Prof. Dr. U. Stadtmüller
M. Zampiceni

Winter 2012/2013
26.11.2012
Blatt 07

Survival Analysis

(Besprechung: Mo., 03.12.2012)

1. Betrachte wieder folgende (rechtszensierte) Überlebenszeiten,

$$\{(2,0),(1,1),(10,0),(3,1),(10,1),(13,1),(4,1),(5,1),(15,0),(1,0),\\ (3,1),(2,0),(12,1),(6,0),(10,1),(3,1),(14,1),(3,1),(6,0),(8,0)\}$$

   (a) Berechne (per Hand) die Werte $\hat{H}_{NA}(t_j)$ und schreibe $\widehat{H}_{NA}(t)$, den (Nelson-Aalen) Schätzer für die kumulierte Hazardfunktion, auf.

   (b) Berechne den (asymptotischen) Schätzer für die Varianz von $\widehat{S}_{KM}(t)$ und für $\hat{H}_{NA}(t)$ an der Stellen $t_0 = \{1, 3, 5\}$ (per Hand).

   (c) Nehmen wir an, dass die Daten Überlebenszeiten (in Jahren) darstellen. In welchen Bereichen liegt (asymptotisch) die Wahrscheinlichkeit, dass ein Patient nach 5 Jahren noch lebt? Bestimme hierzu den 99%-Konfidenzintervall.

(1+2+1 Punkte)

2. Sei W eine Weibull-verteilten Zufallsvariable mit Parametern $\alpha = 1.5$ und $\lambda = 0.4$, und C eine exponentialverteilte Zufallsvariable mit Parameter γ. W stellt die (echte) Überlebenszeit und C die (zufällige) Rechtszensierung. Betrachte die zensierte Überlebenszeit $T = \min\{W, C\}$.

   (a) Simuliere 100 Realisierungen von T, für 3 verschiedene Fälle: $\gamma = 0.1, 0.5, 1$. Stelle die Kaplan-Meier- Schätzern für die 3 Fälle in einem gemeinsamen Plot dar, zusammen mit der theoretischen Survivalfunktion. Erkläre die Ergebnisse.
   Vergleiche dann graphisch die Schätzerfunktionen $\widehat{S}_{KM}(t)$ mit $\widehat{S}_{NA}(t)$, und $\widehat{H}_{KM}(t)$ mit $\widehat{H}_{KM}(t)$ (nur für $C \sim exp(1)$).

   (b) Berechne (und vergleiche graphisch) die Schätzerfunktionen $\widehat{S}_{KM}(t)$ und $\widehat{S}_{NA}(t)$ für $n = 25, 100, 500, 2000$ Realisierungen von T. Was kann man erwarten?

   Hinweis: *mit dem Befehl* `survfit(...)` *wird die Survivalfunktion geschätzt. Um die kumulierte Hazardfunktion graphisch darzustellen, kann man* `plot(..., fun='cumhaz')` *schreiben.*

(2.5+1.5 Punkte)

3. Der Nelson-Aalen- Schätzer für die kumulierte Hazardfunktion ist asympthotisch normalverteilt mit Mittelwert $H(t)$ (die kumulierte Hazardfunktion) und Varianz $\sigma^2_{ANA}(t)$, die durch $\widehat{\sigma^2}_{ANA}(t)$ geschätzt werden kann.

   (a) Sei φ eine streng monoton wachsende Funktion, $\varphi \in C^1((0, +\infty))$. Zeige, dass $\varphi(\widehat{H}_{NA}(t))$ asymptotisch normalverteilt ist mit Mittelwert $\varphi(H(t))$ und Varianz $(\varphi'(H(t)))^2\sigma^2_{ANA}(t)$ (die sich durch $(\varphi'(\widehat{H}_{NA}(t)))^2\widehat{\sigma^2}_{ANA}(t)$ schätzen lässt).

   (b) Gib das $(1-\alpha)-$Konfidenzintervall für $\varphi(H(t_0))$ an einer Stelle $t_0 > 0$ an.

(c) Sei $\varphi(x) = \log(x)$. Leite das sogenannte log-Konfidenzintervall zum Niveau $1-\alpha$ für $H(t_0)$ her: Bilde erstmals das $(1-\alpha)-$Konfidenzintervall für $\log H(t_0)$, und dann den darauf basierten Konfidenzintervall für $H(t_0)$.

(2+1+1 Punkte)