# Hazard Model and partial Likelihood II

Stefanie Zürcher und Fabian Schnelli

20. Juni 2006

## 1 Einleitung

Im multiplikativen Dichte-Modell ist der Dichteprozess für einen Zählprozess $N_i$ zur Zeit t gegeben durch $l_i(t) = Y_i(t)\lambda\{t|\mathbf{Z}_i(t)\}$, wobei $\mathbf{Z}_i$ ein Kovariablenvektor-Prozess, $Y_i$ ein previsibler $0, 1-$wertiger Prozess und

$$\lambda(t|\mathbf{Z}_i(t))dt = exp\{\beta^T \mathbf{Z}_i(t)\}d\Lambda_0(t) \tag{1}$$

für eine nicht spezifizierte kumulative Hazardfunktion $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$ und einen Vektor von Regressionskoeffizienten $\beta$. Wir haben gesehen, dass dann

$$M_i(t) = N_i(t) - \int_0^t Y_i(s)exp\{\beta^T \mathbf{Z}_i(s)\}d\Lambda_0(s)$$

ein lokales Martingal ist.

Wir betrachten nun einen Spezialfall des multiplikativen Dichte-Modells : das *'Cox proportional hazards model'* mit zeitunabhängigen Kovariablen. Dieses entsteht, wenn zusätzlich folgende Bedingung erfüllt ist:

$$\mathbf{Z}_i(t) = \mathbf{Z}_i \qquad \forall t \geq 0$$

und folgendes gilt:

$$Y_i(t) = \mathbb{1}_{\{X_i \geq t\}} \qquad N_i(t) = \mathbb{1}_{\{X_i \leq t; \delta_i = 1\}}$$

wobei $X_i = T_i \wedge U_i$ ($T_i$ = Ausfalls-Zufallsvariable und $U_i$ = Zensierungszeit-Zufallsvariable) und $\delta_i = \mathbb{1}_{\{X_i = T_i\}}$

Wie wir gesehen haben, wird der Vektor der Regressionsparameter $\beta$ üblicherweise mit dem Maximum-Partial-Likelihood-Schätzer $\hat{\beta}$ und die kumulative Hazardfunktion $\Lambda_0(t)$ mit dem Breslow-Schätzer

$$\hat{\Lambda}_0(t) = \sum_{i=1}^n \int_0^t \frac{dN_i(s)}{\sum_{k=1}^n Y_k(s) \exp\{\hat{\beta}^T \mathbf{Z}_k(s)\}} \tag{2}$$

geschätzt.

## 2 Martingal-Residuen

Wir möchten nun ein Diagnosetool für solche Regressionsanalysen einführen und zwar die *Martingal-Residuen*. Diese sind folgendermassen definiert:

$$\widehat{M}_i(t) := N_i(t) - \int_0^t Y_i(s) \exp\{\hat{\beta}^T \mathbf{Z}_i(s)\} d\hat{\Lambda}_0(s) \qquad i = 1, \ldots, n$$

und $\widehat{M}_i(\infty) := \widetilde{M}_i$.
Im 'Cox proportional hazards model' ist dies äquivalent mit:

$$\widetilde{M}_i = \delta_i - \int_0^{X_i} \exp\{\hat{\beta}^T \mathbf{Z}_i(s)\} d\hat{\Lambda}_0(s) \qquad i = 1, \ldots, n$$

Das heisst, das Martingal-Residuum eines Individuums i, $\widehat{M}_i(t)$, ist für jedes t die Differenz zwischen der beobachteten und der erwarteten Anzahl von Ereignissen über [0,t].

Die Martingal-Residuen haben die folgenden schönen Eigenschaften:

1. $\sum_{i=1}^n \widehat{M}_i(t) = 0$ für alle t $\in [0, \infty]$
2. $\mathbb{E}[\widehat{M}_i] = 0$
3. $\mathrm{Cov}(\widehat{M}_i, \widehat{M}_j) = 0$ für grosse Proben

Neben den Martingal-Residuen gibt es auch noch andere Residuen, z.B. Score-Residuen, Deviance-Residuen, Schönfeld-Residuen,.... Wir werden aber neben den Martingal-Residuen nur kurz auf die Score-Residuen eingehen. Die Übrigen behandeln wir hier nicht.

Diese Residuen sind gute Diagnosetools. Mit ihrer Hilfe können insbesondere die folgenden vier Punkte beurteilt werden:

- Transformationen einer Kovariable in einem Modell, welches für alle übrigen Kovariablen bereits angepasst ist, d.h. es kann die am besten passende Funktionenform für diese zusätzliche Kovariable bestimmt werden.
- Einfluss, den jedes einzelne Individuum auf die Parameterschätzung hat.
- Genauigkeit eines Modells bei der Vorhersage der Ausfallsrate für einen gegebenen Fall.
- Gültigkeit der 'Proportional Hazards'-Bedingung.

Die Martingal-Residuen eignen sich nicht für alle oben erwähnten Anwendungen. Dies liegt vor allem an der verzerrten Verteilung. Wenn $N_i(\infty) \leq 1$, wie z.B. beim Cox-Modell, ist der maximale Wert, den $\widehat{M}_i$ annehmen kann 1 und der minimale $-\infty$.

Wir werden nun einige Anwendungen der Martingal-Residuen genauer ansehen.

# 3 Transformation einer Kovariablen

In diesem Kapitel wollen wir die Theorie etwas verallgemeinern. Daher betrachten wir nun das Modell

$$\lambda\{t|X, \mathbf{Z}(t)\}dt = h(X)exp\{\beta^T \mathbf{Z}(t)\}d\Lambda_0(t) := exp\{f(X)\}exp\{\beta^T \mathbf{Z}(t)\}d\Lambda_0(t) \tag{3}$$

dabei sei $exp\{\beta^T \mathbf{Z}(t)\}$ bekannt, die Funktion $h(X)$ (bzw. $f(X)$) jedoch nicht. Die Variable $X$ sei zudem Zeitunabhängig.
Sei $\widehat{M}$(t) das Martingalresiduum das aus dem Modell (3) entsteht, wenn X ausgelassen wird:

$$\begin{aligned} \lambda\{t|\mathbf{Z}(t)\}dt &= exp\{\tilde{\beta}^T \mathbf{Z}(t)\}d\widetilde{\Lambda}_0(t) \\ \mathbf{Z}(t) &:= \mathbf{Z} \end{aligned}$$

$$\begin{aligned} &\Rightarrow \mathbb{E}[\widehat{M}(t)|X] = \mathbb{E}[N(t)|X] - \mathbb{E}\left[\int_0^t Y(s)e^{\widehat{\beta}^T \mathbf{Z}}d\widehat{\Lambda}_0(s)|X\right] \\ &= \mathbb{E}[N(t)|X] + \underbrace{\mathbb{E}\left[-\int_0^t Y(s)e^{\beta^T \mathbf{Z}}\bar{h}(s,\mathbf{Z})d\Lambda_0(s)|X\right]}_{:=T_2} + \underbrace{\mathbb{E}\left[\int_0^t Y(s)\{d\Lambda(s|\mathbf{Z}) - e^{\hat{\beta}^T \mathbf{Z}}d\widehat{\Lambda}_0(s)|X\right]}_{:=T_3} \end{aligned} \tag{4}$$

wobei $d\Lambda(s|\mathbf{Z}) = e^{\beta^T \mathbf{Z}} \frac{\mathbb{E}\left[e^{f(X)}Y(s)|\mathbf{Z}\right]}{\mathbb{E}[Y(s)|\mathbf{Z}]} d\Lambda_0(s) := e^{\beta^T \mathbf{Z}}\bar{h}(s,\mathbf{Z})d\Lambda_0(s)$ die Ausscheidungsrate von N(s) ist, wenn X ignoriert wird.

$$\Rightarrow T_3 = -\mathbb{E}\left[\sum_{i=1}^n \int_0^t \frac{Y(s)e^{\hat{\beta}^T \mathbf{Z}}}{\sum_{j=1}^n Y_j(s)e^{\hat{\beta}^T \mathbf{Z}_j}} \left\{dN_i(s) - Y_i(s)\frac{e^{\beta^T \mathbf{Z}}}{e^{\hat{\beta}^T \mathbf{Z}}} e^{\hat{\beta}^T \mathbf{Z}_i}\bar{h}(s,\mathbf{Z})d\Lambda_0(s)\right\}\middle| X\right]$$

wegen der Martingalstruktur konvergiert $T_3$ in Wahrscheinlichkeit gegen 0, falls

$$\frac{e^{\beta^T \mathbf{Z}}}{e^{\hat{\beta}^T \mathbf{Z}}} e^{\hat{\beta}^T \mathbf{Z}_i}\bar{h}(s,\mathbf{Z}) \approx e^{\beta^T \mathbf{Z}_i}\bar{h}(s,\mathbf{Z}_i)$$

Sei nun $\bar{h}(s) = \mathbb{E}[\bar{h}(s,\mathbf{Z})]$ und $t_0$ ein fixierter Zeitpunkt.

$$\begin{aligned} \Rightarrow T_2 &= -\mathbb{E}\left[\int_0^t Y(s)e^{\beta^T \mathbf{Z}}(\bar{h}(t_0) - \bar{h}(t_0) + \bar{h}(s,\mathbf{Z}))d\Lambda_0(s)|X\right] \\ &= -\mathbb{E}\left[\int_0^t Y(s)e^{\beta^T \mathbf{Z}}\frac{\bar{h}(t_0)}{h(X)}h(X)d\Lambda_0(s)|X\right] + \underbrace{\mathbb{E}\left[\int_0^t Y(s)e^{\beta^T \mathbf{Z}}(\bar{h}(t_0) - \bar{h}(s,\mathbf{Z}))d\Lambda_0(s)|X\right]}_{:=R(t,X)} \\ &= -\frac{\bar{h}(t_0)}{h(X)}\mathbb{E}[N(t)|X] + R(t,X) \end{aligned}$$

$$\Rightarrow \mathbb{E}[\widehat{M}(t)|X] \approx \left(1 - \frac{\bar{h}(t_0)}{h(X)}\right)\mathbb{E}[N(t)|X] + R(t,X) \tag{5}$$

Das heisst also, falls man $R(t,X)$ als Restterm betrachtet, dass $\mathbb{E}[\widehat{M}(t)|X]$ etwa gleich (1 - Hazardrate) mal der erwarteten Anzahl Ereignisse ist.

Aus (5) erhält man

$$
\begin{aligned}
-log\left(1-\frac{\mathbb{E}[\widehat{M}(t)|X]}{\mathbb{E}[N(t)|X]}\right) &\approx (f(X)-log[\bar{h}(t_0)])-log\left(1-\frac{h(X)R(t,X)}{\bar{h}(t_0)\mathbb{E}[N(t)|X]}\right)\\
&= f(X)-log\left(\frac{\mathbb{E}\left[\int_0^t \bar{h}(s,\mathbf{Z})Y(s)h(X)e^{\beta^T\mathbf{Z}}d\Lambda_0(s)|X\right]}{\mathbb{E}\left[\int_0^t Y(s)h(X)e^{\beta^T\mathbf{Z}}d\Lambda_0(s)|X\right]}\right)\\
&:= f(X)-log\ \tilde{h}\\
&:= f(X)-\bar{f}
\end{aligned}
$$

Falls nun $\bar{f}$ nur wenig variiert im Vergleich zu f, dann haben wir einen akzeptablen Schätzer für f gefunden. Wenn $X$ und $\mathbf{Z}$ nur schwach korreliert sind, ist dies meist der Fall.

# 4 Überprüfung von Modellannahmen

In einem Modell werden jeweils gewisse Annahmen gemacht. Im folgenden Abschnitt geht es darum Methoden zu zeigen, mit welchen man diese Annahmen überprüfen kann.
Insbesondere wollen wir uns mit Methoden zur Überprüfung der 'Proportional Hazards'-Annahme beschäftigen, da diese eine der grundlegenden Annahmen für das Cox-Modell ist.
Die *'Proportional Hazards'*-Annahme besagt folgendes:

$$
\frac{\lambda_i(t)}{\lambda_j(t)}=\frac{\lambda_0(t)\exp\{\beta^T\mathbf{Z}_i\}}{\lambda_0(t)\exp\{\beta^T\mathbf{Z}_j\}}=\frac{\exp\{\beta^T\mathbf{Z}_i\}}{\exp\{\beta^T\mathbf{Z}_j\}}=k(\mathbf{Z}_i,\mathbf{Z}_j) \qquad i\neq j \tag{6}
$$

d.h., die sogenannte *Hazard Ratio* ist unabhängig von der Zeit.

## 4.1 Erste Methode: log(-log S)-Plots

Für das Cox-Modell mit zeitunabhängigen Kovariablen gilt:

$$
\begin{aligned}
\lambda(t|\mathbf{Z}_i) &= \lambda_0(t)\exp\{\beta^T\mathbf{Z}_i{}^{p\times 1}\}\\
S(t|\mathbf{Z}_i=0) &:= S_0(t)
\end{aligned}
$$

$$
\Rightarrow S(t|\mathbf{Z}_i)=(S_0(t))^{\exp\{\beta^T\mathbf{Z}_i\}}
$$

und

$$
\begin{aligned}
log(-log(S|\mathbf{Z}_i)) &= \beta^T\mathbf{Z}_i+log(-logS_0(t))\\
&= \beta^T\mathbf{Z}_i+log(\Lambda_0(t))
\end{aligned} \tag{7}
$$

Wir nehmen an, dass wir die 'Proportional Hazards'-Annahme für eine diskrete Variable $Z_{i,p+1}$ in einem Cox-Modell mit den Variablen $\mathbf{Z}_i^{p\times 1}=(Z_{i,1},\ldots,Z_{i,p})^T$ überprüfen möchten. Weiter nehmen wir an, dass die Variable $Z_{i,p+1}$ nur Q verschiedene Werte annimmt. Indem man nun in diese Q Stufen (Strata) von $Z_{i,p+1}$ unterteilt, kommt man zum stratifizierten Modell

$$
\begin{aligned}
\lambda_q(t|\mathbf{Z}_i) &= \lambda_{0q}(t)\exp\{\beta^T\mathbf{Z}_i\}\\
S_q(t|\mathbf{Z}_i) &= (S_{0q}(t))^{exp\{\beta^T\mathbf{Z}_i\}}
\end{aligned}
$$

Das heisst, in diesem Modell gibt es stratum-spezifizierte Baseline Hazardfunktionen $\lambda_{0q}$, der Vektor der Regressionskoeffizienten $\beta$ ist hingegen global.

Mit Hilfe des Schätzers $\hat{S}_{0q}(t) = \exp\{-\hat{\Lambda}_{0q}(t)\}$, wobei $\hat{\Lambda}_{0q}$ die geschätzte kumulative Baseline Hazardfunktion in der q-ten Stufe ist, kann $\{log(-log(\hat{S}_{0q}(t)) : t \geq 0\}$, $q = 1, \ldots, Q$, geplottet werden. Aus (7) folgt, dass $Z_{i+1}$ die 'Proportional Hazards'-Annahme erfüllt, wenn die Q Plots annähernd parallel sind.
Diese Methode kann auch etwas abgeändert werden, indem man $log(\hat{\Lambda}_{0q(t)}) - log(\hat{\Lambda}_{01(t)})$ gegen t plottet, für $q = 2, \ldots, Q$. Wenn die 'Proportional Hazards'-Bedingung erfüllt ist, sollte jede Kurve ungefähr konstant sein.
Beide Plots geben uns nur die Information, dass die Baseline Hazardfunktionen für jedes Stratum nicht proportional sind. Sie geben aber keine detaillierte Information von welcher Art die Abweichung ist.

Bermerkung:
Diese zwei Methoden können auch für eine stetige Kovariable $Z_{i,p+1}$ angewandt werden, indem man die Kovariable in Q disjunkte Strata, $G_1, G_2, \ldots, G_Q$, unterteilt.

## 4.2 Zweite Methode: Methode basierend auf Score Residuen

Wir haben gesehen, dass der *Score-Vektor* $U(\beta)$ folgendermassen definiert ist:

$$U_i(\beta) = \frac{\partial}{\partial \beta_i} log(L(\beta)) \qquad i = 1, \ldots, p$$

Das heisst, für das multiplikative Dichte-Modell mit Dichtefunkion $l_i(t) = Y_i(t)\lambda\{t|\mathbf{Z}(t)\}$ und bedingter Hazardfunktion $\lambda(t|\mathbf{Z}_i(t))dt = exp\{\beta^T\mathbf{Z}_i(t)\}d\Lambda_0(t)$ und p Kovariablen ist der Scorevektor (wobei nur die Information bis zur Zeit t benutzt wird) folgendermassen definiert:

$$U^{p\times 1}(\beta, t) = \sum_{i=1}^{n}\int_0^{\infty}\left[\mathbf{Z}_i(s) - \frac{\sum_{l=1}^{n}\mathbf{Z}_l(s)Y_l(s)\exp\{\beta^T\mathbf{Z}_l(s)\}}{\sum_{l=1}^{n}Y_l(s)\exp\{\beta^T\mathbf{Z}_l(s)\}}\right]dM_i(s)$$

Nun definieren wir:

$$\bar{\mathbf{Z}}(t) := \frac{\sum_{i=1}^{n}\mathbf{Z}_i(t)Y_i(t)\exp\{\hat{\beta}^T\mathbf{Z}_i(t)\}}{\sum_{i=1}^{n}Y_i(t)\exp\{\hat{\beta}^T\mathbf{Z}_i(t)\}}$$

und

$$S_i(t) := \int_0^t (\mathbf{Z}_i(s) - \bar{\mathbf{Z}}(s))d\widehat{M}_i(s)$$

$S_i(t)$ ist das p-dimensionale *Score Vektor Residuum* zur Zeit t für das Individuum i.

Es gilt nun, dass die 'Proportional Hazards'-Bedingung fraglich ist, wenn wir für $\sup_t \sum_{i=1}^{n} S_i(t)$ einen grossen negativen oder positiven Wert erhalten.

Eine approximative Verteilung von $\sup_t \sum_{i=1}^{n} S_i(t)$ kann man herleiten, indem man beobachtet, dass

$$\sum_{i=1}^{n} S_i(t) = U(\hat{\beta}, t)$$

gilt. Da $\hat{\beta}$ der Maximum-Partial-Likelihood-Schätzer basierend auf der Information über [0,∞] ist, gilt $\sum_{i=1}^{n} S_i(t) = 0$ für t=0 und t=∞.

**Satz.** *Sei* $\boldsymbol{U(\beta,t)}$ *der Score Vektor Prozess,* $\hat{\beta}$ *der Maximum Likelihood Schätzer von* $\beta$ *und* $\mathcal{I}(\beta,t)$ *die beobachtete Informationsmatrix, definiert durch:*

$$\mathcal{I} = \sum_{i=1}^{n} \int_0^t \left\{ \frac{\sum_{j=1}^n Y_j(s) \boldsymbol{Z}_j \boldsymbol{Z}_j^T e^{\beta^T \boldsymbol{Z}_j}}{\sum_{j=1}^n Y_j(s) e^{\beta^T \boldsymbol{Z}_j}} - \left( \frac{\sum_{j=1}^n Y_j(s) \boldsymbol{Z}_j e^{\beta^T \boldsymbol{Z}_j}}{\sum_{j=1}^n Y_j(s) e^{\beta^T \boldsymbol{Z}_j}} \right) \left( \frac{\sum_{j=1}^n Y_j(s) \boldsymbol{Z}_j e^{\beta^T \boldsymbol{Z}_j}}{\sum_{j=1}^n Y_j(s) e^{\beta^T \boldsymbol{Z}_j}} \right)^T \right\} dN_i(s)$$

*wobei* $N_i, Y_i, \boldsymbol{Z}_i, i = 1, ..., n$ *unabhängige, identisch verteilte Prozesse sind, welche die Definition des multiplikativen Dichte Modells erfüllen, so dass* $n^{-1}\mathcal{I}(\beta,t)$ *in Wahrscheinlichkeit gegen* $\Sigma(t) = \int_0^t \boldsymbol{v}(\beta,x) s^{(0)}(\beta,x) \lambda_0(x) dx$ *konvergiert.* $s^{(0)}$ *und* $(\boldsymbol{v})_{jk}$ *bezeichnen dabei die Grenzwerte von* $S^{(0)}$ *bzw.* $(\boldsymbol{V})_{jk}$*, in Wahrscheinlichkeit für* $j,k \in \{1,...,p\}$

$$S^{(0)}(\beta,t) = n^{-1} \sum_{i=1}^{n} Y_i(t) exp\{\beta^T Z_i(t)\}$$
$$\boldsymbol{V}(\beta,t) = \frac{n^{-1} \sum_{i=1}^n [\boldsymbol{Z}_i(t) - \boldsymbol{E}(\beta,t)] [\boldsymbol{Z}_i(t) - \boldsymbol{E}(\beta,t)]^T Y_i(t) exp\{\beta^T Z_i(t)\}}{S^{(0)}(\beta,t)}$$
$$\boldsymbol{E}(\beta,t) = \frac{\sum_{i=1}^n \boldsymbol{Z}_i(t) Y_i(t) exp\{\beta^T Z_i(t)\}}{\sum_{i=1}^n Y_i(t) exp\{\beta^T Z_i(t)\}}$$

1. *Sei* $\boldsymbol{B(t)}$ *ein zentrierter Gauss-Prozess mit unabhängigen Zuwächsen und Kovarianzmatrix* $\Sigma(t)$.

$$\textit{Dann:} \quad \frac{1}{\sqrt{n}} \boldsymbol{U}(\hat{\beta},t) \xrightarrow{schwach} \boldsymbol{B}(t) - \Sigma(t)\{\Sigma(\infty)\}^{-1}\boldsymbol{B}(\infty) \tag{8}$$

2. *Falls* $\{\boldsymbol{v}(\beta,t)\}_{jk}$ *proportinal zu* $\{\boldsymbol{v}(\beta,t)\}_{jj}$ *ist für alle k, dann:*

$$\{\mathcal{I}(\hat{\beta},\infty)_{jj}\}^{-1/2}\{\boldsymbol{U}(\hat{\beta},t)\}_j \xrightarrow{schwach} W^0(\sigma_{jj}(t)/\sigma_{jj}(\infty)) \tag{9}$$

   *wobei* $\sigma_{jj}(t) := \{\Sigma(t)\}_{jj}$ *und* $\{W^0(t) : 0 \leq t \leq 1\}$ *die Verteilung einer Brownschen Brücke besitzt.*

Wenn die j-te Komponente des Kovariablenvektors die 'Proportional Hazards'-Annahme(6) efüllt, folgt aus (9), dass die Teststatistik

$$[\mathcal{I}(\hat{\beta}\infty)_{jj}]^{-1/2} \sup_t \sum_{i=1}^{n} S_{ij}(t)$$

asymptotisch dieselbe Verteilung hat wie $\sup_{0 \leq t \leq 1} W^{(0)}(t)$, solange $\{\mathbf{v}(\beta,t)_{jk}\}/\{\mathbf{v}(\beta,t)_{jj}\}$ nicht von t abhängt.