Prof. Dr. U. Stadtmüller
M. Zampiceni

Winter 2012/2013
21.01.2013
Blatt 13

Survival analysis

(Besprechung: Mo., 28.01.2013)

1. Gegeben sei den Datensatz $D = (T_i^*, \delta_i, Z_i)_{i=1}^n$. Die Kovariable Z ist eindimensional (und nimmt nur die Werte 0 und 1 an).

   (a) Sei $r(Z, \beta) = e^{Z\beta}$.
   Bestimme die partielle log-likelihood-Funktion $l(\beta)$, den Score-(Vektor) $U(\beta)$ und die Information(smatrix) $I(\beta)$ in Abhängigkeit von:

$$\begin{aligned} d^{(1)} &= \sum_{j=1}^{R} Z_{\ell_j} = \sum_{i=1}^{n} Z_i \\ R_j^{(0)} &= \#\{i \,|\, T_i^* \geq t_j, Z_i = 0\} \qquad j = 1, \ldots, R \\ R_j^{(1)} &= \#\{i \,|\, T_i^* \geq t_j, Z_i = 1\} \qquad j = 1, \ldots, R \end{aligned}$$

   (b) Lade den Datensatz `dataset.txt` aus der Vorlesungshomepage herunter. Bestimme mithilfe der R-Funktion `coxph(...)` den partiellen Maximum-Likelihood-Schätzer für β und das zugehörige 98%-Konfidenzintervall.
   (Die Variable `coxph(...)$var` könnte nützlich sein)

   (c) Um welchen Faktor unterscheidet sich die 'Sterblichkeit' unter den Patienten in der Gruppe 2 im Vergleich zur Gruppe 1? Kann man mit 98% Sicherheit sagen, dass die gefragte Sterblichkeit höher liegt als die 'Referenzsterblichkeit'?

(2,5+1+1 Punkte)

2. Betrachte wieder das Cox-Modell $h(t|\vec{Z}) = h_0(t)r(\vec{Z}, \vec{\beta})$. Wir wollen die Partial- Likelihoodfunktion anders als in der Vorlesung herleiten.
   Dabei seien rechtszensierte Daten $\{(T_i^*, \delta_i, \vec{Z}_i)_{i=1}^n\}$ gegeben und R_i bezeichne die Menge der Individuen, die zum i-ten Beobachtungszeitpunkt t_i noch unter Risiko stehen. Es wird zusätzlich angenommen, dass keine Bindungen vorliegen.
   Gehe folgendermaßen vor:

   (a) Nimm an, dass die baseline Hazardrate folgende Form hat:

$$h_0(u) = \begin{cases} 0 & u \neq t_j, \\ h_{0j} > 0 & u = t_j, \end{cases}$$

   wobei $0 = t_0 < t_1 < \ldots < t_R$ die Ereigniszeitpunkte seien ($R \leq n$).
   Betrachte die (volle) Likelihoodfunktion im Cox-Modell $L(\vec{\beta}, \vec{h}_0)$: Zeige, dass sie gleich

$$L(\vec{\beta}, \vec{h}_0) = \prod_{j=1}^{R} h_{0j} r(\vec{Z}_{\ell_j}, \vec{\beta}) \exp\left(-h_{0j} \sum_{m \in R_j} r(\vec{Z}_{\ell_j}, \vec{\beta}) \right)$$

   ist. Fixiere nun $\vec{\beta}$ in der Likelihoodfunktion und bestimme die Maximum-Likelihood-Schätzer für $h_{01}, \ldots, h_{0R}$.

(b) Zeige, dass man durch den Schätzer $\hat{h}_0 = \left(\hat{h}_{01}, \ldots, \hat{h}_{0R}\right)^T$ folgenden Schätzer für die (kumulative) baseline Hazardfunktion $H_0(t)$ bekommt:

$$\widehat{H_0}(t) = \sum_{t_i \leq t} \frac{1}{\sum_{m \in \mathcal{R}_i} r(\vec{\beta}^T \vec{Z}_m)} = \sum_{j=1}^{n} \frac{\mathbb{1}\{T_j^* \leq t\} \cdot \delta_j}{\sum_{m \in \mathcal{R}_j} r(\vec{\beta}^T \vec{Z}_m)}.$$

(c) Zeige, dass man den partiellen Maximum-Likelihood-Schätzer $\hat{\vec{\beta}}$ erhält, wenn man $L(\vec{\beta}, \hat{h}_0)$ bezglich $\vec{\beta}$ maximiert, d.h. es gilt

$$\text{argmax}_{\vec{\beta}} L(\vec{\beta}, \widehat{h_0}) = \text{argmax}_{\vec{\beta}} L_P(\vec{\beta}).$$

wobei $L(\vec{\beta})$ die partielle Likelihoodfunktion aus der Vorlesung bezeichnet.
Wie kann man mithilfe von $\hat{H}_0(t)$ und $\hat{\vec{\beta}}$ einen Schätzer für $S(t|\vec{Z})$ konstruieren?

(7,5 Punkte)