= \mathbb{E} \sup_{z \in B\left(r\frac{\delta}{\sqrt{n}}, \frac{\delta}{\sqrt{n}}\right)} \left| h \left(X_{\alpha_1}, \dots, X_{\alpha_k}; \lambda + z \right) - h \left(X_{\alpha_1}, \dots, X_{\alpha_k}; \lambda + r\frac{\delta}{\sqrt{n}} \right) \right. \\ & \quad \left. - \mathbb{E} \left\{ h \left(X_{\alpha_1}, \dots, X_{\alpha_k}; \lambda + z \right) - h \left(X_{\alpha_1}, \dots, X_{\alpha_k}; \lambda + r\frac{\delta}{\sqrt{n}} \right) \right\} \right| \\ &\leq \mathbb{E} \sup_{z \in B\left(r\frac{\delta}{\sqrt{n}}, \frac{\delta}{\sqrt{n}}\right)} \left| h \left(X_{\alpha_1}, \dots, X_{\alpha_k}; \lambda + z \right) - h \left(X_{\alpha_1}, \dots, X_{\alpha_k}; \lambda + r\frac{\delta}{\sqrt{n}} \right) \right| \\ & \quad + \mathbb{E} \sup_{z \in B\left(r\frac{\delta}{\sqrt{n}}, \frac{\delta}{\sqrt{n}}\right)} \left| \mathbb{E} \left[h \left(X_{\alpha_1}, \dots, X_{\alpha_k}; \lambda + z \right) - \right. \right. \\ & \quad \left. \left. h \left(X_{\alpha_1}, \dots, X_{\alpha_k}; \lambda + r\frac{\delta}{\sqrt{n}} \right) \right] \right| \leq \frac{2K_1 \delta}{\sqrt{n}}. \end{aligned}$$

(ii) Für hinreichend großes n liegt $B\left(\lambda + r\frac{\delta}{\sqrt{n}}, \frac{\delta}{\sqrt{n}}\right)$ in $K(\lambda)$. Mittels (3.3.8) erhält man dann für beliebiges $\alpha \in (n, k)$ unter Verwendung der Ungleichung

$$|a + b|^2 \leq 4(|a|^2 + |b|^2):$$

$$\begin{aligned}
& \mathbb{E} \left[H_{r,n} \left(X_{\alpha_1}, \dots, X_{\alpha_k}, \frac{\delta}{\sqrt{n}} \right) \right]^2 \\
&= \mathbb{E} \left[\sup_{z \in B(r \frac{\delta}{\sqrt{n}}, \frac{\delta}{\sqrt{n}})} \left| h \left(X_{\alpha_1}, \dots, X_{\alpha_k}; \lambda + z \right) - h \left(X_{\alpha_1}, \dots, X_{\alpha_k}; \lambda + r \frac{\delta}{\sqrt{n}} \right) \right. \right. \\
&\quad \left. \left. - \mathbb{E} \left\{ h \left(X_{\alpha_1}, \dots, X_{\alpha_k}; \lambda + z \right) - h \left(X_{\alpha_1}, \dots, X_{\alpha_k}; \lambda + r \frac{\delta}{\sqrt{n}} \right) \right\} \right| \right]^2 \\
&\leq 4 \mathbb{E} \left[\sup_{z \in B(r \frac{\delta}{\sqrt{n}}, \frac{\delta}{\sqrt{n}})} \left| h \left(X_{\alpha_1}, \dots, X_{\alpha_k}; \lambda + z \right) - h \left(X_{\alpha_1}, \dots, X_{\alpha_k}; \lambda + r \frac{\delta}{\sqrt{n}} \right) \right| \right]^2 \\
&\quad + 4 \mathbb{E} \left[\sup_{z \in B(r \frac{\delta}{\sqrt{n}}, \frac{\delta}{\sqrt{n}})} \mathbb{E} \left| h \left(X_{\alpha_1}, \dots, X_{\alpha_k}; \lambda + z \right) \right. \right. \\
&\quad \quad \left. \left. - h \left(X_{\alpha_1}, \dots, X_{\alpha_k}; \lambda + r \frac{\delta}{\sqrt{n}} \right) \right| \right]^2 \\
&< 4K_2 \frac{\delta}{\sqrt{n}} + 4 \left(K_1 \frac{\delta}{\sqrt{n}} \right)^2 \leq K_3 \frac{\delta}{\sqrt{n}}.
\end{aligned}$$

Nun schätzt man mittels der Cauchy-Schwarz'schen Ungleichung ab:

$$\begin{aligned}
& \mathbb{E} \left| \left[H_{r,n} \left(X_{\alpha_1}, \dots, X_{\alpha_k}, \frac{\delta}{\sqrt{n}} \right) - \mathbb{E} H_{r,n} \left(X_{\alpha_1}, \dots, X_{\alpha_k}, \frac{\delta}{\sqrt{n}} \right) \right] \right. \\
&\quad \left. \cdot \left[H_{r,n} \left(X_{\beta_1}, \dots, X_{\beta_k}, \frac{\delta}{\sqrt{n}} \right) - \mathbb{E} H_{r,n} \left(X_{\beta_1}, \dots, X_{\beta_k}, \frac{\delta}{\sqrt{n}} \right) \right] \right| \\
&\leq \left(\mathbb{E} \left| H_{r,n} \left(X_{\alpha_1}, \dots, X_{\alpha_k}, \frac{\delta}{\sqrt{n}} \right) - \mathbb{E} H_{r,n} \left(X_{\alpha_1}, \dots, X_{\alpha_k}, \frac{\delta}{\sqrt{n}} \right) \right|^2 \right. \\
&\quad \left. \cdot \mathbb{E} \left| H_{r,n} \left(X_{\beta_1}, \dots, X_{\beta_k}, \frac{\delta}{\sqrt{n}} \right) - \mathbb{E} H_{r,n} \left(X_{\beta_1}, \dots, X_{\beta_k}, \frac{\delta}{\sqrt{n}} \right) \right|^2 \right)^{1/2} \\
&\leq \left(K_3 \frac{\delta}{\sqrt{n}} \cdot K_3 \frac{\delta}{\sqrt{n}} \right)^{1/2} = K_3 \frac{\delta}{\sqrt{n}}.
\end{aligned}$$

(iii) Zunächst ist (unter der 2-Norm auf \mathbb{R}^p)

$$\begin{aligned} & \mathbb{E} |Q_{r,n}(t)|^2 \\ = & n \binom{n}{k}^{-2} \sum_{\alpha, \beta} \mathbb{E} \left\{ \left[W \left(X_{\alpha_1}, \dots, X_{\alpha_k}; \frac{t}{\sqrt{n}} \right) - W \left(X_{\alpha_1}, \dots, X_{\alpha_k}; r \frac{\delta}{\sqrt{n}} \right) \right] \right. \\ & \left. \cdot \left[W \left(X_{\beta_1}, \dots, X_{\beta_k}; \frac{t}{\sqrt{n}} \right) - W \left(X_{\beta_1}, \dots, X_{\beta_k}; r \frac{\delta}{\sqrt{n}} \right) \right] \right\}. \end{aligned} \quad (3.3.21)$$

Für jeden der $\binom{n}{k}^2$ Ausdrücke folgt aus der Cauchy-Schwarz'schen Ungleichung:

$$\begin{aligned} & \left| \mathbb{E} \left\{ \left[W \left(X_{\alpha_1}, \dots, X_{\alpha_k}; \frac{t}{\sqrt{n}} \right) - W \left(X_{\alpha_1}, \dots, X_{\alpha_k}; r \frac{\delta}{\sqrt{n}} \right) \right] \right. \right. \\ & \quad \left. \cdot \left[W \left(X_{\beta_1}, \dots, X_{\beta_k}; \frac{t}{\sqrt{n}} \right) - W \left(X_{\beta_1}, \dots, X_{\beta_k}; r \frac{\delta}{\sqrt{n}} \right) \right] \right\} \right| \\ & \leq \left(\mathbb{E} \left| W \left(X_{\alpha_1}, \dots, X_{\alpha_k}; \frac{t}{\sqrt{n}} \right) - W \left(X_{\alpha_1}, \dots, X_{\alpha_k}; r \frac{\delta}{\sqrt{n}} \right) \right|^2 \right. \\ & \quad \left. \cdot \mathbb{E} \left| W \left(X_{\beta_1}, \dots, X_{\beta_k}; \frac{t}{\sqrt{n}} \right) - W \left(X_{\beta_1}, \dots, X_{\beta_k}; r \frac{\delta}{\sqrt{n}} \right) \right|^2 \right)^{1/2}. \end{aligned}$$

Die Faktoren können nun für genügend großes n unter Verwendung von

(3.3.8) weiter abgeschätzt werden durch

$$\begin{aligned}
& \mathbb{E} \left| W \left(X_{\alpha_1}, \dots, X_{\alpha_k}, \frac{t}{\sqrt{n}} \right) - W \left(X_{\alpha_1}, \dots, X_{\alpha_k}, r \frac{\delta}{\sqrt{n}} \right) \right|^2 \\
&= \mathbb{E} \left| h \left(X_{\alpha_1}, \dots, X_{\alpha_k}; \lambda + \frac{t}{\sqrt{n}} \right) - h \left(X_{\alpha_1}, \dots, X_{\alpha_k}; \lambda + r \frac{\delta}{\sqrt{n}} \right) \right. \\
&\quad \left. - \mathbb{E} \left[h \left(X_{\alpha_1}, \dots, X_{\alpha_k}; \lambda + \frac{t}{\sqrt{n}} \right) - h \left(X_{\alpha_1}, \dots, X_{\alpha_k}; \lambda + r \frac{\delta}{\sqrt{n}} \right) \right] \right|^2 \\
&\leq \mathbb{E} \left| h \left(X_{\alpha_1}, \dots, X_{\alpha_k}; \lambda + \frac{t}{\sqrt{n}} \right) - h \left(X_{\alpha_1}, \dots, X_{\alpha_k}; \lambda + r \frac{\delta}{\sqrt{n}} \right) \right|^2 \\
&\quad + \left| \mathbb{E} \left[h \left(X_{\alpha_1}, \dots, X_{\alpha_k}; \lambda + \frac{t}{\sqrt{n}} \right) - h \left(X_{\alpha_1}, \dots, X_{\alpha_k}; \lambda + r \frac{\delta}{\sqrt{n}} \right) \right] \right|^2 \\
&\leq 2 \mathbb{E} \left(\sup_{s \in B(r\delta, \delta)} \left| h \left(X_{\alpha_1}, \dots, X_{\alpha_k}; \lambda + \frac{s}{\sqrt{n}} \right) - \right. \right. \\
&\qquad \qquad \qquad \left. \left. h \left(X_{\alpha_1}, \dots, X_{\alpha_k}; \lambda + r \frac{\delta}{\sqrt{n}} \right) \right|^2 \right) \\
&\leq 2K_2 \frac{\delta}{\sqrt{n}}
\end{aligned}$$

Faßt man nun in (3.3.21) Summanden zusammen, die in c , $c = 0, 1, \dots, k$, Argumenten $X_{\alpha_1}, \dots, X_{\alpha_k}$ und $X_{\beta_1}, \dots, X_{\beta_k}$ von h übereinstimmen (vergleiche mit (3.1.4)), dann folgt

$$\begin{aligned}
\mathbb{E} |Q_{r,n}(t)|^2 &\leq n \binom{n}{k}^{-2} \sum_{c=1}^k \binom{n}{k} \binom{k}{k-c} \binom{n-k}{k-c} K_2 \frac{\delta}{\sqrt{n}} \\
&\leq \sum_{c=1}^k n^{-c+1} A \frac{\delta}{\sqrt{n}}.
\end{aligned}$$

für eine Konstante A . Man beachte, daß für $c = 0$ die Ausdrücke in (3.3.21) unabhängig sind und deren Erwartungswert daher verschwindet. Der nächste Index $c = 1$ liefert das Ergebnis, während die restlichen Ausdrücke von höherer negativer Ordnung in n sind. Für genügend großes n existiert daher eine Konstante A_0 , so daß (3.3.20) erfüllt ist. \square

Es gilt

$$Q_n(t) = Q_{r,n}(t) - Q_{0,n}(r\delta). \quad (3.3.22)$$

Für genügend großes n erhält man wegen (3.3.18):

$$\begin{aligned}
& \sup_{t \in B(r \frac{\delta}{\sqrt{n}}, \frac{\delta}{\sqrt{n}})} |Q_{r,n}(t)| \\
& \leq \sqrt{n} \binom{n}{k}^{-1} \sum_{\alpha} H_{r,n} \left(X_{\alpha_1}, \dots, X_{\alpha_k}, \frac{\delta}{\sqrt{n}} \right) \\
& = \sqrt{n} \binom{n}{k}^{-1} \sum_{\alpha} \left[\left[H_{r,n} \left(X_{\alpha_1}, \dots, X_{\alpha_k}, \frac{\delta}{\sqrt{n}} \right) \right] \right. \\
& \quad \left. - \mathbb{E} H_{r,n} \left(X_{\alpha_1}, \dots, X_{\alpha_k}, \frac{\delta}{\sqrt{n}} \right) \right] \\
& \quad + \sqrt{n} \binom{n}{k}^{-1} \sum_{\alpha} \mathbb{E} H_{r,n} \left(X_{\alpha_1}, \dots, X_{\alpha_k}, \frac{\delta}{\sqrt{n}} \right) \\
& \leq \sqrt{n} \binom{n}{k}^{-1} \sum_{\alpha} \left| H_{r,n} \left(X_{\alpha_1}, \dots, X_{\alpha_k}, \frac{\delta}{\sqrt{n}} \right) - \mathbb{E} H_{r,n} \left(X_{\alpha_1}, \dots, X_{\alpha_k}, \frac{\delta}{\sqrt{n}} \right) \right| \\
& \quad + 2K_1 \delta \\
& = D_1 + 2K_1 \delta
\end{aligned}$$

D_1 ist durch den darüberstehenden Ausdruck definiert.

Nun berechnet man mittels (3.3.19)

$$\begin{aligned}
\mathbb{E} D_1^2 & = n \binom{n}{k}^{-2} \sum_{\alpha, \beta} \mathbb{E} \left\{ \left[H_{r,n} \left(X_{\alpha_1}, \dots, X_{\alpha_k}, \frac{\delta}{\sqrt{n}} \right) - \right. \right. \\
& \quad \left. \left. \mathbb{E} H_{r,n} \left(X_{\alpha_1}, \dots, X_{\alpha_k}, \frac{\delta}{\sqrt{n}} \right) \right] \right. \\
& \quad \cdot \left. \left[H_{r,n} \left(X_{\beta_1}, \dots, X_{\beta_k}, \frac{\delta}{\sqrt{n}} \right) - \mathbb{E} H_{r,n} \left(X_{\beta_1}, \dots, X_{\beta_k}, \frac{\delta}{\sqrt{n}} \right) \right] \right\} \\
& \leq n \binom{n}{k}^{-2} \sum_{c=1}^k \binom{n}{k} \binom{k}{k-c} \binom{n-k}{k-c} K_3 \frac{\delta}{\sqrt{n}} \\
& \leq \sum_{c=1}^k n^{-c+1/2} A_0 \delta \leq A'_0 \frac{\delta}{\sqrt{n}}.
\end{aligned} \tag{3.3.23}$$

Dabei wurde verwendet, daß die Summe nach der Anzahl c der gemeinsamen Argumente in den Faktoren angeordnet werden kann. Für $c = 0$ sind die Faktoren unabhängig und haben daher Kovarianz 0.

Aufgrund der Čebyšev'schen Ungleichung gilt für genügend großes n und $0 < \delta < \epsilon/(8K_1)$:

$$\begin{aligned} \mathbb{P} \left(\sup_{t \in B(r\frac{\delta}{\sqrt{n}}, \frac{\delta}{\sqrt{n}})} |Q_{r,n}(t)| > \frac{\epsilon}{2} \right) &\leq \mathbb{P} \left(|D_1| > \frac{\epsilon}{2} - 2K_1\delta \right) \\ &\leq \frac{\mathbb{E} D_1^2}{(\epsilon/2 - 2K_1\delta)^2} \\ &\leq \frac{16}{\epsilon^2} A'_0 \frac{\delta}{\sqrt{n}}. \end{aligned}$$

Für den zweiten Teil der Zerlegung (3.3.22) von $Q_n(t)$ findet man unter Verwendung von (3.3.20), daß für genügend großes n

$$\mathbb{P} \left(|Q_{0,n}(r\delta)| > \frac{\epsilon}{2} \right) \leq \frac{\mathbb{E} |Q_{0,n}(r\delta)|^2}{\epsilon^2/4} \leq \frac{4}{\epsilon^2} A_0 \frac{|r|\delta}{\sqrt{n}}$$

Somit ist die Behauptung (3.3.15) bewiesen. \square

Kapitel 4

Schätzung von marginalen Modellen mit der Cox'schen partiellen Likelihood

4.1 Modellformulierung

Es werden multivariate Zählprozesse betrachtet:

$$N = (N_i; i = 1, \dots, n) \quad (4.1.1)$$

mit Kompensator

$$\Lambda^\theta = (\Lambda_i^\theta; i = 1, \dots, n) \quad (4.1.2)$$

bezüglich einer Filtrierung (\mathcal{F}_t) und einer Wahrscheinlichkeit P_{θ_ϕ} . Der Kompensator soll jedoch nur mehr von θ abhängen. Auch ein allfälliger Zensierungsprozeß soll darin schon berücksichtigt sein. N ist also die vereinfachte Schreibweise für N^c und Λ^θ für $\Lambda^c(\cdot, \theta)$, den Kompensator bezüglich (\mathcal{F}_t^c) (hier (\mathcal{F}_t)), der in Abschnitt 2 zur Unterscheidung vom unzensierten Prozeß N an mehreren Stellen eingeführt wurde. Die *partielle Likelihoodfunktion* ist dann (siehe (2.2.13) und (2.3.2)) gegeben durch (Andersen et al. 1993, (7.1.1)):

$$L_\tau(\theta) = \prod_t \left\{ \prod_i d\Lambda_i(t, \theta)^{\Delta N_i(t)} (1 - d\Lambda(t, \theta))^{1 - \Delta N(t)} \right\} \quad (4.1.3)$$

In Regressionsmodellen wird der Kompensator $\Lambda_i^\theta(t) = \Lambda_i(\cdot, \theta)$ zu einem Vektor von Kovariablen $Z_i(t)$, $i = 1, \dots, n$, in Beziehung gesetzt. Dabei sind

fundamentale Annahmen, daß dieser Kovariablen-Prozeß (\mathcal{F}_t) -*vorhersagbar* (siehe 1.2) und *lokal beschränkt* (siehe 2) ist.

Weiters soll Λ_i^θ absolut stetig sein, d. h. es soll also ein *Intensitätsprozeß* $\lambda_i^\theta(t)$ mit

$$\Lambda_i^\theta(t) = \int_0^t \lambda_i^\theta(u) \, du$$

existieren.

Konkret soll der Intensitätsprozeß folgende multiplikative Struktur haben.

$$\lambda_i^\theta(t) = Y_i(t) \alpha_i^\theta(t, Z_i(t)), \quad i = 1 = 1, \dots, n.$$

Es wird vorausgesetzt, daß der Prozeß $Y_i(\cdot)$ *vorhersagbar* und nicht von θ abhängig ist. Üblicherweise ist $Y_i(t)$ eine Indikatorfunktion dafür, daß ein Sprung von N_i unmittelbar vor t beobachtbar ist. In den nachfolgenden Anwendungen genügt es, eine $Y_i(t)$ als Indikatorfunktion für eine endliche Vereinigung von Intervallen E_I der Form

$$E_I = \bigcup_{j=1}^r (V_{ji}, U_{ji}]$$

anzunehmen, wobei $0 \leq V_{1i} \leq U_{1i} \leq \dots \leq V_{ri} \leq U_{ri} \leq \tau$ (vergleiche mit (2.3.1)).

Die Funktion $\alpha_i^\theta(\cdot, \cdot)$ ist die *Hasardrate*. Deren Aufbau kann *additiv* und *multiplikativ* sein. In dieser Arbeit wird nur auf den multiplikativen Typ eingegangen:

$$\alpha_i^\theta(t, Z_i(t)) = \alpha_0(t, \gamma) r(\beta^\top Z_i(t)), \quad i = 1 = 1, \dots, n$$

Dabei ist $\theta = (\gamma, \beta)$, $\beta \in \mathbb{R}^p$ und r eine positive Funktion. Derartige Modelle werden auch als *Relatives-Risiko-Modelle* bezeichnet, da im Quotienten der Hasardraten der Ausdruck $\alpha_0(t, \gamma)$ wegfällt. Im folgenden beschränkt man sich auf $r(x) = \exp(x)$, d. h. die Modelle mit *proportionalen Hasardfunktionen*, die in (Cox 1972) eingeführt wurden. Obwohl das Modell ursprünglich nicht für Zählprozesse formuliert worden war, lassen sich sowohl das Modell, als auch ein Großteil der Ergebnisse unmittelbar auf Zählprozesse verallgemeinern (Andersen und Gill 1982).

Im *semiparametrischen Modell* ist $\alpha_0(t, \gamma) = \alpha_0(t)$, die *zugrundeliegende Hasardfunktion*¹, vollkommen frei wählbar, bis auf $\alpha_0(t) > 0$ und

$$A_0(t) = \int_0^t \alpha_0(u) \, du < \infty \quad \text{für alle } t \in \mathcal{T}. \quad (4.1.4)$$

¹engl.: baseline hazard function bzw. underlying hazard function

Die (partielle) Likelihood 4.1.3 („multinomiale Form“, vergleiche mit (1.9.8) beziehungsweise (2.3.2)) ist für dieses Modell proportional zu

$$\prod_{t \in \mathcal{T}} \left\{ \prod_i (dA_0(t) \exp \{\beta^\top Z_i(t)\})^{\Delta N_i(t)} (1 - dA_0(t) S^{(0)}(\beta, t))^{1 - \Delta N_i(t)} \right\} \quad (4.1.5)$$

mit $S^{(0)}(\beta, t) = \sum_{i=1}^n \exp \{\beta^\top Z_i(t)\} Y_i(t)$.

Die Schätzung dieses Modells wird im folgenden genauer untersucht und auf Haufendaten verallgemeinert werden.

4.2 Die partielle Likelihood von Cox

Es soll die Schätzung des *semiparametrischen* proportionalen Hasardratenmodells mit

$$\alpha_i^\theta(t, Z_i(t)) = \alpha_0(t) \exp \{\beta^\top Z_i(t)\} \quad (4.2.1)$$

behandelt werden.

Seien die skalar-, vektor- und matrixwertigen Prozesse $S^{(d)}(\beta, t)$ für $d = 0, 1, 2$ definiert durch²

$$S^{(d)}(\beta, t) = \sum_{i=1}^n Y_i(t) Z_i(t)^{\otimes d} \exp \{\beta^\top Z_i(t)\}. \quad (4.2.2)$$

Die Likelihood (4.1.5) in „Poisson-Form“ (vergleiche (1.9.7)) ist

$$\prod_t \left\{ \prod_i (dA_0(t) \exp \{\beta^\top Z_i(t)\})^{\Delta N_i(t)} \right\} \exp \left\{ - \int_0^\tau S^{(0)}(\beta, u) dA_0(u) \right\}. \quad (4.2.3)$$

Bei festem β kann Likelihood (4.1.5) beziehungsweise (4.2.3) im Sinn von (Andersen et al. 1993, pp 221-224) durch einen Schätzer vom Nelson-Aalen-Typ maximiert werden:³

$$\hat{A}_0(\beta, t) = \int_0^t \frac{J(u)}{S^{(0)}(\beta, u)} dN(u) \quad (4.2.4)$$

²Schreibweise: $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$ und $a^{\otimes 2} = aa^\top$

³Vereinbarung: $0/0 = 0$

mit $Y = Y_1 + \dots + Y_n$, $J(u) = I(Y(u) > 0)$ und $N = N_1 + \dots + N_n$.

Nun setzt man (4.2.4) in (4.1.5) beziehungsweise (4.2.3) ein und erhält die folgenden partiell maximierten (partiellen) Likelihoods (Likelihood Profile) für β :

$$L(\beta) \prod_{t \in \mathcal{T}} \Delta N(t)^{\Delta N(t)} (1 - dN.(t))^{1 - \Delta N.(t)} \quad (4.2.5)$$

beziehungsweise

$$L(\beta) \prod_{t \in \mathcal{T}} \Delta N(t)^{\Delta N(t)} \exp(-N.(\tau)). \quad (4.2.6)$$

Dabei ist

$$L(\beta) = \prod_{t \in \mathcal{T}, i \in \mathbb{N}} \left(\frac{\exp\{\beta^\top Z_i(t)\}}{S^{(0)}(\beta, t)} \right)^{\Delta N_i(t)}. \quad (4.2.7)$$

Diese, erstmals in (Cox 1972, Cox 1975) bei zensierten Überlebensdaten auf ganz andere Weise hergeleitete Funktion, ist also ein partiell maximiertes partielles Likelihoodprofil, das nur mehr von β abhängt.

Der *logarithmierte Likelihood-Prozeß* ist

$$C(\beta, t) = \log L(\beta) = \sum_{i=1}^n \int_0^t \beta^\top Z_i(u) dN_i(u) - \int_0^t \log S^{(0)}(\beta, u) dN(u) \quad (4.2.8)$$

Die logarithmierte Likelihood ist dann $C(\beta, \tau)$. Der Wert von β , der $C(\beta, \tau)$ maximiert, wird mit $\hat{\beta}$ bezeichnet (falls er existiert). Der *Breslow-Schätzer* der kumulativen Hasardrate wird mittels (4.2.4) als $\hat{A}_0(\hat{\beta}, t)$ berechnet (Breslow 1972).

Mit Hilfe des Vektors

$$E(\beta, t) = \frac{S^{(1)}(\beta, t)}{S^{(0)}(\beta, t)}$$

schreibt man den Score-Prozeß an als

$$U(\beta, t) = \frac{\partial C(\beta, t)}{\partial \beta} = \sum_{i=1}^n \int_0^t Z_i(u) dN_i(u) - \int_0^t E(\beta, u) dN(u) \quad (4.2.9)$$

Mittels der Matrix

$$V(\beta, t) = \frac{S^{(2)}(\beta, t)}{S^{(0)}(\beta, t)} - E(\beta, t)^{\otimes 2} \quad (4.2.10)$$

ist die Matrix der zweiten partiellen Ableitungen der Likelihood gegeben durch

$$\begin{aligned} \mathcal{I}(\beta, t) &= \frac{\partial^2 C(\beta, t)}{\partial \beta^{\otimes 2}} = \int_0^t \left\{ \frac{S^{(2)}(\beta, t)}{S^{(0)}(\beta, t)} - E(u, \beta)^{\otimes 2} \right\} dN(u) \\ &= \int_0^t V(\beta, u) dN(u). \end{aligned} \quad (4.2.11)$$

4.3 Asymptotische Eigenschaften

Die Asymptotik beruht auf Folgen von Modellen, die durch n indiziert sind. Es sind dies die Prozesse $N_i^{(n)}$, $Y_i^{(n)}$ und $Z_i^{(n)}$ auf $(\Omega^{(n)}, \mathcal{F}^{(n)}, \mathcal{P}^{(n)})$, angepaßt an eine Filtrierung $(\mathcal{F}_t^{(n)})$. Die „wahren“ Parameter β_0 und α_0 sind für alle n gleich. Der Index (n) wird zur Erleichterung der Lesbarkeit weggelassen.

Die Daten (N_i, Y_i, Z_i) bezüglich verschiedener Indices seien voneinander unabhängig.

Die Norm $|\cdot|$ von Vektoren und Matrizen ist als Supremum der Komponenten definiert. Wenn das Supremum der Norm über eine Menge \mathcal{B} von Argumenten einer Funktion gemeint ist, dann wird das mit $|\cdot|_{\mathcal{B}}$ bezeichnet. Diese Schreibweise wird nur zur übersichtlichen Darstellung verwendet, wenn die Bedeutung klar ist. Konvergenz in Wahrscheinlichkeit und Konvergenz in Verteilung sind Grenzwerte bezüglich $n \rightarrow \infty$.

Die asymptotischen Eigenschaften von $\hat{\beta}$ beruhen weitgehend auf der Darstellung des Score-Prozesses in β_0 , dem wahren Parameter, als Martingal (Satz 1.4.1)

$$U(\beta_0, t) = \sum_{i=1}^n \int_0^t [Z_i(u) - E(\beta_0, u)] dM_i(u), \quad (4.3.1)$$

wobei M_i die eindeutig bestimmten Martingale aus der Doob-Meyer-Zerlegung

$$N_i(t) = M_i(t) + \int_0^t Y_i(u) \exp \{ \beta_0^\top Z_i(u) \} \alpha_0(u) du \quad (4.3.2)$$

sind. (4.3.1) gilt wegen

$$\begin{aligned} & \sum_{i=1}^n [Z_i(u) - E(\beta_0, u)] Y_i(u) \exp \{ \beta_0^\top Z_i(u) \} \\ &= S^{(1)}(\beta_0, u) - S^{(0)}(\beta_0, u) \frac{S^{(1)}(\beta_0, u)}{S^{(0)}(\beta_0, u)} = 0 \end{aligned} \quad (4.3.3)$$

Darüberhinaus sind die Kompensatoren absolut stetig, sodaß nach (1.5.3) die M_i orthogonal sind mit

$$\langle M_i \rangle (t) = \int_0^t Y_i(u) \exp \{ \beta_0^\top Z_i(u) \} \alpha_0(u) du. \quad (4.3.4)$$

Für die nachstehenden Sätze werden folgende Bedingungen benötigt:

C 1. Es existiert eine offene Umgebung \mathcal{B} von β_0 und skalarwertige, p -vektorwertige und $p \times p$ -matrixwertige Funktionen $s^{(0)}(\beta, t)$, $s^{(1)}(\beta, t) = \frac{\partial s^{(0)}(\beta, t)}{\partial \beta}$, $s^{(2)}(\beta, t) = \frac{\partial^2 s^{(0)}(\beta, t)}{\partial \beta^2}$ mit Definitionsbereich $\mathcal{B} \times \mathcal{T}$, sodaß für $m = 0, 1, 2$

$$\left| \frac{1}{n} S^{(m)}(\beta, t) - s^{(m)}(\beta, t) \right|_{\mathcal{B} \times \mathcal{T}} \xrightarrow{P} 0 \quad (4.3.5)$$

C 2. $s^{(m)}(\beta, t)$ ist für $m = 0, 1, 2$ eine gleichmäßig in $t \in \mathcal{T}$ stetige Funktion von $\beta \in \mathcal{B}$, und beschränkt auf $\mathcal{B} \times \mathcal{T}$.

C 3. $s^{(0)}(\beta_0, \cdot)$ ist auf \mathcal{T} von 0 weg beschränkt.

C 4. $\Sigma_\tau = \int_0^\tau v(\beta_0, t) s^{(0)}(\beta_0, t) \alpha_0(t) dt$ ist positiv definit. Dabei ist $v = s^{(2)}/s^{(0)} - e^{\otimes 2}$ und $e = s^{(1)}/s^{(0)}$.

Satz 4.3.1. Seien C 1 bis C 4 und $A_0(\tau) < \infty$, sowie $\alpha_0(\cdot) > 0$ erfüllt. Dann konvergiert die Wahrscheinlichkeit, daß die Gleichung $U(\beta, \tau) = 0$ eine eindeutige Lösung $\hat{\beta}$ hat, gegen 1 für $n \rightarrow \infty$ und $\hat{\beta} \xrightarrow{P} \beta_0$

Beweis. (Andersen et al. 1993, p497f): Man betrachtet folgenden Prozeß

$$\begin{aligned} X(\beta, t) &= \frac{1}{n} (C(\beta, t) - C(\beta_0, t)) \\ &= \frac{1}{n} \sum_{i=1}^n \int_0^t (\beta - \beta_0)^\top Z_i(s) dN_i(s) - \frac{1}{n} \int_0^t \log \left\{ \frac{S^{(0)}(\beta, s)}{S^{(0)}(\beta_0, s)} \right\} dN(s) \end{aligned} \quad (4.3.6)$$

Zu diesem Ausdruck kann man aufgrund der Zerlegung (4.3.2) und der Vorhersagbarkeit und lokalen Beschränktheit der $Z_i(t)$ und der $Y_i(t)$ unter Anwendung von Satz 1.4.1 den Kompensator angeben als

$$\begin{aligned} & \tilde{X}(\beta, t) \\ &= \frac{1}{n} \int_0^t \left[(\beta - \beta_0)^\top S^{(1)}(\beta_0, s) - \log \left\{ \frac{S^{(0)}(\beta, s)}{S^{(0)}(\beta_0, s)} \right\} S^{(0)}(\beta_0, s) \right] \alpha_0(s) ds. \end{aligned} \quad (4.3.7)$$

$X(\beta, \cdot) - \tilde{X}(\beta, \cdot)$ ist folglich ein lokal quadratisch integrierbares Martingal. Die vorhersagbare Variation ist aufgrund von Satz 1.4.1 gegeben durch

$$\begin{aligned} & \left\langle X(\beta, \cdot) - \tilde{X}(\beta, \cdot) \right\rangle (t) \\ &= \frac{1}{n^2} \sum_{i=1}^n \int_0^t \left[(\beta - \beta_0)^\top Z_i(s) - \log \left\{ \frac{S^{(0)}(\beta, s)}{S^{(0)}(\beta_0, s)} \right\} \right]^2 \\ & \quad \cdot Y_i(s) \exp \{ \beta_0^\top Z_i(s) \} \alpha_0(s) ds. \end{aligned} \quad (4.3.8)$$

Wegen $A_0(\tau) < \infty$ wegen C 1, C 2 und C 3 konvergiert $\tilde{X}(\beta, \tau)$ für ein $\beta \in \mathcal{B}$ in Wahrscheinlichkeit gegen eine Funktion

$$f(\beta) = \int_0^\tau \left[(\beta - \beta_0)^\top s^{(1)}(\beta_0, s) - \log \left\{ \frac{s^{(0)}(\beta, s)}{s^{(0)}(\beta_0, s)} \right\} s^{(0)}(\beta_0, s) \right] \alpha_0(s) ds \quad (4.3.9)$$

und auch $n \left\langle X(\beta, \cdot) - \tilde{X}(\beta, \cdot) \right\rangle (\tau)$ konvergiert gegen eine beschränkte Funktion von β . Nun kann die Lenglart'sche Ungleichung (1.7.3) angewendet werden. Wegen $\left\langle X(\beta, \cdot) - \tilde{X}(\beta, \cdot) \right\rangle (\tau) \xrightarrow{P} 0$ gilt auch

$$X(\beta, \tau) \xrightarrow{P} f(\beta). \quad (4.3.10)$$

Wegen $A_0(\tau) < \infty$ und C 1 bis C 3 gilt

$$\frac{\partial}{\partial \beta} f(\beta) = \int_0^\tau \{ e(\beta_0, s) - e(\beta, s) \} s^{(0)}(\beta_0, s) \alpha_0(s) ds \quad (4.3.11)$$

und

$$\frac{\partial^2}{\partial \beta^{\otimes 2}} f(\beta) = \int_0^\tau v(\beta, s) s^{(0)}(\beta_0, s) \alpha_0(s) ds \quad (4.3.12)$$

ist positiv semidefinit und positiv definit für $\beta = \beta_0$ aufgrund von C 4. $X(\beta, \tau)$ konvergiert daher in \mathcal{B} punktweise in Wahrscheinlichkeit gegen eine konkave Funktion $f(\beta)$ mit eindeutigem Maximum bei $\beta = \hat{\beta}$. Die zufällige Funktion $X(\beta, \tau)$ ist konkav mit Maximum bei $\beta = \beta_0$, wenn dieses existiert. Nach Satz 1.7.2 ist die Konvergenz von $X(\beta, \tau)$ gleichmäßig in einer Umgebung B von β_0 und nach Folgerung 1.7.3 konvergiert die Wahrscheinlichkeit, daß $X(\beta, \tau)$ ein Maximum bei $\hat{\beta}$ annimmt, gegen 1, und die maximierenden Werte $\hat{\beta}$ konvergieren in Wahrscheinlichkeit gegen den maximierenden Wert β_0 von $f(\beta)$. \square

Zum Beweis der asymptotischen Normalverteilung von $\hat{\beta}$ ist eine stärkere Beschränktheitsbedingung für den Kovariablenvektor erforderlich. Die Beschränktheit von $E|Z|^r$ für ein $r > 2$ impliziert die nachstehende Bedingung vom Lindeberg-Typ:

C 5. *Es existiert ein $\delta > 0$, so daß*

$$n^{-1/2} \sup_{i \in n, t \in \mathcal{T}} |Z_i(t)| Y_i(t) I(\beta_0^\top Z_i(t) > -\delta |Z_i(t)|) \xrightarrow{P} 0 \quad (4.3.13)$$

Der folgende Satz stammt von Andersen und Gill (1982):

Satz 4.3.2. *Seien C 1 bis C 5 und $A_0(\tau) < \infty$, sowie $\alpha_0(\cdot) > 0$ erfüllt. Dann gilt*

$$n^{1/2}(\hat{\beta} - \beta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_\tau^{-1}) \quad (4.3.14)$$

und

$$\left| n^{-1} \mathcal{I}_t(\hat{\beta}) - \Sigma_t \right|_{\mathcal{T}} \xrightarrow{P} 0 \quad (4.3.15)$$

Der Beweis ist technisch aufwendig (Andersen et al. 1993, p. 499ff). Da diese Details später nicht mehr benötigt werden, wird auf dessen Abdruck verzichtet. Stattdessen wird ein Resultat gegeben, das auf Haufendaten verallgemeinerbar ist. Die Bedingung C 5 vom Lindebergtyp wird abgeschwächt zur Existenz einer globalen Schranke für den Kovariablenvektor. Es ist bekannt, daß unbeschränkte Kovariable insbesondere die Robustheitseigenschaften des Schätzers ungünstig beeinflussen können (Andersen et al. 1993, p501).

Lemma 4.3.3. *Unter C 1, C 2 und C 3 gilt für eine Umgebung \mathcal{B} von β_0*

$$|E(\cdot, \cdot) - e(\cdot, \cdot)|_{\mathcal{B} \times \mathcal{T}} = o_p(1). \quad (4.3.16)$$

und

$$\left| \frac{S^{(0)}(\beta, t)}{S^{(0)}(\beta_0, t)} - \frac{s^{(0)}(\beta, t)}{s^{(0)}(\beta_0, t)} \right|_{\mathcal{B} \times \mathcal{T}} \xrightarrow{P} 0 \quad \text{für } \beta_0 \in \mathcal{B}, \text{ fest} \quad (4.3.17)$$

Beweis. Aufgrund von C 1, C 2 und C 3 gilt

$$\begin{aligned}
|E - e|_{\mathcal{B} \times \mathcal{T}} &= \left| \frac{n^{-1}S^{(1)}}{n^{-1}S^{(0)}} - \frac{s^{(1)}}{s^{(0)}} \right|_{\mathcal{B} \times \mathcal{T}} \\
&\leq \left| \frac{n^{-1}S^{(1)}}{n^{-1}S^{(0)}} - \frac{n^{-1}S^{(1)}}{s^{(0)}} \right|_{\mathcal{B} \times \mathcal{T}} + \left| \frac{n^{-1}S^{(1)}}{s^{(0)}} - \frac{s^{(1)}}{s^{(0)}} \right|_{\mathcal{B} \times \mathcal{T}} \\
&\leq \frac{|n^{-1}S^{(1)}|_{\mathcal{B} \times \mathcal{T}}}{|n^{-1}S^{(0)}s^{(0)}|_{\mathcal{B} \times \mathcal{T}}} |n^{-1}S^{(0)} - s^{(0)}|_{\mathcal{B} \times \mathcal{T}} + \frac{1}{|s^{(0)}|_{\mathcal{B} \times \mathcal{T}}} |n^{-1}S^{(1)} - s^{(1)}|_{\mathcal{B} \times \mathcal{T}} \\
&\leq |n^{-1}S^{(1)}|_{\mathcal{B} \times \mathcal{T}} |nS^{(0)-1}|_{\mathcal{B} \times \mathcal{T}} |s^{(0)-1}|_{\mathcal{B} \times \mathcal{T}} |n^{-1}S^{(0)} - s^{(0)}|_{\mathcal{B} \times \mathcal{T}} + o_p(1).
\end{aligned}$$

Weiters gilt

$$|nS^{(0)-1}|_{\mathcal{B} \times \mathcal{T}} = \left(|nS^{(0)-1}|_{\mathcal{B} \times \mathcal{T}} - \delta \right)^{-1} + o_p(1) = O_p(1).$$

Damit ist (4.3.16) bewiesen. Beim Beweis von (4.3.17) geht man analog vor. \square

Für die im nachstehenden Satz hergeleitete asymptotische Verteilung von $U(\beta_0, \tau)$ werden folgende Definitionen benötigt: Sei $B_n = \frac{1}{n} \sum_{i=1}^n w_i^{\otimes 2}$ und $\hat{B}_n = \frac{1}{n} \sum_{i=1}^n \hat{w}_i^{\otimes 2}$, wobei

$$w_i = \int_0^\tau (Z_i(t) - e(\beta_0, t)) \{dN_i(t) - Y_i(t) \exp\{\beta_0^\top Z_i(t)\} \alpha_0(t) dt\} \quad (4.3.18)$$

und unter Verwendung des Breslow-Schätzers $\hat{A}_0(\cdot, \cdot)$ aus (4.2.4)

$$\hat{w}_i = \int_0^\tau (Z_i(t) - E(\hat{\beta}, t)) \left\{ dN_i(t) - Y_i(t) \exp\{\hat{\beta}^\top Z_i(t)\} d\hat{A}_0(\hat{\beta}, t) \right\} \quad (4.3.19)$$

Satz 4.3.4. Seien C 1 bis C 4 und $A_0(\tau) < \infty$, sowie $\alpha_0(\cdot) > 0$ erfüllt. Darüberhinaus sei $|Z_i(t)| \leq Z_0$ für $i = 1, \dots, n$ und $t \in \mathcal{T}$. Dann existiert eine positiv definite Matrix $B = \Sigma_\tau$, so daß

$$n^{-1/2}U(\beta_0, \tau) \xrightarrow{\mathcal{D}} \mathcal{N}(0, B) \quad (4.3.20)$$

und

$$\hat{B}_n \xrightarrow{P} B \quad (4.3.21)$$

Beweis. Wegen (4.3.1) und (4.3.2) gilt

$$\begin{aligned}
& n^{-1/2}U(\beta_0, \tau) \\
&= n^{-1/2} \sum_{i=1}^n \int_0^\tau (Z_i - E(\beta_0, t)) (dN_i(t) - Y_i(t) \exp\{\beta_0^\top Z_i(t)\} \alpha_0(t) dt) \\
&= n^{-1/2} \sum_{i=1}^n \int_0^\tau (Z_i - e(\beta_0, t)) \\
&\quad \cdot (dN_i(t) - Y_i(t) \exp\{\beta_0^\top Z_i(t)\} \alpha_0(t) dt) + o(1),
\end{aligned}$$

denn die vorhersagbare Variation von

$$\begin{aligned}
& n^{-1/2} \sum_{i=1}^n \int_0^\tau \{E(\beta_0, t) - e(\beta_0, t)\} (dN_i(t) - Y_i(t) \exp\{\beta_0^\top Z_i(t)\} \alpha_0(t) dt) \\
&= n^{-1/2} \int_0^\tau \{E(\beta_0, t) - e(\beta_0, t)\} \{dN(t) - S^{(0)}(\beta_0, t) \alpha_0(t) dt\}
\end{aligned}$$

ist nach (1.5.4) gegeben durch

$$\begin{aligned}
& n^{-1} \int_0^\tau \{E(\beta_0, t) - e(\beta_0, t)\}^2 S^{(0)}(\beta_0, t) \alpha_0(t) dt \\
& \leq |E(\beta_0, \cdot) - e(\beta_0, \cdot)|_{\mathcal{T}}^2 \{ |s^{(0)}(\beta_0, \cdot)|_{\mathcal{T}} + o_p(1) \} \int_0^\tau \alpha_0(t) dt
\end{aligned}$$

und konvergiert wegen Lemma 4.3.3 in Wahrscheinlichkeit gegen 0.

Aufgrund der Lengart'schen Ungleichung (1.7.3) konvergiert der gesamte Prozeß gleichmäßig in Wahrscheinlichkeit gegen 0, wenn die vorhersagbare Variation in τ in Wahrscheinlichkeit gegen 0 konvergiert.

$n^{-1/2}U(\beta_0, t)$ ist also asymptotisch äquivalent zu einer Summe von unabhängigen Zufallsvariablen mit Erwartungswert 0. Nach dem zentralen Grenzwertsatz 1.7.1 ist dieser Ausdruck gaußverteilt, wenn die Kovarianzmatrizen der Folge der gegen eine positiv definite Matrix B konvergieren.

Deshalb betrachtet man

$$\begin{aligned}
& \langle n^{-1/2} U(\beta_0, \tau) \rangle \\
&= \left\langle n^{-1/2} \sum_{i=1}^n \int_0^\tau (Z_i(t) - E(\beta_0, t)) \right. \\
&\quad \left. \cdot \{ dN_i(t) - Y_i(t) \exp \{ \beta_0^\top Z_i(t) \} \alpha_0(t) dt \} \right\rangle \\
&= \frac{1}{n} \sum_{i=1}^n \int_0^\tau \{ Z_i(t) - E(\beta_0, t) \}^{\otimes 2} Y_i(t) \exp \{ \beta_0^\top Z_i(t) \} \alpha_0(t) dt \\
&= \frac{1}{n} \sum_{i=1}^n \int_0^\tau \{ Z_i(t)^{\otimes 2} - E(\beta_0, t) Z_i(t)^\top - Z_i(t) E(\beta_0, t)^\top + E(\beta_0, t)^{\otimes 2} \} \\
&\quad \times Y_i(t) \exp \{ \beta_0^\top Z_i(t) \} \alpha_0(t) dt \\
&= \frac{1}{n} \int_0^\tau \left\{ S^{(2)}(\beta_0, t) - \frac{S^{(1)}(\beta_0, t)}{S^{(0)}(\beta_0, t)} S^{(1)}(\beta_0, t)^\top \right. \\
&\quad \left. - S^{(1)}(\beta_0, t) \frac{S^{(1)}(\beta_0, t)^\top}{S^{(0)}(\beta_0, t)} + E(\beta_0, t)^{\otimes 2} S^{(1)}(\beta_0, t) \right\} \alpha_0(t) dt \\
&\xrightarrow{P} \int_0^\tau v(\beta_0, t) s^{(0)}(\beta_0, t) \alpha_0(t) dt = \Sigma_\tau.
\end{aligned}$$

Das gilt aufgrund von C 1 bis C 3. Wegen C 4 ist Σ_τ positiv definit. Eine Folgerung daraus ist:

$$\mathbb{E} B_n = \mathbb{E} n^{-1} \sum_{i=1}^n w_i^{\otimes 2} = n^{-1} \sum_{i=1}^n \mathbb{E} \langle w_i \rangle \xrightarrow{P} \mathbb{E} \Sigma_\tau = B$$

Nun ist noch die Konsistenz von \hat{B}_n zu zeigen.

$$\begin{aligned}
\left| \hat{B}_n - B_n \right| &\leq \frac{1}{n} \sum_{i=1}^n |\hat{w}_i^{\otimes 2} - w_i^{\otimes 2}| \\
&\leq \frac{1}{n} \sum_{i=1}^n |\hat{w}_i - w_i| (|\hat{w}_i| + |w_i|) \quad (4.3.22)
\end{aligned}$$

Im Beweis von Satz 4.4.3 wird in einer allgemeineren Situation gezeigt, daß diese Summe in Wahrscheinlichkeit gegen 0 konvergiert. \square

In der Schreibweise wurde B von Σ_τ unterschieden, um die Analogien zu marginalen Modellen für Haufendaten hervorzuheben, bei denen im allgemeinen B nicht mit Σ_τ übereinstimmt.

Satz 4.3.5. Seien C 1 bis C 4 und $A_0(\tau) < \infty$, sowie $\alpha_0(\cdot) > 0$ erfüllt. Darüberhinaus sei $|Z_i(t)| \leq Z_0$ für alle $i = 1, \dots, n$ und $t \in \mathcal{T}$. Dann gilt

$$n^{1/2}(\hat{\beta} - \beta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_\tau^{-1} B \Sigma_\tau^{-1}) \quad (4.3.23)$$

und

$$\left| n^{-1} \mathcal{I}_t(\hat{\beta}) - \Sigma_t \right|_{\mathcal{T}} \xrightarrow{P} 0. \quad (4.3.24)$$

Beweis. (4.3.24) kann mit der Beweismethode in Satz 4.4.4 hergeleitet werden (vergleiche mit (Andersen et al. 1993, p. 500f)).

Um (4.3.23) zu zeigen, wird zuerst $U(\beta, \tau)$ komponentenweise nach dem Mittelwertsatz um β_0 entwickelt:

$$U_j(\beta, \tau) - U_j(\beta_0, \tau) = -\mathcal{I}_\tau^j(\beta_j^*)(\beta - \beta_0) \quad j = 1, \dots, p. \quad (4.3.25)$$

Dabei ist $\beta_j^* = \lambda_j \beta + (1 - \lambda_j) \beta_0$, $\lambda_j \in (0, 1)$, und $-\mathcal{I}_\tau^j$ ein Zeilenvektor von $-\mathcal{I}_\tau$ mit Komponenten

$$\frac{\partial}{\partial \beta_l} U_j(\beta, \tau), \quad l = 1, \dots, p.$$

An der Stelle $\beta = \hat{\beta}$ erhält man wegen $U_j(\hat{\beta}, \tau) = 0$

$$n^{-1/2} U_j(\beta_0, \tau) = \{n^{-1} \mathcal{I}_\tau^j(\beta_j^*)\} n^{1/2} (\hat{\beta} - \beta_0). \quad (4.3.26)$$

Aufgrund von C 4 ist Σ_τ positiv definit. $\mathcal{I}_\tau(\hat{\beta})$ konvergiert wegen (4.3.24) in Wahrscheinlichkeit gegen Σ_τ . Da $\mathcal{I}_\tau^j(\beta)$ wegen C 1 und C 3 stetig ist in β und somit auch für β aus einer (konvexen) Umgebung $B(\beta_0, \delta)$, $\delta > 0$, positiv definit ist, ist für $\hat{\beta} \in B(\beta_0, \delta)$ die Matrix I , die aus den Zeilen $\{n^{-1} \mathcal{I}_\tau^j(\beta_j^*)\}$ besteht, positiv definit und daher invertierbar und

$$n^{-1/2} I^{-1} U(\beta_0, \tau) = n^{1/2} (\hat{\beta} - \beta_0). \quad (4.3.27)$$

Wegen der Konsistenz von $\hat{\beta}$ gilt $\hat{\beta} \in B(\beta_0, \delta)$ und $\beta^* \in B(\beta_0, \delta)$ mit beliebig großer Wahrscheinlichkeit. Aufgrund der Stetigkeit von $\mathcal{I}_t(\beta)$ in $B(\beta_0, \delta)$ konvergiert $I \xrightarrow{P} \Sigma_\tau^{-1}$. Wegen Satz 4.3.4 gilt

$$n^{-1/2} U(\beta_0, \tau) \xrightarrow{\mathcal{D}} \mathcal{N}(0, B).$$

Aus Folgerung 1.6.2 und Satz 1.6.3 folgt nun bezüglich (4.3.27)

$$I^{-1} n^{-1/2} U(\beta_0, \tau) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_\tau^{-1} B \Sigma_\tau^{-1}).$$

□

4.4 Das Cox-Modell für Haufendaten

Im Cox-Modell sind die Daten der Individuen unabhängig verteilt. Häufig hat man jedoch die Situation, daß die Daten durch die Umstände der Stichprobenziehung gruppiert sind (siehe 2.5 und 2.6). Beispielsweise ist der Zeitpunkt der Erkrankung eines Auges an Retinopathia diabetica abhängig von der Erkrankung des anderen Auges am selben Patienten. Auch die Haltbarkeit eines Zahnes ist von den anderen Zähnen des Gebisses abhängig. In diesem Abschnitt wird die Schätzung eines Cox-Modells bei derartigen Abhängigkeiten in der Stichprobe behandelt.

Es wird angenommen, daß alle Haufen vollständig sind, d.h. daß die Haufengröße von allen Haufen gleich K ist. Unvollständige Haufen können durch den Zensierungsindikator Y_{ik} , der gleich 0 gesetzt wird, berücksichtigt werden. Es treten nun Schätzfunktionen auf, die keine Martingale mehr sind, denn der Kompensator bezüglich der gemeinsam erzeugten Filtrierung eines Haufens kann im marginalen Modell nicht als bekannt angenommen werden. Dann wäre nämlich die gesamte multivariate Verteilung bekannt. Aber gerade das will man vermeiden; man will vielmehr die Modellannahmen sparsam einsetzen, da sie in den Anwendungen zumeist nicht ausreichend begründet oder verifiziert werden können.

Seien die Daten $N_{ik}(t)$, $Y_{ik}(t)$ und $Z_{ik}(t)$, $t \in \mathcal{T}$, von Individuen $i1, \dots, iK$ aus Haufen $i = 1, \dots, n$ gegeben. Die Daten von verschiedenen Haufen seien unabhängig bezüglich der σ -Algebra der Kovariablen

$$\mathcal{Z} = \sigma(Z) = \sigma(Z_{ik}(t), i = 1, \dots, n, k = 1, \dots, K; t \in \mathcal{T}).$$

D. h. bezüglich $\mathbb{P}_{Z=z}$ sind für $i \neq i'$, $k, k' = 1, \dots, K$ die Vektoren von Zufallsvariablen $(N_{ik}(t), Y_{ik}(t), z_{ik}(t))$ und $(N_{i'k'}(t), Y_{i'k'}(t), z_{i'k'}(t))$ unabhängig. Ab nun soll $\mathbb{P}_{Z=z}$ die zugrundeliegende Wahrscheinlichkeit sein.

Jedes Haufenelement erzeugt eine Filtrierung

$$\mathcal{F}_{t,ik} = \sigma(N_{ik}(s), Y_{ik}(s), Z_{ik}(s) : 0 \leq s \leq t). \quad (4.4.1)$$

Die Beobachtungen sollen aus einem Zählprozeß N_{ik} stammen, dessen Intensität bezüglich $\mathcal{F}_{t,ik}$ und \mathbb{P} gegeben ist durch

$$\Lambda_{ik}(t) = Y_{ik}(t)\alpha_{ik}(t, Z_{ik}(t)) = Y_{ik}(t) \exp\{\beta_0^\top Z_{ik}(t)\} \alpha_0(t) \quad (4.4.2)$$

mit $i = 1, \dots, n$ und $k = 1, \dots, K$. Wie in Abschnitt 4.1 ist darin ein Zensierungs- und Filterungs-Prozeß schon berücksichtigt.

Es ist sinnvoll, die für einen Index k unabhängigen Filtrierungen zu $\mathcal{F}_{t,k} = \bigvee_{i=1}^n \mathcal{F}_{t,ik}$ zusammenzufassen. Genauso faßt man die Zählprozesse für den Index k zusammen zu $N^{(k)}(t) = (N_{1k}(t), \dots, N_{nk}(t))$. $N^{(k)}(t)$ ist ein multivariater Zählprozeß bezüglich $\mathcal{F}_{t,k}$ (siehe 1.5.4). Der Kompensator der Komponente $N_{ik}(t)$ bezüglich $\mathcal{F}_{t,k}$ ist nach Konstruktion $\Lambda_{ik}(t)$. Der Kompensator von $N_{.k} = \sum_{i=1}^n N_{ik}$ bezüglich $\mathcal{F}_{t,k}$ ist folglich $\Lambda_{.k} = \sum_{i=1}^n \Lambda_{ik}$.

Λ_{ik} soll absolut stetig sein. Der Intensitätsprozeß soll einem *semiparametrischen* Modell mit *proportionalen* Hasardfunktionen genügen:

$$\alpha_{ik}^\theta(t, Z_{ik}(t)) = \alpha_0(t) \exp \{ \beta_0^\top Z_{ik}(t) \}. \quad (4.4.3)$$

Für die nachstehenden Berechnungen werden die folgenden Ausdrücke benötigt: Seien die skalar-, vektor- und matrixwertigen Prozesse $S_k^{(d)}(\beta, t)$ für $d = 0, 1, 2$ und $k = 1, \dots, K$ definiert durch

$$S_k^{(d)}(\beta, t) = \sum_{i=1}^n Z_{ik}(t)^{\otimes d} Y_{ik}(t) \exp \{ \beta^\top Z_{ik}(t) \} dN_{.k}(t), \quad (4.4.4)$$

sowie $S^{(d)}(\beta, t) = \sum_{k=1}^K S_k^{(d)}(\beta, t)$, $E_k(\beta, t) = S_k^{(1)}(\beta, t)/S_k^{(0)}(\beta, t)$, $E_K(\beta, t) = \sum_{k=1}^K E_k(\beta, t)$ und $E(\beta, t) = S^{(1)}(\beta, t)/S^{(0)}(\beta, t)$.

In (Lee, Wei und Amato 1992) wird aufbauend auf (Wei et al. 1989, Lin und Wei 1989) festgestellt, daß unter gewissen Regularitätsbedingungen die fehlspezifizierte Cox-Likelihood⁴

$$L(\beta) = \prod_{t \in \mathcal{T}, i, k} \left(\frac{\exp \{ \beta^\top Z_{ik}(t) \}}{S^{(0)}(\beta, t)} \right)^{\Delta N_{ik}(t)} \quad (4.4.5)$$

die konsistente Schätzung von β_0 ermöglicht (vergleiche mit (4.2.7)). Sie ist proportional zum Produkt der partiellen marginalen Profile-Likelihoods (4.2.5) für $k = 1, \dots, K$.

Der *logarithmierte (fehlspezifizierte) Likelihoodprozeß* ist (mit $N = \sum_{k=1}^K N_{.k}$)

$$C(\beta, t) = \sum_{i=1}^n \sum_{k=1}^K \int_0^t \beta^\top Z_{ik}(u) dN_{ik}(u) - \int_0^t \log S^{(0)}(\beta, u) dN(u) \quad (4.4.6)$$

Der Schätzer $\hat{\beta}$ ist wiederum als Nullstelle des (fehlspezifizierten) Score-Prozesses

$$U(\beta, t) = \frac{\partial C(\beta, t)}{\partial \beta} = \sum_{i=1}^n \sum_{k=1}^K \int_0^t Z_{ik}(u) dN_{ik}(u) - \int_0^t E(\beta, u) dN(u) \quad (4.4.7)$$

⁴im Sinn von (Boos 1990, Lin und Wei 1989): Die Likelihood ist nicht gültig, da das Wahrscheinlichkeitsmodell für die Daten nicht zutrifft.

definiert. Der (fehlspezifizierte) Informationsprozeß ist

$$\mathcal{I}(\beta, t) = \frac{\partial^2 C(\beta, t)}{\partial \beta^{\otimes 2}} = \int_0^t \left\{ \frac{S^{(2)}(\beta, u)}{S^{(0)}(\beta, u)} - E(u, \beta)^{\otimes 2} \right\} dN(u) \quad (4.4.8)$$

Lee et al. (1992) gaben darüberhinaus einen konsistenten Schätzer für die Kovarianz von $\hat{\beta}$ an.

In den Publikationen, die sich auf den Sandwich-Schätzer der asymptotischen Kovarianz (Lin und Wei 1989) beziehen, gibt es zumeist nur Beweisskizzen oder lückenhafte Beweise. Verschiedene Aspekte von marginalen oder fehlspezifizierten Cox-Modellen (Wei et al. 1989, Lin und Wei 1989, Lin 1991, Lee et al. 1992, Lin und Ying 1993, Pepe und Cai 1993, Cai und Prentice 1995, Spiekerman und Lin 1996) behandelt. Im übrigen beziehen sich diese Arbeiten nur auf unabhängig identisch verteilte Haufendaten und nur Zählprozesse mit maximal einem Ereignis (Einsprung-Prozesse) sind zugelassen.

4.4.1 Konsistenz von $\hat{\beta}$

Der (fehlspezifizierte) Score-Prozeß $U(\beta_0, t)$ ist kein Martingal, aber es gibt einen in asymptotischen ähnlichen Prozeß, auf den die Argumente aus der Asymptotik des Cox-Modells angewendet werden können:

$$\begin{aligned} U_k(\beta_0, t) &= \sum_{k=1}^K \sum_{i=1}^n \int_0^t [Z_{ik}(u) - E_k(\beta_0, u)] dN_{ik}(u) \\ &= \sum_{k=1}^K \sum_{i=1}^n \int_0^t [Z_{ik}(u) - E_k(\beta_0, u)] dM_{ik}(u) \end{aligned} \quad (4.4.9)$$

wobei M_{ik} die eindeutig bestimmten Martingale aus der Doob-Meyer-Zerlegung

$$N_{ik}(t) = M_{ik}(t) + \int_0^t Y_{ik}(u) \exp \{ \beta_0^\top Z_i(u) \} \alpha_0(u) du \quad (4.4.10)$$

bezüglich $\mathcal{F}_{t,k}$ sind. (4.4.9) gilt wegen

$$\begin{aligned} &\sum_{i=1}^n [Z_{ik}(u) - E_k(\beta, u)] Y_{ik}(u) \exp \{ \beta^\top Z_{ik}(u) \} \\ &= S_k^{(1)}(\beta, u) - S_k^{(0)}(\beta, u) S_k^{(1)}(\beta, u) / S_k^{(0)}(\beta, u) = 0 \end{aligned}$$

Darüberhinaus sind die Kompensatoren absolut stetig, sodaß nach (1.5.3) die M_{ik} für ein k orthogonal sind mit

$$\langle M_{ik} \rangle (t) = \int_0^t Y_{ik}(u) \exp \{ \beta_0^\top Z_{ik}(u) \} \alpha_0(u) du \quad (4.4.11)$$

Schließlich definiert man in Analogie zu (4.2.10)

$$V_k(\beta, t) = \frac{S_k^{(2)}(\beta, t)}{S_k^{(0)}(\beta, t)} - E_k(\beta, t)^{\otimes 2}. \quad (4.4.12)$$

Folgende Bedingungen werden für die nachfolgenden Ergebnisse benötigt:

L 1. Es existiert eine offene Umgebung \mathcal{B} von β_0 und skalarwertige, p -vektorwertige und $p \times p$ -matrixwertige Funktionen $s^{(0)}(\beta, t)$, $s^{(1)}(\beta, t) = \frac{\partial s^{(0)}(\beta, t)}{\partial \beta}$, $s^{(2)}(\beta, t) = \frac{\partial^2 s^{(0)}(\beta, t)}{\partial \beta^{\otimes 2}}$ mit Definitionsbereich $\mathcal{B} \times \mathcal{T}$, sodaß für $m = 0, 1, 2$

$$\left| \frac{1}{n} S_k^{(d)}(\beta, t) - s^{(d)}(\beta, t) \right|_{\mathcal{B} \times \mathcal{T}} \xrightarrow{P} 0 \quad \text{für} \quad k = 1, \dots, K. \quad (4.4.13)$$

L 2. $s^{(m)}(\cdot)$ für $m = 0, 1, 2$ ist eine stetige Funktion von $\beta \in \mathcal{B}$, gleichmäßig in $t \in \mathcal{T}$ und beschränkt auf $\mathcal{B} \times \mathcal{T}$.

L 3. $s^{(0)}(\beta_0, \cdot)$ ist auf \mathcal{T} von 0 weg beschränkt.

L 4. $\Sigma_\tau = \int_0^\tau v(\beta_0, t) s^{(0)}(\beta_0, t) \alpha_0(t) dt$ ist positiv definit. Dabei ist $v = s^{(2)}/s^{(0)} - e^{\otimes 2}$ und $e = s^{(1)}/s^{(0)}$.

Satz 4.4.1. Sei L 1 bis L 4 und $A_0(\tau) < \infty$, sowie $\alpha_0(\cdot) > 0$ erfüllt. Darüberhinaus sei $|Z_{ik}(t)| \leq Z_0$. Dann konvergiert die Wahrscheinlichkeit, daß die Gleichung $U(\beta, \tau) = 0$ eine eindeutige Lösung $\hat{\beta}$ hat, gegen 1 für $n \rightarrow \infty$ und $\hat{\beta} \xrightarrow{P} \beta_0$

Beweis. Die fehlspezifizierte logarithmierte Likelihoodfunktion (4.4.6) kann

folgendermaßen dargestellt werden:

$$\begin{aligned}
& X(\beta, t) \\
&= \frac{1}{nK} (C(\beta, t) - C(\beta_0, t)) \\
&= \frac{1}{nK} \sum_{k=1}^K \sum_{i=1}^n \int_0^t (\beta - \beta_0)^\top Z_{ik}(s) dN_{ik}(s) - \frac{1}{n} \int_0^t \log \left\{ \frac{S^{(0)}(\beta, s)}{S^{(0)}(\beta_0, s)} \right\} dN_{.k}(s)
\end{aligned} \tag{4.4.14}$$

$$\begin{aligned}
&= \frac{1}{nK} \sum_{k=1}^K \sum_{i=1}^n \int_0^t (\beta - \beta_0)^\top Z_{ik}(s) dN_{ik}(s) \\
&\quad - \frac{1}{nK} \int_0^t \log \left\{ \frac{S_k^{(0)}(\beta, s)}{S_k^{(0)}(\beta_0, s)} \right\} dN_{.k}(s) + R(\beta, t)
\end{aligned} \tag{4.4.15}$$

mit

$$R(\beta, t) = \frac{1}{nK} \sum_{k=1}^K \int_0^t \left[\log \left\{ \frac{S_k^{(0)}(\beta, s)}{S_k^{(0)}(\beta_0, s)} \right\} - \log \left\{ \frac{S^{(0)}(\beta, s)}{S^{(0)}(\beta_0, s)} \right\} \right] dN_{.k}(s).$$

Der Restterm konvergiert gleichmäßig in (β, t) gegen 0 mit folgender Begründung:

$$\left| \frac{1}{nK} S^{(0)}(\beta, t) - s^{(0)}(\beta, t) \right|_{\mathcal{B} \times \mathcal{T}} \xrightarrow{P} 0 \tag{4.4.16}$$

gilt aufgrund von L 1 und somit ist (4.3.17) aus Lemma 4.3.3 gültig. Daher konvergiert

$$\left| \sum_{k=1}^K \log \left\{ \frac{S_k^{(0)}(\beta, s)}{S_k^{(0)}(\beta_0, s)} \right\} - \log \left\{ \frac{S^{(0)}(\beta, s)}{S^{(0)}(\beta_0, s)} \right\} \right|_{\mathcal{B} \times \mathcal{T}} \xrightarrow{P} 0. \tag{4.4.17}$$

Darüberhinaus liefert die Lengart'sche Ungleichung (1.7.2)

$$\mathbb{P} \left(\frac{1}{n} \left| \int_0^\tau dN_{.k}(s) \right| > \eta \right) \leq \frac{\delta}{\eta} + \mathbb{P} \left(\int_0^\tau \frac{1}{n} S_k^{(0)}(\beta_0, t) \alpha_0(t) dt > \delta \right) \tag{4.4.18}$$

Für $\delta > \frac{1}{n} \int_0^\tau s_k^{(0)}(\beta_0, t) \alpha_0(t) dt$ konvergiert (4.4.18) für η gegen unendlich gegen 0 (Wegen L 1, L 2 und $A_0(\tau) < \infty$). Somit ist $\frac{1}{n} \int_0^\tau dN_{.k}(s)$ in Wahrscheinlichkeit beschränkt und die Behauptung, daß $R(\beta, t)$ in Wahrscheinlichkeit gegen 0 konvergiert, gezeigt.

Mit den Argumenten aus dem Beweis von Satz 4.3.1 kann man nun die Konsistenz von $\hat{\beta}$ zeigen, indem man mit Hilfe von (4.4.14) zeigt, daß die konkave Funktion (4.4.14) punktweise gegen eine konkave Funktion konvergiert (Andersen et al. 1993, p497f). Man betrachtet den k -ten Summanden von (4.4.15) (ohne den Restterm):

$$X_k(\beta, t) = \frac{1}{n} \sum_{i=1}^n \int_0^t (\beta - \beta_0)^\top Z_{ik}(s) dN_{ik}(s) - \frac{1}{n} \int_0^t \log \left\{ \frac{S_k^{(0)}(\beta, s)}{S_k^{(0)}(\beta_0, s)} \right\} dN_{\cdot k}(s).$$

Zu diesem Ausdruck kann man aufgrund der Zerlegung (4.4.10) und der Vorhersagbarkeit und lokalen Beschränktheit der $Z_{ik}(t)$ und der $Y_{ik}(t)$ den Kompensator bezüglich $\mathcal{F}_{t,k}$ angeben als (Satz 1.4.1)

$$\begin{aligned} & \tilde{X}_k(\beta, t) \\ &= \frac{1}{n} \int_0^t \left[(\beta - \beta_0)^\top S_k^{(1)}(\beta_0, s) - \log \left\{ \frac{S_k^{(0)}(\beta, s)}{S_k^{(0)}(\beta_0, s)} \right\} S_k^{(0)}(\beta_0, s) \right] \alpha_0(s) ds. \end{aligned} \quad (4.4.19)$$

$X_k(\beta, \cdot) - \tilde{X}_k(\beta, \cdot)$ ist ein lokal quadratisch integrierbares Martingal (Satz 1.4.1). Die vorhersagbare Variation ist daher aufgrund von Satz 1.4.1

$$\begin{aligned} & \left\langle X_k(\beta, \cdot) - \tilde{X}_k(\beta, \cdot) \right\rangle (t) \\ &= \frac{1}{n^2} \sum_{i=1}^n \int_0^t \left[(\beta - \beta_0)^\top Z_{ik}(s) - \log \left\{ \frac{S_{ik}^{(0)}(\beta, s)}{S_{ik}^{(0)}(\beta_0, s)} \right\} \right]^2 \\ & \quad \cdot Y_{ik}(s) \exp \{ \beta_0^\top Z_{ik}(s) \} \alpha_0(s) ds. \end{aligned} \quad (4.4.20)$$

Wegen $A_0(\tau) < \infty$ und L 1, L 2 und L 3 konvergiert $\tilde{X}_k(\beta, \tau)$ für ein $\beta \in \mathcal{B}$ in Wahrscheinlichkeit gegen eine Funktion

$$f(\beta) = \int_0^\tau \left[(\beta - \beta_0)^\top s^{(1)}(\beta_0, s) - \log \left\{ \frac{s^{(0)}(\beta, s)}{s^{(0)}(\beta_0, s)} \right\} s^{(0)}(\beta_0, s) \right] \alpha_0(s) ds. \quad (4.4.21)$$

und auch $n \left\langle X_k(\beta, \cdot) - \tilde{X}_k(\beta, \cdot) \right\rangle (\tau)$ konvergiert gegen eine beschränkte Funktion von β . Nun kann die Lengart'sche Ungleichung (1.7.3) angewendet werden. Wegen $\left\langle X_k(\beta, \cdot) - \tilde{X}_k(\beta, \cdot) \right\rangle (\tau) \xrightarrow{P} 0$ gilt auch

$$X_k(\beta, \tau) \xrightarrow{P} f(\beta). \quad (4.4.22)$$

Wegen $A_0(\tau) < \infty$ und wegen L 2 und L 3 gilt

$$\frac{\partial}{\partial \beta} f(\beta) = \int_0^\tau \{e(\beta_0, s) - e(\beta, s)\} s^{(0)}(\beta_0, s) \alpha_0(s) ds \quad (4.4.23)$$

und

$$\frac{\partial^2}{\partial \beta^{\otimes 2}} f(\beta) = \int_0^\tau v(\beta, s) s^{(0)}(\beta_0, s) \alpha_0(s) ds \quad (4.4.24)$$

ist positiv semidefinit und positiv definit für $\beta = \beta_0$ aufgrund von L 4. $X_k(\beta, \tau)$ konvergiert daher in \mathcal{B} punktweise in Wahrscheinlichkeit gegen eine konkave Funktion $f(\beta)$ mit eindeutigem Maximum bei $\beta = \beta_0$. Die zufällige Funktion $X_k(\beta, \tau)$ ist konkav. Auch $X(\beta, \tau) = \sum_{k=1}^K X_k(\beta, \tau)$ ist konkav und konvergiert punktweise gegen $f(\beta)$. Nach Satz 1.7.2 ist die Konvergenz von $X(\beta, \tau)$ gleichmäßig in einer Umgebung B von β_0 und nach Folgerung 1.7.3 konvergiert die Wahrscheinlichkeit, daß $X(\beta, \tau)$ ein Maximum bei $\hat{\beta}$ annimmt, gegen 1, und die maximierenden Werte $\hat{\beta}$ konvergieren in Wahrscheinlichkeit gegen den maximierenden Wert β_0 von $f(\beta)$. \square

4.4.2 Asymptotische Verteilung von $\hat{\beta}$

Für die im nachstehenden Satz hergeleitete asymptotische Verteilung von $U(\beta_0, \tau)$ werden folgende Definitionen benötigt: Sei

$$B_n = \frac{1}{n} \sum_{i=1}^n w_i^{\otimes 2},$$

wobei

$$w_i = \sum_{k=1}^K \int_0^\tau (Z_{ik}(t) - e(\beta_0, t)) \{dN_{ik}(t) - Y_{ik}(t) \exp\{\beta_0^\top Z_{ik}(t)\} \alpha_0(t) dt\}. \quad (4.4.25)$$

Sei

$$\hat{B}_n = \frac{1}{n} \sum_{i=1}^n \hat{w}_i^{\otimes 2},$$

wobei unter Verwendung des Breslow-Schätzers $\hat{A}(\cdot, \cdot)$ aus (4.2.4) definiert ist

$$\hat{w}_i = \sum_{k=1}^K \int_0^\tau (Z_{ik}(t) - E(\hat{\beta}, t)) \{dN_{ik}(t) - Y_{ik}(t) \exp\{\hat{\beta}^\top Z_{ik}(t)\} d\hat{A}_0(\hat{\beta}, t)\} \quad (4.4.26)$$

Satz 4.4.2. Seien L 1 bis L 3 und $A_0(\tau) < \infty$, sowie $|\alpha_0(\cdot)|_{\mathcal{T}} > 0$ erfüllt. Dann gilt für $t \in [0, \tau]$

$$\sup_{t \in \mathcal{T}} \left| \hat{A}_0(\hat{\beta}, t) - \int_0^t \alpha_0(u) du \right| \xrightarrow{P} 0. \quad (4.4.27)$$

Beweis. Sei $J_k(t) = I(S_k^{(0)}(\beta_0(t) > 0)$. $n^{-1}N_{\cdot k}(t)$ ist in Wahrscheinlichkeit beschränkt und $\int_0^t (J(u) - J_k(u)) n^{-1} dN_{\cdot k}(u)$ in Wahrscheinlichkeit gegen 0 (gleichmäßig in $t \in \mathcal{T}$!). Wegen

$$\begin{aligned} \hat{A}_0(\hat{\beta}, t) &= \int_0^t \frac{J(u)}{S^{(0)}(\hat{\beta}, u)} dN(u) = \sum_{k=1}^K \int_0^t \frac{J(u)}{S^{(0)}(\hat{\beta}, u)} dN_{\cdot k}(u) \\ &= K^{-1} \sum_{k=1}^K \int_0^t \frac{J_k(u)}{S_k^{(0)}(\beta_0, u)} dN_{\cdot k}(u) \\ &\quad + K^{-1} \sum_{k=1}^K \int_0^t (J(u) - J_k(u)) \cdot K S^{(0)}(\hat{\beta}, u)^{-1} dN_{\cdot k}(u) \\ &\quad + K^{-1} \sum_{k=1}^K \int_0^t J_k(u) \cdot \left(K S^{(0)}(\hat{\beta}, u)^{-1} - S_k^{(0)}(\beta_0, u)^{-1} \right) dN_{\cdot k}(u) \\ &= K^{-1} \sum_{k=1}^K \int_0^t \frac{J_k(u)}{S_k^{(0)}(\beta_0, u)} dN_{\cdot k}(t) + o_p(1) \\ &= K^{-1} \sum_{k=1}^K \int_0^t \frac{J_k(u)}{S_k^{(0)}(\beta_0, u)} dM_{\cdot k}(t) + \int_0^t J_k(u) \alpha_0(u) du + o_p(1). \end{aligned}$$

Wegen L 1 bis L 3 und $\int_0^t J(u) - J_k(u) dN_{\cdot k}(u) = 0$ konvergiert der Restterm gleichmäßig in \mathcal{T} . Das Resultat enthält das Integral eines vorhersagbaren $(\mathcal{F}_{t,k})$ -Prozesses bezüglich eines quadratisch integrierbaren Martingals (bezüglich $(\mathcal{F}_{t,k})$). Wegen Satz 1.4.1 und nach der Lenglart'schen Ungleichung kann das quadratisch integrierbare Martingal durch einen Ausdruck abgeschätzt werden, der gleichmäßig für $t \in \mathcal{T}$ in Wahrscheinlichkeit gegen 0 geht. Schließlich ist wegen L 1 und L 3 $|J_h - 1|_{\mathcal{T}} = 0$ mit beliebig hoher Wahrscheinlichkeit für $n \rightarrow \infty$. Daher konvergiert auch $\int_0^t J_k(u) \alpha_0(u) du$ gleichmäßig in Wahrscheinlichkeit gegen $\int_0^t \alpha_0(u) du$. \square

Vom folgenden Satz sind bisher nur Versionen für unabhängig identisch verteilte Zufallsvariable und Einsprung-Prozesse angegeben. Der Beweis von Lee et al. (1992) ist unvollständig. Die Anfrage bei der Korrespondenzadresse ergab, daß der vollständige Beweis nicht auffindbar ist.

Satz 4.4.3. Seien L 1 bis L 4 und $A_0(\tau) < \infty$, sowie $\alpha_0(\cdot) > 0$ erfüllt. Darüberhinaus sei $Z_{ik}(t)$ gleichmäßig beschränkt, d.h. $|Z_{ik}|_{\mathcal{T}} \leq Z_0$ für $i = 1, \dots, n, k = 1, \dots, K$.

Wenn $B = \lim_{n \rightarrow \infty} \mathbb{E} B_n$ existiert, dann gilt

$$n^{-1/2}U(\beta_0, \tau) \xrightarrow{\mathcal{D}} \mathcal{N}(0, B). \quad (4.4.28)$$

Wenn darüberhinaus $Z_{ik}(t)$ linksseitig stetig und von gleichmäßig beschränkter Variation ist, d.h. $|Z_{ik}|_{\mathcal{T}} + \int_0^{\tau} |dZ_{ik}|(t) dt \leq Z_0$ für $i = 1, \dots, n, k = 1, \dots, K$, und auch $e(\beta_0, t)$ von beschränkter Variation ist, dann

$$\hat{B}_n \xrightarrow{P} B. \quad (4.4.29)$$

Beweis. Zur Demonstration von (4.4.28) wird die Scorefunktion als Summe unabhängiger Zufallsvariablen plus einem Restterm, der in Wahrscheinlichkeit gegen 0 konvergiert, dargestellt werden. Auf diese Summe wird dann der zentrale Grenzwertsatz angewendet werden.

Wegen (4.4.7) und (4.4.10) gilt:

$$\begin{aligned} & n^{-1/2}U(\beta_0, \tau) \\ &= n^{-1/2} \sum_{k=1}^K \sum_{i=1}^n \int_0^{\tau} (Z_{ik}(t) - E(\beta_0, t)) dM_{ik}(t) \\ &= n^{-1/2} \sum_{k=1}^K \sum_{i=1}^n \int_0^{\tau} (Z_{ik}(t) - e(\beta_0, t)) dM_{ik}(t) + R_1 \quad (4.4.30) \\ &= n^{-1/2} \sum_{k=1}^K \sum_{i=1}^n \int_0^{\tau} (Z_{ik}(t) - E_k(\beta_0, t)) dM_{ik}(t) + R_2 \\ &= n^{-1/2} \sum_{i=1}^n w_i(\beta_0) + R_1 \end{aligned}$$

Dabei ist

$$R_1 = n^{-1/2} \sum_{k=1}^K \int_0^{\tau} (E(\beta_0, t) - e(\beta_0, t)) dM_{.k}(t). \quad (4.4.31)$$

und

$$\begin{aligned} R_2 &= n^{-1/2} \sum_{k=1}^K \int_0^{\tau} (E(\beta_0, t) - E_k(\beta_0, t)) dM_{.k}(t) \\ &= n^{-1/2} \sum_{k=1}^K \int_0^{\tau} (E(\beta_0, t) - e(\beta_0, t)) - (E_k(\beta_0, t) - e(\beta_0, t)) dM_{.k}(t). \end{aligned}$$

Wenn man gezeigt hat, daß R_1 in Wahrscheinlichkeit gegen 0 konvergiert, dann folgt sogleich, daß auch R_2 in Wahrscheinlichkeit gegen 0 konvergiert, da $R_2 \xrightarrow{P} 0$ auf den Spezialfall $R_1 \xrightarrow{P} 0$ für $K = 1$ zurückgeführt werden kann. Darüberhinaus ist die Konvergenz jedes der Summanden mit $k = 1, \dots, K$ von R_1 gegen 0 hinreichend.

Es gilt für $k = 1, \dots, K$ die Abschätzung

$$\begin{aligned} & n^{-1/2} \left| \int_0^\tau (E(\beta_0, t) - e(\beta_0, t)) \, dM_{.k}(t) \right| \\ & \leq \sup_{t \in \mathcal{T}} |E(\beta_0, t) - e(\beta_0, t)| n^{-1/2} \int_0^\tau |dM_{.k}(t)| \\ & \leq \sup_{t \in \mathcal{T}} |E(\beta_0, t) - e(\beta_0, t)| n^{-1/2} (N_{.k}(\tau) + \Lambda_{.k}(\tau)). \end{aligned} \quad (4.4.32)$$

$\sup_{t \in \mathcal{T}} |E(\beta_0, t) - e(\beta_0, t)|$ konvergiert aufgrund von (4.3.16) aus Lemma 4.3.3 in Wahrscheinlichkeit gegen 0 (Aus L 1 bis L 3 folgt C 1 bis C 3). Aus dem angegebenen Beweis ist ersichtlich, daß die Aussage auch dann gilt, wenn E der Quotient zweier Summen von Ausdrücken $S_k^{(d)}$, $d = 1, 0$ ist.). Der zweite Faktor $n^{-1/2} (N_{.k}(\tau) + \Lambda_{.k}(\tau))$ von (4.4.32) ist in Wahrscheinlichkeit beschränkt. Denn aufgrund der Čebyšev'schen Ungleichung ist dafür die Beschränktheit des zweiten Moments ausreichend, was aufgrund von

$$\begin{aligned} & \mathbb{E} n^{-1} (N_{.k}(\tau) + \Lambda_{.k}(\tau))^2 \\ & = \mathbb{E} n^{-1} (M_{.k}(\tau) + 2\Lambda_{.k}(\tau))^2 \\ & \leq n^{-1} 2 (\mathbb{E} M_{.k}^2(\tau) + 4 \mathbb{E} \Lambda_{.k}^2(\tau)) \\ & = 2n^{-1} (\mathbb{E} \Lambda_{.k}(\tau) + 4 \mathbb{E} \Lambda_{.k}^2(\tau)) \\ & \leq 2n^{-1} \left(\mathbb{E} n \exp \left\{ |\beta_0|_1^\top Z_0 \right\} A_0(\tau) + 4 \mathbb{E} n \exp \left\{ 2 |\beta_0|_1^\top Z_0 \right\} A_0^2(\tau) \right) \\ & = 2 \left(\exp \left\{ |\beta_0|_1^\top Z_0 \right\} A_0(\tau) + 4 \exp \left\{ 2 |\beta_0|_1^\top Z_0 \right\} A_0^2(\tau) \right) \end{aligned}$$

und der Voraussetzungen des Satzes erfüllt ist.

Daher ist $n^{-1/2} U(\beta_0)$ asymptotisch äquivalent zu $n^{-1/2} \sum_{i=1}^n w_i(\beta_0)$, einer Summe von unabhängigen Zufallsvariablen. Es gilt $\mathbb{E} w_i = 0$ und $\frac{\text{Var}}{w}_i = B$. Nach dem zentralen Grenzwertsatz Satz 1.7.1 ist dieser Ausdruck asymptotisch gaußverteilt.

Nun ist noch die (4.4.29) zu zeigen.

B_n ist ein Mittelwert von unabhängigen Zufallsvariablen und konvergiert aufgrund des Varianzkriteriums für Reihen von Khinchin und Kolmogorov

(Lemma 1.7.4) gegen B , da die benötigten Momente von $N(\tau)$ aufgrund von Folgerung 1.5.3 existieren und gleichmäßig beschränkt sind.

Es genügt daher,

$$\begin{aligned}
\left| \hat{B}_n - B_n \right| &\leq \frac{1}{n} \sum_{i=1}^n \left| \hat{w}_i^{\otimes 2} - w_i(\beta_0)^{\otimes 2} \right| \\
&\leq \frac{1}{n} \sum_{i=1}^n |\hat{w}_i - w_i(\beta_0)| \cdot (|\hat{w}_i| + |w_i(\beta_0)|) \\
&\leq \left(\frac{1}{n} \sum_{i=1}^n |\hat{w}_i - w_i|^2 \right)^{-1/2} \left(\frac{1}{n} \sum_{i=1}^n |\hat{w}_i + w_i|^2 \right)^{-1/2} \quad (4.4.33)
\end{aligned}$$

zu betrachten. Es gilt

$$\begin{aligned}
\sup_{i \in \mathbb{N}} |w_i|^2 &\leq 2 \sup_{i \in \mathbb{N}} \left| Z_{ik}(t) - E(\hat{\beta}, t) \right|_{\mathcal{T}}^2 \\
&\quad \cdot \left(N_i(\tau)^2 + \sup_{i \in \mathbb{N}} \exp \left\{ 2\hat{\beta}^\top Z_{ik}(t) \right\} \hat{A}_0(\hat{\beta}, \tau)^2 \right).
\end{aligned}$$

Dementsprechend kann der zweite Faktor von (4.4.33) in eine Summe von Produkten aus einem Faktor, der in Wahrscheinlichkeit beschränkt ist, und einer Reihe aus unabhängigen Summanden, die nach dem Varianzkriterium von Khinchin und Kolmogorov 1.7.4 abgeschätzt werden kann, dargestellt werden. Der Beitrag dieses Faktors zu (4.4.33) ist also in Wahrscheinlichkeit beschränkt.

Nun wird gezeigt, daß der erste Faktor von (4.4.33) in Wahrscheinlichkeit gegen 0 konvergiert.

Nun wird partielle Integration angewendet (Fleming und Harrington 1991, p. 320):

$$\begin{aligned}
& \sum_{k=1}^K \int_0^\tau (Z_{ik}(t) - e(\beta_0, t)) Y_{ik}(t) \exp \{ \beta_0^\top Z_{ik}(t) \} \{ d\hat{A}_0(\hat{\beta}, t) - \alpha_0 dt \} \\
&= \sum_{k=1}^K \left(\hat{A}_0(\hat{\beta}, \tau) - A_0(\beta_0, \tau) \right) - \sum_{k=1}^K \int_0^\tau \left(\hat{A}_0(\hat{\beta}, t-) - A_0(\beta_0, t) \right) \\
&\quad d \{ Z_{ik}(t) - e(\beta_0, t) \} Y_{ik}(t) \exp \{ \beta_0^\top Z_{ik}(t) \} \\
&\leq \sum_{k=1}^K \left| \hat{A}_0(\hat{\beta}, t-) - A_0(\beta_0, t) \right|_{\mathcal{T}} \\
&\quad \cdot \left(1 + \int_0^\tau |d \{ Z_{ik}(t) - e(\beta_0, t) \} Y_{ik}(t) \exp \{ \beta_0^\top Z_{ik}(t) \}| \right).
\end{aligned} \tag{4.4.36}$$

Aufgrund von Satz 4.4.2 und der gleichmäßig beschränkten Variation der $Z_{ik}(t)$ konvergiert dieser Ausdruck gleichmäßig in i in Wahrscheinlichkeit gegen 0. Somit konvergiert auch der Beitrag von (4.4.35) zu (4.4.33) in Wahrscheinlichkeit gegen 0. \square

Satz 4.4.4. *Seien C 1 bis C 4 und $A_0(\tau) < \infty$, sowie $\alpha_0(\cdot) > 0$ erfüllt. Darüberhinaus sei $|Z_{ik}(t)| \leq Z_0$ für alle $i = 1, \dots, n$ und $t \in \mathcal{T}$. Dann*

$$n^{1/2}(\hat{\beta} - \beta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_\tau^{-1} B \Sigma_\tau^{-1}) \tag{4.4.37}$$

und

$$\left| n^{-1} \mathcal{I}_t(\hat{\beta}) - \Sigma_t \right|_{\mathcal{T}} \xrightarrow{P} 0 \tag{4.4.38}$$

Beweis. (4.4.38) wird zur Vervollständigung der Darstellung gezeigt, obwohl sich die Argumente auf (Andersen et al. 1993, p500f) fast direkt übertragen lassen.

Für beliebiges β_* gilt:

$$\begin{aligned}
 \left| \frac{1}{n} \mathcal{I}_t(\beta_*) - \Sigma_t \right|_{\mathcal{T}} &\leq \left| \int_0^t (V(\beta_*, u) - v(\beta_*, u)) \frac{dN(u)}{n} \right|_{\mathcal{T}} \\
 &+ \left| \int_0^t (v(\beta_*, u) - v(\beta_0, u)) \frac{dN(u)}{n} \right|_{\mathcal{T}} \\
 &+ \left| \int_0^t v(\beta_0, u) \left\{ \frac{dN(u)}{n} - \frac{1}{n} S^{(0)}(\beta_0, u) \right\} \alpha_0(u) du \right|_{\mathcal{T}} \\
 &+ \left| \int_0^t v(\beta_0, u) \left(\frac{1}{n} S^{(0)}(\beta_0, u) - s^{(0)}(\beta_0, u) \right) \alpha_0(u) du \right|_{\mathcal{T}} \\
 &= D_1 + D_2 + D_3 + D_4.
 \end{aligned} \tag{4.4.39}$$

Aus L 1 bis L 3 folgt

$$|V(\beta, t) - v(\beta, t)|_{\mathcal{B} \times \mathcal{T}} \xrightarrow{P} 0$$

und für $\beta_* \xrightarrow{P} \beta_0$ gilt daher auch $|V(\beta_*, t) - v(\beta_*, t)|_{\mathcal{T}} \xrightarrow{P} 0$. Mittels der Lenglar't'schen Ungleichung (1.7.2) erhält man

$$P \left\{ \frac{1}{n} N_{.k}(\tau) > \eta \right\} \leq \frac{\delta}{\eta} + P \left\{ \int_0^\tau \frac{1}{n} S_k^{(0)}(\beta_0, t) \alpha_0(t) dt > \delta \right\}$$

Wenn man nun $\delta > \int_0^\tau s_k^{(0)}(\beta_0, t) \alpha_0(t) dt$ wählt, dann geht dieser Ausdruck in Wahrscheinlichkeit gegen 0. Daher ist $n^{-1} N_{.k}(\tau)$ in Wahrscheinlichkeit beschränkt und D_1 konvergiert in Wahrscheinlichkeit gleichmäßig gegen 0. Für $\beta_* \xrightarrow{P} 0$ gilt das auch für D_2 .

Zur Abschätzung von D_3 erhält man mittels der Lenglar't'schen Ungleichung (1.7.3)

$$\begin{aligned}
 P \left\{ \left| \int_0^t v(\beta_0, u) \frac{1}{n} dM_{.k} \right|_{\mathcal{T}} > \eta \right\} &\leq \frac{\delta}{\eta^2} \\
 &+ P \left\{ \frac{1}{n^2} \int_0^\tau v^2(\beta_0, t) S_k^{(0)}(\beta_0, t) \alpha_0(t) dt > \delta \right\}
 \end{aligned}$$

Wegen L 1 bis L 3 konvergiert auch D_3 in Wahrscheinlichkeit gegen 0 für $\beta_* \xrightarrow{P} 0$. Analog kann man für D_4 vorgehen.

Zum Beweis von (4.4.37) wird $U(\beta, \tau)$ komponentenweise nach dem Mittelwertsatz um β_0 entwickelt:

$$U_j(\beta, \tau) - U_j(\beta_0, \tau) = -\mathcal{I}_\tau^j(\beta_j^*)(\beta - \beta_0) \quad j = 1, \dots, p \tag{4.4.40}$$

Dabei ist $\beta_j^* = \lambda_j \beta + (1 - \lambda_j) \beta_0$, $\lambda_j \in (0, 1)$, und $-\mathcal{I}_\tau^j$ ein Zeilenvektor von $-\mathcal{I}_\tau$ mit Komponenten

$$\frac{\partial}{\partial \beta_l} U_j(\beta, \tau), \quad l = 1, \dots, p.$$

An der Stelle $\beta = \hat{\beta}$ erhält man wegen $U_j(\hat{\beta}, \tau) = 0$

$$n^{-1/2} U_j(\beta_0, \tau) = \{n^{-1} \mathcal{I}_\tau^j(\beta_j^*)\} n^{1/2} (\hat{\beta} - \beta_0) \quad (4.4.41)$$

Aufgrund von L 4 ist Σ_τ positiv definit. $\mathcal{I}_\tau(\hat{\beta})$ konvergiert wegen (4.4.38) in Wahrscheinlichkeit gegen Σ_τ . Da $\mathcal{I}_\tau^j(\beta)$ wegen L 1 und L 3 stetig ist in β und somit auch für β aus einer (konvexen) Umgebung $B(\beta_0, \delta)$, $\delta > 0$, positiv definit ist, ist für $\hat{\beta} \in B(\beta_0, \delta)$ die Matrix I , die aus den Zeilen $\{n^{-1} \mathcal{I}_\tau^j(\beta_j^*)\}$ besteht, positiv definit und invertierbar und

$$n^{-1/2} I^{-1} U(\beta_0, \tau) = n^{1/2} (\hat{\beta} - \beta_0) \quad (4.4.42)$$

Wegen der Konsistenz von $\hat{\beta}$ gilt $\hat{\beta} \in B(\beta_0, \delta)$ und $\beta^* \in B(\beta_0, \delta)$ mit beliebig großer Wahrscheinlichkeit. Aufgrund der Stetigkeit von $\mathcal{I}_t(\beta)$ in $B(\beta_0, \delta)$ konvergiert $I \xrightarrow{P} \Sigma_\tau^{-1}$. Wegen Satz 4.4.3 gilt

$$n^{-1/2} U(\beta_0, \tau) \xrightarrow{\mathcal{D}} \mathcal{N}(0, B).$$

Aus Folgerung 1.6.2 und Satz 1.6.3 folgt nun bezüglich (4.4.42)

$$I^{-1} n^{-1/2} U(\beta_0, \tau) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_\tau^{-1} B \Sigma_\tau^{-1})$$

□

Kapitel 5

Schätzung von marginalen Modellen durch U -Statistiken

5.1 Der Schätzer

Sei ein marginales Modell für einen Zählprozeß gegeben wie in 4.4.

In Liang et al. (1993) wird ein Schätzer für das marginale Cox-Modell vorgeschlagen, der die paarweise Unabhängigkeit der Beobachtungen aus verschiedenen Haufen verwertet. In diesem Ansatz wird die Likelihood (4.1.3) jeweils paarweise für ein ij und lk aus dem Haufen i beziehungsweise l gebildet, wobei $i \neq l$. Unter Beibehaltung der paarweisen Betrachtung schreibt man in Analogie zu (4.2.7)

$$L_{ik;l_j}(\beta) = \prod_{t \in \mathcal{T}} \left\{ \frac{\exp \{ \beta^\top Z_{ik}(t) \}}{S_{ik;l_j}^{(0)}(\beta, t)} \right\}^{\Delta N_{ik}(t)}. \quad (5.1.1)$$

Dabei ist für $d = 0, 1, 2$ wiederum definiert durch (vgl. (4.2.2))

$$S_{ik;l_j}^{(d)}(\beta, t) = Y_{ik}(t) Z_{ik}(t)^{\otimes d} \exp \{ \beta^\top Z_{ik}(t) \} + Y_{l_j}(t) Z_{l_j}(t)^{\otimes d} \exp \{ \beta^\top Z_{l_j}(t) \}.$$

Weiter unten wird auch $E_{ik;l_j}(\beta, t) = S_{ik;l_j}^{(1)}(\beta, t) / S_{ik;l_j}^{(0)}(\beta, t)$ benötigt. ¹

Man kann zeigen, daß das Produkt dieser Likelihood-Elemente eine konsistente Schätzung von β ermöglichen würde. Allerdings ist die Gewichtung der Zeitpunkte durch die Paarbildung ungünstig, denn frühe Ereignisse mit

¹Ab nun wird $0/0 = 0$ vereinbart

vielen Individuen in der Risikomenge lieferten einen zu großen Beitrag. Liang et al. (1993) schlugen ein geometrisches Mittel

$$\prod_{t \in \mathcal{T}} \left\{ \frac{\exp \{ \beta^\top Z_{ik}(t) \}}{S_{ik;l_j}^{(0)}(\beta, t)} \right\}^{\Delta N_{ik}(t) Y_{lj}(t) / n_i(t)}.$$

vor (Auch das arithmetische Mittel wurde dort erwogen). Die Gewichtung mit $n_i(t)^{-1}$ ist umgekehrt proportional zur Anzahl der Individuen in den Risikomenge, die nicht im Haufen i sind. Nach Vereinbarung der Schreibweise $Y_{\cdot}(t) = \sum_{i=1}^n Y_i(t)$ und $Y_i(t) = \sum_{k=1}^K Y_{ik}(t)$ kann das als $n_i(t) = Y_{\cdot}(t) - Y_i(t)$ angeschrieben werden. Der logarithmierte Pseudo-Likelihoodprozeß ist daher (vgl. (4.2.8))

$$C_{PL}(\beta) = \sum_{\substack{i,l=1 \\ l \neq i}}^n \sum_{k,j=1}^K \int_0^\tau I[n_i(t) > 0] n_i(t)^{-1} Y_{lj}(t) \cdot \left\{ Y_{ik}(t) \beta^\top Z_{ik}(t) - \log \left(S_{ik;l_j}^{(0)}(\beta, t) \right) \right\} dN_{ik}(t) \quad (5.1.2)$$

Die Ableitung nach β ergibt die folgende Schätzfunktion

$$\begin{aligned} S(\beta) &= \sum_{\substack{i,l=1 \\ l \neq i}}^n \sum_{k,j=1}^K \int_0^\tau I[n_i(t) > 0] \frac{Y_{lj}(t)}{n_i(t)} \{ Z_{ik}(t) - E_{ik;l_j}(\beta, t) \} dN_{ik}(t) \quad (5.1.3) \\ &= \sum_{i=1}^n \sum_{k=1}^K \int_0^t I[n_i(t) > 0] \left\{ Z_{ik}(t) - n_i(t)^{-1} \sum_{\substack{l=1 \\ l \neq i}}^n \sum_{j=1}^K Y_{lj}(t) E_{ik;l_j}(\beta, t) \right\} dN_{ik}(t). \quad (5.1.4) \end{aligned}$$

Die letzte Umformung gilt deshalb, weil zur Zeit T eines Ereignisses mit $\Delta N_{ik}(T) \neq 0$ ein Individuum lj mit $l \neq i$ genau dann einen Summanden ungleich 0 beiträgt, wenn $Y_{lj}(t) = 1$. Deren Anzahl ist $Y_{\cdot}(t) - Y_i(t) = n_i(t)$. Der Ausdruck (5.1.4) erlaubt die Interpretation der Nullstelle $\hat{\beta}$ der Schätzgleichung (5.1.3) als denjenigen Wert, der die mittlere Abweichung des Regressors $Z_{ij}(t)$ vom gewichteten „Mittel“ der Regressoren (mit Ausnahme der Regressoren des i -ten Haufens) in der Risikomenge zur Zeit des Ereignisses minimiert (vgl. dazu den Aufbau von (4.2.9)). Dies motiviert die Wahl der Gewichtung $n_i(t)$. Der Ausdruck (5.1.3) ermöglicht die Interpretation

der Schätzfunktion als U -Statistik vom Grad 2 durch

$$U_n(\beta) = \binom{n}{2}^{-1} \sum_{\substack{i,l=1 \\ l>i}}^n \phi_n(\beta; X_i, X_l) = \binom{n}{2}^{-1} n S(\beta) \quad (5.1.5)$$

Dabei ist X_i (beziehungsweise X_l) das K -Tupel $((N_{ik}, Z_{ik}, Y_{ik}))_{k=1}^K$ und

$$\begin{aligned} \phi_n^*(\beta; X_i, X_l) = \sum_{k,j=1}^K \int_0^\tau \frac{I[n_i(t) > 0]}{n_i(t)/n} \\ \cdot Y_{lj}(t) \{Z_{ik}(t) - E_{ik;l j}(\beta, t)\} dN_{ik}(t). \end{aligned} \quad (5.1.6)$$

Die symmetrisierte Form ist gegeben durch

$$\phi_n(\beta; X_i, X_l) = \phi_n^*(\beta; X_i, X_l) + \phi_n^*(\beta; X_l, X_i). \quad (5.1.7)$$

Allerdings hängt ϕ_n^* über die Gewichtsfunktion $n_i(t)/n$ von der gesamten Stichprobe ab. Deshalb ist U_n keine U -Statistik im Sinn von (3.1.1) in Kapitel 3.

5.2 Asymptotische Verteilung

Der Kern ϕ_n hängt mittels Gewichtsfunktion noch immer von der gesamten Stichprobe ab. Um den zentralen Grenzwertsatz für U -Statistiken direkt anwenden zu können und in Hinblick auf Verallgemeinerbarkeit, wird darin $n_i(t)/n$ durch eine Gewichtsfunktion W ersetzt, $W(t) \geq c > 0$ für alle $t \in \mathcal{T}$. Man definiert (in Verallgemeinerung von (5.1.5))

$$U(\beta) = \binom{n}{2}^{-1} \sum_{i=1}^n \sum_{\substack{i,l=1 \\ l>i}}^n \phi(\beta; X_i, X_l) \quad (5.2.1)$$

mit

$$\phi(\beta; X_i, X_l) = \phi^*(\beta; X_i, X_l) + \phi^*(\beta; X_l, X_i) \quad (5.2.2)$$

und

$$\phi^*(\beta; X_i, X_l) = \sum_{k,j=1}^K \int_0^\tau W(t)^{-1} \{Z_{ik}(t) - E_{ik;l j}(\beta, t)\} dN_{ik}(t) \quad (5.2.3)$$

Die Stammfunktion von $U(\beta)$ ist

$$C(\beta) = \binom{n}{2}^{-1} \sum_{\substack{l=1 \\ l \neq i}}^n \sum_{k,j=1}^K \int_0^\tau W(t)^{-1} \left\{ Y_{ik}(t) \beta^\top Z_{ik}(t) \right. \quad (5.2.4)$$

$$\left. - \log S_{ik;l_j}^{(0)}(\beta, t) \right\} dN_{ik}(t) \quad (5.2.5)$$

$$= \binom{n}{2}^{-1} \sum_{i=1}^n \sum_{\substack{l=1 \\ l \neq i}}^n \Phi^*(\beta; X_i, X_l) \quad (5.2.6)$$

Dadurch ist wiederum eine U -Statistik gegeben. Genauso ist $H(\beta) = -\partial U(\beta)/\partial \beta^{\otimes 2}$ als U -Statistik darstellbar. Die nachstehenden Bedingungen werden zur Herleitung der Eigenschaften der Schätzer benötigt:

PL 1. *Der Kovariablenvektor ist gleichmäßig beschränkt, d.h. $|Z_{ik}(t)|_{\mathcal{T}} \leq Z_0$ ist für alle $t \in \mathcal{T}$ und $ik, i = 1, \dots, n, k = 1, \dots, K$. Darüberhinaus gilt $A_0(\tau) = \int_0^\tau \alpha_o(t) dt < \infty$.*

PL 2. *Es existiert eine Funktion \bar{H} , die in einer (offenen) Umgebung \mathcal{B} des wahren Parameters β_0 stetig und positiv definit ist, und für jedes β in \mathcal{B} gilt:*

$$-\frac{\partial}{\partial \beta^\top} U(\beta) = H(\beta) \xrightarrow{P} \bar{H}(\beta)$$

PL 3. *Sei $\phi_{1,i;j}(\beta; x) = \mathbb{E} \phi(\beta; x, X_j)$. Sei $\phi_{1,i}(\beta; x) = \frac{1}{n-1} \sum_{j=1, j \neq i}^n \phi_{1,i;j}(\beta; x)$. Sei $S_n = \frac{K}{n} \sum_{i=1}^n \phi_{1,i}(\beta_0; X_i)$. Dann soll S_n einen zentralen Grenzwertsatz mit Erwartungswert 0 erfüllen, d.h. es gibt eine invertierbare Kovarianzmatrix Σ mit $n \text{Cov } S_n \rightarrow \Sigma$ und*

$$n^{1/2} S_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma) \quad (5.2.7)$$

Zunächst stellt man die Erwartungstreue der Schätzgleichung fest:

Proposition 5.2.1. *Sei PL 1 erfüllt. Dann konvergiert die Varianz von $C(\beta)$, $U(\beta)$ und $H(\beta)$ jeweils gegen 0 und*

$$\mathbb{E} U(\beta_0) = 0. \quad (5.2.8)$$

Beweis. Aufgrund der Überlegungen, die zu (3.1.5) führten, konvergieren die Varianzen von $C(\beta)$, $U(\beta)$ und $H(\beta)$ jeweils gegen 0, wenn Erwartungswert und Varianz der Summanden der U -Statistiken gleichmäßig beschränkt

sind. Dazu wird exemplarisch die Existenz der entsprechenden Konstanten für $H(\beta)$ gezeigt. Die Varianz der Matrix $H(\beta)$ kann über ein inneres Produkt definiert werden. Dann ist hinreichend für die Beschränktheit der Varianz, daß das Kroneckerprodukt $H(\beta) \otimes H(\beta)$ endlichen Erwartungswert hat. Die Elemente dieser Matrix sind die Erwartungswerte der Produkte der Elemente von $H(\beta)$, die komponentenweise durch die Erwartungswerte der Quadrate der Faktoren abgeschätzt werden können (Cauchy-Schwarz-Bunjakowsky-Ungleichung). Diese Ausdrücke werden weiter komponentenweise abgeschätzt durch

$$\begin{aligned}
& \left| \mathbb{E} \frac{\partial}{\partial \beta} \phi^*(\beta; X_i, X_l)^2 \right| \\
&= \left| \mathbb{E} \left(\sum_{k,j=1}^K \int_0^\tau \frac{Y_{lj}(t)}{W(t)} \left(\frac{S_{ik;l_j}^{(2)}(\beta, t)}{S_{ik;l_j}^{(0)}(\beta, t)} - E_{ik;l_j}(\beta, t)^{\otimes 2} \right) dN_{ik}(t) \right)^2 \right| \\
&\leq 2^{K^2} \sum_{k,j=1}^K \left| \mathbb{E} \left[\int_0^\tau \frac{Y_{lj}(t)}{W(t)} \left(\frac{S_{ik;l_j}^{(2)}(\beta, t)}{S_{ik;l_j}^{(0)}(\beta, t)} - E_{ik;l_j}(\beta, t)^{\otimes 2} \right) dN_{ik}(t) \right]^2 \right| \\
&\leq 2^{K^2} \sum_{k,j=1}^K \left| \frac{1}{W(t)} \left(\frac{S_{ik;l_j}^{(2)}(\beta, t)}{S_{ik;l_j}^{(0)}(\beta, t)} - E_{ik;l_j}(\beta, t)^{\otimes 2} \right) \right|_{\mathcal{T}}^2 N_{ik}(t)^2 \\
&\leq K^2 2^{K^2} \frac{(Z_0^2 + Z_0^2)^2}{c^2} \exp \{ |\beta| Z_0 \} A_0(\tau).
\end{aligned}$$

Somit ist die Varianz von $\frac{\partial}{\partial \beta} \phi^*(\beta; X_i, X_l)^\top$ beschränkt durch einen Ausdruck, der nicht von den Indizes abhängt. Wegen (3.1.5) konvergiert die Varianz dieser U -Statistik in Wahrscheinlichkeit gegen 0.

Nun ist noch (5.2.8) zu beweisen. Man stellt fest, daß es genügt, wenn für jeden Summanden $\mathbb{E} \phi(\beta_0; X_i, X_l) = 0$ mit $l > i$ gilt. Für diese Aussage genügt es wiederum zu zeigen, daß sie für jeden Summanden aus $\phi(\beta_0; X_i, X_l)$, der einem Paar (ik, lj) , $k, j = 1, \dots, K$ entspricht, gilt. Dafür genügt es wiederum zu zeigen, daß dieser Ausdruck (gemeint ist (5.2.9)) die Summe zweier Martingale ist.

Bezüglich der gemeinsamen Filtrierung $\mathcal{F}_{ik} \vee \mathcal{F}_{lj}$ der unabhängigen Zählprozesse N_{ik} und N_{lj} ist die eindeutige Doob-Meyer-Zerlegung in ein Martingal und einen vorhersagbaren Prozeß $N_{ik}(t) = M_{ik}(t) + \int_0^t Y_{ik}(u) \exp \{ \beta_0^\top Z_{ik}(u) \} \alpha_0(u) du$ und $N_{lj}(t) = M_{lj}(t) + \int_0^t Y_{lj}(u) \exp \{ \beta_0^\top Z_{lj}(u) \} \alpha_0(u) du$. Man stellt wiederum fest, daß der vorhersagbare Anteil aufgrund der Bauart der Schätzfunktion verschwindet

(vergleiche (4.3.3)):

$$\begin{aligned}
& \int_0^t W(u)^{-1} \{Z_{ik}(u) - E_{ik;l_j}(\beta_0, u)\} dN_{ik}(u) \\
& \quad + \int_0^t W(u)^{-1} \{Z_{l_j}(u) - E_{ik;l_j}(\beta_0, u)\} dN_{l_j}(u) \\
& = \int_0^t W(u)^{-1} \{Z_{ik}(u) - E_{ik;l_j}(\beta_0, u)\} dM_{ik}(u) \\
& \quad + \int_0^t W(u)^{-1} \{Z_{l_j}(u) - E_{ik;l_j}(\beta_0, u)\} dM_{l_j}(u)
\end{aligned} \tag{5.2.9}$$

Die Integranden sind jeweils vorhersagbar bezüglich der zugrundeliegenden Filtrierung und lokal beschränkt. Die Martingal-Integratoren sind lokal quadratisch integrierbar und von endlicher Variation (Der Kompensator ist absolut stetig und es gibt fast sicher endlich viele Sprünge). Somit sind die Voraussetzungen von Satz 1.4.1 erfüllt und der Ausdruck (5.2.9) ist daher die Summe zweier Martingale. \square

Die Funktionen $H(\beta, t)$ sind *positiv semidefinit* in β : Da in $U(\beta, t)$ die $E_{ik;l_j}(\beta, t)$ die einzigen Ausdrücke sind, die von β abhängen, genügt es, deren Ableitung zur Zeit $t \in \mathcal{T}$ zu betrachten.

Zunächst definiert man $I = \{ij, lk\}$, $S_I^{(d)}(\beta, t) = S_{ik;l_j}^{(d)}(\beta, t)$, $d = 0, 1, 2$ und $E_I(\beta, t) = E_{ik;l_j}(\beta, t)$. Dann stellt man fest, daß die folgenden Ausdrücke positiv semidefinit sind:

$$\begin{aligned}
& \sum_{\iota \in I} [Z_\iota(t) - E_I(\beta, t)]^{\otimes 2} Y_\iota \exp \{ \beta^\top Z_\iota(t) \} \\
& = \sum_{\iota \in I} \left[Z_\iota(t)^{\otimes 2} - Z_\iota(t) E_I(\beta, t)^\top \right. \\
& \quad \left. - E_I(\beta, t) Z_\iota(t)^\top + E_I(\beta, t)^{\otimes 2} \right] Y_\iota \exp \{ \beta^\top Z_\iota(t) \} \\
& = S_I^{(2)}(\beta, t) - S_I^{(1)}(\beta, t) E_I(\beta, t)^\top - E_I(\beta, t) S_I^{(1)}(\beta, t)^\top + E_I(\beta, t)^{\otimes 2} S_I^{(0)}(\beta, t) \\
& = S_I^{(2)}(\beta, t) - E_I(\beta, t)^{\otimes 2} S_I^{(0)}(\beta, t)
\end{aligned}$$

Das ist aber bis auf einen negativen Proportionalitätsfaktor die Matrix der

Ableitungen:

$$\begin{aligned} \frac{\partial}{\partial \beta^\top} (Z_{ik}(t) - E_I(\beta, t)) &= -\frac{\partial}{\partial \beta^\top} \frac{S_I^{(1)}(\beta, t)}{S_I^{(0)}(\beta, t)} \\ &= -\frac{S_I^{(2)}(\beta, t)}{S_I^{(0)}(\beta, t)} + \left[\frac{S_I^{(1)}(\beta, t)}{S_I^{(0)}(\beta, t)} \right]^{\otimes 2} \\ &= -S_I^{(0)}(\beta, t)^{-1} \left(S_I^{(2)}(\beta, t) - E_I(\beta, t) S_I^{(0)}(\beta, t) \right) \end{aligned}$$

Nach Integration über \mathcal{T} bezüglich eines positiven Maßes

$$\int_0^\tau W(t)^{-1} \left[\frac{S_I^{(2)}(\beta, t)}{S_I^{(0)}(\beta, t)} - E_I(\beta, t) \right] dN_{ik}(t)$$

bleibt das Ergebnis positiv semidefinit.

$H(\beta)$ ist genau dann positiv definit, wenn es zumindest in einem Haufen i zu einer Zeit t ein Ereignis bei Individuum ik und ein Individuum $i'k'$ in einem anderen Haufen, mit $Y_{i'k'}(t) = 1$ gibt und dessen Kovariablenvektor $Z_{ik}(t)$ ungleich $Z_{i'k'}(t)$ ist. Es ist leicht, Bedingungen anzugeben, so daß $H(\beta)$ fast sicher positiv definit ist. Insbesondere genügt $H(\beta_0) \xrightarrow{P} \bar{H}(\beta_0)$ und $\bar{H}(\beta_0)$ positiv definit. Denn dann ist $H(\beta_0)$ fast sicher positiv definit. Aber dann ist $H(\beta)$ auch für jedes andere β positiv definit.

Satz 5.2.2. *Sei PL 1 und PL 2 erfüllt. Dann konvergiert die Wahrscheinlichkeit, daß die Funktion $U(\beta, \tau)$ eine eindeutige Lösung $\hat{\beta}$ hat, gegen 1 für $n \rightarrow \infty$ und es gilt $\hat{\beta} \xrightarrow{P} \beta_0$*

Beweis. Wegen PL 2 und aufgrund der dem Satz vorangestellten Bemerkungen ist $f_n(\beta) = C(\beta) - C(\beta_0)$ fast sicher konkav. Die Folge von zufälligen Funktionen konvergiert aufgrund von Satz 5.2.1 punktweise gegen eine Funktion f . Daher kann Satz 1.7.2 angewendet werden und die Folge konvergiert daher in Wahrscheinlichkeit gleichmäßig auf kompakten Mengen gegen die konkave Funktion f .

Nun ist zu zeigen, daß der f ein lokales Maximum bei β_0 hat; das impliziert ein globales Maximum. Zum indirekten Beweis nimmt man an, daß es in \mathcal{B} ein x gibt, mit $f(x) - f(\beta_0) = c > 0$. Ohne Einschränkung der Allgemeinheit des Beweises kann man annehmen, daß die Verbindungslinie von x und β_0 auf der ersten Koordinate liegt. Ab nun werden in diesem und dem darauf folgenden Absatz alle Punkte und Funktionen eingeschränkt auf diese Verbindungslinie

betrachtet. Angenommen $x > \beta_0$. Die Ableitung f'_n ist aufgrund von PL 2 fast sicher nicht fallend. Es gilt daher:

$$\begin{aligned} 0 < c &= f(x) - f(\beta_0) \\ &< (f(x) - f_n(x)) + (f_n(\beta_0) - f(\beta_0)) + f_n(\beta) - f_n(\beta_0) \\ &= (f(x) - f_n(x)) + (f_n(\beta_0) - f(\beta_0)) + f'_n(\xi_n)(x - \beta_0) \\ &< (f(x) - f_n(x)) + (f_n(\beta_0) - f(\beta_0)) + f'_n(\beta_0)(x - \beta_0) \end{aligned}$$

Die Abschätzung gilt fast sicher. Die ersten beiden Summanden konvergieren jedoch in Wahrscheinlichkeit gegen 0. Auch $f'_n(\beta_0) = U(\beta_0)$ konvergiert wegen Satz 5.2.1 und Satz 3.1.2 in Wahrscheinlichkeit gegen 0. Dies ist ein Widerspruch zur Voraussetzung $c > 0$.

Eine Nullstelle $\hat{\beta}$ der Ableitung von f_n ist fast sicher ein Maximum von f_n , sodaß $\hat{\beta}$ aufgrund von Satz 1.7.3 in Wahrscheinlichkeit gegen β_0 konvergiert. \square

Satz 5.2.3. *Sei PL 1 bis PL 3 erfüllt. Sei $\hat{\beta}$ die nach Satz 5.2.2 fast sicher eindeutig bestimmte Lösung von $U(\beta) = 0$. Dann gilt*

$$n^{1/2}\hat{\beta} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \bar{H}^{-1}\Sigma\bar{H}^{-1}) \quad (5.2.10)$$

$\bar{H} = \bar{H}(\beta_0)$ kann durch $H(\hat{\beta})$ konsistent geschätzt werden. Σ ist die asymptotische Kovarianzmatrix von $U(\beta_0)$. Diese kann konsistent geschätzt werden durch

$$\hat{\Sigma} = \frac{4(n-1)}{(n-2)^2} \sum_{i=1}^n V_n^{(1)}(X_i)^{\otimes 2} \quad (5.2.11)$$

Dabei ist definiert

$$V_n^{(1)}(X_i) = \frac{1}{n-1} \sum_{i \neq l}^n \left(\phi^*(\hat{\beta}; X_i, X_l) + \phi^*(\hat{\beta}; X_l, X_i) \right) \quad (5.2.12)$$

Beweis. Zunächst stellt man fest, daß $|\phi(\beta_0; X_i, X_l)| < A$ für alle $1 \leq i < l \leq n$ (siehe Proposition 5.2.1). Daher kann Satz 3.1.2 angewendet werden. Aufgrund von PL 3 ist $U_n(\beta_0)$ asymptotisch so verteilt wie eine Summe von unabhängigen Zufallsvariablen, die den zentralen Grenzwertsatz mit Erwartungswert 0 (siehe Proposition 5.2.1) und invertierbarer Kovarianzmatrix Σ erfüllt:

$$n^{1/2}U(\beta_0) \xrightarrow{\mathcal{D}} N(0, \Sigma). \quad (5.2.13)$$

(5.2.10) wird bewiesen, indem die Argumente im Beweis von Satz 4.3.5 auf $U(\hat{\beta})$ und $H(\hat{\beta})$ angewendet werden. Eine notwendige Bedingung für dieses Argument ist, daß

$$H(\hat{\beta}) \xrightarrow{P} \bar{H}(\beta_0) \quad (5.2.14)$$

erfüllt ist.

Zum Beweis von (5.2.14) wird das Teilfolgenprinzip für Konvergenz in Wahrscheinlichkeit benützt, das besagt, daß eine hinreichende Bedingung für Konvergenz in Wahrscheinlichkeit gegeben ist, wenn jede Teilfolge eine Teilfolge besitzt, die fast sicher konvergiert.

Zur Vereinfachung der Notation werden Teilfolgen gleich bezeichnet wie die ursprüngliche Folge.

Sei eine beliebige Teilfolge von H und $\hat{\beta}$ gegeben. Aufgrund von Satz 5.2.2 und PL 2 gibt es eine kompakte Umgebung \mathcal{B} von β_0 und eine Teilfolge, so daß $\hat{\beta}$ fast sicher in \mathcal{B} liegt und $H(\beta)$ für $\beta \in \mathcal{B}$ in Wahrscheinlichkeit gegen $\bar{H}(\beta)$ konvergiert. Sei D eine abzählbare Menge, die in \mathcal{B} dicht liegt. Sei mit dem Verfahren aus dem Beweis von Satz II.1 aus (Andersen und Gill 1982, p1116) eine fast sicher konvergente Teilfolge von $\hat{\beta}$ und $H(\beta)$ gewählt, sodaß $H(\beta)$ für β aus D und $\hat{\beta}$ punktweise fast sicher gegen $\bar{H}(\beta)$ beziehungsweise fast sicher gegen β_0 konvergiert.

Um zu zeigen, daß auf dieser Teilfolge $H(\hat{\beta})$ fast sicher gegen $\bar{H}(\beta_0)$ konvergiert, wird auf das Teilfolgenprinzip auf \mathbb{R}^p zurückgegriffen. Sei nun $\omega \in \Omega$ und eine Teilfolge von $\hat{\beta}(\omega)$ beliebig gewählt. Aufgrund der gleichmäßigen Stetigkeit von \bar{H} in \mathcal{B} kann eine Teilfolge $\beta_*(\omega)$ in D derart gewählt werden, daß $\beta_*^{(n)} \rightarrow \beta_0$,

$$\left| H(\hat{\beta})(\omega) - H(\beta_*)(\omega) \right| < \frac{1}{n}$$

und

$$\left| H(\beta_*)(\omega) - \bar{H}(\beta_*)(\omega) \right| < \frac{1}{n}$$

Auf dieser Teilfolge gilt daher:

$$\begin{aligned} & \left| H(\hat{\beta})(\omega) - \bar{H}(\beta_0) \right| \\ & \leq \left| H(\hat{\beta})(\omega) - H(\beta_*)(\omega) \right| \\ & \quad + \left| H(\beta_*)(\omega) - \bar{H}(\beta_*)(\omega) \right| + \left| \bar{H}(\beta_*)(\omega) - \bar{H}(\beta_0) \right| \\ & \leq \frac{1}{n} + \frac{1}{n} + \left| \bar{H}(\beta_*)(\omega) - \bar{H}(\beta_0) \right| \end{aligned}$$

Wegen der Stetigkeit von \bar{H} in β_0 konvergiert auf dieser Teilfolge $H(\hat{\beta})(\omega)$ gegen $\bar{H}(\beta_0)$. Da mit einer beliebigen Teilfolge $\hat{\beta}(\omega)$ begonnen wurde, konvergiert aufgrund des Teilfolgenprinzips auf \mathbb{R}^K auch die ursprüngliche Folge. Diese Konstruktion ist auf ganz Ω bis auf eine Nullmenge möglich. Folglich konvergiert eine Teilfolge von $H(\hat{\beta})$ fast sicher gegen $\bar{H}(\beta_0)$. Aufgrund des Teilfolgenprinzips für Konvergenz in Wahrscheinlichkeit ist (5.2.14) gezeigt. Nun kann der Beweis von (5.2.10) fortgesetzt werden, indem $U(\beta)$ komponentenweise nach dem Mittelwertsatz um β_0 entwickelt wird:

$$U_j(\beta) - U_j(\beta_0) = -H^j(\beta_j^*)(\beta - \beta_0) \quad j = 1, \dots, p \quad (5.2.15)$$

Dabei ist $\beta_j^* = \lambda_j\beta + (1 - \lambda_j)\beta_0$, $\lambda_j \in (0, 1)$, und $-H^j$ ein Zeilenvektor von $-H$ mit Komponenten

$$\frac{\partial}{\partial \beta_l} U_j(\beta), \quad l = 1, \dots, p.$$

An der Stelle $\beta = \hat{\beta}$ erhält man wegen $U_j(\hat{\beta}) = 0$

$$U_j(\beta_0) = H^j(\hat{\beta}_j^*) (\hat{\beta} - \beta_0) \quad (5.2.16)$$

Aufgrund von PL 2 ist \bar{H} positiv definit und wegen (5.2.14) konvergiert $H(\hat{\beta})$ in Wahrscheinlichkeit gegen \bar{H} . Da aus $H(\beta_0)$ positiv definit folgt, daß $H(\beta)$ positiv definit ist für $\beta \in \mathbb{R}^p$, ist die Matrix I , die aus den Zeilen $H(\hat{\beta}_j^*)$ besteht, fast sicher positiv definit und invertierbar, und

$$I^{-1}U(\beta_0) = (\hat{\beta} - \beta_0). \quad (5.2.17)$$

Aufgrund der Stetigkeit von $H(\beta)$ konvergiert $I \xrightarrow{P} \bar{H}^{-1}$. Wegen (5.2.13), Folgerung 1.6.2 und Satz 1.6.3 folgt nun bezüglich (5.2.17) und uns somit bezüglich (5.2.10):

$$n^{1/2}I^{-1}U(\beta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \bar{H}^{-1}\Sigma\bar{H}^{-1}).$$

Zuletzt ist (5.2.11) zu zeigen. $n^{-1/2}\hat{\Sigma}$ stimmt bis auf $\hat{\beta}$ statt β_0 mit dem Jackknife-Varianzschätzer von $U_n(\beta_0)$ überein. Auf Seite 74 wird gezeigt, daß der Jackknife-Varianzschätzer von U -Statistiken selbst eine Summe von U -Statistiken ist. Durch Anwendung von Satz 3.3.3 folgt nun (5.2.11). Wegen der asymptotischen Normalverteilung von $n^{1/2}(\hat{\beta} - \beta_0)$ ist die Bedingung der Straffheit erfüllt. Die Stetigkeit des Erwartungswerts der Kernfunktion beziehungsweise von deren Quadrat in β sind die Bedingungen von Satz 3.3.3 erfüllt. \square

5.3 Gewichtungsfunktionen

In den Anwendungen muß die Gewichtungsfunktion W durch eine geschätzte Gewichtungsfunktion \hat{W} ersetzt werden. Unter der folgenden Bedingung ändert sich die asymptotische Verteilung der U -Statistik nicht:

PL 4.

$$n^{1/2} (\hat{W} - W) \quad (5.3.1)$$

ist gleichmäßig straff bezüglich der Supremumsnorm.

Es gilt dann der folgende Satz:

Satz 5.3.1. *Seien PL 1 bis PL 4 erfüllt. Dann ist Satz 5.2.3 auch dann gültig, wenn W durch eine empirische Gewichtungsfunktion \hat{W} ersetzt wird.*

Beweis. Die Aussage folgt aus Satz 3.3.3, dessen Bedingungen somit zu erfüllen sind.

PL 4 fordert genau Bedingung 3.3.1.

(3.3.7) und (3.3.2) aus Bedingung 3.3.2 ist erfüllt, denn für beliebige Gewichtungsfunktionen $|W|_{\mathcal{T}} < c$ und beschränktes β , Z_{ik} und A_0 gilt für $r \leq c/2$, sowie für $s = 1, 2$, und es existieren Konstanten K_s :

$$\begin{aligned} & \mathbb{E} \sup_{W' \in B(W, r)} |\phi^*(\beta, W'; X_i, X_j) - \phi^*(\beta, W; X_i, X_j)|^s \\ & \leq \sum_{k=1}^K \sum_{j=1}^K \mathbb{E} \sup_{W' \in B(W, r)} \left| \int_0^\tau (W'(t)^{-1} - W(t)^{-1}) (Z_{ik}(t) - E_{ik;l_j}(\beta, t)) dN_{ik}(t) \right|^s \\ & \leq \left(\frac{2r}{c^2} \right)^s \sum_{k=1}^K \sum_{j=1}^K \mathbb{E} |Z_{ik}(t) - E_{ik;l_j}(\beta, t)|_{\mathcal{T}}^s N_{ik}(\tau)^s \\ & \leq K_s r \end{aligned}$$

□

Die Gewichtungsfunktion $Y_{..}(t)/n$ erfüllt PL 4 für $W_{PL} = \lim_{n \rightarrow \infty} \mathbb{E} Y_{..}(t)/n$ im Allgemeinen nicht. Diese Bedingung muß durch weitere Regularitätsbedingungen für den Zensierungs- und Kovariablenprozeß erzwungen werden. In Abschnitt 1.6.2 werden Sätze angegeben, die zur Konstruktion von nichttrivialen Beispielen verwendet werden können.

Aufgrund von Folgerung 3.1.1 kann der Kern einer U -Statistik U_n um Ausdrücke, die in Wahrscheinlichkeit gegen 0 konvergieren, geändert werden, ohne daß damit die asymptotische Verteilung von $\sqrt{n}U_n$ verändert wird. Insbesondere gilt das für die Ersetzung von W_{PL} durch $W_{PL} - Y_i/n$ im entsprechenden Kern.

In der Anwendung muß W_{PL} durch $Y_{..}(t)/n$ ersetzt werden. Trotzdem gilt Satz 3.3.3 für Kerne mit der empirischen Gewichtsfunktion $n_i(t) = (Y_{..}(t) - Y_i(t))/n$ (vergleiche mit (5.1.3)) mit folgender Begründung: Für $0 < w < c$, $|w - w'| < r < c/4$, $k \leq K$ und $n_0 > 4K/c$ gilt

$$\sup_{n \geq n_0} \left| \frac{1}{w + k/n} - \frac{1}{w' + k/n} \right| = \sup_{n \geq n_0} \left| \frac{w' - w}{(w + k/n)(w' + k/n)} \right| \leq \frac{4}{c^2} r.$$

Daher kann die Bedingung 3.3.2 für genügend großes $n \geq n_0$ unabhängig von n erfüllt werden (mit der gleichen Begründung wie in Satz 5.3.1). Im Beweis von Satz 3.3.3 muß die Kernfunktion h durch eine Funktion h_n , die von n abhängt, ersetzt werden. Diese Ersetzung ändert die Gültigkeit der getroffenen Aussagen an keiner Stelle, sodaß der Satz auch mit dieser Modifikation gilt.

Die ungewichtete Scorefunktion ($W = 1$) bewertet das früheste Ereignis etwa n mal so stark wie das späteste Ereignis, denn es kommt in etwa ebenso vielen Paaren von Wahrscheinlichkeitselementen mit Beiträgen vor, die von 0 verschieden sind. Im Gegensatz dazu werden im Modell der proportionalen Hasardraten in der Score-Gleichung, die auf der Cox'schen partiellen Likelihood beruht, alle Ereignisse gleich gewichtet. Aus der Theorie der Schätzgleichungen (Godambe 1985, Chang und Hsiung 1990) weiß man, daß diese Gewichtung den Schätzer mit minimaler Varianz ergibt. Score-Funktionen zur Cox'schen partiellen Likelihood, in die Gewichtsfunktionen eingefügt werden, werden in (Lin 1991, Sasieni 1993) betrachtet. Da der Schätzer der Regressionskoeffizienten unter dem Modell für beliebige Gewichtsfunktionen erwartungstreu und konsistent ist, kann man aus der Differenz der gewichteten Score-Funktionen sogar eine Teststatistik für die Abweichung von der Annahme der Proportionalität der Hasardraten konstruieren (Lin 1991).

5.4 Der Schätzer von Liang et al. (1993)

(Liang et al. 1993) schlagen zur Schätzung eine U -Statistik vom Grad 3 vor. Deren (nicht symmetrisierter) Kern ist gegeben durch

$$\phi_{\text{Liang3}}^*(\beta; X_i, X_l, X_u) = \sum_{k=1}^K \sum_{j=1}^K \int_0^\tau W_{u;L3}^{-1}(t) (Z_{ij}(t) - E_{ij;lk}(\beta, t)) dN_{ij}(t). \quad (5.4.1)$$

Dabei ist die (inverse) empirische Gewichtsfunktion definiert durch

$$W_{u;L3}^{-1}(t) = \frac{n}{Y_{..}(t)} \left[\frac{2n}{n-2} - \frac{nY_u(t)}{Y_{..}(t)} \right]. \quad (5.4.2)$$

Die Schätzgleichung ist dann gegeben durch

$$U_{\text{Liang3}}(\beta) = \binom{n}{3}^{-1} \sum_{\substack{i,l,u=1 \\ i \neq l, l \neq u, u \neq i}}^n \phi_{\text{Liang3}}^*(\beta; X_i, X_l, X_u) \quad (5.4.3)$$

Diese Darstellung weicht in mehreren Punkten von (Liang et al. 1993) ab. Zunächst wird die nicht symmetrisierte Form des Kerns verwendet, so daß im Kern die Reihenfolge der Argumente relevant ist. Darüberhinaus wird hier eine feste Anzahl K von Individuen pro Haufen angenommen. Dadurch muß die Beschränktheit der Anzahl der Individuen pro Haufen nicht zusätzlich gefordert werden; ungleiche Haufengrößen können durch Einfügen von Individuen mit Indikatorfunktion $Y_{ik} \equiv 0$ berücksichtigt werden. Der Erwartungswert $\mathbb{E} Y_{..}(t)$ in (9) von (Liang et al. 1993) muß bei der tatsächlichen Schätzung ohnehin durch $Y_{..}(t)$ ersetzt werden. Da $\mathbb{E} Y_{..}$ in Gleichung (9) vorkommt, ist der Kern von (n) abhängig.

Der Index u kommt in $\phi_{\text{Liang3}}^*(\beta; X_i, X_l, X_u)$ nur in der Gewichtsfunktion vor. Summiert man diese über den Index u , so erhält man eine neue Gewichtsfunktion. Für $i \neq l$ gilt

$$\begin{aligned} \sum_{\substack{u=1 \\ u \neq i, u \neq l}}^n W_{u;L3}^{-1}(t) &= \sum_{\substack{u=1 \\ u \neq i, u \neq l}}^n \frac{n^2}{Y_{..}(t)} \left[\frac{2}{n-1} - \frac{Y_u(t)}{Y_{..}(t)} \right] \\ &= \frac{n^2}{Y_{..}(t)} \left[\frac{2(n-2)}{n-1} - \frac{Y_{..}(t) - Y_i(t) - Y_l(t)}{Y_{..}(t)} \right] \\ &= (n-2) W_{il;L2}^{-1}(t) \end{aligned}$$

mit

$$W_{il;L2}^{-1}(t) = \frac{n^2}{Y_{..}(t)} \left[\frac{2}{n-1} - \frac{Y_{..}(t) - Y_i(t) - Y_l(t)}{(n-2)Y_{..}(t)} \right]. \quad (5.4.4)$$

Mit dieser Gewichtsfunktion definiert man den neuen Kern vom Grad 2

$$\phi_{\text{Liang2}}^*(\beta; X_i, X_l) = 3 \sum_{k=1}^K \sum_{j=1}^K \int_0^\tau W_{il;L2}^{-1}(t) (Z_{ij}(t) - E_{ij;lk}(\beta, t)) dN_{ij}(t); \quad (5.4.5)$$

Und erhält die äquivalente U -Statistik mit Kern vom Grad 2 durch

$$U_{\text{Liang2}}(\beta) = \binom{n}{2}^{-1} \sum_{\substack{i,l=1 \\ i \neq l}}^n \phi_{\text{Liang2}}^*(\beta; X_i, X_l). \quad (5.4.6)$$

Sie erfüllt

$$\begin{aligned} U_{\text{Liang2}}(\beta) &= \binom{n}{2}^{-1} \sum_{\substack{i,l=1 \\ i \neq l}}^n \phi_{\text{Liang2}}^*(\beta; X_i, X_l) \\ &= \frac{(n-2)}{3} \binom{n}{3}^{-1} \sum_{\substack{i,l=1 \\ i \neq l}}^n \phi_{\text{Liang2}}^*(\beta; X_i, X_l) \\ &= U_{\text{Liang3}}(\beta) \end{aligned}$$

$U_{\text{Liang3}}(\beta)$ kann also auch als U -Statistik vom Grad 2 aufgefaßt werden. Sie unterscheidet sich von $U(\beta)$ aus (5.2.1), abgesehen von der Gewichtsfunktion, nur durch den Faktor 3, der die Eigenschaften der Schätzfunktion nicht beeinflußt und daher unerheblich ist.

Die Darstellung impliziert auch, daß (5.4.3) durch einen Algorithmus mit einem Rechenaufwand der Ordnung n^2 berechnet werden kann.

Die Gewichtsfunktionen $W_{il;L2}(t)$ und $n_i(t)/n$ unterscheiden sich nur um einen Faktor $O(n^{-1})$. Es gilt nämlich

$$\begin{aligned} &W_{il;L2}^{-1}(t) - \frac{n}{Y_{..}(t) - Y_i(t)} \\ &= \frac{[n^2 2(n-2)Y_{..}(t) - n^2(n-1)(Y_{..}(t) - Y_i(t) - Y_l(t))] \cdot (Y_{..}(t) - Y_i(t) - nY_{..}^2(t)(n-1)(n-2))}{Y_{..}^2(t)(n-1)(n-2)(Y_{..}(t) - Y_i(t))} \end{aligned}$$

Die Ausdrücke im Nenner von der gleichen Ordnung wie der Nenner fallen weg:

$$n^2 2nY_{..}^2(t) - n^2 nY_{..}(t)Y_{..}(t) - nY_{..}^2(t)nn = 0$$

Die übrigen Ausdrücke sind gleichmäßig in t beschränkt und von der Ordnung n^{-1} .

Daher impliziert die Gültigkeit von PL 4 für $\hat{W}(t) = Y_{..}(t)/n$ die asymptotischen Eigenschaften der Schätzung für $W_{il;L2}(t)$ und $W_{iju;L3}(t)$.

Zuletzt sei noch bemerkt, daß zwar die U -Statistiken U_{Liang3} und U_{Liang2} übereinstimmen und somit auch die Jackknife-Varianzschätzer. Die Abweichung des Jackknife-Varianz-Schätzers (3.2.11) nach (Puri und Sen 1971) vom Jackknife-Varianz-Schätzer (3.2.2) nimmt jedoch mit dem Grad des Kerns der U -Statistik zu. Die Korrektur des Jackknife-Varianz-Schätzers von (Liang et al. 1993) (siehe Seite 76), wirkt diesem Effekt entgegen.

5.5 Der Logrank-Test als U -Statistik

Das marginale Cox-Modell 4.4 kann mittels des marginalen partiellen Likelihood-Prozesses und U -Statistiken konsistent geschätzt und getestet werden. Die Ähnlichkeiten zwischen den Schätzgleichungen (4.4.7) und (5.1.4) sind offensichtlich. In diesem Abschnitt wird gezeigt, daß unter den Voraussetzungen des Logrank-Tests die Teststatistiken tatsächlich asymptotisch äquivalent sind.

5.5.1 k -Stichproben Tests

(Andersen, Borgan, Gill und Keiding 1982) ist eine grundlegende Arbeit, in der k -Stichproben Tests für Überlebensdaten auf Zählprozesse verallgemeinert und im Rahmen einer Klasse von Tests behandelt werden. Die Darstellung folgt (Andersen et al. 1993).

Sei $N = (N_1, \dots, N_k)$, $k \geq 2$, ein k -dimensionaler Zählprozeß mit Intensitätsprozeß $\lambda_h(t) = \alpha_h(t)Y_h(t)$. Y_h ist eine lokal beschränkte, vorhersagbare Funktion. In den hier behandelten Anwendungen genügt es, unter Y_h die Anzahl der Individuen in der Risikomenge in der Stichprobe h zu verstehen. Die Nullhypothese ist

$$H_0 : \alpha_1 = \dots = \alpha_k. \quad (5.5.1)$$

Das Komplement dazu ist die Alternativhypothese. Die Teststatistiken beruhen auf dem Nelson-Aalen-Schätzer für die kumulativen Hasardraten

$$\hat{A}_h(t) = \int_0^t \frac{J_h(s)}{Y_h(s)} dN_h(s). \quad (5.5.2)$$

Dabei ist $J_h(t) = I(Y_h(t) > 0)$ und $0/0 = 0$ definiert.

Unter der Nullhypothese kann eine gemeinsame kumulative Hasardrate A angenommen werden, die mit Hilfe von $N. = \sum_h N_h$ und $Y. = \sum_h Y_h$ geschätzt wird durch

$$\hat{A}(t) = \int_0^t Y.(s)^{-1} dN.(s) \quad (5.5.3)$$

Für den Zweck des Vergleichs der gruppenspezifischen mit der gemeinsamen Hasardrate schränkt man die gemeinsame Hasardrate mittels der Funktion $J_h(t)$ auf den Bereich positiver Werte von $\hat{A}_h(t)$ ein:

$$\tilde{A}_h(t) = \int_0^t \frac{J_h(s)}{Y.(s)} dN.(s) \quad (5.5.4)$$

Sei $M_h(t) = N_h(t) - \int_0^t Y_h(s)\alpha_h(s) ds$ und $M.(t) = \sum_h M_h(t)$. Unter der Nullhypothese ist die Differenz der kumulativen Hasardraten ein quadratisch integrierbares Martingal

$$\hat{A}_h(t) - \tilde{A}_h(t) = \int_0^t \frac{J_h(s)}{Y_h(s)} dM_h(s) - \int_0^t \frac{J_h(s)}{Y.(s)} dM.(s) \quad (5.5.5)$$

Aus (5.5.5) erhält man eine Familie von Teststatistiken durch Gewichtung der Differenzen in den Zuwächsen durch einen lokal beschränkten und vorhersagbaren Prozeß K_h

$$Z_h(t) = \int_0^t K_h(s) d(\hat{A}_h - \tilde{A}_h)(s) \quad (5.5.6)$$

Überlicherweise wird eine Gewichtung der Form $K_h(t) = Y_h(t)K(t)$ gewählt. Außerdem soll $K(t)$ nur von $Y.(t)$ und $N.(t)$ abhängen. Daher gilt

$$Z_h(t) = \int_0^t K(s) dN_h(s) - \int_0^t K(s) \frac{Y_h(s)}{Y.(s)} dN.(s), \quad (5.5.7)$$

sodaß $\sum_h Z_h(t) = 0$. Unter der Nullhypothese gilt dann nach (5.5.5)

$$Z_h(t) = \sum_{l=1}^k \int_0^t K(s) J_h(s) \left(\delta_{hl} - \frac{Y_h(s)}{Y.(s)} \right) dM_l(s) \quad (5.5.8)$$

Dabei ist δ_{hl} ein Kroneckerdelta. Wegen Satz 1.4.1 ist $Z_h(t)$ ein quadratisch integrierbares Martingal und

$$\langle Z_h, Z_j \rangle(t) = \int_0^t K^2(s) J_h(s) J_j(s) \frac{Y_h(s)}{Y(s)} \left(\delta_{hj} - \frac{Y_j(s)}{Y(s)} \right) \alpha(s) Y(s) ds \quad (5.5.9)$$

Das lokale Martingal (5.5.8) ist quadratisch integrierbar, wenn $\mathbb{E} \langle Z_h \rangle(t) < \infty$ (siehe Seite 7). Notwendig und hinreichend für die Existenz der ersten und zweiten Momente von $Z_h(t)$ für alle h ist daher $\mathbb{E} \int_0^t \{K^2(s) Y(s)\} \alpha(s) ds < \infty$. Unter der Nullhypothese ist $\mathbb{E} Z_h(t) = 0$ und

$$\text{Cov}(Z_h(t), Z_j(t)) = \mathbb{E} \langle Z_h, Z_j \rangle(t). \quad (5.5.10)$$

Ein erwartungstreuer Schätzer für die Kovarianz von $Z(t)$ ist gegeben durch

$$\hat{\sigma}_{hj}^2(t) = \int_0^t K^2(s) J_h(s) J_j(s) \frac{Y_h(s)}{Y(s)} \left(\delta_{hj} - \frac{Y_j(s)}{Y(s)} \right) dN(s), \quad (5.5.11)$$

da die Differenz von (5.5.9) und (5.5.11) wiederum ein lokal quadratisch integrierbares Martingal ist. Sei $\hat{\Sigma}(t)$ die $k \times k$ -Matrix bestehend aus den Elementen $\hat{\sigma}_{hj}^2(t)$. Die naheliegende Teststatistik ist $X^2(t) = Z(t)^\top \hat{\Sigma}(t)^- Z(t)$ bei $t = \tau$ oder für eine „große“ Zeit t . $\hat{\Sigma}(t)^-$ ist eine verallgemeinerte Inverse ($\hat{\Sigma}$ hat maximal Rang $k - 1$). Der maximale Rang wird dann angenommen, wenn es eine Kette von Paaren $\{h, j\}$ gibt, die die Punkte $\{1, \dots, k\}$ verbindet, sodaß eine Zeit t gibt, bei der N springt und K , Y_h und Y_j positiv sind. Sind die Bedingungen für maximalen Rang erfüllt und nimmt man eine Komponente von $Z(t)$ weg, so ist die resultierende Kovarianzmatrix invertierbar. In jedem Fall ist die Teststatistik asymptotisch χ^2 -verteilt mit $k - 1$ Freiheitsgraden.

Im Fall $k = 2$ erhält man

$$X^2(t) = (Z_1(t))^2 / \hat{\sigma}_{11}^2(t)$$

Den Logrank-Test (Mantel 1966, Cox 1972, Peto und Peto 1972) für die Nullhypothese (5.5.1) erhält man für $K(t) = I(Y(t) > 0)$. Es gilt dann

$$Z_h(\tau) = \int_0^\tau I(Y(t) > 0) dN_h(s) - \int_0^\tau \frac{Y_h(s)}{Y(s)} dN(s) \quad (5.5.12)$$

Der Varianzschätzer ist

$$\hat{\sigma}_{hj}^2(\tau) = \int_0^\tau \frac{Y_h(s)}{Y(s)} J_h(s) J_j(s) \left(\delta_{hj} - \frac{Y_j(s)}{Y(s)} \right) dN(s). \quad (5.5.13)$$

5.5.2 Das Cox-Modell für k Stichproben

Angenommen n Individuen aus den Gruppen $h = 1, \dots, k$ sollen auf Unterschiede in der Überlebensverteilung getestet werden. Dazu kodiert man h -te Komponente der Kovariablen des i -ten Individuums als

$$Z_i^h = I(i \text{ in Stichprobe } h), \quad h = 1, \dots, k-1 \quad i = 1, \dots, n$$

Die Hasardfunktion des multiplikativen Regressionsmodells (siehe 4.1) ist

$$\alpha_i(t) = \alpha_0(t) \exp \left\{ \sum_{h=1}^{k-1} \beta_h Z_i^h \right\}$$

Für den Score-Test der Nullhypothese

$$\beta_1 = \dots = \beta_{k-1} = 0 \quad (5.5.14)$$

wird die Scorefunktion (4.2.9) mit den Parameterwerten der Nullhypothese ausgewertet:

$$U_\tau^h(0) = \bar{N}_h(\tau) - \int_0^\tau -\frac{\bar{Y}_h(t)}{Y(t)} dN(t) \quad (5.5.15)$$

für $h = 1, \dots, k-1$. Dabei ist $\bar{N}_h(t) = \sum_i Z_i^h N_i(t)$. $Y_i(t)$ ist ein Indikator dafür, ob Individuum i , $i = 1, \dots, n$, zur Zeit t in der Risikomenge ist. $\bar{Y}_h = \sum_i Z_i^h Y_i$, $N = \sum_h \bar{N}_h$ und $Y = \sum_h \bar{Y}_h$. Die Nullhypothese (5.5.14) spezifiziert das gleiche Modell wie (5.5.1). (5.5.15) stimmt mit (5.5.12) überein.

Die Matrix der zweiten partiellen Ableitungen ist

$$\mathcal{I}_\tau^{hl}(0) = \int_0^\tau \frac{\bar{Y}_h(t)}{Y(t)} \left(\delta_{hl} - \frac{\bar{Y}_l(t)}{Y(t)} \right) dN(t) \quad (5.5.16)$$

Auch (5.5.16) stimmt mit dem entsprechenden Ausdruck (5.5.13) überein. Daher ist der Logrank-Test für k Gruppen äquivalent zum Score-Test des Cox-Modells.

5.5.3 Der Pseudolikelihoodschätzer

Es wird gezeigt, daß der „Score-Test“ basierend auf dem Pseudolikelihoodschätzer für unabhängig verteilte Stichprobe (Haufengröße $K = 1$) asymptotisch mit dem Logrank-Test übereinstimmt.

Seien die k Gruppen wie in Abschnitt 5.5.2 in den Regressoren Z_i^h , $h = 1, \dots, k-1$ kodiert (die Notation wird von dort übernommen). Die h -te Komponente Score-Statistik zur Schätzgleichung (5.1.4) für $\beta = 0$ (vergleiche (5.5.15)) ist gegeben durch

$$S^h(0) = \sum_{i=1}^n \int_0^\tau Z_i^h - \frac{I(Y.(t) > 1)}{Y.(t) - 1} \sum_{\substack{j=1 \\ j \neq i}}^n Y_j(t) \frac{Z_i^h Y_i(t) + Z_j^h Y_j(t)}{Y_i(t) + Y_j(t)} dN_i(t). \quad (5.5.17)$$

für $h = 1, \dots, k-1$. Der Summand Z_i^h im Integral liefert insgesamt $\sum_{i=1}^n \int_0^\tau Z_i^h dN_i(t) = \int_0^\tau d\bar{N}_h(t) = \bar{N}_h(\tau)$.

Für den zweiten Summanden im Integral (5.5.17) kann man in der h -ten Komponente der Schätzfunktion folgende Fälle unterscheiden:

$$Z_i^h = 1, Z_j^h = 1 \text{ ergibt } \int_0^\tau \frac{\bar{Y}_h(t) - 1}{Y.(t) - 1} d\bar{N}_h(t)$$

$$Z_i^h = 1, Z_j^h = 0 \text{ ergibt } \int_0^\tau \frac{Y.(t) - \bar{Y}_h(t)}{2(Y.(t) - 1)} d\bar{N}_h(t)$$

$$Z_i^h = 0, Z_j^h = 1 \text{ ergibt } \int_0^\tau \frac{\bar{Y}_h(t)}{2(Y.(t) - 1)} d(N. - \bar{N}_h)(t)$$

$$Z_i^h = 0, Z_j^h = 0 \text{ ergibt } 0$$

und somit

$$\begin{aligned} S^h(0) &= \int_0^\tau d\bar{N}_h(t) - \int_0^\tau \frac{\bar{Y}_h(t) - 1}{Y.(t) - 1} + \frac{\bar{Y}.(t) - \bar{Y}_h(t)}{2(Y.(t) - 1)} d\bar{N}_h(t) \\ &\quad - \int_0^\tau \frac{\bar{Y}_h(t)}{2(Y.(t) - 1)} d(N. - \bar{N}_h)(t) \\ &= \int_0^\tau d\bar{N}_h(t) - \int_0^\tau \frac{2\bar{Y}_h(t) - 2 - 2\bar{Y}_h(t) + Y.(t)}{2(Y.(t) - 1)} d\bar{N}_h(t) \\ &\quad - \int_0^\tau \frac{\bar{Y}_h(t)}{2(Y.(t) - 1)} dN.(t) \\ &= \frac{1}{2} \left\{ \int_0^\tau d\bar{N}_h(t) - \int_0^\tau \frac{\bar{Y}_h(t)}{Y.(t)} dN.(t) \right\} \end{aligned} \quad (5.5.18)$$

$$- \frac{1}{2} \int_0^\tau \frac{-2 + Y.(t) - Y.(t) + 1}{Y.(t) - 1} d\bar{N}_h(t) \quad (5.5.19)$$

$$- \frac{1}{2} \int_0^\tau \frac{\bar{Y}_h(t)}{Y.(t) - 1} - \frac{\bar{Y}_h(t)}{Y.(t)} dN.(t) \quad (5.5.20)$$

(5.5.18) entspricht (5.5.15). Unter den Regularitätsbedingungen L 1 bis L 4 und PL 1 bis PL 4 gilt für dieses Modell Satz 4.3.4 und Satz 5.2.3, in denen jeweils die asymptotischen Verteilungen der Score-Funktionen bei β_0 angegeben werden. (5.5.19) und (5.5.20) kann in der asymptotischen Verteilung von $n^{-1/2}S^h(0)$ vernachlässigt werden. Daher stimmen die beiden Schätzer in ihren asymptotischen Verteilungen überein.

Kapitel 6

Simulationen

Asymptotisch gültige Resultate können für den Einsatz in der angewandten Statistik zu wenig aussagekräftig sein. Sie müssen daher an endlichen Stichproben validiert werden. Die Stichproben können durch Simulationen erzeugt werden oder aus realen Datensätzen stammen.

6.1 Bivariate Überlebensfunktionen

Ein möglicher Ansatz zur Erzeugung von bivariaten Überlebensverteilungen sind Gebrechlichkeits-Modelle (siehe 2.5). Dies ist die Mischung einer bedingten univariaten Verteilung mit einer Gebrechlichkeits-Verteilung. In diesem Zusammenhang ist der Begriff der Kopula von Verteilungsfunktionen hilfreich (engl: copula). Nach Genest und MacKay (1986) und Shih und Louis (1995) ist eine Kopula eine bivariate Verteilungsfunktion C auf $[0, 1]^2$. Sei (T_1, T_2) eine bivariate Überlebenszeit mit marginalen Überlebensfunktionen S_1 und S_2 . Wenn (T_1, T_2) von der Kopula C kommt, dann ist die Überlebensfunktion gegeben durch

$$S(t_1, t_2) = C(S_1(t_1), S_2(t_2)) \quad \text{für} \quad t_1, t_2 \geq 0. \quad (6.1.1)$$

Archimedische Kopulae ((Schweizer und Sklar 1983) zitiert nach (Genest 1987)) sind definiert durch

$$C(u, v) = \phi[\phi^{-1}(u) + \phi^{-1}(v)] \quad \text{für} \quad 0 \leq u, v \leq 1. \quad (6.1.2)$$

Dabei gilt $0 \leq \phi \leq 1$, $\phi(0) = 1$, $\phi' < 0$, $\phi'' > 0$.

Wenn ϕ die Laplace-Transformierte einer Verteilung ist, dann entspricht die Kopula einem proportionalen Gebrechlichkeitsmodell auf folgende Weise:

Seien H_i , $i = 1, 2$, zwei bedingte kumulative Hasardraten für $Z = 1$ in einem Gebrechlichkeitsmodell. Es gilt

$$H_i(t_i) = -\log(S_i(t_i|Z = 1)) \quad i = 1, 2.$$

Dann ist unter Annahme von bedingter Unabhängigkeit die bedingte Überlebensfunktion bezüglich $Z = z$ gegeben durch

$$\begin{aligned} S(t_1, t_2|Z = z) &= S_1(t_1|Z = z) \cdot S_2(t_2|Z = z) \\ &= \exp^{-z\{H_1(t_1)+H_2(t_2)\}}. \end{aligned} \quad (6.1.3)$$

Die bivariate Überlebensfunktion ist dann

$$\begin{aligned} S(t_1, t_2) &= \mathbb{E} \exp\{-Z\{H_1(t_1) + H_2(t_2)\}\} \\ &= \phi(H_1(t_1) + H_2(t_2)). \end{aligned} \quad (6.1.4)$$

Die bivariate Überlebensfunktion ist also die Laplace-Transformierte ϕ der Gebrechlichkeit Z bei $H_1(t_1) + H_2(t_2)$. Analog dazu sind die marginalen Überlebensfunktionen gegeben durch

$$S_i(t_i) = \phi(H_i(t_i)) \quad \text{für} \quad t_i \geq 0, \quad i = 1, 2. \quad (6.1.5)$$

Bei der Erzeugung von bivariaten Überlebenszeiten aufgrund der bivariaten und marginalen Überlebensfunktion kann folgendermaßen vorgegangen werden: Zunächst wird eine Zufallszahl U aus der Gleichverteilung auf $[0, 1]$ erzeugt. $T_1 = S_1^{-1}(U)$ hat dann Überlebensfunktion S_1 . Sei t_1 eine derart erhaltene Realisierung von T_1 . Bezüglich t_1 ist die Überlebensfunktion von T_2

$$S_2(t_2|T_1 = t_1) = \frac{\partial S(t_1, t_2)/\partial t_1}{\partial S_1(t_1)/\partial t_1}. \quad (6.1.6)$$

Mittels einer zweiten gleichverteilten Zufallsvariablen V erzeugt man eine Zufallszahl T_2 mit Überlebensfunktion (6.1.6). (t_1, t_2) ist dann die Realisierung einer bivariaten Zufallsvariablen mit gemeinsamer Überlebensfunktion $S(\cdot, \cdot)$.

Bei der Erzeugung von bivariaten Überlebenszeiten mit Hilfe einer Gebrechlichkeits-Verteilung geht man folgendermaßen vor:

Zunächst berechnet man durch Umkehrung von (6.1.5) bedingte kumulative Hasardraten für $Z = 1$:

$$H_i(t_i) = \phi^{-1}(S_i(t_i)) \quad \text{für} \quad i = 1, 2. \quad (6.1.7)$$

Sei z eine Realisierung einer Zufallsvariablen Z . Dann ist die bedingte Überlebensfunktion bezüglich $Z = z$ nach (6.1.3) gegeben durch

$$S_i(t_i|Z = z) = e^{-zH_i(t_i)} \quad \text{für} \quad i = 1, 2. \quad (6.1.8)$$

Mittels der Transformation

$$T_i = H_i^{-1} \left(-\frac{1}{z} \log U_i \right), \quad U_i \sim U(0, 1) \quad i = 1, 2 \quad (6.1.9)$$

erhält man wiederum eine Realisierung einer Zufallsvariablen mit der gemeinsamen Überlebensfunktion (6.1.3).

In einem Modell mit proportionalen marginalen Hasardraten sind die Überlebensfunktionen von der Form

$$S_i(t_i) = \exp \{ -\tau_i A_0(t_i) \} \quad \text{für} \quad i = 1, 2. \quad (6.1.10)$$

Dabei ist τ_i ein Proportionalitätsfaktor für die Hasardfunktion und A_0 eine gemeinsame kumulative Hasardfunktion. Zumeist wird für $S_i(\cdot)$, $i = 1, 2$, die Exponentialverteilung mit kumulativer Hasardfunktion $A_0(t_i) = t_i$ verwendet.

6.1.1 Positiv stabile Gebrechlichkeits-Verteilungen

Die Laplace-Transformierte einer positiv stabilen Verteilung ist

$$\phi(s) = e^{-s^\alpha} \quad \text{für} \quad 0 < \alpha < 1. \quad (6.1.11)$$

Kleine Werte von α induzieren starke Korrelationen zwischen T_1 und T_2 .

Hougaard (1986) bemerkte, daß die positiv stabile Verteilung die einzige Gebrechlichkeits-Verteilung ist, bei der sowohl die bedingten, als auch die marginalen Verteilungen proportionale Hasardraten haben. Diese Eigenschaft kann folgendermaßen begründet werden: Sei die bedingte Überlebensfunktion $S(t|Z = z) = \exp \{ -z\theta H_0(t) \}$. Dabei ist $H_0(t)$ eine kumulative zugrundeliegende Hasardfunktion und θ ein Proportionalitätsfaktor. Die marginale Überlebensfunktion ist dann gegeben durch

$$S(t) = \phi(\theta H_0(t)) \stackrel{(6.1.11)}{=} \exp \{ -[\theta H_0(t)]^\alpha \}.$$

Somit erfüllt die marginale Überlebensfunktion ein Modell mit proportionalen Hasardraten mit zugrundeliegender Hasardfunktion $H_0(t)^\alpha$ und Proportionalitätsfaktor θ^α .

Lee et al. (1992) verwendeten in ihrer Simulationsstudie positiv stabile Verteilungen zur Validierung ihres Schätzers für das marginale Cox-Modell.

Für die Simulation von Überlebenszeiten mittels der Gebrechlichkeitsverteilung hat diese Verteilungsfamilie die Eigenschaft, daß die bedingten Verteilungen nicht wie in (6.1.7) und (6.1.8) durch Anwendung der inversen Laplace-Transformation und der Exponentialfunktion berechnet werden, sondern daß einfache Verteilungen wie beispielsweise die Exponentialverteilung als bedingte Verteilungen bezüglich $Z = z$ verwendet werden können.

6.1.2 Die Clayton-Familie

Diese Familie von bivariaten Überlebensfunktionen wurde in Clayton (1978) eingeführt. Die Bezeichnung geht auf Shih und Louis (1995) zurück.

Zunächst wird festgestellt, daß die Laplace-Transformierte der Gammaverteilung mit Gestaltparameter ν und Skalierungsparameter η durch

$$\phi(t) = \left(\frac{\eta}{\eta + t} \right)^\nu \quad (6.1.12)$$

gegeben ist. Im Gebrechlichkeits-Modell können Skalierungen auch in der bedingten Hasardfunktion durch einen Proportionalitätsfaktor berücksichtigt werden. Daher ist es zweckmäßig, sich auf Gammaverteilungen mit Erwartungswert 1, d. h. $1/\xi = \eta = \nu$, zu beschränken. Dann erhält man die Laplacetransformierte $(1 + \xi t)^{-1/\xi}$.

Eine andere Möglichkeit der Einschränkung des Parameterraumes wurde von Oakes (1982) eingeführt. Er setzte $\eta = 1$ und $\theta = 1 + \nu^{-1}$. Die Laplacetransformation dieser Verteilung ist dann gegeben durch

$$\phi(s) = (1 + s)^{1/(1-\theta)}, \quad \theta > 1, \quad (6.1.13)$$

deren inverse Funktion ist

$$\phi^{-1}(u) = z^{1-\theta} - 1, \quad \theta > 1.$$

$\theta > 1$ induziert positive Korrelation zwischen T_1 und T_2 , für $\theta \rightarrow 1+$ erhält man Unabhängigkeit. Die Clayton-Familie wird aus den marginalen Verteilungen und der durch ϕ in (6.1.13) gegebenen archimedischen Kopula generiert. Die gemeinsame Überlebensfunktion ist nach (6.1.1) und (6.1.2)

$$S(t_1, t_2) = (S_1(t_1)^{(1-\theta)} + S_2(t_2)^{(1-\theta)} - 1)^{1/(1-\theta)}, \quad \theta > 1. \quad (6.1.14)$$

unter Verwendung des Zwischenergebnisses

$$\begin{aligned} C(u, v) &= \phi(\phi^{-1}(u) + \phi^{-1}(v)) = \phi(u^{1-\theta} + v^{1-\theta} - 2) \\ &= (u^{1-\theta} + v^{1-\theta} - 1)^{\frac{1}{1-\theta}}. \end{aligned} \quad (6.1.15)$$

Sei $a_1(t_1|t_2)$ die Hasardfunktion von T_1 bezüglich $T_2 = t_2$ und sei $b_1(t_1|t_2)$ die Hasardfunktion der Verteilung von T_2 bezüglich $T_2 \geq t_2$. Dann hat nur (siehe Clayton, 1978) die Clayton-Familie die Eigenschaft, daß

$$\frac{a_1(t_1|t_2)}{b_1(t_1|t_2)} = \theta \quad \text{für} \quad t_1, t_2 > 0. \quad (6.1.16)$$

Zur Erzeugung von bivariaten Zufallszahlen mit der Überlebensfunktion (6.1.14) benötigt man aufgrund von (6.1.6) zunächst

$$\begin{aligned} \frac{\partial S(t_1, t_2)}{\partial t_1} &= \frac{\partial}{\partial t_1} [S_1(t_1)^{1-\theta} + S_2(t_2)^{1-\theta} - 1]^{-1/(\theta-1)} \\ &= [S_1(t_1)^{1-\theta} + S_2(t_2)^{1-\theta} - 1]^{-1/(\theta-1)-1} S_1(t_1)^{-\theta} \frac{\partial S_1(t_1)}{\partial t_1} \end{aligned}$$

und schließlich

$$\begin{aligned} S_2(t_2|T_1 = t_1) &= \frac{\partial S(t_1, t_2)/\partial t_1}{\partial S_1(t_1)/\partial t_1} \\ &= [S_1(t_1)^{1-\theta} + S_2(t_2)^{1-\theta} - 1]^{\theta/(1-\theta)} S_1(t_1)^{-\theta}. \end{aligned}$$

Die für die Erzeugung von Zufallszahlen benötigte inverse bedingte Überlebensfunktion für konstante marginale Hasardraten ist gegeben durch

$$S_2^{-1}(U_2|T_1 = t_1) = S_2^{-1} \left(\left\{ [U_2 S_1(t_1)^\theta]^{\frac{1-\theta}{\theta}} - S_1(t_1)^{1-\theta} + 1 \right\}^{1/(1-\theta)} \right). \quad (6.1.17)$$

Darauf basiert der folgende XLISP-STAT-Code, der n Paare von Überlebenszeiten nach der obenstehenden Methode erzeugt:

```
(defun margph-cond-rand (n theta ebt1 ebt2)
  "Args: (n theta ebt1 ebt2)
  Erzeuge n Paare von Zufallszahlen mit marginal konstanten
  Hasardraten, die durch EBT1 bzw. EBT2 gegeben sind und
  durch Mischung mit einer Gammaverteilung mit Parameter
  THETA zusammenhaengen. Dabei wird die erste Komponente nach
```

der marginalen Verteilung erzeugt und die zweite Komponente durch die bedingte Verteilung bezueglich ersten Komponente."

```
(let* ((theta-1 (when (not (equal theta 1))
                     (1- theta)))
      (logSt1 (log (uniform-rand n))); Exponentialvert.
      (t1 (- (/ logSt1 ; marginale Hasardrate
              ebt1))) ; konstant
      (t2 (if theta-1
              ;theta ungleich 1
              (/ (log (- (exp
                          (* (+ (log (uniform-rand n))
                                (* theta logSt1))
                                (/ (- 1 theta) theta)))
                  (exp (* (1- theta) logSt1)
                        -1))
                (* theta-1 ebt2))
              ;theta gleich 1 - Unabhaengigkeit
              (- (/ (log (uniform-rand n)) ebt2))))))
      (return (append t1 t2))))
```

Bei der Simulation von bivariaten Überlebenszeiten mit proportionalen marginalen kumulativen Hasardraten — nach dem Clayton-Modell durch Mischung mit einer Gebrechlichkeits-Verteilung — spezifiziert man zunächst die marginale Überlebensfunktion

$$S_i(t) = e^{-t\tau_i} \quad \text{für} \quad i = 1, 2.$$

Dabei ist τ_i der Proportionalitätsfaktor der Hasardfunktionen und die kumulative zugrundeliegende Hasardfunktion ist $A_0(t) = t$. Die bedingten kumulativen Hasardraten für $Z = 1$ sind aufgrund von (6.1.5)

$$H_i(t) = e^{(\theta-1)\tau_i} - 1 \quad \text{für} \quad i = 1, 2. \quad (6.1.18)$$

Dementsprechend ist die bedingte Überlebensfunktion für $Z = z$ gegeben durch

$$S_i(t|Z = z) = \exp\{-z\tau_i(\theta - 1) - 1\} \quad (6.1.19)$$

Der nachstehende XLISP-STAT-Code erzeugt n Paare von Überlebenszeiten nach dieser Methode:

```
(defun marg-prop-haz-gamma-frailty (n theta ebt1 ebt2)
```

```

"Args: (n theta ebt1 ebt2)
Erzeuge n Paare von Zufallszahlen mit marginal konstanten
Hasardraten, die durch EBT1 bzw. EBT2 gegeben sind und die
durch Mischung mit einer Gammaverteilung mit Parameter THETA
zusammenhaengen."
(if (equal theta 1)
  ; Unabh angigkeit - Exponentialverteilung
  (/ (- (log (uniform-rand (* 2 n)))) ebt)
  ; Abh angigkeit - Mischverteilung
  (let* ((theta-1 (- theta 1))
         (anz (repeat (gamma-rand anz (/ 1 theta-1)) 2)))
        ; Mischverteilung mal (1-theta)
        (/ (log (- 1 (/ (log (uniform-rand (* anz 2)) z)))
            (* theta-1 ebt)))))

```

6.1.3 Transformation der Normalverteilung

In (Loughin und Koehler 1997) werden bivariate  berlebenszeiten durch Transformation von Paaren unabh angiger standardnormalverteilter Zufallszahlen erzeugt. Zur Parametrisierung dieser Verteilungsfamilie wird der Kendall'sche Korrelationskoeffizient verwendet. F ur 2 unabh angige Paare von Zufallsvariablen (X, Y) und (X', Y') mit derselben bivariaten Verteilung ist der Kendall'sche Korrelationskoeffizient nach (Cox und Hinkley 1974, p204) definiert als

$$\tau = \mathbb{E} \operatorname{sign}(X - X')(Y - Y'). \quad (6.1.20)$$

Er ist unter monotonen Transformationen invariant. Bei bivariat normalverteilten Zufallsvariablen gilt folgender Zusammenhang zum Pearson'schen Korrelationskoeffizienten ρ :

$$\rho = \sin \left[\tau \frac{\pi}{2} \right]. \quad (6.1.21)$$

Ein normalverteilter Vektor (Y_1, Y_2) mit Kovarianzmatrix $\Sigma = \begin{pmatrix} 1 & \rho^2 \\ \rho^2 & 1 \end{pmatrix}$ kann aus zwei unabh angig standardnormalverteilten Zufallsvariablen (X_1, X_2) durch Transformation $(Y_1, Y_2)^\top = \sqrt{\Sigma} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ mit der Matrix $\sqrt{\Sigma} = \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1 - \rho^2} \end{pmatrix}$ erzeugt werden. Durch die Transformation $(U_1, U_2) = (\Phi(Y_1), \Phi(Y_2))$ mit der Verteilungsfunktion Φ der Standard-

normalverteilung erhält man einen gleichverteilten Zufallsvektor mit Kendall'schem Korrelationskoeffizienten $\tau = 2 \arcsin(\rho)/\pi$.

Abschließend können durch Transformation mit der inversen Verteilungsfunktion beliebige marginale Überlebensfunktionen $S_1(t_1)$ und $S_2(t_2)$ simuliert werden. Da der Kendall'sche Korrelationskoeffizient unter monotonen Transformationen invariant ist, haben T_1 und T_2 Kendall'sche Korrelation τ .

6.2 Schema der Simulationsstudien

Die nachstehenden Simulationen wurden mittels XLISP-STAT 2.1 Release 3.45 (Beta) unter Windows NT 4.0 auf Pentium-Prozessor beziehungsweise XLISP-STAT 2.1 Release 3.44 (Beta) unter Digital UNIX V4.0B (Rev. 564) mit Alpha-Prozessor gerechnet (Tierney 1990). Das Programm kann über ftp von `umnstat.stat.umn.edu` bezogen werden.

Es wurden die darin eingebauten Pseudozufallszahlengeneratoren verwendet. Diese basieren auf dem modifizierten Wichman-Hill-Generator zur Erzeugung von gleichverteilten Zufallszahlen (Wichman und Hill 1982, L'Ecuyer 1986).

Der *Keim* der *Simulationen* wurde durch eine Abfrage des Zustandes des Pseudozufallszahlengenerators festgelegt mit

```
(setf initial-seed-0 '#$(1 #(2147483562 0 11906 54714))).
```

In einer der beiden folgenden Simulationsstudien werden Design-Vektoren aus einer Gleichverteilung verwendet, die für alle Wiederholungen dieselben sind. In der anderen Studie müssen die Zensierungzeiten vor Beginn der eigentlichen Simulation durch einen Simulationsvorlauf bestimmt werden. `initial-seed-0` ist jeweils der Keim für die darin benötigten Zufallszahlen.

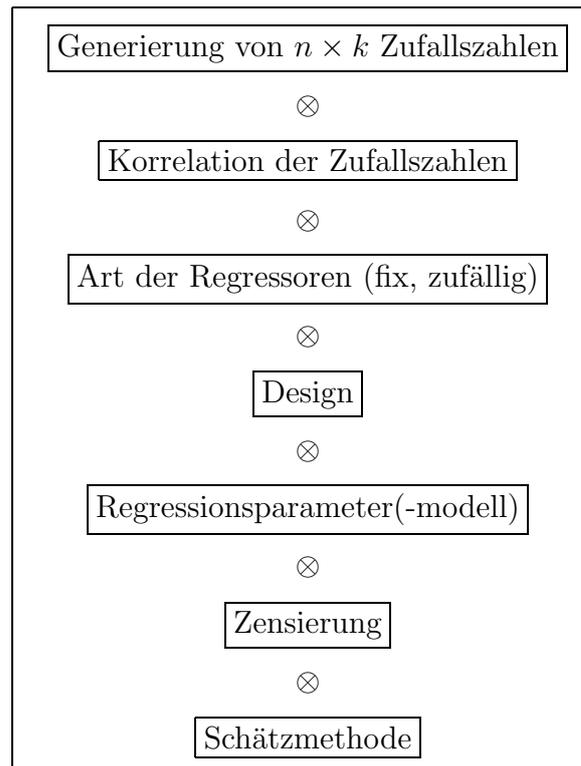
Ein *zweiter Keim* wurde durch 10000 Aufrufe des Zufallsgenerators für gleichverteilte Zufallszahlen ausgehend von obenstehender Saat erzeugt. Das Ergebnis ist der Initialisierungskeim

```
(setf initial-seed
'#$(1 #(2147483562 1639197836 1679793079 40595243))).
```

Dieser Keim ist jeweils der Ausgangspunkt für die eigentliche Simulation.

Es bestehen keine wesentlichen Unterschiede zwischen den angegebenen Implementierungen. Insbesondere sind die verwendeten Pseudozufallszahlengeneratoren unabhängig von der Implementierung. Ein Unterschied besteht in der Rechengenauigkeit. Alle Gleitkommaoperationen sind in XLISP-STAT

Abbildung 6.1: Aufbau der Simulation. Ausgehend von einer Stichprobe vom Umfang n von jeweils k Zufallszahlen werden neue Zellen durch Anwendung der Faktoren der Simulation auf das vorhergehende Ergebnis erzeugt. (Die Schätzmethode und die Art der Regressoren wurden jedoch in getrennten Simulationsläufen untersucht).



von doppelter Genauigkeit. Die Variable `machine-precision`, die die kleinste Gleitkommazahl enthält, für die der Ausdruck `(not (= 1 (+ 1 x)))` das Ergebnis T (true) ergibt, ist $1.0842021724855044E-19$ für Windows NT 4.0 und $2.220446049250313E-16$ für Digital UNIX V4.0B.

Die Simulationen werden nach der Methode von Stine (1992) organisiert (siehe Abbildung 6.1). Die *Zellen* ergeben sich durch die sequentielle Anwendung der in der Simulation zu variierenden Faktoren auf das Ergebnis des vorhergehenden Faktors. Wenn der Faktor nicht nur aus einer Stufe besteht, werden dadurch neue Zellen erzeugt. Die Bezeichnung der Zelle entsteht durch die Aneinanderreihung der Bezeichnungen der durchlaufenen Faktorstufen. Durch diesen systematischen Zugang wird die Konzeption und Auswertung von statistischen Simulationsstudien unterstützt. In jeder Wiederholung wer-

den die Simulationsergebnisse in den entsprechenden Zellen ergänzt. Auf das so entstandene *Simulationsojekt* (in Sinn von Datenstruktur) können Auswertefunktionen angewendet werden.

Der Aufbau der Studie ist wie folgt: Zuerst werden die für die Wiederholung benötigten Zufallszahlen generiert. (Mit Ausnahme der zufälligen Regressoren.) Sodann werden in dieser Reihenfolge eine Korrelations-, Design-, Regressionskoeffizienten-, Zensierungs- und Schätzer-Funktion angewendet. Alle Ergebnisse werden als XLISP-STAT-Objekte abgespeichert.

Dennoch war es aufgrund des Umfangs der Studie notwendig, die Simulation nach der Art der Regressoren (fix oder zufällig) und nach der Schätzmethode aufzuteilen.

6.3 Die Schätzer

Auf die simulierten Daten werden jeweils 4 Schätzer für den Regressionsparameter für das marginale Modell der proportionalen Hazardraten (siehe Abschnitt 4.4.2) bei Haufendaten angewendet. Es sind dies zunächst der *Schätzer von Cox (1972)*, der die Unabhängigkeit der Beobachtungen voraussetzt. Dann wird der *Schätzer von Lee et al. (1992)* (siehe Satz 4.4.1 und Satz 4.4.3) angewendet, der sich vom vorhergehenden Schätzer nur dadurch unterscheidet, daß er die Kovarianzmatrix des Schätzers für den Regressionskoeffizienten auch bei abhängigen Beobachtungen konsistent schätzt. Schließlich werden zwei Pseudolikelihood-Schätzer (Liang et al. 1993) nach Satz 5.2.2, Satz 5.2.3 und Satz 5.3.1 berechnet. Der erste *Pseudolikelihood-Schätzer* verwendet die Gewichtsfunktion $\hat{W}_i(t) = n_i(t)/n$, während der zweite *Schätzer nach (Liang et al. 1993)* die Gewichtsfunktion $\hat{W}_{i;l;L2}(t)$ mit

$$\hat{W}_{i;l;L2}^{-1}(t) = \frac{n^2}{Y_{..}(t)} \left[\frac{2}{n-1} - \frac{Y_{..}(t) - Y_i(t) - Y_l(t)}{(n-2)Y_{..}(t)} \right] \quad (6.3.1)$$

verwendet, die in (5.4.4) definiert wird. In (Liang et al. 1993) wird ein modifizierter Jackknife-Schätzer der Varianz angegeben, der jedoch bis auf den Faktor $(n-2)(n-3)/[(n-1)(n-4)] > 1$ mit dem Varianzschätzer aus Satz 5.2.3 übereinstimmt. Für $n = 50$ Haufen ist dieser Faktor 1,00089 und somit vernachlässigbar.

Dabei wurde folgende Überlegung angestellt: Der Jackknife-Schätzer für die

Varianz einer U -Statistik ist aufgrund von (3.2.2) für $U_n = 0$ gegeben durch

$$\hat{V}_{\text{Jack}}(U_n) = \binom{n}{k}^{-2} \frac{n(n-1)}{(n-k)^2} \sum_{i=1}^n \left[\sum_{(n-1, k-1)}^{(-i)} \phi(X_i, X_{i_2}, \dots, X_{i_k}) \right]^2. \quad (6.3.2)$$

Der entsprechende Jackknife-Varianzschätzer für $k = 3$ nach (Liang et al. 1993) ist gegeben durch

$$\hat{V}_{\text{Jack-Liang}}(U_n) = c \binom{n}{3}^{-2} \frac{n}{(n-1)} \sum_{i=1}^n \left[\sum_{(n-1, 2)}^{(-i)} \phi(X_i, X_{i_2}, X_{i_3}) \right]^2 \quad (6.3.3)$$

mit Endlichkeitskorrektur $c = (n-1)(n-2)/[(n-3)(n-4)]$. Es gilt somit

$$\hat{V}_{\text{Jack-Liang}}(U_n) = \frac{(n-2)(n-3)}{(n-1)(n-4)} \hat{V}_{\text{Jack}}(U_n). \quad (6.3.4)$$

Die Lösungen der Schätzgleichungen wurden mit dem Newton-Raphson-Verfahren gefunden. Die entsprechenden Programme wurden vom Autor in XLISP-STAT programmiert und sind auf Anfrage erhältlich.

6.4 Ein Clayton-Modell

6.4.1 Plan der Simulationsstudie

Diese Simulation wird durchgeführt, um einen direkten Vergleich der Schätzmethoden nach (Liang et al. 1993) und nach (Lee et al. 1992) zu erhalten. Ein derartiger Vergleich wurde bereits von Lin (1994) vorgeschlagen. Das möglichst gleichartige Design zu (Liang et al. 1993) wird deshalb gewählt, da die Ergebnisse in (Liang et al. 1993) für invers abhängiges Design und $\theta = 3$ nicht plausibel erscheinen und somit auch eine Überprüfung dieses Ergebnisses möglich ist (siehe 6.4.3).

In der Studie werden 1000 Datensätze von jeweils 50 Paaren von Überlebenszeiten nach dem Clayton-Modell unter Verwendung der bivariaten Überlebensfunktion

$$S(t_1, t_2) = [e^{-(\theta-1)t_1 \exp\{\beta_{z_1}\}} + e^{-(\theta-1)t_2 \exp\{\beta_{z_2}\}} - 1]^{1/(1-\theta)} \quad (6.4.1)$$

erzeugt.

Der Designvektor z_1 beziehungsweise z_2 hat nur eine Komponente. In der Variante mit *fixen* Regressoren werden vor der Simulation zwei mal 50 Zufallszahlen aus $U[-0.5, 0.5]$ mit Initialisierungskeim `initial-seed-0` erzeugt (genauer: An der Stelle, an der der Designvektor benötigt wird, wird der gegenwärtige Keim zwischengespeichert, der Designvektor wird ausgehend vom Keim `initial-seed-0` berechnet, sodann wird wieder der zwischengespeicherte Keim gesetzt), während der Initialisierungskeim für den Rest der Simulation `initial-seed` ist. In der Variante mit *zufälligen* Regressoren werden diese Zufallszahlen in jeder Wiederholung der Simulation neu erzeugt. Die beiden Vektoren von Zufallszahlen werden z_1 und z_2 zugeordnet. Nur in den Simulationszellen mit *unabhängigem* Design werden alle 100 Zufallszahlen verwendet. Bei *abhängigem* Design wird stattdessen für das zweite Element des Dupels (Paares) (z_1, z_2) jeweils $z_2 = z_1$ gesetzt. Bei *invers abhängigem* Design wird für das zweite Element des Dupels $z_2 = -z_1$ gesetzt. Das Schema der Simulation ist in Tabelle 6.1 zusammengefaßt. Die $2 \times 3 \times 2 \times 2 \times 4 = 96$ Kombinationen der Faktoren definieren die Zellen der Simulation.

Es wird zu festen Zeiten zensiert. Wie in (Liang et al. 1993) werden Überlebenszeiten größer als 1,9 für $\beta = 0$ und Überlebenszeiten größer als 1,96 für $\beta = 1$ durch eine Zensierungszeit 1,90 beziehungsweise 1,96 ersetzt. Das ergibt jeweils etwa 15% Zensierungen.

6.4.2 Ergebnisse

Die Ergebnisse sind in Tabelle 6.4 bis 6.7 angegeben. In Tabelle 6.2 werden die Bezeichnungen zusammengefaßt. Es gibt getrennte Tabellen für die 4 Kombinationen der Faktoren *Art der Regressoren* (fix und zufällig) und *Regressionsparameter* ($\beta = 0$ und $\beta = 1$). Die Tabellen enthalten jeweils für jede Kombination aus Design und Zusammenhangsparameter Kennzahlen über das Verhalten der Art der Schätzung. Die Kennzahlen sind der Mittelwert ($\hat{\beta}$) und die Varianz ($\text{Var}\hat{\beta}$) von $\hat{\beta}$, sowie der Mittelwert ($\widehat{\text{Var}\beta}$) und die geschätzte Standardabweichung ($\hat{\sigma}_{\widehat{\text{Var}\beta}}$) der Schätzung der Varianz von $\hat{\beta}$. Darüberhinaus ist der Anteil der Simulationen angegeben, bei denen β im 95%-Konfidenzintervall (KI95%) enthalten ist, beziehungsweise bei denen β kleiner ist als die Untergrenze des Konfidenzintervalls von $\hat{\beta}$ (UG95%) oder bei denen β größer ist als die Obergrenze des Konfidenzintervalls (OG95%). Schließlich sind auch die Standardfehler (SE) von $\hat{\beta}$, $\text{Var}\hat{\beta}$ und $\widehat{\text{Var}\beta}$ enthalten.

Der Standardfehler von $\text{Var}\hat{\beta}$ ergibt sich wie folgt: Unter der Annahme, daß

Regressoren	fix zufällig
Design	unabhängig abhängig invers
Regressionsparameter	$\beta = 0$ $\beta = 1$
Korrelationsparameter	$\theta = 1$ (keine Korrelation) $\theta = 3$ ($\rho=0.83$)
Zensierung	15%, einfach, vom Typ I (s. Seite 42)
Schätzmethode für β	Cox (1972) Lee et al. (1992) Paarweise Pseudolikelihood Liang et al. (1993)

Tabelle 6.1: Übersicht über die Simulation nach Liang et al. (1993).

die Stichprobe normalverteilt ist mit Varianz σ^2 , hat der aus der Stichprobenvarianz S^2 abgeleitete Ausdruck $Y = (n-1)S^2/\sigma^2$ eine χ^2 -Verteilung mit $n-1$ Freiheitsgraden. Daher ist $\text{Var } Y = 2(n-1)$ und somit $\widehat{\text{Var}}(S^2) = 2S^4/(n-1)$.

Der systematische Fehler von $\hat{\beta}$

Die Ergebnisse bei zufälligen Regressoren sind Mittelwerte über verschiedene Designs. Sie sind in diesem Sinn allgemeiner gültig als die Ergebnisse bei fixen Regressoren. Die Unterschiede im systematischen Fehler von $\hat{\beta}$ zwischen fixen und zufälligen Regressoren sind bisweilen in der Größenordnung des Standardfehlers. $\hat{\beta}$ ist für $\beta = 1$ bei fixen Regressoren kleiner als bei zufälligen Regressoren und für $\beta = 0$ größer als bei zufälligen Regressoren (außer bei abhängigem Design - hier ist jedoch auch der systematische Fehler negativ).

Für $\beta = 0$ übersteigt der systematische Fehler niemals den doppelten Standardfehler. Für $\beta = 1$ und zufällige Regressoren hat der Pseudolikelihood-Schätzer bei unabhängigem Design einen systematischen Fehler, der den doppelten Standardfehler übersteigt. Für $\beta = 1$ und abhängigem Design ist der systematische Fehler positiv und an der Grenze des zweifachen Standardfehlers. Bei *invers abhängigem* Design wird diese Grenze für $\theta = 3$ von der Methode nach (Cox 1972), (Lee et al. 1992) und dem Pseudolikelihood-Schätzer überschritten. Der systematische Fehler des Pseudolikelihood-Schätzers übertrifft in dieser Simulationszelle alle anderen. Er beträgt 33,8% der Standardabweichung von $\hat{\beta}$. Auch für $\theta = 1$ ist der systematische Fehler der Pseudolikelihood-Schätzung viel größer als jene der anderen Schätzmethoden. Er beträgt hier aber lediglich 14,3% der Standardabweichung. Im Gegensatz dazu ist für $\theta = 3$ der systematische Fehler des Schätzers nach (Liang et al. 1993) besonders gering.

Die Varianz von $\hat{\beta}$ ($\text{Var}\hat{\beta}$)

Für $\theta = 3$ ist die Varianz bei abhängigem Design größer und bei invers abhängigem Design geringer als bei unabhängigem Design. Die Varianz der Schätzung ist bei zufälligen Regressoren tendenziell größer als bei fixen Regressoren. Es gibt jedoch Ausnahmen, was darauf hinweist, daß diese Eigenschaft der Schätzung sehr stark vom Design abhängt. Es gibt keine großen Unterschiede zwischen den Schätzmethoden mit Ausnahme der Pseudolikelihood-Schätzung, die zumeist die höchste Varianz aufweist.

n	Anzahl der Wiederholungen
i	Index $i = 1, \dots, n$
$\hat{\beta}_i$	Schätzung von β
$\hat{\beta}$	$n^{-1} \sum_{i=1}^n \hat{\beta}_i$
$\text{Var}\hat{\beta}$	$(n-1)^{-1} \sum_{i=1}^n (\hat{\beta}_i - \hat{\beta})^2$
$\widehat{\text{Var}}\beta_i$	Schätzung von $\text{Var}\hat{\beta}_i$
$\widehat{\text{Var}}\beta$	$n^{-1} \sum_{i=1}^n \widehat{\text{Var}}\beta_i$
$\hat{\sigma}_{\widehat{\text{Var}}\beta}$	$(n-1)^{-1/2} \sqrt{\sum_{i=1}^n (\widehat{\text{Var}}\beta_i - \widehat{\text{Var}}\beta)^2}$
α	Konfidenzkoeffizient (hier: 95%)
z_α	$-\Phi^{-1}((1-\alpha)/2)$
$\text{KI}_\alpha(\hat{\beta}_i)$	$\left[\hat{\beta}_i - z_\alpha \widehat{\text{Var}}\beta_i^{1/2}, \hat{\beta}_i + z_\alpha \widehat{\text{Var}}\beta_i^{1/2} \right]$
$\text{KI}_{95\%}$	$100 \left \left\{ i \beta \in \text{KI}_{95\%}(\hat{\beta}_i) \right\} \right / n$
$\text{UG}_{95\%}$	$100 \left \left\{ i \beta < \min \text{KI}_{95\%}(\hat{\beta}_i) \right\} \right / n$
$\text{OG}_{95\%}$	$100 \left \left\{ i \beta > \max \text{KI}_{95\%}(\hat{\beta}_i) \right\} \right / n$
$\text{SE}(\hat{\beta})$	$(\text{Var}\hat{\beta}/n)^{1/2}$
$\text{SE}(\text{Var}\hat{\beta})$	$\sqrt{2/(n-1)} \text{Var}\hat{\beta}$
$\text{SE}(\widehat{\text{Var}}\beta)$	$n^{-1/2} \hat{\sigma}_{\widehat{\text{Var}}\beta}$

Tabelle 6.2: Legende der Symbole zur Simulation nach Liang et al. (1993).

Die Schätzung der Varianz von $\hat{\beta}$ ($\widehat{\text{Var}}\beta$)

Die Schätzung nach der Methode von (Cox 1972) kann durch $\theta = 3$ induzierten Abhängigkeiten zwischen Überlebenszeiten nicht berücksichtigen. Daher sind starke Verzerrungen dieses Schätzers bei abhängigem und invers abhängigem Design für $\theta = 3$ nicht unerwartet.

Der systematische Fehler $(\widehat{\text{Var}}\beta - \text{Var } \hat{\beta}) / \text{Var } \hat{\beta}$ beträgt bisweilen über 50%. Für $\theta = 1$ ist der systematische Fehler unter 10%.

Im Gegensatz dazu ist die Methode von Lee et al. (1992) auch für $\theta \neq 1$ einsetzbar. Der systematische Fehler beträgt bei zufälligen Regressoren und $\beta = 0$, $\theta = 1$, bei abhängigem und unabhängigem Design -6,6% und bei invers abhängigem Design -6,5%. Für $\theta = 3$ wird bei abhängigem Design sogar ein systematischer Fehler von -11,0% gefunden.

Auch beim Pseudolikelihood-Schätzer gibt es systematische Fehler von mehr als 10%. Für $\beta = 1$ erreicht dieser bei abhängigem Design, zufälligen Regressoren und $\theta = 1$ den Wert 11,2%. Für $\theta = 1$ erreicht er bei invers abhängigem Design für $\beta = 0$ und $\beta = 1$ jeweils den Wert 10,0%. Ein systematischer Fehler von 11,0% wurde für die Methode von (Liang et al. 1993) bei invers abhängigem Design, zufälligen Regressoren und $\beta = 1$, $\theta = 1$ gefunden.

Die Variabilität der Varianzschätzung von $\hat{\beta}$

Der Standardfehler $\hat{\sigma}_{\widehat{\text{Var}}\beta}$ ist für zufällige Regressoren generell größer als für fixe Regressoren. Für $\theta = 1$ ist er für die Schätzung nach (Cox 1972) am geringsten. Auch der Standardfehler der Methode nach (Lee et al. 1992) ist in allen Simulationszellen um etwa 20% geringer als bei den Methoden von (Liang et al. 1993) und der Pseudolikelihood-Schätzung. Die beiden letztgenannten Methoden liefern ähnliche Werte.

Die Konfidenzintervalle von $\hat{\beta}$

Zur Erleichterung der Interpretation der Ergebnisse wurden in Tabelle 6.3 Konfidenzintervalle für ausgewählte relative Häufigkeiten die für $n = 1000$ angegeben. Umgekehrt ist aufgrund der Binomialverteilung der kritische Bereich für $H_0 : 100 \cdot \pi = 95$ und $\alpha = 0,05$ gegeben durch das Komplement von $[93,5; 96,3]$ und für $H_0 : 100 \cdot \pi = 2,5$ und $\alpha = 0,05$ gegeben durch das Komplement von $[1,6; 3,4]$.

Tabelle 6.3: Ausgewählte 95%-Vertrauensbereiche für den Parameter 100π der Binomialverteilung für $n = 1000$ (nach Sachs (1997), p. 434 f.).

$100 \cdot \pi$	95%-Vertrauensber.
1	0,48 - 1,84
2	1,22 - 3,08
3	2,03 - 4,26
4	2,87 - 5,41
5	3,73 - 6,54
6	4,61 - 7,66
7	5,50 - 8,76
8	6,40 - 9,86
9	7,30 - 10,95
10	8,21 - 12,03
15	12,84 - 17,37

Die geschätzten Überdeckungswahrscheinlichkeiten des 95%-Vertrauensbereichs lagen bei zufälligen Regressoren nur bei der Methode von Cox (1972) in einer Simulationszelle (nur $\theta = 1$ wurde gewertet) im unteren Teil des kritischen Bereichs für $\alpha = 0,05$. Bei der Methode nach (Lee et al. 1992) war das in 4 Simulationszellen der Fall. Die Methode nach (Liang et al. 1993) und der Pseudolikelihood-Schätzer lagen jeweils nur mit einem Simulationsergebnis im oberen Teil des kritischen Bereichs.

Die einseitigen Überdeckungswahrscheinlichkeiten des kritischen Bereichs für $\alpha = 0,05$ lagen bei zufälligen Regressoren für die Methode von Cox (1972) (nur $\theta = 1$ wurde gewertet) niemals und die Methode nach (Lee et al. 1992) 5 mal im oberen Teil des kritischen Bereichs. Die Methode nach (Liang et al. 1993) lag einmal im unteren (für die Untergrenze) Teil des kritischen Bereichs und der Pseudolikelihood-Schätzer lag einmal im oberen und einmal im unteren Teil des kritischen Bereichs.

6.4.3 Vergleich mit dem Ergebnis von (Liang et al. 1993)

Simulationen in (Liang et al. 1993) wurden mit fixen Regressoren gerechnet. Die Beschreibung schließt aber ein zufällige Regressoren nicht ausdrücklich aus. Der Vektor der Regressoren stammt aus der selben Gleichverteilung

wie in unserer Studie. Allerdings wurde nicht angegeben, wie die Pseudozufallszahlen generiert wurden. Deshalb sind Abweichungen von unseren Ergebnissen zu erwarten.

In der Zelle *Design 1* (das entspricht unabhängigem Design), $\beta = 0$ und $\theta = 1$, findet man für den Mittelwert der $\hat{\beta}$ den Wert -0,018. In der entsprechenden Zelle unserer Simulation (Tabelle 6.6 unter Schätzmethode (Liang et al. 1993), zufällige Regressoren) findet man den Wert 0,020. Unter Verwendung des Standardfehlers von $\hat{\beta}$ (0,0123)¹ errechnet man ein 95%-Konfidenzintervall der Differenz 0,038 der Mittelwerte von [0,0038; 0,072]. Der Unterschied ist also knapp signifikant. Ähnliche Ergebnisse findet man in vielen anderen Simulationszellen.

Bei Design 3 (invers abhängiges Design) sind die Unterschiede viel größer. Hier ist für $\theta = 1$, $\beta = 0$, der Quotient der Varianzen von $\hat{\beta}$ ($\text{Var}\hat{\beta}$) aus den beiden Simulationsstudien $0.039/0.155 = 0,25$ (zufällige Regressoren). Für $\theta = 3$, $\beta = 1$ sind die entsprechenden Werte des Quotienten $0,0024/0,0058=0,41$. Die Differenz beträgt 0,034, der Standardfehler ist 0,0026 und das Konfidenzintervall ist [0,026; 0,042].

Die oben angeführten Resultate aus (Liang et al. 1993) decken sich also nicht mit unseren Resultaten. Das unterschiedliche Design ist in Hinblick auf die übrigen Resultate, die weitaus ähnlicher sind, keine plausible Erklärung.

Ein Indiz für einen Fehler in den Resultaten von (Liang et al. 1993) ist folgendes: Für $\theta = 1$ sollten sich die Varianzen der 3 Designs nur durch den Simulationsfehler unterscheiden, da die entsprechenden Modelle die selben sind. Die Ergebnisse der entsprechenden Simulationszellen in unserer Studie sind für alle 3 Designs ähnlich, in der Größenordnung 0,15. (Liang et al. 1993) geben in Tabelle I, p.449 als Varianzen in Design 3 die extrem niedrigen Werte von 0,039 und 0,046 an, die nur circa 25% bis 28% der Varianzen in den Designs 1 (0,149, 0,159) und 2 (0,154, 0,168) betragen. Es ist daher anzunehmen, daß für Design 3 offensichtlich falsche Werte angegeben wurden.

¹Hier wird angenommen, daß die Zelle in beiden Studien den gleichen Standardfehler hat.

Tabelle 6.4: Simulation von 1000 Stichproben mit korrelierten Überlebenszeiten mit $\beta = \mathbf{0}$ mit **fixen** Regressoren. Schätzung nach Cox (1972), Lee et al. (1992), mit der Pseudolikelihoodmethode und nach Liang et al (1993). Siehe Legende in Tabelle 6.2.

Design	θ	Schätzer	$\hat{\beta}$	$\widehat{\text{Var}}\hat{\beta}$	$\widehat{\sigma}_{\widehat{\text{Var}}\hat{\beta}}$	KL _{95%}	UG _{95%}	OG _{95%}	Zens.	SE($\hat{\beta}$)	SE($\widehat{\text{Var}}\hat{\beta}$)	SE($\widehat{\sigma}_{\widehat{\text{Var}}\hat{\beta}}$)	
unabh.	1	Cox72	0,022	0,145	0,156	0,013	95,2	1,7	3,1	15,1	0,0120	0,0065	0,0004
		Lee92	0,022	0,145	0,152	0,029	94,7	2,0	3,3	15,1	0,0120	0,0065	0,0009
		PL	0,020	0,151	0,172	0,033	96,1	1,7	2,2	15,1	0,0123	0,0068	0,0010
		Liang93	0,022	0,148	0,165	0,033	95,7	1,7	2,6	15,1	0,0122	0,0066	0,0010
3		Cox72	0,014	0,162	0,156	0,014	94,9	2,4	2,7	15,1	0,0127	0,0073	0,0004
		Lee92	0,014	0,162	0,153	0,032	93,7	3,0	3,3	15,1	0,0127	0,0073	0,0010
		PL	0,011	0,168	0,178	0,039	95,1	2,3	2,6	15,1	0,0129	0,0075	0,0012
		Liang93	0,012	0,165	0,172	0,039	94,4	2,5	3,1	15,1	0,0128	0,0074	0,0012
abh.	1	Cox72	-0,011	0,151	0,141	0,011	94,2	2,7	3,1	14,9	0,0123	0,0068	0,0004
		Lee92	-0,011	0,151	0,133	0,031	92,4	4,0	3,6	14,9	0,0123	0,0068	0,0010
		PL	-0,009	0,149	0,148	0,035	94,6	2,5	2,9	14,9	0,0122	0,0067	0,0011
		Liang93	-0,012	0,151	0,148	0,035	94,1	2,6	3,3	14,9	0,0123	0,0068	0,0011
3		Cox72	0,006	0,236	0,142	0,015	87,6	5,9	6,5	15,1	0,0154	0,0106	0,0005
		Lee92	0,006	0,236	0,238	0,047	94,0	2,7	3,3	15,1	0,0154	0,0106	0,0015
		PL	0,009	0,238	0,270	0,057	95,3	2,0	2,7	15,1	0,0154	0,0106	0,0018
		Liang93	0,006	0,239	0,265	0,057	95,1	2,2	2,7	15,1	0,0154	0,0107	0,0018
invers	1	Cox72	0,002	0,140	0,138	0,011	94,9	2,9	2,2	14,8	0,0118	0,0063	0,0003
		Lee92	0,002	0,140	0,132	0,030	94,0	3,1	2,9	14,8	0,0118	0,0063	0,0010
		PL	0,002	0,150	0,154	0,036	94,9	2,5	2,6	14,8	0,0122	0,0067	0,0011
		Liang93	0,001	0,141	0,141	0,035	94,4	2,8	2,8	14,8	0,0119	0,0063	0,0011
3		Cox72	0,011	0,034	0,138	0,014	100,0	0,0	0,0	15,0	0,0059	0,0015	0,0005
		Lee92	0,011	0,034	0,035	0,014	94,9	2,0	3,1	15,0	0,0059	0,0015	0,0004
		PL	0,011	0,038	0,042	0,016	95,9	1,6	2,5	15,0	0,0061	0,0017	0,0005
		Liang93	0,010	0,033	0,036	0,016	96,3	1,7	2,0	15,0	0,0057	0,0015	0,0005

Tabelle 6.5: Simulation von 1000 Stichproben mit korrelierten Überlebenszeiten mit $\beta = 1$ mit **fixen** Regressoren. Schätzung nach Cox (1972), Lee et al. (1992), mit der Pseudolikelihoodmethode und nach Liang et al (1993). Siehe Legende in Tabelle 6.2.

Design	θ	Schätzer	$\hat{\beta}$	$\widehat{\text{Var}}\hat{\beta}$	$\widehat{\sigma}_{\widehat{\text{Var}}\hat{\beta}}$	KL _{95%}	UG _{95%}	OG _{95%}	Zens.	SE($\hat{\beta}$)	SE(Var $\hat{\beta}$)	SE($\widehat{\text{Var}}\hat{\beta}$)	
unabh.	1	Cox72	1,011	0,170	0,164	0,016	95,0	3,1	1,9	14,5	0,0130	0,0076	0,0005
		Lee92	1,011	0,170	0,158	0,032	94,4	3,2	2,4	14,5	0,0130	0,0076	0,0010
		PL	1,030	0,180	0,185	0,038	95,2	2,9	1,9	14,5	0,0134	0,0080	0,0012
		Liang93	1,018	0,173	0,177	0,037	95,1	3,1	1,8	14,5	0,0132	0,0077	0,0012
3	Cox72	1,010	0,167	0,163	0,016	95,3	2,0	2,7	14,5	0,0129	0,0075	0,0005	
		Lee92	1,010	0,167	0,166	0,038	94,4	2,5	3,1	14,5	0,0129	0,0075	0,0012
		PL	1,041	0,172	0,196	0,046	95,9	1,2	2,9	14,5	0,0131	0,0077	0,0014
		Liang93	1,000	0,168	0,187	0,044	95,6	2,0	2,4	14,5	0,0129	0,0075	0,0014
abh.	1	Cox72	1,002	0,164	0,148	0,014	93,9	3,5	2,6	16,1	0,0128	0,0073	0,0004
		Lee92	1,002	0,164	0,140	0,035	92,3	4,3	3,4	16,1	0,0128	0,0073	0,0011
		PL	0,997	0,162	0,158	0,042	93,8	3,7	2,5	16,1	0,0127	0,0072	0,0013
		Liang93	1,004	0,166	0,158	0,042	93,9	3,5	2,6	16,1	0,0129	0,0074	0,0013
3	Cox72	1,030	0,265	0,149	0,018	86,0	7,0	7,0	16,0	0,0163	0,0119	0,0006	
		Lee92	1,030	0,265	0,243	0,055	93,4	2,5	4,1	16,0	0,0163	0,0119	0,0017
		PL	1,033	0,266	0,282	0,065	95,4	1,9	2,7	16,0	0,0163	0,0119	0,0021
		Liang93	1,033	0,267	0,278	0,065	95,5	2,0	2,5	16,0	0,0164	0,0120	0,0020
invers	1	Cox72	1,003	0,148	0,144	0,013	94,9	2,8	2,3	15,0	0,0122	0,0066	0,0004
		Lee92	1,003	0,148	0,140	0,034	93,4	3,5	3,1	15,0	0,0122	0,0066	0,0011
		PL	1,050	0,167	0,171	0,042	94,9	2,4	2,7	15,0	0,0129	0,0075	0,0013
		Liang93	1,012	0,154	0,155	0,040	94,5	3,4	2,1	15,0	0,0124	0,0069	0,0013
3	Cox72	1,023	0,053	0,145	0,015	99,9	0,0	0,1	15,2	0,0073	0,0024	0,0005	
		Lee92	1,023	0,053	0,052	0,017	94,1	3,6	2,3	15,2	0,0073	0,0024	0,0005
		PL	1,078	0,062	0,068	0,023	96,3	1,6	2,1	15,2	0,0079	0,0028	0,0007
		Liang93	0,997	0,057	0,060	0,021	94,3	4,6	1,1	15,2	0,0075	0,0025	0,0007

Tabelle 6.6: Simulation von 1000 Stichproben mit korrelierten Überlebenszeiten mit $\beta = \mathbf{0}$ mit zufälligen Regressoren. Schätzung nach Cox (1972), Lee et al. (1992), mit der Pseudolikelihoodmethode und nach Liang et al (1993). Siehe Legende in Tabelle 6.2.

Design	θ	Schätzer	$\hat{\beta}$	$\widehat{\text{Var}}\hat{\beta}$	$\hat{\sigma}_{\widehat{\text{Var}}\hat{\beta}}$	KL _{95%}	UG _{95%}	OG _{95%}	Zens.	SE($\hat{\beta}$)	SE($\widehat{\text{Var}}\hat{\beta}$)	SE($\widehat{\text{Var}}\hat{\beta}$)	
unabh.	1	Cox72	0,020	0,151	0,146	0,018	94,5	2,6	2,9	15,0	0,0123	0,0068	0,0006
		Lee92	0,020	0,151	0,141	0,031	93,8	3,0	3,2	15,0	0,0123	0,0068	0,0010
		PL	0,020	0,158	0,162	0,036	94,7	2,6	2,7	15,0	0,0126	0,0071	0,0011
		Liang93	0,020	0,154	0,156	0,036	94,6	2,5	2,9	15,0	0,0124	0,0069	0,0011
3		Cox72	0,011	0,145	0,146	0,019	95,6	2,1	2,3	14,7	0,0121	0,0065	0,0006
		Lee92	0,011	0,145	0,140	0,038	94,4	2,8	2,8	14,7	0,0121	0,0065	0,0012
		PL	0,011	0,150	0,163	0,045	95,1	2,5	2,4	14,7	0,0122	0,0067	0,0014
		Liang93	0,010	0,146	0,157	0,044	95,2	2,2	2,6	14,7	0,0121	0,0065	0,0014
abh.	1	Cox72	-0,004	0,164	0,150	0,025	93,0	3,6	3,4	15,0	0,0128	0,0073	0,0008
		Lee92	-0,004	0,164	0,141	0,039	91,5	4,5	4,0	15,0	0,0128	0,0073	0,0012
		PL	-0,003	0,163	0,157	0,043	93,9	3,6	2,5	15,0	0,0128	0,0073	0,0014
		Liang93	-0,004	0,165	0,157	0,043	93,4	3,8	2,8	15,0	0,0128	0,0074	0,0014
3		Cox72	0,005	0,283	0,150	0,026	85,5	7,0	7,5	14,8	0,0168	0,0127	0,0008
		Lee92	0,005	0,283	0,252	0,061	93,8	2,6	3,6	14,8	0,0168	0,0127	0,0019
		PL	0,003	0,281	0,287	0,073	94,8	2,3	2,9	14,8	0,0168	0,0126	0,0023
		Liang93	0,003	0,282	0,282	0,072	94,9	2,3	2,8	14,8	0,0168	0,0126	0,0023
invers	1	Cox72	0,009	0,154	0,147	0,024	94,7	2,4	2,9	15,1	0,0124	0,0069	0,0007
		Lee92	0,009	0,154	0,144	0,037	93,7	3,1	3,2	15,1	0,0124	0,0069	0,0012
		PL	0,010	0,165	0,166	0,043	95,3	2,5	2,2	15,1	0,0128	0,0074	0,0014
		Liang93	0,010	0,155	0,153	0,041	94,4	2,6	3,0	15,1	0,0125	0,0069	0,0013
3		Cox72	0,007	0,036	0,146	0,026	100,0	0,0	0,0	15,0	0,0060	0,0016	0,0008
		Lee92	0,007	0,036	0,037	0,016	94,6	2,0	3,4	15,0	0,0060	0,0016	0,0005
		PL	0,007	0,040	0,044	0,018	95,7	1,5	2,8	15,0	0,0063	0,0018	0,0006
		Liang93	0,006	0,034	0,037	0,017	95,3	1,5	3,2	15,0	0,0058	0,0015	0,0005

Tabelle 6.7: Simulation von 1000 Stichproben mit korrelierten Überlebenszeiten mit $\beta = 1$ mit **zufälligen** Regressoren. Schätzung nach Cox (1972), Lee et al. (1992), mit der Pseudolikelihoodmethode und nach Liang et al (1993). Siehe Legende in Tabelle 6.2.

Design	θ	Schätzer	$\hat{\beta}$	$\widehat{\text{Var}}\hat{\beta}$	$\widehat{\sigma}_{\widehat{\text{Var}}\hat{\beta}}$	KL _{95%}	UG _{95%}	OG _{95%}	Zens.	SE($\hat{\beta}$)	SE($\widehat{\text{Var}}\hat{\beta}$)	SE($\widehat{\text{Var}}\hat{\beta}$)	
unabh.	1	Cox72	1,018	0,164	0,153	0,020	94,8	2,3	2,9	15,1	0,0128	0,0073	0,0006
		Lee92	1,018	0,164	0,147	0,033	93,9	2,6	3,5	15,1	0,0128	0,0073	0,0011
		PL	1,037	0,169	0,173	0,039	95,7	1,6	2,7	15,1	0,0130	0,0075	0,0012
		Liang93	1,022	0,165	0,164	0,038	95,3	2,0	2,7	15,1	0,0128	0,0074	0,0012
3	Cox72	1,013	0,166	0,154	0,020	94,9	2,3	2,8	15,2	0,0129	0,0074	0,0006	
		Lee92	1,013	0,166	0,148	0,040	93,2	3,0	3,8	15,2	0,0129	0,0074	0,0013
		PL	1,045	0,173	0,179	0,049	94,8	1,9	3,3	15,2	0,0131	0,0077	0,0016
		Liang93	1,004	0,168	0,171	0,048	94,8	2,7	2,5	15,2	0,0130	0,0075	0,0015
abh.	1	Cox72	1,024	0,155	0,157	0,026	94,6	3,0	2,4	15,1	0,0125	0,0069	0,0008
		Lee92	1,024	0,155	0,148	0,040	93,5	3,4	3,1	15,1	0,0125	0,0069	0,0013
		PL	1,018	0,152	0,169	0,049	94,4	3,1	2,5	15,1	0,0123	0,0068	0,0016
		Liang93	1,029	0,155	0,169	0,049	94,8	2,7	2,5	15,1	0,0125	0,0070	0,0016
3	Cox72	1,033	0,306	0,158	0,029	84,7	7,6	7,7	15,1	0,0175	0,0137	0,0009	
		Lee92	1,033	0,306	0,257	0,067	92,5	3,8	3,7	15,1	0,0175	0,0137	0,0021
		PL	1,035	0,305	0,300	0,086	94,5	3,3	2,2	15,1	0,0175	0,0136	0,0027
		Liang93	1,038	0,306	0,295	0,086	94,4	3,3	2,3	15,1	0,0175	0,0137	0,0027
invers	1	Cox72	1,013	0,142	0,153	0,024	96,7	1,7	1,6	14,9	0,0119	0,0064	0,0008
		Lee92	1,013	0,142	0,147	0,040	95,6	1,6	2,8	14,9	0,0119	0,0064	0,0013
		PL	1,054	0,156	0,177	0,049	96,4	1,2	2,4	14,9	0,0125	0,0070	0,0015
		Liang93	1,018	0,145	0,161	0,045	96,5	1,6	1,9	14,9	0,0120	0,0065	0,0014
3	Cox72	1,023	0,056	0,154	0,026	99,5	0,0	0,5	15,4	0,0075	0,0025	0,0008	
		Lee92	1,023	0,056	0,055	0,019	94,6	3,0	2,4	15,4	0,0075	0,0025	0,0006
		PL	1,080	0,064	0,070	0,025	94,9	1,6	3,5	15,4	0,0080	0,0029	0,0008
		Liang93	0,999	0,058	0,063	0,022	94,4	3,9	1,7	15,4	0,0076	0,0026	0,0007

6.5 Vergleich der Gütefunktionen

6.5.1 Plan der Simulationsstudie

Das Vorbild für diese Simulationsstudie ist (Lee et al. 1992). Wiederum ist die Vergleichbarkeit der Resultate das Motiv für die Wahl des Designs. Allerdings sind die positiv stabilen Verteilungen als Mischverteilung allein durch die Eigenschaft der Erhaltung der proportionalen Hasardraten der bedingten Verteilung in der marginalen Verteilung ausgezeichnet. Diese Eigenschaft wird aber nicht benötigt. Schließlich ist der Zusammenhangsparameter α nicht anschaulich interpretierbar.

Stattdessen ist die bivariate Normalverteilung eine anschauliche Grundlage für den Zusammenhang der Überlebenszeiten. Auf die Angabe der Verteilungsfunktion wie in (6.4.1) wird verzichtet.

In jeder Simulationszelle werden 1000 Datensätze von jeweils 50 Paaren von Überlebenszeiten nach dem in 6.1.3 beschriebenen Modell (Methode von Loughin und Koehler (1997)) erzeugt. In diesem Modell ist der Zusammenhangsparameter zugleich der Kendall'sche Korrelationskoeffizient τ aus (6.1.20).

Der Designvektor besteht aus 2 Komponenten. Die zweite Komponente des Designvektors hat die Werte 0 und 1 gleich oft und wird orthogonal zur ersten Komponente gewählt. Da die zweite Komponente als Kovariable aufgefaßt wird, wird deren Schätzung nicht ausgewertet. Die erste Komponente hat ebenfalls nur die Werte 0 oder 1.

Betrachtet man nur die erste Komponente des Designvektors, so sollte bei *unabhängigem* Design jeder Wert an der zweiten Stelle des 2-Tupels (Paares) mit jedem Wert an der ersten Stelle gleich oft vorkommen. Das ist mit dem gewählten Stichprobenumfang nicht durchführbar, da 50 nicht durch 4 teilbar ist. Allerdings ist die Schätzung invariant bezüglich Vertauschung der Komponenten in den 2-Tupeln, sodaß alle ungeordneten Paare von ersten Komponenten des Designvektors gleich oft vorkommen. Die gleiche Bedingung ist für die Kombinationen der zweiten Komponente des Designvektors erfüllt. Beide Komponenten des Designvektors zusammen erfüllen diese Bedingung nicht. Konkret werden in der ersten und zweiten Komponente des 2-Tupels folgende Sequenzen von Designvektoren verwendet:

```
;Erste Komponente
'(#(0 0) #(0 0) #(1 0) #(1 0) #(0 1) #(0 1) #(1 1) #(1 1))
;Zweite Komponente
'(#(0 0) #(0 1) #(0 0) #(0 1) #(1 1) #(1 0) #(1 1) #(1 0))
```

Die Sequenzen werden aneinandergereiht und jeweils die ersten 50 Vektoren verwendet. Bei *abhängigem* Design stimmen die beiden Designvektoren eines 2-Tupels überein. Bei *invers abhängigem* Design ist der Designvektor der zweiten Komponente die Differenz von $\mathbf{1} \# (1 \ 1)$ und dem Designvektor der ersten Komponente.

Der Vektor β der Regressionsparameter ist in der zweiten Komponente $\beta_2 = 1$. β_1 wird in Schritten der Länge 0,2 von 0 bis 0,8 variiert.

Die Überlebenszeiten werden entweder gar nicht oder zu festen Zeiten zensiert. Dabei soll jeweils ein gewisser Anteil von zensierten Beobachtungen erreicht werden. Zur Bestimmung dieser Zensierungszeiten werden am Beginn der Simulation mit dem Initialisierungskeim auf Seite 142 1000 Paare von Überlebenszeiten erzeugt und für diese das entsprechende Quantil bestimmt.²

Das Schema der Simulation ist in Tabelle 6.8 zu finden.

6.5.2 Ergebnisse

Der systematische Fehler des Regressionskoeffizienten

Die Ergebnisse sind in Abbildung 6.2 bis 6.4 dargestellt.

Der systematische Fehler von $\hat{\beta}_1$ steigt in fast allen Simulationen mit β_1 an. Er beträgt maximal 0,0380 für die Methode nach (Liang et al. 1993) in der Zelle mit invers abhängigem Design ohne Zensierung und mit $\tau = 0,9$, $\beta_1 = 0,8$. Das ist wenig im Vergleich zur minimalen Standardabweichung von $\hat{\beta}_1$ über alle Simulationszellen von 0,041, die in der selben Simulationszelle gefunden wurde. Extrapoliert man den Verlauf der Kurven des systematischen Fehlers jedoch auf den nicht untersuchten Bereich $\beta > 0,8$, so sind viel größere systematische Fehler zu vermuten.

In der Darstellung des systematischen Fehlers als Funktion von β_1 hat der Schätzer von Cox (1972) meist einen etwas flacheren Anstieg.

Die Ergebnisse hängen wenig vom Ausmaß der Zensierung und von der Korrelation ab. Eine diesbezügliche Ausnahme sind der Pseudolikelihoodschätzer und der Schätzer nach (Liang et al. 1993) in den Zellen ohne Zensierung und mit Korrelation $\tau = 0,9$ bei invers abhängigem Design. Während der systematische Fehler des Schätzers nach (Cox 1972) besonders niedrig ist, nimmt

²Die Zensierungszeit hätte auch durch Inversion der marginalen Verteilungsfunktion berechnet werden können.

Design	unabhängig abhängig invers
Korrelationsparameter	$\tau = 0,0$ $\tau = 0,5$ $\tau = 0,9$
Zensierung	Keine Zensierung 30% Zensierung 50% Zensierung
Regressionsparameter	$\beta_1 = 0,0$ $\beta_1 = 0,2$... $\beta_1 = 0,8$
Schätzmethode	Cox (1972) Lee et al. (1992) (nur Varianzschätzung) Paarweise Pseudolikelihood Liang et al. (1993)

Tabelle 6.8: Übersicht über die Simulation nach Lee et al. (1992).

der systematische Fehler der beiden anderen Schätzmethoden für $\beta_1 = 0,8$ jeweils das Maximum über alle Simulationszellen an.

Die Varianz von $\hat{\beta}_1$

Die Ergebnisse sind in Abbildung 6.5 bis 6.7 dargestellt.

Die Varianz von $\hat{\beta}_1$ ist weitgehend unabhängig von β_1 (Ausnahmen siehe weiter unten). Die Varianz von $\hat{\beta}_1$ steigt mit dem Anteil der Zensierungen (annähernd proportional zum Anteil der beobachteten Ereignisse).

Bei unabhängigem Design ist die Varianz von $\hat{\beta}_1$ weitgehend unabhängig von der Korrelation τ . Bei abhängigem Design steigt die Varianz mit der Korrelation τ . Gleichzeitig ist für $\tau = 0,9$ und 50% Zensierung eine Abhängigkeit von der Varianz von β festzustellen. Bei inversem Design nimmt die Varianz von $\hat{\beta}_1$ mit der Korrelation τ ab. Auch hier sind Simulationszellen vorhanden, die stark von β_1 abhängen. Es ist dies (Cox 1972) für $\tau = 0,9$ ohne Zensierung. Die Varianz von $\hat{\beta}_1$ erreicht hier das Minimum über alle Zellen mit $V_{\hat{\beta}_1} = 0,0016$.

Die Varianz der Schätzung von $\hat{\beta}$ ist bei der Pseudolikelihoodschätzung und bei der Methode nach (Liang et al. 1993) fast identisch, wenn Zensierung vorhanden ist. Wenn keine zensierten Beobachtungen vorhanden sind, hat der Schätzer nach (Liang et al. 1993) vor allem für kleines β_1 etwas größere Varianz als der Pseudolikelihoodschätzer. Der Schätzer nach (Cox 1972) hat für unabhängiges und abhängiges Design durchwegs niedrigere Varianz als die beiden zuvor behandelten Schätzer. Der Vorteil gegenüber dem Mittelwert aus der Methode nach (Liang et al. 1993) und der Pseudolikelihoodschätzung beträgt für abhängiges Design 5,8% bis 13,2% und für unabhängiges Design 5,9% bis 10,1%. Für invers abhängiges Design ist die Varianz der Schätzung von $\hat{\beta}_1$ nach (Cox 1972) um 6,2% niedriger bis 15,6% höher als bei den anderen beiden Schätzmethoden.

Systematischer Fehler der Varianzschätzung von $\hat{\beta}_1$

Die Ergebnisse sind in Abbildung 6.8 bis 6.10 dargestellt.

Der systematische Fehler von $\hat{V}_{\hat{\beta}_1}$ ist bei unabhängigem Design sehr gering (mit Maximum 0,0138). Bei abhängigem Design ist der systematische Fehler der Varianzschätzung nach (Cox 1972) für paarweise korrelierte Überlebenszeiten nicht konsistent. Die Schätzung der Varianz von $\hat{\beta}_1$ nach

(Lee et al. 1992) ist dann besonders gut, wenn zensierte Beobachtungen vorhanden sind. Bei invers abhängigem Design ist der systematische Fehler durchwegs gering.

Die Gütefunktion des Wald-Tests

Die Ergebnisse sind in Abbildung 6.11 bis 6.13 dargestellt. Die Unterschiede zwischen den konsistenten Schätzern sind gering. Wenn Zensierung vorhanden ist, dann haben Pseudolikelihoodschätzer und der Schätzer nach (Liang et al. 1993) durchwegs etwas geringere Güte als der Schätzer nach (Lee et al. 1992). Wenn keine Zensierung vorhanden ist, dann sind die Ergebnisse deutlich schlechter als beim Schätzer nach (Lee et al. 1992).

Abbildung 6.2: Der **systematische Fehler von $\hat{\beta}_1$** in Abhängigkeit von der Kendall'schen Korrelation τ der Überlebenszeiten und von der Zensierung bei **unabhängigem** Design. (\circ Cox (1972) und Lee et al. (1992), \heartsuit Pseudolikelihood, \diamond Liang et al. (1993)). Die Fehlerbalken geben die asymptotischen 95%-Konfidenzintervalle an ($n = 1000$).

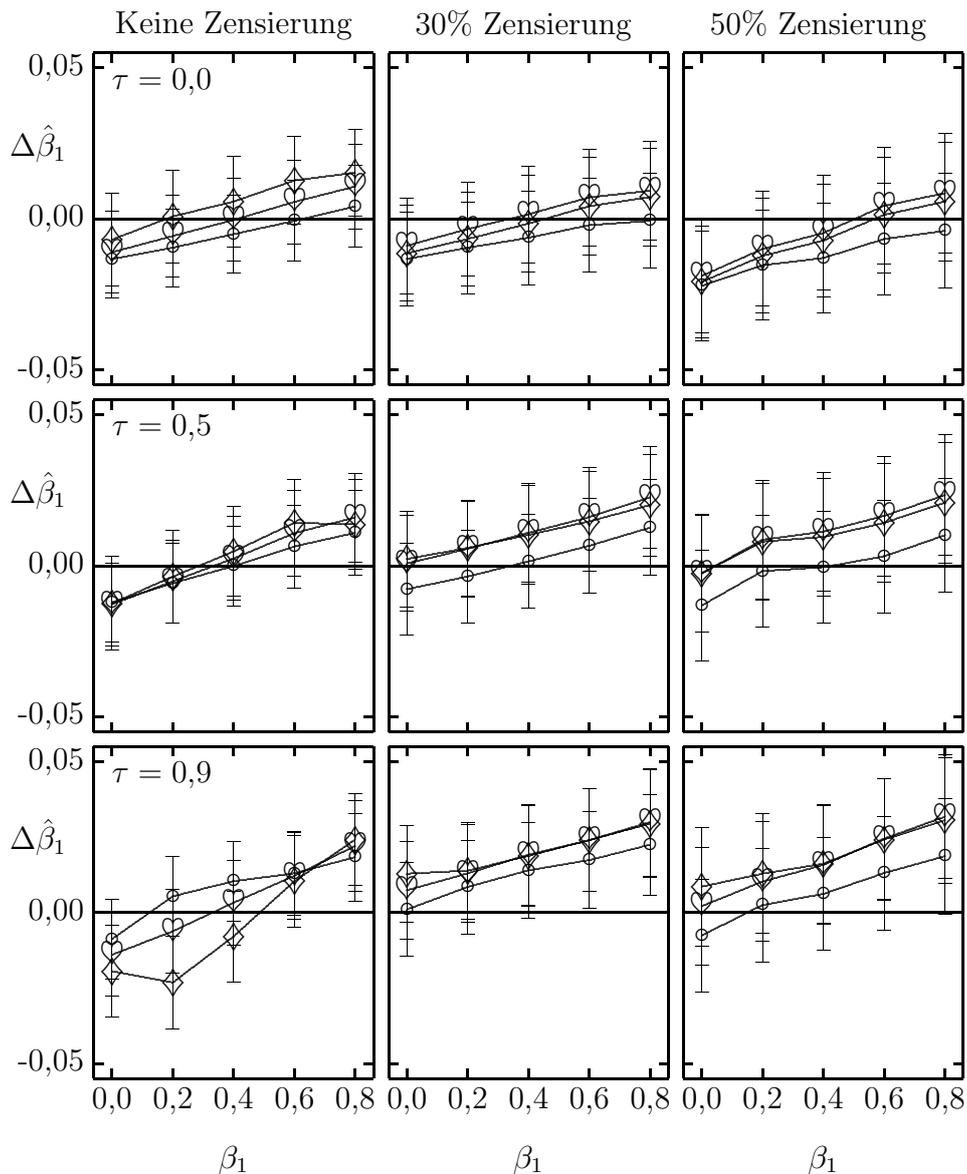


Abbildung 6.3: Der **systematische Fehler von $\hat{\beta}_1$** in Abhängigkeit von der Kendall'schen Korrelation τ der Überlebenszeiten und von der Zensierung bei **abhängigem** Design. (\circ Cox (1972) und Lee et al. (1992), \heartsuit Pseudolikehood, \diamond Liang et al. (1993)). Die Fehlerbalken geben die asymptotischen 95%-Konfidenzintervalle an ($n = 1000$).

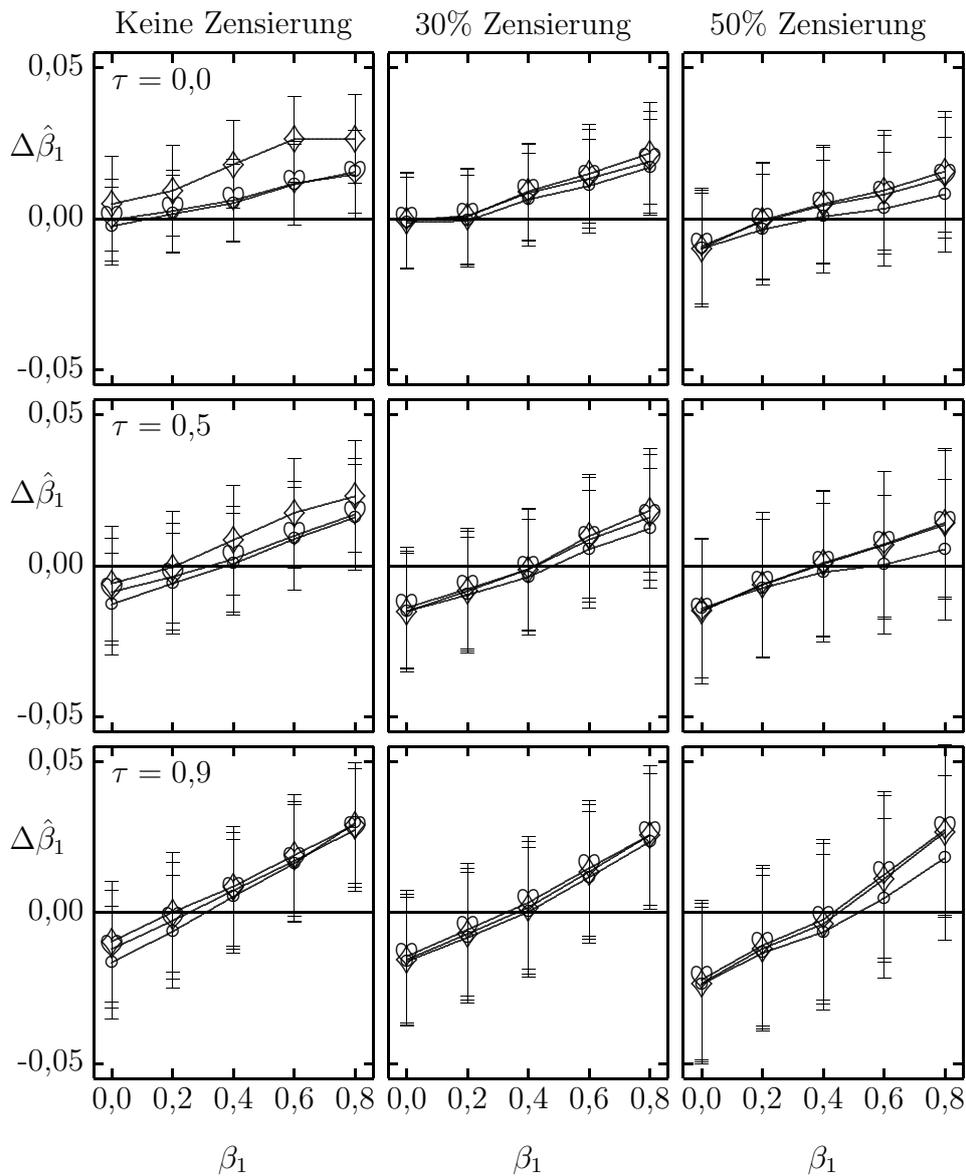


Abbildung 6.4: Der **systematische Fehler** von $\hat{\beta}_1$ in Abhängigkeit von der Kendall'schen Korrelation τ der Überlebenszeiten und von der Zensierung bei **inversem** Design. (\circ Cox (1972) und Lee et al. (1992), \heartsuit Pseudolikelihood, \diamond Liang et al. (1993)). Die Fehlerbalken geben die asymptotischen 95%-Konfidenzintervalle an ($n = 1000$).

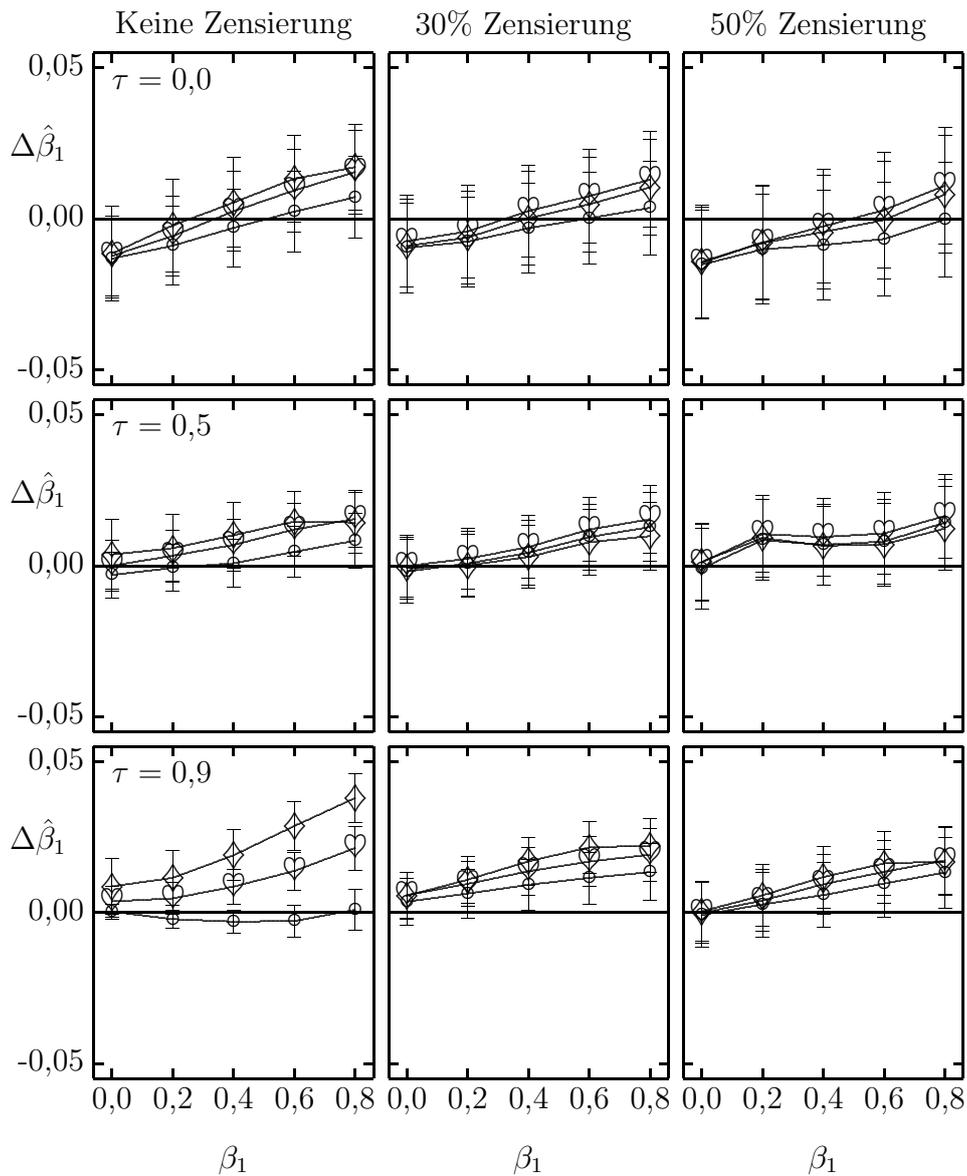


Abbildung 6.5: Die **Varianz von $\hat{\beta}_1$** in Abhängigkeit Kendall'schen Korrelation τ der Überlebenszeiten und von der Zensierung bei **unabhängigem** Design. (\circ Cox (1972), \heartsuit Pseudolikelihood, \diamond Liang et al. (1993)). Die Fehlerbalken geben die asymptotischen 95%-Konfidenzintervalle an ($n = 1000$).

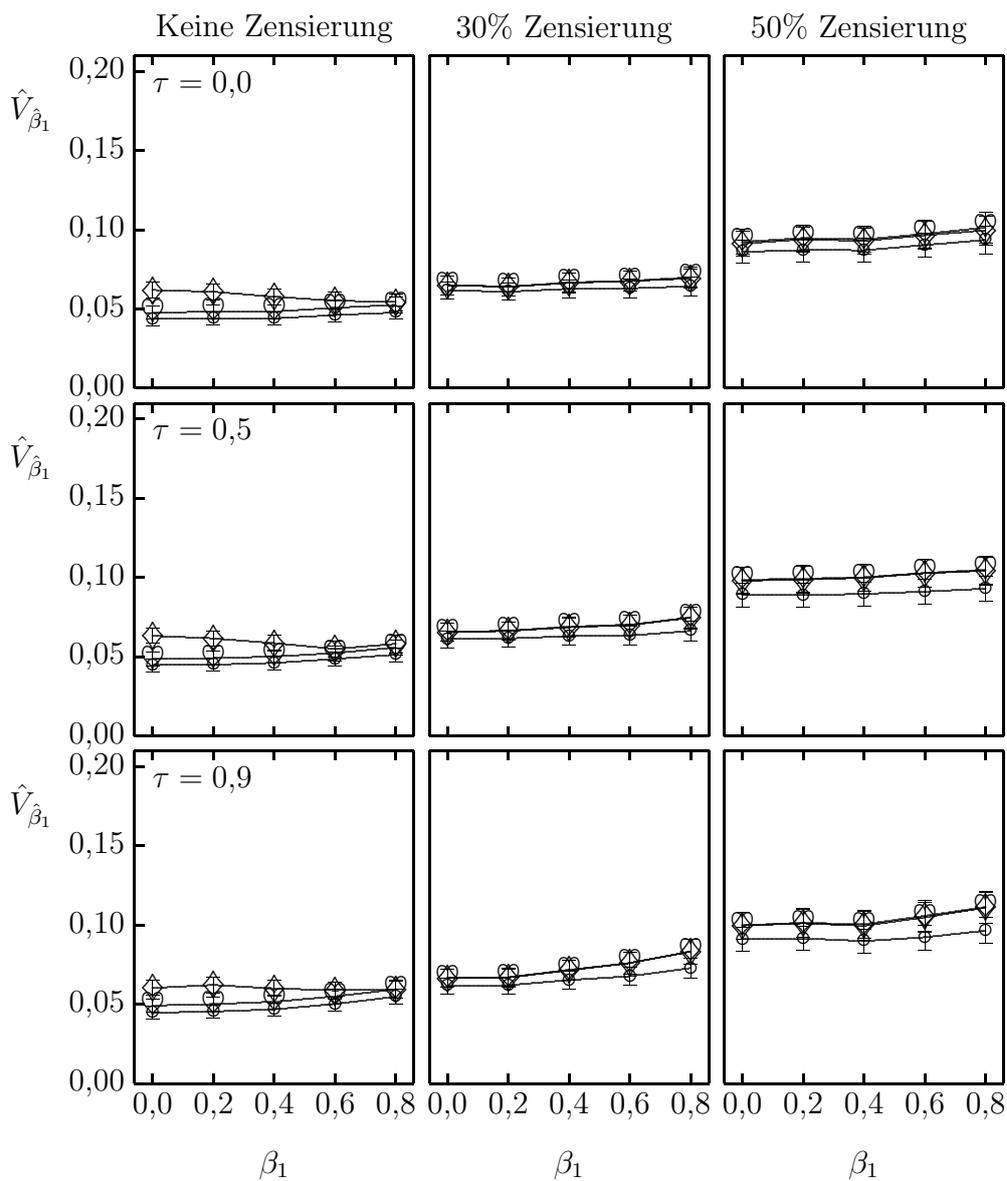


Abbildung 6.6: Die **Varianz von $\hat{\beta}_1$** in Abhängigkeit Kendall'schen Korrelation τ der Überlebenszeiten und von der Zensierung bei **abhängigem** Design. (\circ Cox (1972), \heartsuit Pseudolikelihood, \diamond Liang et al. (1993)). Die Fehlerbalken geben die asymptotischen 95%-Konfidenzintervalle an ($n = 1000$).

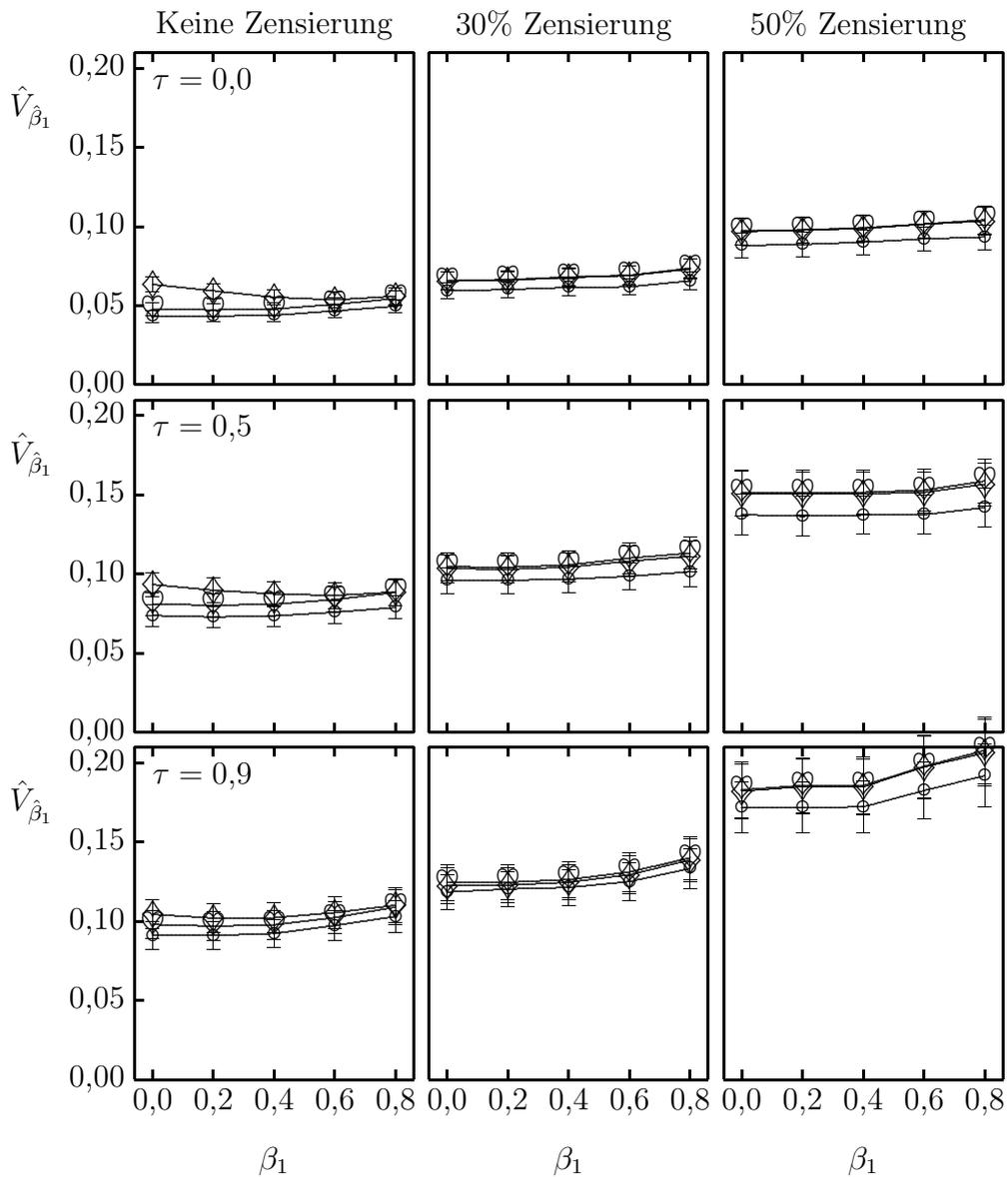


Abbildung 6.7: Die **Varianz von $\hat{\beta}_1$** in Abhängigkeit Kendall'schen Korrelation τ der Überlebenszeiten und von der Zensierung bei **inversem** Design. (\circ Cox (1972), \heartsuit Pseudolikelihood, \diamond Liang et al. (1993)). Die Fehlerbalken geben die asymptotischen 95%-Konfidenzintervalle an ($n = 1000$).

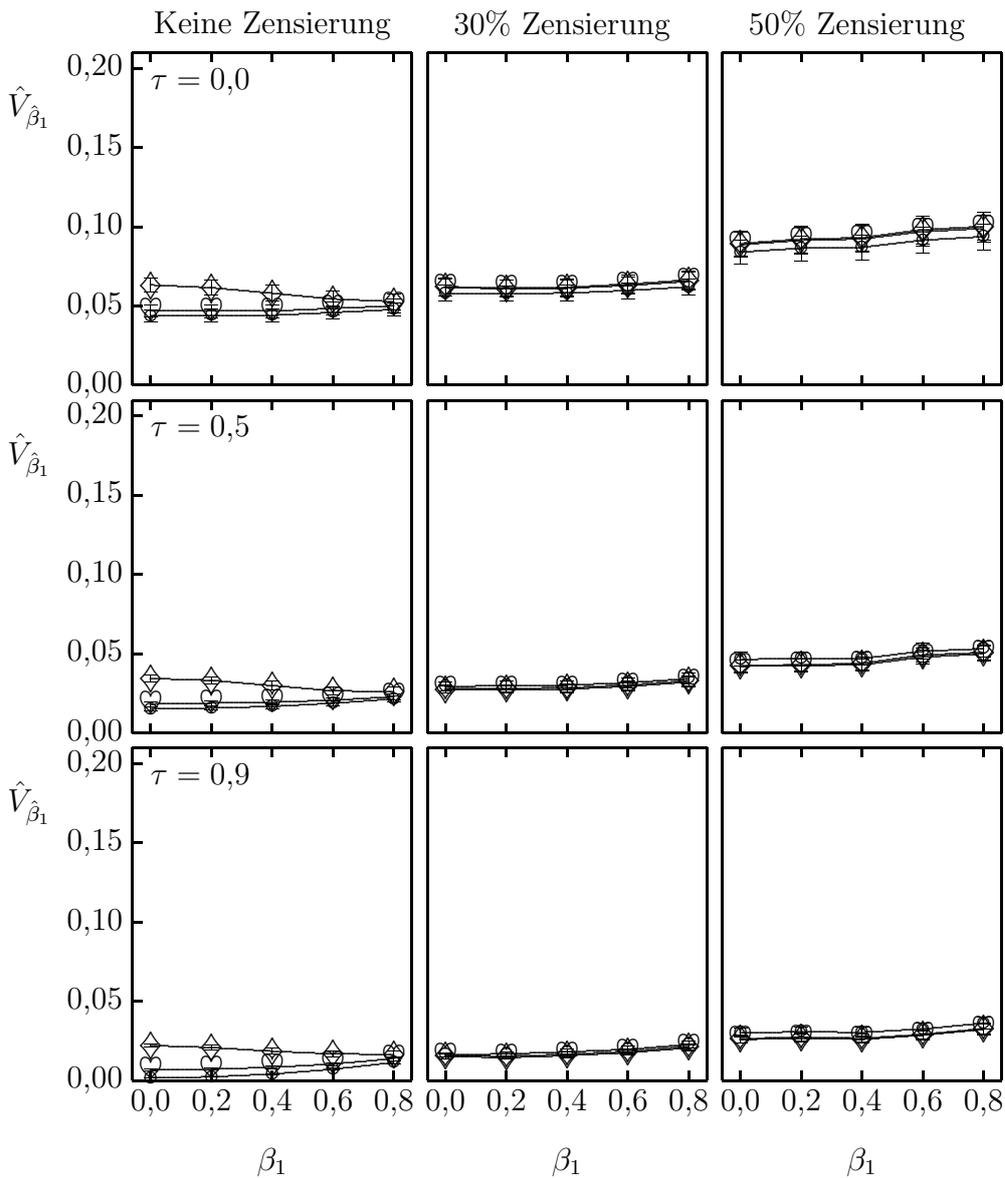


Abbildung 6.8: Der **systematische Fehler von $\hat{V}_{\hat{\beta}_1}$** in Abhängigkeit von der Kendall'schen Korrelation τ der Überlebenszeiten und von der Zensierung bei **unabhängigem** Design. (\circ Cox (1972), \star Lee et al. (1992), \heartsuit Pseudolikelihood, \diamond Liang et al. (1993)). Die Fehlerbalken geben die asymptotischen 95%-Konfidenzintervalle an ($n = 1000$).

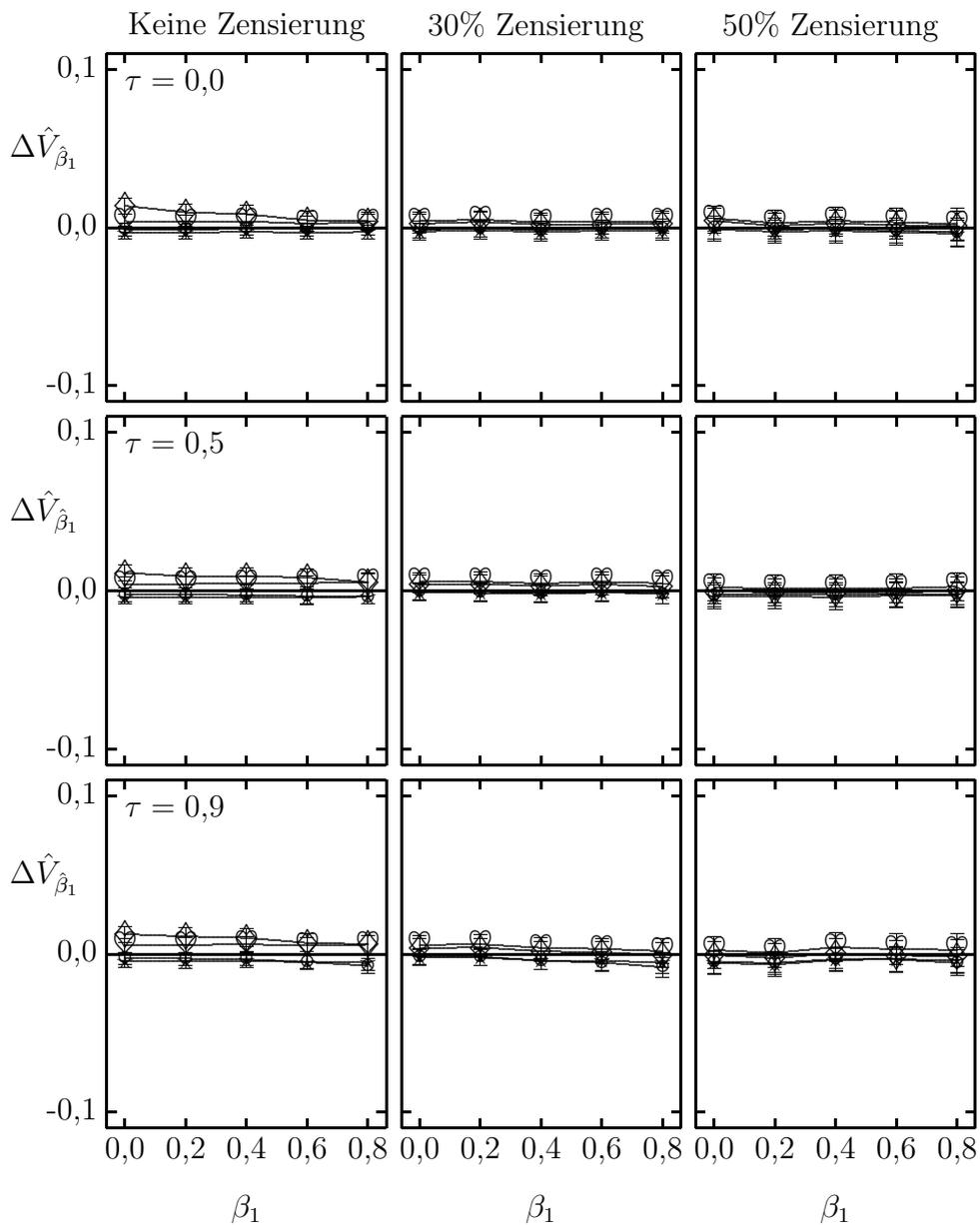


Abbildung 6.9: Der **systematische Fehler** von $\hat{V}_{\hat{\beta}_1}$ in Abhängigkeit von der Kendall'schen Korrelation τ der Überlebenszeiten und von der Zensierung bei **abhängigem** Design. (\circ Cox (1972), \star Lee et al. (1992), \heartsuit Pseudolike-likelihood, \diamond Liang et al. (1993)). Die Fehlerbalken geben die asymptotischen 95%-Konfidenzintervalle an ($n = 1000$).

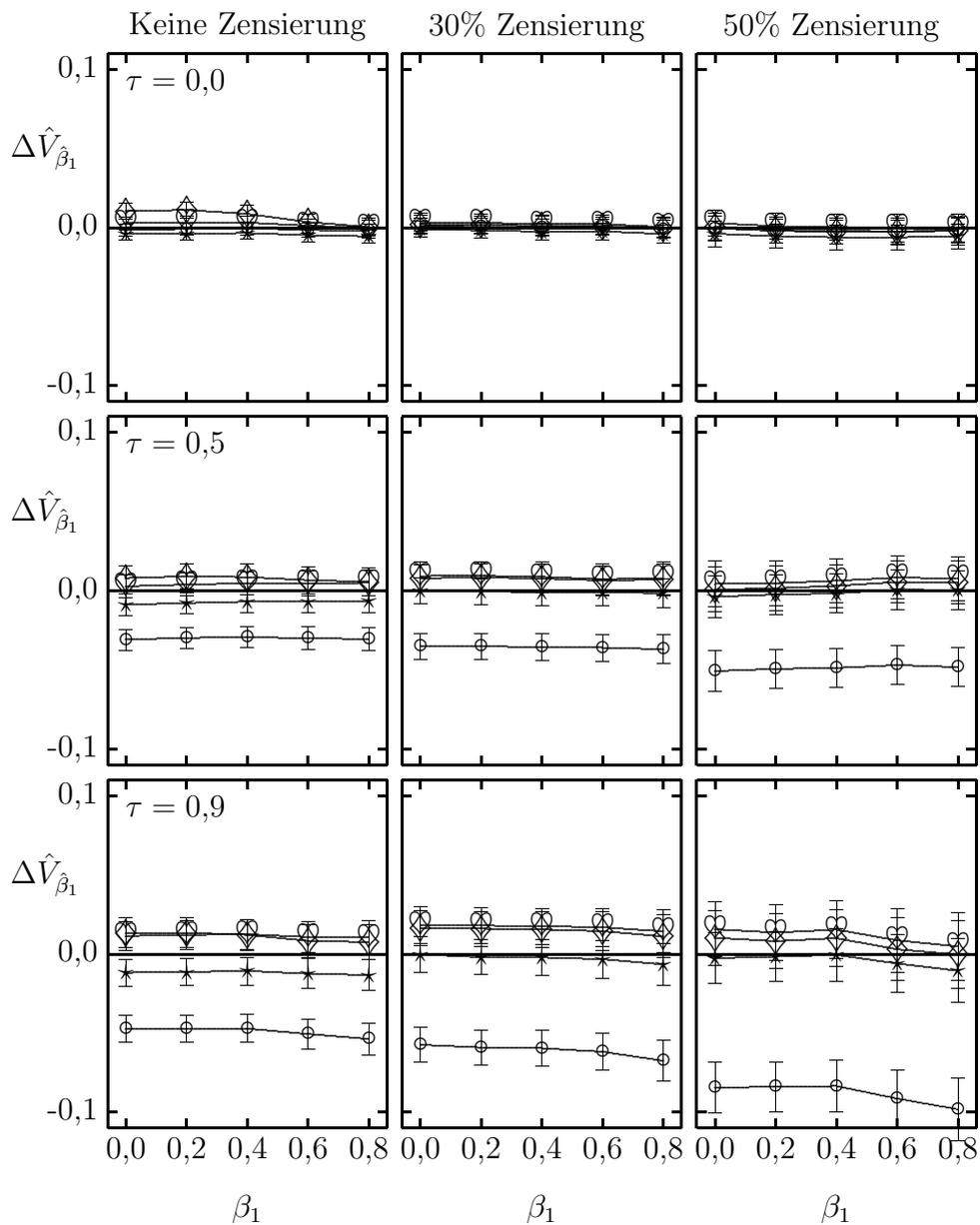


Abbildung 6.10: Der **systematische Fehler** von \hat{V}_{β_1} in Abhängigkeit von der Kendall'schen Korrelation τ der Überlebenszeiten und von der Zensierung bei **inversem** Design. (\circ Cox (1972), \star Lee et al. (1992), \heartsuit Pseudolikehood, \diamond Liang et al. (1993)). Die Fehlerbalken geben die asymptotischen 95%-Konfidenzintervalle an ($n = 1000$).

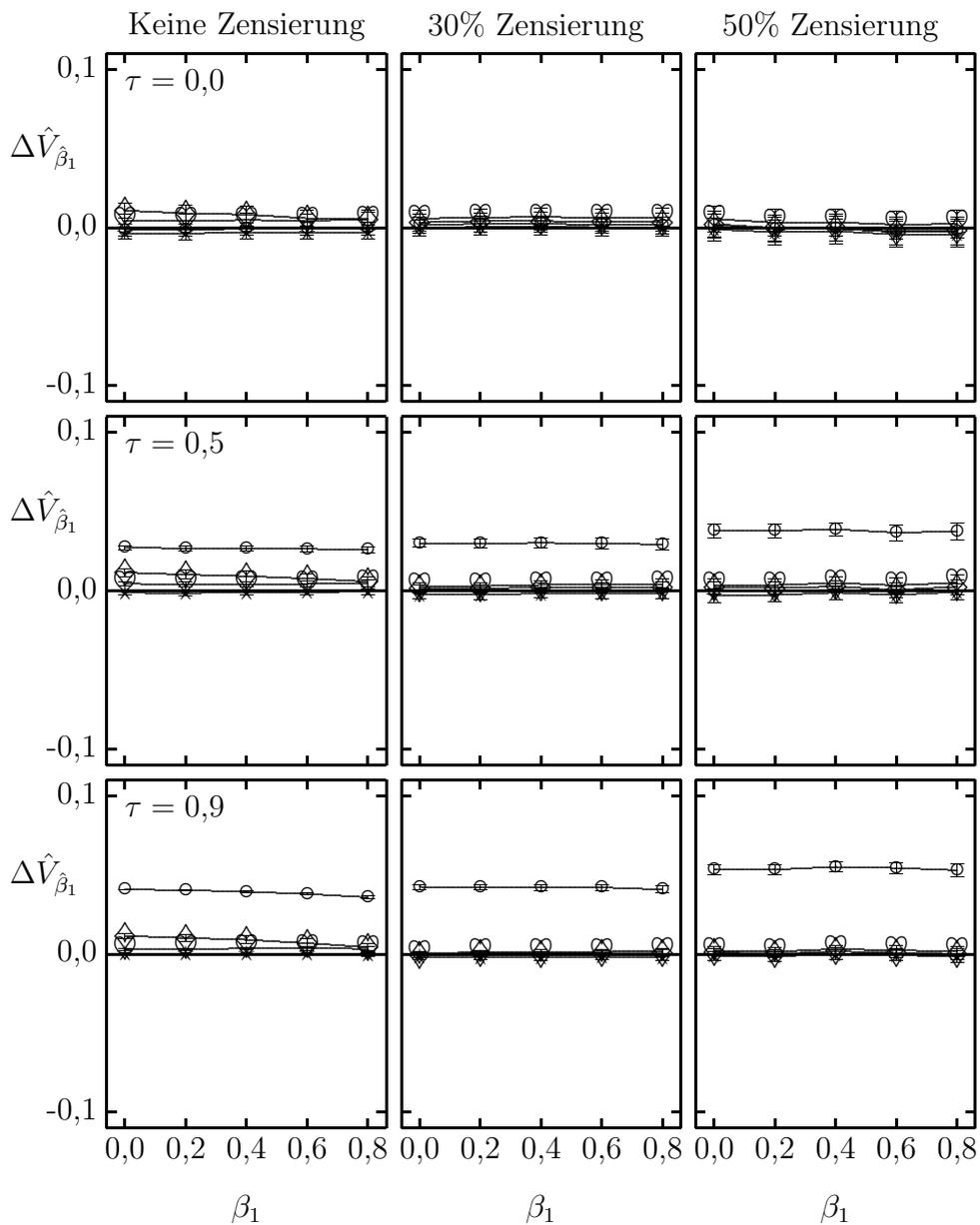


Abbildung 6.11: Anteil der **Wiederholungen im kritischen Bereich** $c_{95\%}$ des Wald-Tests von $H_0 : \beta_1 = 0$ in Abhängigkeit von der Kendall'schen Korrelation τ der Überlebenszeiten und von der Zensierung bei **unabhängigem** Design. (\circ Cox (1972), \star Lee et al. (1992), \heartsuit Pseudolikehood, \diamond Liang et al. (1993)). Die Fehlerbalken geben die asymptotischen 95%-Konfidenzintervalle an ($n = 1000$).

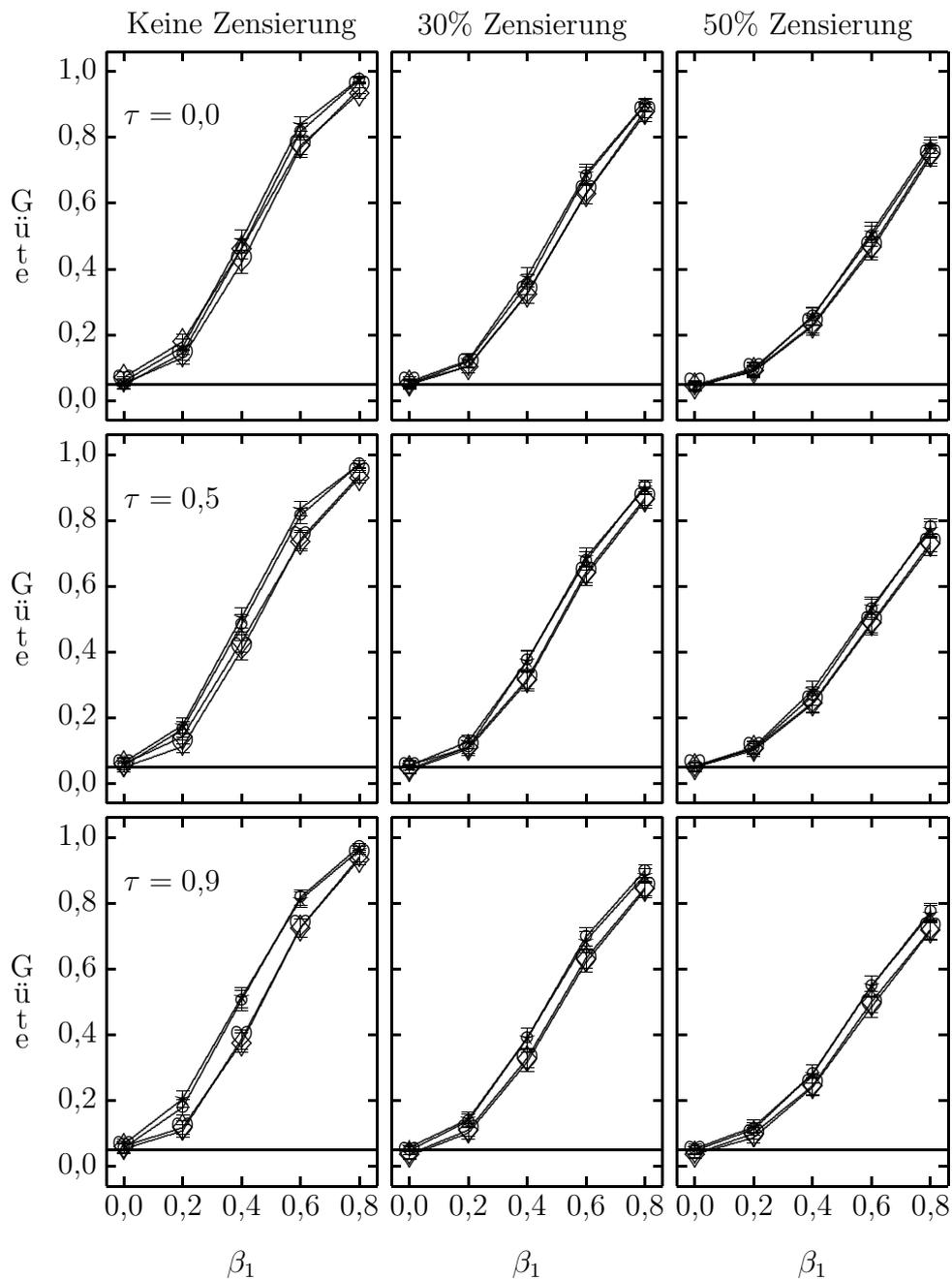


Abbildung 6.12: Anteil der **Wiederholungen im kritischen Bereich** $c_{95\%}$ des Wald-Tests von $H_0 : \beta_1 = 0$ in Abhängigkeit von der Kendall'schen Korrelation τ der Überlebenszeiten und von der Zensierung bei **abhängigem** Design. (\circ Cox (1972), \star Lee et al. (1992), \heartsuit Pseudolikehood, \diamond Liang et al. (1993)). Die Fehlerbalken geben die asymptotischen 95%-Konfidenzintervalle an ($n = 1000$).

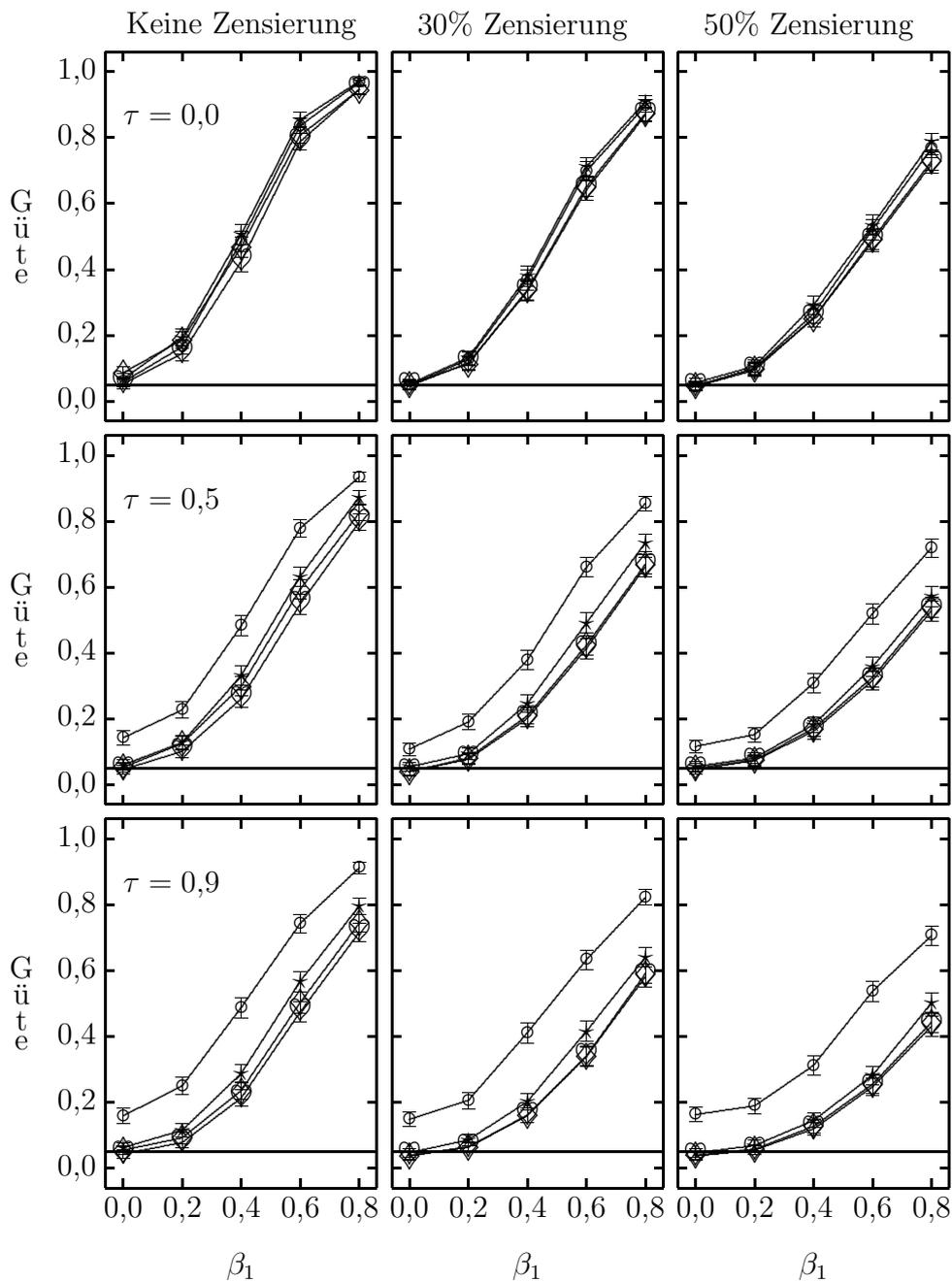
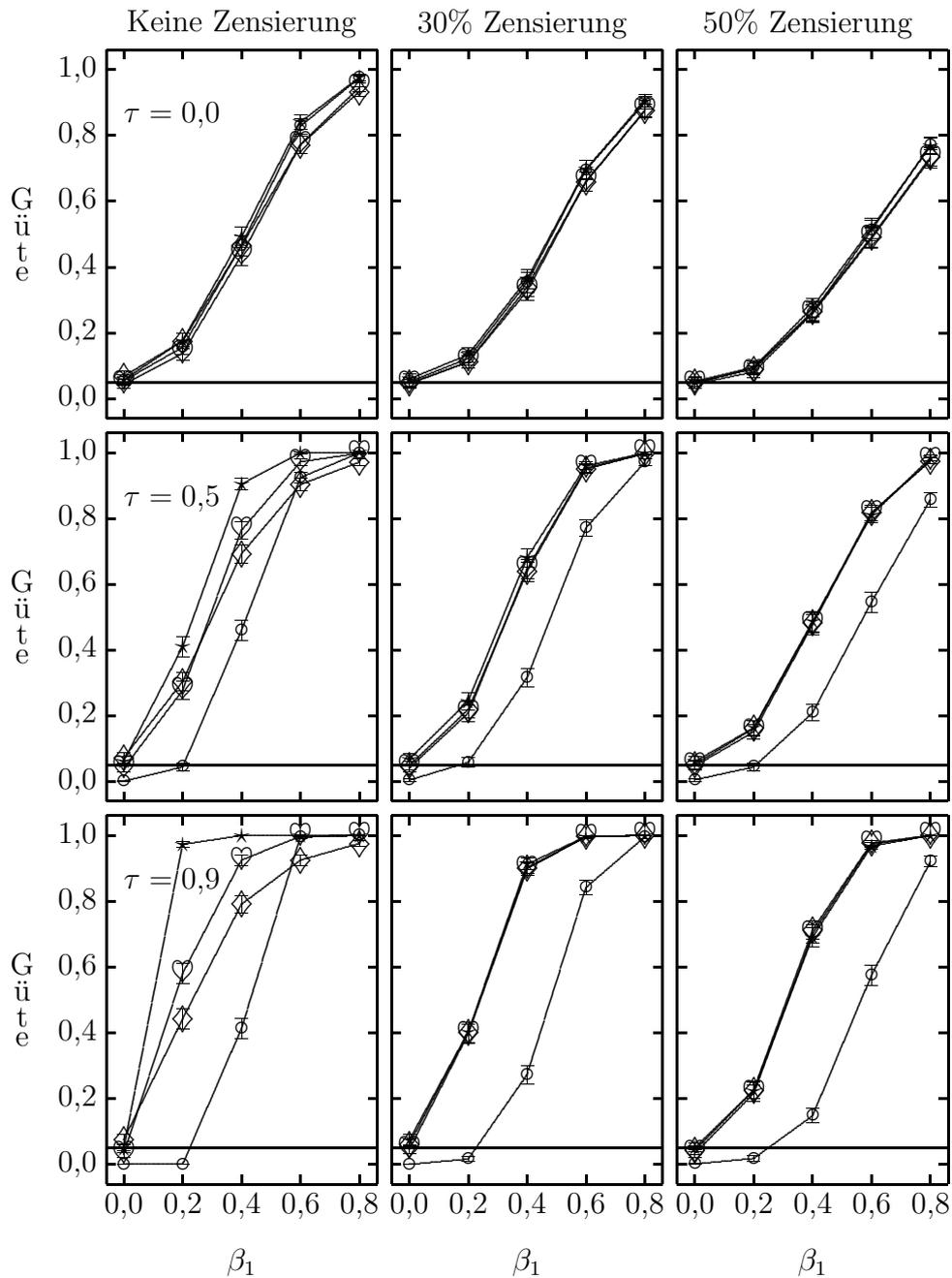


Abbildung 6.13: Anteil der **Wiederholungen im kritischen Bereich** $c_{95\%}$ des Wald-Tests von $H_0 : \beta_1 = 0$ in Abhängigkeit von der Kendall'schen Korrelation τ der Überlebenszeiten und von der Zensierung bei **inversem** Design. (\circ Cox (1972), \star Lee et al. (1992), \heartsuit Pseudolikelihood, \diamond Liang et al. (1993)). Die Fehlerbalken geben die asymptotischen 95%-Konfidenzintervalle an ($n = 1000$).



6.5.3 Zusammenfassung der Simulationsergebnisse

In zwei Simulationsstudien wurden 4 Schätzer für das marginale Modell mit proportionalen Hasardraten für paarweise abhängige Überlebensdaten verglichen. Es waren die Schätzung mittels Cox-Likelihood ohne Berücksichtigung der Abhängigkeit zwischen den Daten, die Schätzung des marginalen Modells mittels der Cox-Likelihood nach (Lee et al. 1992), der Pseudolikelihood-Schätzer mit Gewichtsfunktion $(Y_{..}(t) - Y_{i.}(t))/n$ und einen Pseudolikelihood-Schätzer mit einer weiteren Gewichtsfunktion $Wil; L2$ (siehe (6.3.1)), die von Liang et al. (1993) vorgeschlagen wurde. Die Simulationsdaten stammten einerseits von einem Clayton-Modell und andererseits aus der Transformation einer bivariaten Normalverteilung.

Abhängigkeiten zwischen den Daten ändern die Varianz der Schätzungen des Regressionsparameters in designabhängiger Weise: Wenn die Regressoren innerhalb der Haufen jeweils gleich sind, dann wird die Varianz erhöht („abhängiges Design“ innerhalb der Haufen), wenn die Regressoren jeweils innerhalb der Haufen verschieden sind („invers abhängiges“ Design innerhalb der Haufen), dann wird die Varianz erniedrigt. Diese Auswirkung der Abhängigkeiten in den Haufen werden vom Schätzer nach (Cox 1972) nicht berücksichtigt. Die Verringerung der Varianz des Schätzers des Regressionsparameters geht nicht mit einer Verringerung der Verzerrung einher, sodaß die Bedeutung der Verzerrung als Fehlerquelle zunimmt.

Die Zensierung ist im gegenständlichen Modell ein Nebenparameter. Die Bedingungen für die Validität der asymptotischen Ergebnisse aus Abschnitt 4.4 und Abschnitt 5 fordern jedoch im Fall der Cox-Likelihood explizit in L 1 und im Fall der Schätzung mittels U -Statistik implizit in PL 1 über die Straffheit der datenabhängigen Gewichtsfunktion eine Untergrenze (über die Beobachtungszeit $[0, \tau]$) für den Anteil der Individuen, die noch in der Risikomenge vorhanden sind. Der Schätzer von Lee et al. (1992), die auf der Cox-Likelihood beruht, ist auffallend robust gegen diese Verletzung der Modellannahmen, während Schätzer, die auf U -Statistiken beruhen, insbesondere bei abhängigen Daten und invers abhängigem Design, stark beeinträchtigt werden (Siehe Abbildung 6.13, keine Zensierung, $\tau = 0, 9$).

Im Bereich der Gültigkeit der Modellannahmen sind nur in einzelnen Zellen der Simulationsstudien Unterschiede zwischen den Schätzern festzustellen. Die Schätzer verhalten sich unter dem Einfluß der verschiedenen untersuchten Faktoren sehr ähnlich.

Kapitel 7

Ein Beispiel aus der Zahnmedizin

7.1 Beschreibung der Studie

In einer retrospektiven Kohortenstudie an der Universitätsklinik für Zahn-Mund- und Kieferheilkunde der Karl-Franzens-Universität Graz wurden etwa 450 Patienten, an denen Teleskopkronen gefertigt worden waren, zu einer Nachuntersuchung eingeladen (Jerebic 1996). 196 Patienten folgten dem Aufruf und wurden tatsächlich untersucht und über die Haltbarkeit ihrer Teleskopkronen und ihre Zahnhygiene befragt (siehe Beispiel 2.6.7).

Teleskopkronen, die zum Zeitpunkt der Untersuchung noch bestanden, werden als zensiert betrachtet. Es liegt progressive Zensierung vom Typ I vor (siehe Beispiel 2.2.3). 52 von 652 Teleskopkronen bestanden zum Zeitpunkt der Untersuchung nicht mehr und ihre Liegedauer ist daher bekannt.

Die Beobachtungszeit der Teleskopkronen betrug im Median 4 Jahre. Es waren bis zu 10 Teleskopkronen pro Patient hergestellt worden. Der Median war 3, der Modus 2 Teleskopkronen.

Der Einfluß von *Kiefer, Zahnart, Prothesenart, Anzahl der hergestellten Teleskopkronen, Alter des Patienten* bei der Herstellung der Teleskopkronen, Zahnhygiene (Verwendung von *Zahnbürste, Zahnpasta* und *Wasserstoffperoxid* zur Prothesenreinigung) auf die Liegedauer war Gegenstand der Untersuchung (Tabelle 7.1).

Tabelle 7.1: Liste der Kovariablen

Kiefer	1=Oberkiefer	0=Unterkiefer
Prothese	1=Modellguß	2=Kunststoff
Schneidezahn	1=Ja	0=Nein
Mahl- oder Vormahlzahn	1=Ja	0=Nein
H ₂ O ₂	1=Ja	2=Nein
Zahnpasta	1=Ja	2=Nein
Zahnbürste	1=Ja	2=Nein
Anzahl der Kronen		
Alter		

7.2 Das Modell

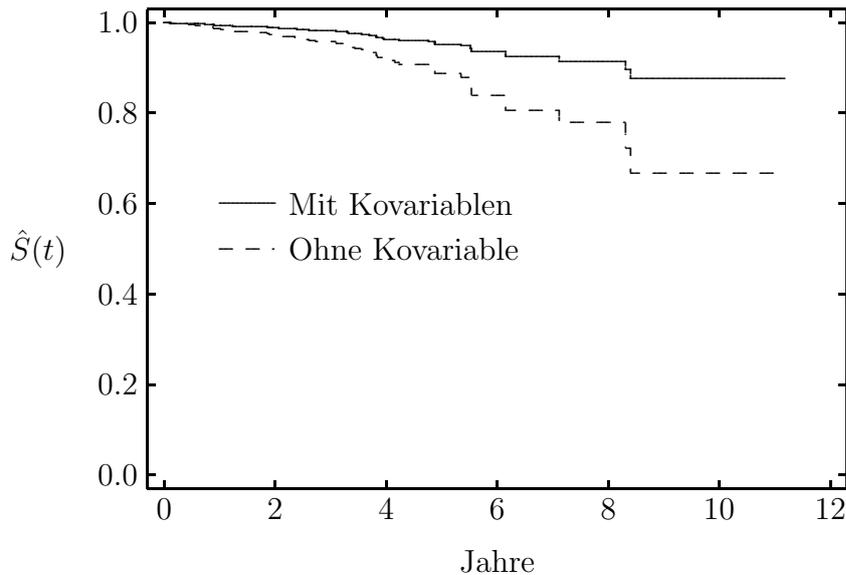
Für Teleskopkronen aus unterschiedlichen Gebissen wird ein multiplikatives Regressionsmodell mit proportionalen Hasardraten angesetzt ((4.2.7), (4.4.5) und (5.1.1)). Der Kovariablenvektor besteht aus den in Tabelle 7.1 angeführten Faktoren. Die Kodierungen wurden von den Originaldaten übernommen. Es wird angenommen, daß die Zensierung nicht informativ ist, und daß die Liegedauern von Teleskopkronen innerhalb eines Gebisses abhängige Zufallsvariable darstellen.

Das Modell beschreibt die Liegedauer von Zahn j in Patient i . Der Ausfall einer Teleskopkrone beziehungsweise der Ersatz durch einen Bona-Anker beendet die Liegedauer. Sie ist das einzige mögliche Ereignis des Zählprozesses. Gleichzeitig wirkt ein progressiver Zensierungsprozeß vom Typ I. Der Zensierung wird durch die Feststellung des Gebißstatus gegeben. Da dieser Zeitpunkt ab Herstellung der Teleskopkrone im Voraus fixiert ist, sind die Zensierungszeiten nicht zufällig. Nachfolgende Ereignisse können nicht mehr berücksichtigt werden. Die Intensität des Zählprozesses der zensierten Liegedauern bezüglich der selbsterzeugten durch die Daten von Zahn ij erzeugten Filtrierung $\mathcal{F}_{t,ij}$ ist

$$Y_{ij}(t) \exp \{ \beta_0^\top Z_{ij} \} \alpha_0(t). \quad (7.2.1)$$

Dabei ist $Y_{ij}(t)$ ein Indikator dafür, daß der Zahn unmittelbar vor t noch beobachtbar war. Es liegt eine Einsprung-Prozeß vor (siehe Beispiel 2.1.2 und Beispiel 2.2.3). $\alpha_0(t)$ ist die zugrundeliegende gemeinsame Hasardfunktion. Z_{ij} ist ein Vektor von p Kovariablen (Regressoren) aus Tabelle 7.1. β_0 enthält

Abbildung 7.1: Wahrscheinlichkeit der Liegedauer von Teleskopkronen, berechnet aus dem Breslow-Schätzer der zugrundeliegenden Hasardfunktion. Im Modell mit $p = 4$ Kovariablen (Kiefer, Art der Prothese, Schneidezahn und H_2O_2) enthält der Kovariablenvektor die Mittelwerte der Kovariablen der Stichprobe.



die zu den Kovariablen gehörigen Regressionskoeffizienten. Mittels Exponentialfunktion wirken diese multiplikativ auf den Intensitätsprozeß. Die Filtrierungen $\mathcal{F}_{t,ij}$ und $\mathcal{F}_{t,lk}$ sind unabhängig für $i \neq j$.

Aufgrund der Abhängigkeit der Liegedauern von Teleskopkronen des gleichen Gebisses kann die Varianz mit der Methode nach (Cox 1972) nicht konsistent geschätzt werden. Dies ist aber mit der Methode nach Lee et al. (1992), Liang et al. (1993) und der Pseudolikelihoodmethode möglich. Die Lösungen der Schätzgleichungen wurden mit dem Newton-Raphson-Verfahren gefunden. Die entsprechenden Programme wurden vom Autor in XLISP-STAT programmiert und auf Anfrage erhältlich.

Zu Vergleichszwecken wird auch ein bedingtes partielles Likelihoodmodell nach Holt und Prentice (1974) beziehungsweise (Kalbfleisch und Prentice 1980, p. 189 ff) berechnet (Unter Verwendung von SAS Software¹).

¹SAS Inc., Cary, USA)

7.3 Ergebnisse

7.3.1 Der Breslow-Schätzer

Die kumulative zugrundeliegende Hasardrate des Cox-Modells ist durch den Breslow-Schätzer $\hat{A}_0(t)$ (siehe (4.2.4)) gegeben.

$$\hat{S}(t, Z_0) = \prod_{[0,t]} \left\{ 1 - \exp \left\{ \hat{\beta}^\top Z_0(u) \right\} d\hat{A}_0(u, \hat{\beta}) \right\} \quad (7.3.1)$$

ist dann eine Schätzung der Überlebensfunktion für den Kovariablenvektor Z_0 (Andersen et al. 1993, p. 509).

Wenn das Modell keine Kovariablen enthält, dann stimmt die Überlebensfunktion mit dem Kaplan-Meier-Schätzer überein (Andersen et al. 1993, p. 256), da dann der Breslow-Schätzer dem Nelson-Aalen-Schätzer entspricht. Im Modell mit Kovariablen wurden vier Merkmale berücksichtigt (Begründung siehe unten). Eine mögliche Wahl für Z_0 ist der Mittelwertvektor der Kovariablenvektoren der Stichprobe. Die Ergebnisse einer solchen Schätzung sind in Abbildung 7.1 dargestellt (Unter Verwendung von SAS Software). Eine Liegedauer von 5 Jahren wird im Modell mit Kovariablen mit Wahrscheinlichkeit 0,952 und 171 Zähnen in der Risikomenge erreicht. Eine Liegedauer von 8 Jahren wird mit Wahrscheinlichkeit 0,915 und 30 Zähnen in der Risikomenge erreicht. Die längste beobachtete Liegedauer betrug 8 Jahre und 150 Tage. Zu diesem Zeitpunkt sind nur noch 13 Zähne in der Risikomenge. Die längste Beobachtungsdauer betrug 11 Jahre und 68 Tage.

7.3.2 Überprüfung der Annahme proportionaler Hasardraten

Das Modell der multiplikativen Intensitäten ist „semiparametrisch“ und somit sparsam in den Modellannahmen. Die verwendeten Tests funktionieren zwar auch bei fehlspezifiziertem Modell: Der Sandwich-Test, der für unkorrelierte Beobachtungen die selbe Asymptotik verwendet wie Satz 4.4.4, ist asymptotisch unverfälscht (Lin und Wei 1989). Dennoch wurde möglicherweise der funktionale Zusammenhang der Kovariablen mit der Zeit vom Modell nicht korrekt erfaßt.

In (Grambsch und Therneau 1994) wird zur Untersuchung dieses funktionalen Zusammenhangs vorgeschlagen *Schoenfeldresiduen*

$$r_e(\beta) = Z_{ij}(t_e) - E(\beta, t_e) \quad (7.3.2)$$

zu betrachten (Schoenfeld 1982). Dabei bezeichnen t_e , $e = 1, \dots, N_e$ die Zeitpunkte, zu denen Ereignisse beobachtet wurden. Dies wird durch den zeitabhängigen Ansatz

$$\beta(t) = \beta_0 + G(t)\theta \quad (7.3.3)$$

mit der zeitabhängigen Diagonalmatrix $G(t)$ und einem weiteren Regressionsparameter θ motiviert. Man erhält das Schoenfeldresiduum als Summe des Schoenfeldresiduums des wahren Modells plus die Differenz der gewichteten Summen der Kovariablen unter den beiden Modellen:

$$r_e(\beta) = [Z_{ij}(t_e) - E(\beta(t_e), t_e)] + [E(\beta(t_e), t_e) - E(\beta, t_e)]. \quad (7.3.4)$$

Nun benötigt man Martingaleigenschaften der Schoenfeldresiduen, weshalb die Betrachtungen auf Haufenelemente mit Index k_1, \dots, k_K eingeschränkt werden. Durch Taylorreihenentwicklung von $E_k(\beta(t_e^k), t_e^k)$ um β erhält man (siehe (4.4.12))

$$\mathbb{E} \{r_{e_k}^k(\beta) | \mathcal{F}_{k, t_e^k}\} = V_k(\beta, t_e^k)G(t_e^k)\theta + R_1. \quad (7.3.5)$$

Darüberhinaus definiert man das *skalierte Schoenfeldresiduum* als

$$r_{e_k}^{*k}(\beta) = V_k^{-1}(\beta, t_e^k)r_{e_k}^k(\beta). \quad (7.3.6)$$

Für dieses Residuum gilt

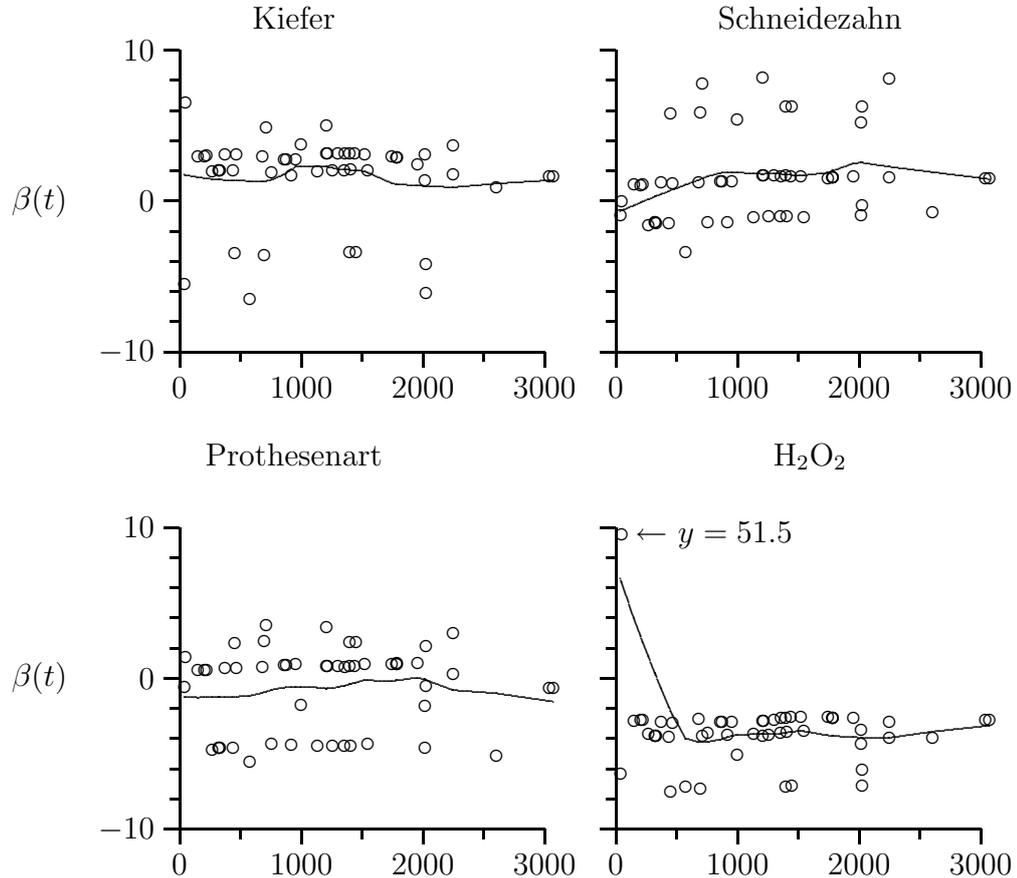
$$\mathbb{E} \{r_{e_k}^{*k}(\beta) | \mathcal{F}_{k, t_e^k}\} = G(t_e^k)\theta + R_2. \quad (7.3.7)$$

Schoenfeldresiduen sind asymptotisch unkorreliert.

Zuletzt benützt man, daß Ausdrücke für Index k durch solche, die über die gesamte Stichprobe berechnet wurden, ersetzt werden können, da sie in Wahrscheinlichkeit gegen den selben Grenzwert konvergieren. Konkret kann unter den Bedingungen L 1 bis L 4 aus Abschnitt 4.4 $E_j(\beta, \cdot)$ durch $E(\beta, \cdot)$ und $V_k(\beta, \cdot)$ durch $V(\beta, \cdot)$ (siehe (4.2.10)) ersetzt werden (Der Rest konvergiert in Wahrscheinlichkeit gegen 0). Schließlich konvergieren die Restterme R_1 und R_2 in Wahrscheinlichkeit gegen 0. Die Definition des skalierten Schoenfeldresiduums lautet schließlich

$$r_e^*(\beta) = V^{-1}(\beta, t_e) (Z_{ij}(t_e) - E(\beta, t_e)). \quad (7.3.8)$$

Abbildung 7.2: Skalierte Schoenfeldresiduen eines Modells mit 4 Kovariablen. Glättung durch LOWESS mit einem Anteil von 0,5 ohne Iteration



$\beta(t) = \beta + G(t_e)\theta$ kann daher mit Hilfe von Glättern aus konsistenten Schätzern für β und r_e^* geschätzt werden. Hier wurde der LOWESS-Glätter von XLISP-STAT verwendet. In Abbildung 7.3.2 sind die Schoenfeldresiduen und deren Glättung für ein Modell mit den 4 Kovariablen Kiefer, Schneidezahn, Art der Prothese und H₂O₂ dargestellt. Das Residuum zum Ereignis nach 47 Tagen in der Komponente H₂O₂ war 51,48. Dieser Wert verursacht den Abfall der geglätteten Werte im entsprechenden Bereich. Dieser Wert kann als „Ausreißer“ betrachtet werden. Der funktionale Zusammenhang sollte nicht von einem einzigen Wert bestimmt werden. Es gibt daher keine Hinweise auf eine Fehlspezifikation des Modells.

Tabelle 7.2: Liegedauer von Teleskopkronen

Methode	Kiefer	Prothese	Schneidezahn	H ₂ O ₂
Holt (1974)	2,26 ± 0,86	-	1,05 ± 0,58	-
Cox (1972)	1,58 ± 0,43	-0,73 ± 0,31	1,46 ± 0,40	-2,80 ± 1,01
Lee et al. (1992)	1,58 ± 0,41	-0,73 ± 0,38	1,46 ± 0,35	-2,80 ± 1,01
Pseudolikelihood	1,62 ± 0,54	-0,78 ± 0,42	1,73 ± 0,44	-2,73 ± 1,04
Liang et al. (1993)	1,62 ± 0,51	-0,83 ± 0,39	1,73 ± 0,41	-2,74 ± 1,03

7.3.3 Das Modell der proportionalen Hasardraten

Bei den Mahlzähnen wurde kein Verlust einer Teleskopkrone beobachtet. In solchen Fällen haben sowohl die Schätzfunktion aufgrund der partiellen Likelihood (Cox 1972), als auch die Schätzfunktionen, die auf der Pseudolikelihood nach Liang et al. (1993) beruhen, keine Nullstellen (Jacobsen 1989). Daher wurden die Vormahlzähne mit den Mahlzähnen zu einer Kategorie zusammengefaßt. Die relevanten Kovariablen werden durch abbauende Merkmalsauswahl (backward variable selection) und durch Anwendung des Wald-Test-Kriterium mit einem Fehler 1. Art von 5% auf das Modell nach (Cox 1972) bestimmt (Unter Verwendung von SAS Software). Dabei blieben die Variablen *Kiefer*, *Schneidezahn*, *Art der Prothese* und H₂O₂ übrig. Insbesondere hat die Anzahl der Teleskopkronen pro Gebiß keinen Einfluß auf die Liegedauer.

Die Ergebnisse des eingeschränkten Regressionsmodells sind in Tabelle 7.2 dargestellt. Teleskopkronen im Oberkiefer haben eine höhere Hasardrate als Teleskopkronen im Unterkiefer. Kunststoffprothesen haben niedrigere Hasardraten als Modellgußprothesen. Teleskopkronen auf Schneidezähnen haben eine höhere Hasardrate als die übrigen Zahnarten. Die Verwendung von Wasserstoffperoxid erniedrigt die Hasardrate. Die Indikatoren für Zahnhygiene (Wasserstoffperoxid, Zahnbürste, Zahnpasta) hängen stark voneinander ab, sodaß nur eine dieser Kovariablen nach der Merkmalsauswahl übriggeblieben ist.

Im Modell nach der Methode von (Holt und Prentice 1974) sind nur die Variablen Kiefer und Schneidezahn enthalten, da nur diese Parameter innerhalb eines Patienten unterschiedlich sind und einen Beitrag zur bedingten Likelihood liefern. Die Schätzung des Regressionsparameters von Kiefer ist höher als diejenige der marginalen Methoden. Derjenige der Schneidezähne ist niedriger.

Bei den marginalen Modellen ist kein großer Unterschied zwischen der Schätzung nach (Cox 1972) und der Schätzung nach (Lee et al. 1992) festzustellen. Daher ist entweder die Abhängigkeit der Liegedauern eines Gebisses nicht stark ausgeprägt, oder es liegt eine Mischung aus invers abhängigem und unabhängigem Design vor. Die Pseudolikelihoodschätzung und die Schätzung nach (Liang et al. 1993) unterscheiden sich ebenfalls kaum von den vorhergehenden Schätzungen. Die Abweichungen liegen innerhalb einer Standardabweichung des Schätzers.

Literaturverzeichnis

- Aalen, O. O. (1978). Nonparametric inference for a family of counting processes, *The Annals of Statistics* **6**: 701–726.
- Andersen, P. K., Borgan, Ø., Gill, R. D. und Keiding, N. (1993). *Statistical Models Based on Counting Processes*, Springer.
- Andersen, P. K., Borgan, Ø., Gill, R. und Keiding, N. (1982). Linear non-parametric tests for comparison of counting processes, with applications to censored survival data, *International Statistical Review* **50**: 219–258.
- Andersen, P. K. und Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study, *The Annals of Statistics* **10**: 1100–1120.
- Boos, D. D. (1990). On generalized score tests, *The American Statistician* **46**(4): 327–333.
- Breslow, N. E. (1972). Discussion of the paper by D.R. Cox, *Journal of the Royal Statistical Society B* **34**: 216–217.
- Cai, J. und Prentice, R. J. (1995). Estimating equations for hazard ratio parameters based on correlated failure time data, *Biometrika* **82**: 151–165.
- 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.
- Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence, *Biometrika* **65**: 141–151.
- Cox, D. und Hinkley, D. (1974). *Theoretical Statistics*, Chapman & Hall, London.

- Cox, D. R. (1972). Regression models and life-tables (with discussion), *Journal of the Royal Statistical Society Ser. B* **34**: 187–220.
- Cox, D. R. (1975). Partial likelihood, *Biometrika* **62**: 269–276.
- Daffer, P. Z. und Taylor, R. L. (1979). Laws of large numbers for $D[0, 1]$, *The Annals of Probability* **7**: 85–95.
- Fleming, T. R. und Harrington, D. P. (1991). *Counting Processes and Survival Analysis*, Wiley.
- Genest, C. und MacKay, R. (1986). The joy of copulas: Bivariate distributions with uniform marginals, *American Statistician* **40**: 280–283.
- Genest, C. (1987). Frank's family of bivariate distributions, *Biometrika* **74**: 549–555.
- Gill, R. D. und Johansen, S. (1990). A survey of product-integration with a view towards application in survival analysis, *Annals of Statistics* **18**: 1501–1555.
- Gill, R. D. (1992). Marginal partial likelihood, *Scandinavian Journal of Statistics* **79**: 133–137.
- Godambe, V. P. (1985). The foundations of finite sample estimation in stochastic processes, *Biometrika* **72**: 419–428.
- Grambsch, P. und Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals, *Biometrika* **81**: 515–526.
- Holt, J. D. und Prentice, R. L. (1974). Survival analysis in twin studies and matched pair experiments, *Biometrika* **61**: 17–30.
- Hougaard, P. (1986). A class of multivariate failure time distributions, *Biometrika* **73**: 671–678.
- Jacobsen, M. (1982). *Statistical Analysis of Counting Processes*, Lecture Notes in Statistics 12, Springer Verlag, New York.
- Jacobsen, M. (1989). Right censoring and martingale methods for failure time data, *The Annals of Statistics* **17**: 1133–1156.
- Jerebic, H. (1996). *Komplikationen bei Teleskopkronen und deren Einfluß auf die Haltbarkeit*, Dissertation, Karl-Franzens-Universität, Graz.

- Kalbfleisch, J. D. und Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data*, Wiley, New York.
- Kallenberg, O. (1997). *Foundations of Modern Probability*, Springer.
- Koziol, J. A. und Green, S. B. (1976). A Cramér-von Mises statistic for randomly censored data, *Biometrika* **63**: 465–474.
- L'Ecuyer, P. (1986). Efficient and portable combined random number generators, *Communications of the ACM* **31**: 742–749.
- Lee, A. J. (1990). *U-Statistics: Theory and Practice*, Marcel Dekker.
- Lee, E. W., Wei, L. J. und Amato, D. A. (1992). Cox-type regression analysis for large numbers of small groups of correlated failure time observations, in J. P. Klein und P. K. Goel (Hrsg.), *Survival Analysis: State of the Art*, Kluwer Academic Publishers, Dordrecht, pp. 237–247.
- Lenglart, E. (1977). Relation de domination entre deux processus, *Annales de l'Institut Henri Poincaré* **13**: 171–179.
- Liang, K. Y., Self, S. G. und Chang, Y. C. (1993). Modelling marginal hazards in multivariate failure time data, *Journal of the Royal Statistical Society Series B* **55**: 441–453.
- Liang, K. Y. und Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models, *Biometrika* **73**: 13–22.
- Lin, D. Y. und Wei, L. J. (1989). The robust inference for the Cox proportional hazards model, *Journal of the American Statistical Association* **84**: 1074–78.
- Lin, D. Y. und Ying, Z. (1993). Cox regression with incomplete covariate measurements, *Journal of the American Statistical Association* **88**: 1341–1349.
- Lin, D. Y. (1991). Goodness-of-fit analysis for the Cox regression model based on a class of parameter estimators, *Journal of the American Statistical Association* **86**: 725–728.
- Lin, D. Y. (1994). Cox regression analysis of multivariate failure time data: the marginal approach, *Statistics in Medicine* **13**: 2233–2247.
- Loughin, T. und Koehler, K. (1997). Bootstrapping regression parameters in multivariate survival analysis, *Lifetime Data Analysis* **3**: 157–177.

- Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration, *Cancer Chemotherapy Reports* **50**: 163–170.
- Oakes, D. (1982). A concordance test for independence in the presence of censoring, *Biometrics* **38**: 451–455.
- Pepe, M. S. und Cai, J. (1993). Some graphical display and marginal regression analyses for recurrent failure times and time dependent covariates, *Journal of the American Statistical Association* **88**: 811–820.
- Peto, R. und Peto, J. (1972). Asymptotically efficient and invariant test procedures (with discussion), *Journal of the Royal Statistical Society A* **135**: 185–206.
- Pollard, D. (1984). *Convergence of Stochastic Processes*, Springer-Verlag, New York.
- Pschyrembel, W. (Hrsg.) (1994). *Klinisches Wörterbuch*, 257. Aufl., de Gruyter, Berlin.
- Puri, M. L. und Sen, P. K. (1971). *Nonparametric Methods in Multivariate Analysis*, Wiley.
- Quehenberger, F. (1997). Comparison of two estimators for the marginal proportional intensity model for clustered event history data, in C. Minder und H. Friedl (Hrsg.), *Proceedings of the 12th International Workshop on Statistical Modelling Biel/Bienne*, Band 5, Schriftenreihe der Österreichischen Statistischen Gesellschaft, pp. 265–269.
- Randles, R. H. (1982). On the asymptotic normality of statistics with estimated parameters, *The Annals of Statistics* **10**: 462–474.
- Sachs, L. (1997). *Angewandte Statistik. Anwendung statistischer Methoden.*, 8. Aufl., Springer Verlag, Berlin.
- Sasieni, P. (1993). Maximum weighted partial likelihood estimators for the Cox model, *Journal of the American Statistical Association* **88**: 144–152.
- Schoenfeld, D. (1982). Partial residuals for the proportional hazards regression model, *Biometrika* **69**: 239–241.
- Schweizer, P. und Sklar, A. (1983). *Probabilistic Metric Spaces*, North-Holland, Amsterdam.

- Shih, J. und Louis, A. (1995). Inferences on the association parameter in copula models for bivariate survival data, *Biometrics* **51**: 1384–1399.
- Shorack, G. R. und Wellner, J. A. (1986). *Empirical Processes with Applications*, Wiley, New York.
- Širjaev, A. (1988). *Wahrscheinlichkeit*, VEB Deutscher Verlag der Wissenschaften, Berlin.
- Spiekerman, C. F. und Lin, D. Y. (1996). Checking the marginal Cox model for correlated failure time data, *Biometrika* **83**(1): 143–156.
- Stine, R. A. (1992). Automated simulations, *ASA Proceedings of the Statistical Computing Section* pp. 62–67.
- Sukhatme, B. V. (1958). Testing the hypotheses that two populations differ only in location, *Annals of Mathematical Statistics* **29**: 60–78.
- Tierney, L. (1990). *Lisp-Stat: an object-oriented environment for statistical computing and dynamic graphics*, Wiley, New York.
- Vaupel, J. W., Manton, K. G. und Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality, *Demography* **16**: 439–454.
- Wei, L. J., Lin, D. Y. und Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modelling marginal distributions, *Journal of the American Statistical Association* **84**: 1065–1073.
- Wichman, B. A. und Hill, I. (1982). An efficient and portable pseudo-random number generator, *Applied Statistics* **74**: 549–555. (Corr:V33 p123).