

Flexible parametrische Survivalmodelle

Seminararbeit

Anna Theresa Stüber

Institut für Statistik, LMU

13.Juni 2017

1 Einleitung

2 Grundlagen

Hazardrate und Survivalfunktion

Bekannte Verteilungsmodelle

Regressionsmodelle: AFT- und Cox-Modell(e)

3 Flexible parametrische Survivalmodelle

Vom Cox-Modell zu flexiblen parametrischen Modellen

Exkurs: (Restringsierte) kubische Splines

Spline-basierte parametrische Survivalmodelle

Bestimmung der Knotenanzahl und -lokation der Splinefunktion

Schätzmethode

Erweiterung um Zeit-abhängige Variablen

4 Datenbeispiel

5 Fazit

6 Literatur und Anhang

1 Einleitung

2 Grundlagen

Hazardrate und Survivalfunktion

Bekannte Verteilungsmodelle

Regressionsmodelle: AFT- und Cox-Modell(e)

3 Flexible parametrische Survivalmodelle

Vom Cox-Modell zu flexiblen parametrischen Modellen

Exkurs: (Restringsierte) kubische Splines

Spline-basierte parametrische Survivalmodelle

Bestimmung der Knotenanzahl und -lokation der Splinefunktion

Schätzmethode

Erweiterung um Zeit-abhängige Variablen

4 Datenbeispiel

5 Fazit

6 Literatur und Anhang

- Survivalanalyse: Zusammenfassung unterschiedlicher Methoden zur Analyse der Zielvariable „Zeitpunkt des Eintretens eines interessierenden Ereignisses“
- Bekanntestes Modell für derartige Untersuchungen: Proportional-Hazards-Modell (PH)

„The success of Cox regression has perhaps had the unintended side-effect that practitioners too seldomly invest efforts in studying the baseline hazard. [...] A parametric version [of the Cox model], ... if found to be adequate, would lead to more precise estimation of survival probabilities and ... concurrently contribute to a better understanding of the phenomenon under study.“ (Hjort, 1992)

- Erstmalige Einführung flexibler parametrischer Survivalmodelle durch P. Royston und M. K. B. Parmar

① Einleitung

② Grundlagen

Hazardrate und Survivalfunktion

Bekannte Verteilungsmodelle

Regressionsmodelle: AFT- und Cox-Modell(e)

③ Flexible parametrische Survivalmodelle

Vom Cox-Modell zu flexiblen parametrischen Modellen

Exkurs: (Restringsierte) kubische Splines

Spline-basierte parametrische Survivalmodelle

Bestimmung der Knotenanzahl und -lokation der Splinefunktion

Schätzmethode

Erweiterung um Zeit-abhängige Variablen

④ Datenbeispiel

⑤ Fazit

⑥ Literatur und Anhang

Annahme: T als nicht-negative, stetige Zufallsvariable

- Dichtefunktion: $f_T(t) := f(t) = \frac{\partial F_T(t)}{\partial t}$
- Verteilungsfunktion: $F_T(t) := F(t) = P(T \leq t)$
- Survivalfunktion: $S_T(t) := S(t) = P(T \geq t) = 1 - F(t)$
- Hazardrate: $\lambda_T(t) := \lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} P(t \leq t + \Delta t | T)$
- Kum. Hazardrate: $\Lambda_T(t) := \Lambda(t) = \int_0^t \lambda(u) du$

Hazardrate und Survivalfunktion

Annahme: T als nicht-negative, stetige Zufallsvariable

- Dichtefunktion: $f_T(t) := f(t) = \frac{\partial F_T(t)}{\partial t}$
 $f(t) = -\frac{\partial S(t)}{\partial t}$
- Verteilungsfunktion: $F_T(t) := F(t) = P(T \leq t)$
- Survivalfunktion: $S_T(t) := S(t) = P(T \geq t) = 1 - F(t)$
 $S(t) = \exp(-\Lambda(t)) = \exp(-\int_0^t \lambda(u) du)$
- Hazardrate: $\lambda_T(t) := \lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} P(t \leq t + \Delta t | T)$
 $\lambda(t) = \frac{f(t)}{1-F(t)} = \frac{f(t)}{S(t)}$
- Kum. Hazardrate: $\Lambda_T(t) := \Lambda(t) = \int_0^t \lambda(u) du$

① Einleitung

② Grundlagen

Hazardrate und Survivalfunktion

Bekannte Verteilungsmodelle

Regressionsmodelle: AFT- und Cox-Modell(e)

③ Flexible parametrische Survivalmodelle

Vom Cox-Modell zu flexiblen parametrischen Modellen

Exkurs: (Restringsierte) kubische Splines

Spline-basierte parametrische Survivalmodelle

Bestimmung der Knotenanzahl und -lokation der Splinefunktion

Schätzmethode

Erweiterung um Zeit-abhängige Variablen

④ Datenbeispiel

⑤ Fazit

⑥ Literatur und Anhang

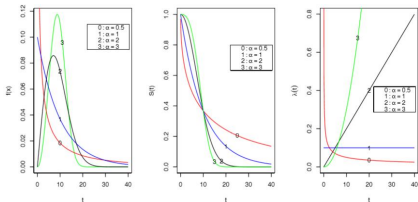
Weibull-Verteilung

$$S(t) = \exp(-(\lambda t)^\alpha)$$

$$f(t) = \alpha \lambda \cdot (\lambda t)^{\alpha-1} \exp(-(\lambda t)^\alpha)$$

$$\lambda(t) = \alpha \lambda \cdot (\lambda t)^{\alpha-1}$$

$$\Lambda(t) = (\lambda t)^\alpha$$



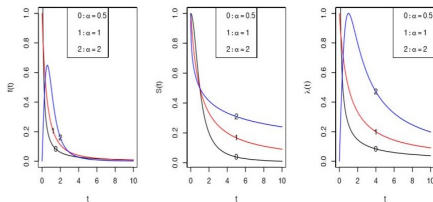
log-logistische Verteilung

$$S(t) = \frac{1}{1+(\lambda t)^\alpha}$$

$$f(t) = \frac{\alpha \lambda \cdot (\lambda t)^{\alpha-1}}{[1+(\lambda t)^\alpha]^2}$$

$$\lambda(t) = \frac{\alpha \lambda \cdot (\lambda t)^{\alpha-1}}{1+(\lambda t)^\alpha}$$

$$\Lambda(t) = \log[1 + (\lambda t)^\alpha]$$



① Einleitung

② Grundlagen

Hazardrate und Survivalfunktion

Bekannte Verteilungsmodelle

Regressionsmodelle: AFT- und Cox-Modell(e)

③ Flexible parametrische Survivalmodelle

Vom Cox-Modell zu flexiblen parametrischen Modellen

Exkurs: (Restringsierte) kubische Splines

Spline-basierte parametrische Survivalmodelle

Bestimmung der Knotenanzahl und -lokation der Splinefunktion

Schätzmethode

Erweiterung um Zeit-abhängige Variablen

④ Datenbeispiel

⑤ Fazit

⑥ Literatur und Anhang

Cox-Modell (PH)

- Beinhaltet als semi-parametrisches Verfahren eine unspezifizierte Baseline-Hazard $\lambda_0(t)$
- Cox-Modell als multiplikatives Hazardraten-Modell der folgenden Form:

$$\lambda(t, \mathbf{x}) = \lambda_0(t) \cdot \exp(\mathbf{x}^T \boldsymbol{\beta}) = \lambda_0(t) \cdot \exp(\beta_1 x_1) \dots \exp(\beta_p x_p) \quad (1)$$

- Kennzeichnende Eigenschaft: Proportionalität der Hazardraten $\lambda(t, \mathbf{x}_1)$ und $\lambda(t, \mathbf{x}_2)$ zweier Individuen mit Kovariablen \mathbf{x}_1 und \mathbf{x}_2

$$\frac{\lambda(t, \mathbf{x}_1)}{\lambda(t, \mathbf{x}_2)} = \frac{\lambda_0(t) \cdot \exp(\mathbf{x}_1^T \boldsymbol{\beta})}{\lambda_0(t) \cdot \exp(\mathbf{x}_2^T \boldsymbol{\beta})} = \exp((\mathbf{x}_1 - \mathbf{x}_2)^T \boldsymbol{\beta}) \quad (2)$$

→ PH-Modell: auf log-Skala paralleler Verlauf der Hazardraten

① Einleitung

② Grundlagen

Hazardrate und Survivalfunktion

Bekannte Verteilungsmodelle

Regressionsmodelle: AFT- und Cox-Modell(e)

③ Flexible parametrische Survivalmodelle

Vom Cox-Modell zu flexiblen parametrischen Modellen

Exkurs: (Restringsierte) kubische Splines

Spline-basierte parametrische Survivalmodelle

Bestimmung der Knotenanzahl und -lokation der Splinefunktion

Schätzmethode

Erweiterung um Zeit-abhängige Variablen

④ Datenbeispiel

⑤ Fazit

⑥ Literatur und Anhang

- Weitere Spezifizierung der Baseline-Hazard des Cox-Modells durch begründete Annahme einer speziellen Form/ einer bekannten Verteilung (z.B. Weibullverteilung) → parametrisches Modell
- Beispielhafte Annahme der Weibull-Verteilung als Baseline-Hazard:
 - Nachfolgend zumeist Betrachtung der log-kumulativen Hazardskala; im Falle der Weibull-Verteilung ergibt sich:

$$\ln[\Lambda(t)] = \ln\{-\ln[S(t)]\} = \alpha \cdot [\ln(\lambda) + \ln(t)] \quad (3)$$

- Entsprechend gilt für das Weibull-PH-Modell:

$$\ln[\Lambda(t|\mathbf{x})] = \alpha \cdot [\ln(\lambda) + \ln(t)] + \mathbf{x}^T \boldsymbol{\beta} \quad (4)$$

- Lineare Wirkung der Baseline ggf. zu wenig flexibel

→ Modellierung der Baseline-Hazard durch (restringierte) kubische Splines

- 1 Einleitung
- 2 Grundlagen
 - Hazardrate und Survivalfunktion
 - Bekannte Verteilungsmodelle
 - Regressionsmodelle: AFT- und Cox-Modell(e)
- 3 **Flexible parametrische Survivalmodelle**
 - Vom Cox-Modell zu flexiblen parametrischen Modellen
 - Exkurs: (Restringsierte) kubische Splines**
 - Spline-basierte parametrische Survivalmodelle
 - Bestimmung der Knotenanzahl und -lokation der Splinefunktion
 - Schätzmethode
 - Erweiterung um Zeit-abhängige Variablen
- 4 Datenbeispiel
- 5 Fazit
- 6 Literatur und Anhang

Exkurs: (Restringsierte) kubische Splines

- Spline als stückweise polynomiale Funktion durch vorgegebene Knotenpunkte mit zusätzlichen Glattheitsanforderungen
- Stückweise polynomiale Funktion in diesem Fall kubische Parabel
- Restringsierter kubischer Spline impliziert Linearität vor dem ersten und nach dem letzten Knotenpunkt
- Konstruktion restringierter kubischer Splinefunktion mit Knoten k_1, \dots, k_K über $K - 1$ Basisfunktionen

$$s(q) = \alpha_0 + \alpha_1 z_1(q) + \alpha_2 z_2(q) + \dots + \alpha_{K-1} z_{K-1}(q), \quad (5)$$

$$z_1(q) = q \quad \text{und} \quad z_j = (q - k_j)_+^3 - \phi_j (q - k_1)_+^3 - (1 - \phi_j) (q - k_K)_+^3$$

$$\phi_j = \frac{k_K - k_j}{k_K - k_1} \quad \text{und} \quad j = 2, \dots, K - 1$$

- 1 Einleitung
- 2 Grundlagen
 - Hazardrate und Survivalfunktion
 - Bekannte Verteilungsmodelle
 - Regressionsmodelle: AFT- und Cox-Modell(e)
- 3 Flexible parametrische Survivalmodelle**
 - Vom Cox-Modell zu flexiblen parametrischen Modellen
 - Exkurs: (Restringsierte) kubische Splines
 - Spline-basierte parametrische Survivalmodelle**
 - Bestimmung der Knotenanzahl und -lokation der Splinefunktion
 - Schätzmethode
 - Erweiterung um Zeit-abhängige Variablen
- 4 Datenbeispiel
- 5 Fazit
- 6 Literatur und Anhang

- Spezifiziere Baseline-Hazard durch restringierte kubische Splinefunktion für $q = \ln(t)$
- Konstruktion des log-kumulativen Hazardmodells bzw. - durch Transformation - der Survival- und Hazardfunktion nachfolgenden Form:

$$\ln[\Lambda(t|\mathbf{x})] = \eta = s[\ln(t)|\alpha, \mathbf{k}_0] + \mathbf{x}^T \beta \quad (6)$$

$$S(t|\mathbf{x}) = \exp[-\exp(\eta)] \quad (7)$$

$$\lambda(t|\mathbf{x}) = \frac{\partial s[\ln(t)|\alpha, \mathbf{k}_0]}{\partial t} \cdot \exp(\eta) \quad (8)$$

- Beispiel: Flexibles PH-Modell auf log-kumulativer Hazardskala mit $K = 4$ Knoten:

$$\ln[\Lambda(t|\mathbf{x})] = \eta = \alpha_0 + \alpha_1 z_1(\ln(t)) + \alpha_2 z_2(\ln(t)) + \alpha_3 z_3(\ln(t)) + \mathbf{x}^T \beta$$

- Verwendung weiterer flexibler Modelle auf anderen Skalen: Proportional-Hazards-, Proportional-Odds-, Probit-Modell
- Basis dieser Klasse von Modellen durch Transformation der Survivalfunktion durch eine Link-Funktion $g(\cdot)$:

$$g[S(t|\mathbf{x})] = g[S_0(t)] + \mathbf{x}^T \boldsymbol{\beta} \quad (9)$$

$S_0(t)$: Baseline-Survivalfunktion entsprechend gewählter Skala

- log-kumulative Hazardfunktion des PH-Modells:

$$g[S(t|\mathbf{x})] = \ln\{-\ln[S(t|\mathbf{x})]\} = \ln[H_0(t)] + \mathbf{x}^T \boldsymbol{\beta} = s(\mathbf{x}) + \mathbf{x}^T \boldsymbol{\beta} \quad (10)$$

- log-kumulativen Odds des PO-Modells:

$$g[S(t|\mathbf{x})] = \ln[S(t|\mathbf{x})^{-1} - 1] = \ln[O_0(t)] + \mathbf{x}^T \boldsymbol{\beta} = s(\mathbf{x}) + \mathbf{x}^T \boldsymbol{\beta} \quad (11)$$

- 1 Einleitung
- 2 Grundlagen
 - Hazardrate und Survivalfunktion
 - Bekannte Verteilungsmodelle
 - Regressionsmodelle: AFT- und Cox-Modell(e)
- 3 Flexible parametrische Survivalmodelle**
 - Vom Cox-Modell zu flexiblen parametrischen Modellen
 - Exkurs: (Restringsierte) kubische Splines
 - Spline-basierte parametrische Survivalmodelle
 - Bestimmung der Knotenanzahl und -lokation der Splinefunktion**
 - Schätzmethode
 - Erweiterung um Zeit-abhängige Variablen
- 4 Datenbeispiel
- 5 Fazit
- 6 Literatur und Anhang

Bestimmung der Knotenanzahl und -lokation der Splinefunktion

- Problematik hinsichtlich Anzahl gewählter Knoten:
Bias-Varianz-Trade Off
→ In der Praxis zumeist Wahl von zwei bis fünf Stützstellen
- Bestimmung der Lokation häufig anhand von äquidistanten Knotenpunkten, die sich durch die s.g. Perzentil-basierte Positionierung ergeben

Knoten	Freiheitsgrade	Perzentile/Stützstellen				
1	2	50				
2	3	33	67			
3	4	25	50	75		
4	5	20	40	60	80	
5	6	17	33	50	67	83

- 1 Einleitung
- 2 Grundlagen
 - Hazardrate und Survivalfunktion
 - Bekannte Verteilungsmodelle
 - Regressionsmodelle: AFT- und Cox-Modell(e)
- 3 Flexible parametrische Survivalmodelle**
 - Vom Cox-Modell zu flexiblen parametrischen Modellen
 - Exkurs: (Restringsierte) kubische Splines
 - Spline-basierte parametrische Survivalmodelle
 - Bestimmung der Knotenanzahl und -lokation der Splinefunktion
 - Schätzmethode**
 - Erweiterung um Zeit-abhängige Variablen
- 4 Datenbeispiel
- 5 Fazit
- 6 Literatur und Anhang

- Basierend auf voller Maximum-Likelihood-Methode (Full-Information-Maximum-Likelihood $\hat{=}$ FIML)
- Sample aus n unabh. Beobachtungen der Form $\{t_i, \delta_i, \mathbf{x}_i\}_{i=1, \dots, n}$
 δ_i als Zensierungsindikator
- Likelihoodfunktion flexibles PH-Modell:

$$L_i = \begin{cases} \frac{1}{t_i} \cdot \frac{\partial s[\ln(t_i)|\alpha, \mathbf{k}_0]}{\partial \ln(t_i)} \cdot \exp[\eta_i - \exp(\eta_i)], & \text{unzensiert} \\ \exp[-\exp(\eta_i)], & \text{sonst} \end{cases} \quad (12)$$

- Likelihoodfunktion flexibles PO-Modell:

$$L_i = \begin{cases} \frac{1}{t_i} \cdot \frac{\partial s[\ln(t_i)|\alpha, \mathbf{k}_0]}{\partial \ln(t_i)} \cdot \exp(\eta_i)[1 + \exp(\eta_i)]^{-2}, & \text{unzensiert} \\ [1 + \exp(\eta_i)]^{-1}, & \text{sonst} \end{cases} \quad (13)$$

- 1 Einleitung
- 2 Grundlagen
 - Hazardrate und Survivalfunktion
 - Bekannte Verteilungsmodelle
 - Regressionsmodelle: AFT- und Cox-Modell(e)
- 3 Flexible parametrische Survivalmodelle**
 - Vom Cox-Modell zu flexiblen parametrischen Modellen
 - Exkurs: (Restringsierte) kubische Splines
 - Spline-basierte parametrische Survivalmodelle
 - Bestimmung der Knotenanzahl und -lokation der Splinefunktion
 - Schätzmethode
 - Erweiterung um Zeit-abhängige Variablen
- 4 Datenbeispiel
- 5 Fazit
- 6 Literatur und Anhang

Erweiterung um Zeit-abhängige Variablen

- Erweiterung durch multiplikative Interaktion zwischen der jeweiligen Kovariable \mathbf{x} und der für diese angenommene Abhängigkeitsstruktur $f(t)$
- Zur flexiblen Ermittlung der Zeitabhängigkeit wiederum Verwendung von Splinefunktionen
- Verwendung gleicher Knotenzahl für Modellierung Zeit-abhängiger Effekte wie zur Schätzung der Baseline-Effekte führt häufig zur Überparameterisierung ersterer
- log-kumulativen Hazardrate bei gleichzeitiger Berücksichtigung von M Zeit-abhängigen Variablen:

$$\ln[\Lambda(t|\mathbf{x})] = \begin{cases} s[\ln(t)|\alpha, \mathbf{k}_0] + \mathbf{x}^T \beta, & \text{für } M = 0 \\ s[\ln(t)|\alpha, \mathbf{k}_0] + \sum_{m=1}^M s[\ln(t)|\gamma_k, \mathbf{k}_m] \cdot \mathbf{x}_m \\ \quad + \mathbf{x}^T \beta, & \text{für } M \geq 1 \end{cases}$$

Gliederung

- 1 Einleitung
- 2 Grundlagen
 - Hazardrate und Survivalfunktion
 - Bekannte Verteilungsmodelle
 - Regressionsmodelle: AFT- und Cox-Modell(e)
- 3 Flexible parametrische Survivalmodelle
 - Vom Cox-Modell zu flexiblen parametrischen Modellen
 - Exkurs: (Restringsierte) kubische Splines
 - Spline-basierte parametrische Survivalmodelle
 - Bestimmung der Knotenanzahl und -lokation der Splinefunktion
 - Schätzmethode
 - Erweiterung um Zeit-abhängige Variablen
- 4 Datenbeispiel
- 5 Fazit
- 6 Literatur und Anhang

- „Node-positive primary breast cancer“ Studie, Durchführung von German Breast Cancer Study Group
- Rekrutierung (Juli 1984 bis Dezember 1989) von 720 Patienten für die umfangreiche Kohorten-Studie zu nodalpositivem Brustkrebs
- Vollständige Daten von 686 Patienten hinsichtlich Rückfall-freier Ereignisdauer
- Anhand vorangehender Analysen durch Sauerbrei und Royston Unterteilung in drei Prognosegruppen (Good, Medium, Poor) möglich

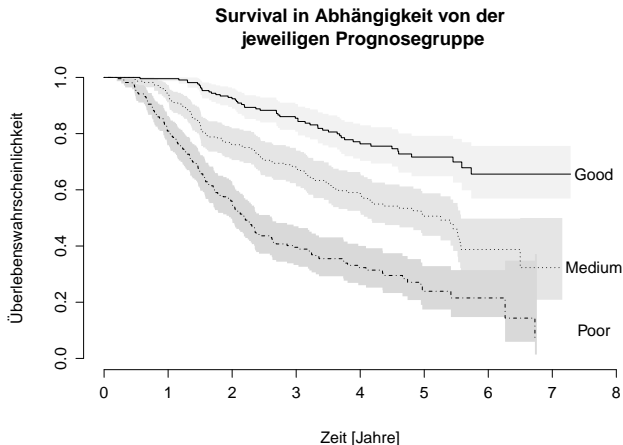


Abbildung: BC-Studie: Kaplan-Meier Survival-Kurven für die drei Prognosegruppen „Good“, „Medium“ und „Poor“; grau hinterlegt: punktweises Konfidenzintervall

Datenbeispiel

- Anwendung Cox-Modell: Referenzkat. Prognosegruppe „Good“
 $HR_{Medium} = 0.84(SE : 0.17)$, $HR_{Poor} = 1.62(SE : 0.16)$
- Modellgleichung des flexiblen PH-Modells:

$$\ln[\Lambda(t|\mathbf{x})] = \eta = \alpha_0 + \sum_{j=1}^K \alpha_j z_j(\ln(t)) + \mathbf{x}_{Medium}^T \beta_{Medium} + \mathbf{x}_{Poor}^T \beta_{Poor}$$

d.f.	PH-Modell			PO-Modell		
	$\hat{\beta}_{Medium}$	$\hat{\beta}_{Poor}$	AIC	$\hat{\beta}_{Medium}$	$\hat{\beta}_{Poor}$	AIC
1	0.847	1.672	1631.9	1.093	2.260	1606.1
2	0.834	1.612	1595.7	1.052	2.171	1588.0
3	0.834	1.612	1595.0	1.052	2.162	1588.5
4	0.836	1.613	1594.0	1.054	2.159	1587.6

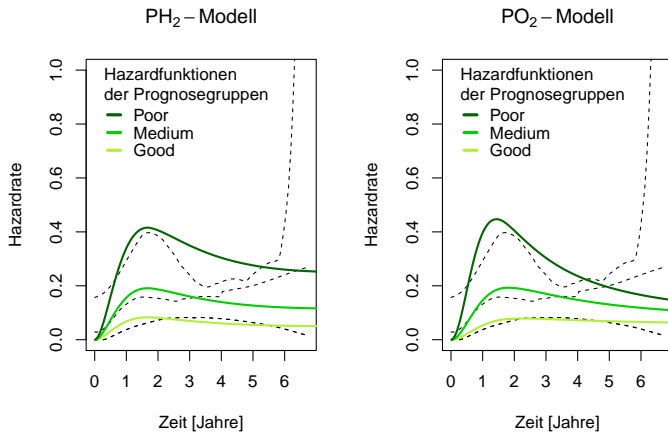


Abbildung: BC-Studie: Hazardfunktionen für die drei Prognosegruppen „Good“, „Medium“ und „Poor“ jeweils für das PH₂-, sowie das PO₂-Modell; gestrichelt: non-parametrische Schätzung

Gliederung

- 1 Einleitung
- 2 Grundlagen
 - Hazardrate und Survivalfunktion
 - Bekannte Verteilungsmodelle
 - Regressionsmodelle: AFT- und Cox-Modell(e)
- 3 Flexible parametrische Survivalmodelle
 - Vom Cox-Modell zu flexiblen parametrischen Modellen
 - Exkurs: (Restringsierte) kubische Splines
 - Spline-basierte parametrische Survivalmodelle
 - Bestimmung der Knotenanzahl und -lokation der Splinefunktion
 - Schätzmethode
 - Erweiterung um Zeit-abhängige Variablen
- 4 Datenbeispiel
- 5 Fazit
- 6 Literatur und Anhang

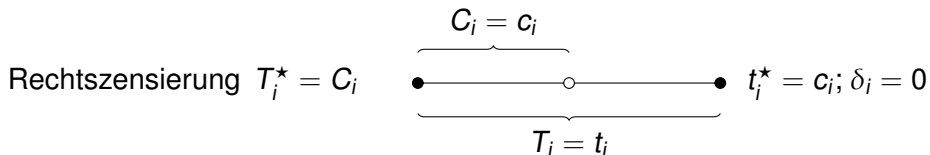
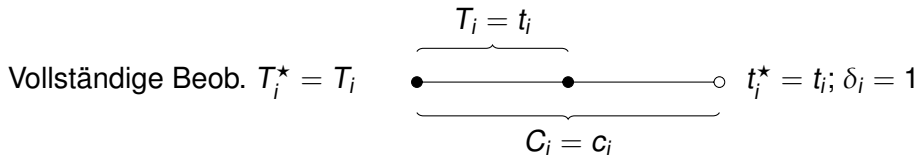
- Innerhalb der Survivalanalyse eine breite Palette an Methoden für die Untersuchung von Ereigniszeiten gegeben
→ Insbesondere breite Auswahl an Modelle (parametrischer, semi-parametrischer und non-parametrischer Art)
- Durch die Einführung flexibler parametrischer Modelle „Verfeinerungen“ bereits existierender Modelle
- Insbesondere durch das flexible PH-Modell Behebung der häufig kritisierten Schwachstelle des Cox-Modells
- V.a. hinsichtlich der Baseline-Hazard sind zusätzliche Analysen mittels der flexiblen parametrischen Survivalmodelle empfehlenswert

- 1 Einleitung
- 2 Grundlagen
 - Hazardrate und Survivalfunktion
 - Bekannte Verteilungsmodelle
 - Regressionsmodelle: AFT- und Cox-Modell(e)
- 3 Flexible parametrische Survivalmodelle
 - Vom Cox-Modell zu flexiblen parametrischen Modellen
 - Exkurs: (Restringsierte) kubische Splines
 - Spline-basierte parametrische Survivalmodelle
 - Bestimmung der Knotenanzahl und -lokation der Splinefunktion
 - Schätzmethode
 - Erweiterung um Zeit-abhängige Variablen
- 4 Datenbeispiel
- 5 Fazit
- 6 Literatur und Anhang

- P. Royston und M. K. B. Parmar (2002). *Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects*. *Statistics in Medicine* 21(15), 2175-2197.
- P. C. Lambert und P. Royston (2009). *Further development of flexible parametric models for survival analysis*. *The Stata Journal* 9(2), 265-290.
- J. P. Klein und M- L- Moeschberger (2003). *Survival Analysis - Techniques for Censored and Truncated Data* (2.Auflage). *Statistics for Biology and Health*. New York: Springer-Verlag
- G. Kauermann (2014). *Vorlesungsskript: Analyse von Lebensdauern - Wintersemester 2014/15 (Version vom 23.01.2015)*.
- L. Fahrmeier L. und T. Kneib und L.S. (2009). *Regression - Modelle, Methoden und Anwendung*. *Statistik und ihre Anwendungen*. Heidelberg: Springer-Verlag.

(Rechts-) Zensurierung

- Wesentliche Unterscheidung: Links-, Intervall- und Rechtszensurierung
- Rechtszensurierung:
 - Interessierendes Ereignis ist bis zum Ende des Beobachtungszeitraumes (noch) nicht eingetreten
 - Definition der Größen Lebensdauer T_i und Zensierungszeit C_i für ein Individuum $i = 1, \dots, n \rightarrow T_i^* = \min(T_i, C_i)$
 - Zufallsvariable $\delta_i = I\{T_i \leq C_i\}$ als Indikator von Unzensiertheit



AFT-Modelle (Accelerated-Failure-Time)

- Modellierung der Lebensdauer $T > 0$ mittels transformierter Zielvariable $\mathbf{y} = g(T)$, sodass:

$$\mathbf{y} = g(T) = \mathbf{x}^T \boldsymbol{\beta} + \sigma \cdot \epsilon \quad \text{bzw.} \quad \mathbf{y} = g(T) = f(\mathbf{x}, \boldsymbol{\beta}) + \sigma \cdot \epsilon \quad (14)$$

- Betrachte insbesondere die Transformation $\mathbf{y} = g(T) = \log(T)$, womit eine Umwandlung von (14) - für den linearen Fall - vorgenommen werden kann:

$$\begin{aligned} \mathbf{y} = \log(T) \Leftrightarrow T &= \exp(\mathbf{x}^T \boldsymbol{\beta}) \cdot \exp(\sigma \epsilon) \\ &= \exp(\beta_0) \exp(\beta_1 x_1) \dots \exp(\beta_p x_p) \cdot \exp(\sigma \epsilon) \end{aligned} \quad (15)$$

- Fehlverteilung von ϵ muss bekannt sein \rightarrow resultierend aus angenommener Verteilung von ϵ : unterschiedliche AFT-Modelle

Verteilung von ϵ bzw. $\log(T)$

$\epsilon \sim N(0, 1)$ normal
Extremwertverteilung
logistisch
log-Gamma
log-gen. Gamma

Verteilung von T

lognormal
Weibull
log-logistisch
Gamma
generalisierte Gamma

- Basierend auf voller Maximum-Likelihood-Methode (Full-Information-Maximum-Likelihood $\hat{=}$ FIML)
- Sample aus n unabh. Beobachtungen der Form $\{t_i, \delta_i, \mathbf{x}_i\}_{i=1, \dots, n}$
 δ_i als Zensierungsindikator
- Likelihoodfunktion flexibles PH-Modell:

$$L_i = \begin{cases} \frac{1}{t_i} \cdot \frac{\partial s[\ln(t_i)|\alpha, \mathbf{k}_0]}{\partial \ln(t_i)} \cdot \exp[\eta_i - \exp(\eta_i)], & \text{unzensiert} \\ \exp[-\exp(\eta_i)], & \text{sonst} \end{cases} \quad (16)$$

- log-Likelihoodfunktion flexibles PH-Modell:

$$l_i = \ln(L_i) = \delta_i \cdot \left[-\ln(t_i) + \ln\left\{ \frac{\partial s[\ln(t_i)|\alpha, \mathbf{k}_0]}{\partial \ln(t_i)} \right\} + \eta \right] - \exp(\eta_i) \quad (17)$$

→ log-Likelihood des gesamten Samples: $l = \sum_{i=1}^n l_i$

- Beurteilung der Modellanpassung anhand des Akaiken Informationskriteriums (AIC) sinnvoll:
 $(-2) \cdot \log - \text{Likelihood} + 2 \cdot \text{geschätzteParameter}$
→ Bevorzugung desjenigen Modells mit niedrigstem AIC-Wert
- Problematik:
 - Komplexität der Modelle durch höhere Knotenzahl führt ggf. zur Bevorzugung zu sparsamer Modelle
 - AIC kein absolutes Gütemaß
 - Innerhalb einer Klasse: kleinere Sets an Knoten nicht unbedingt Untermenge der größeren Sets
→ nicht alle Sub-Modelle einer Klasse geschachtelt innerhalb eines größeren Sets (non-nested)