

Survival analysis

(Besprechung: Mo., 11.02.2013)

1. Ein Verfahren zur Erzeugung von Zufallszahlen aus einer stetigen Verteilung mit invertierbarer Verteilungsfunktion F ist das folgende: Ist U eine auf $[0, 1]$ gleichverteilte Zufallsvariable, so gilt $F^{-1}(U) \sim F$.

- (a) Verwende die Beziehung zwischen der kumulierten Hazardrate $H(t)$ und der Verteilungsfunktion $F(t)$, um ein Verfahren zur Erzeugung von Zufallszahlen aus einer Verteilung mit Hazardrate $h(t) = t$ herzuleiten.
- (b) Verwende die Beziehung zwischen der kumulierten Hazardrate $H(t)$ und der Verteilungsfunktion $F(t)$, um ein Verfahren zur Erzeugung von Zufallszahlen aus einer Verteilung mit Hazardrate $h(t) = t \cdot g(\vec{Z}, \vec{\beta})$ herzuleiten.
- (c) Wir möchten nun rechtszensierte Überlebenszeiten für ein Cox-Modell simulieren (mit R). Setze erstmals den Seed mit `set.seed(42)`. Erzeuge mithilfe der Überlegungen aus Teilaufgabe (b) eine Stichprobe vom Umfang $n = 500$ mit Hazardrate:

$$h(t|\vec{Z}) = t \cdot e^{Z_1^2 + 0.5Z_2 - Z_3/3}.$$

Simuliere die Kovariablen Z_1 und Z_2 als gleichverteilt auf $[-3, 3]$, respektiv $[2, 4]$, $Z_3 \sim \text{Bin}(1, 2/3)$ und die Zensierungszeiten als gleichverteilt auf $[0, 6]$.

- (d) Wir möchten für den Datensatz aus (c) ein klassisches Cox-Regressionmodell anpassen: Sind die drei Kovariablen signifikant ($\alpha = 5\%$)? Benutze die `stepAIC` Funktion aus dem Paket `MASS`: Welches Modell wird empfohlen?
- (e) Betrachte das 'empfohlene' Modell (das eine Kovariable nicht enthält), seine Martingal-residuen und seine Deviance-Residuen. Man bekommt die Residuen zum Beispiel mit der Funktion `residuals(mein.modell, type=...)`. Vergleiche graphisch die Residuen mit den 'gefitteten' Werten. Sehen sie um die Null gestreut, heteroschedastisch, einigermaßen unabhängig aus?
- (f) Plote nun die Residuen gegen der anderen Kovariable. Füge jedem Bild auch eine Regressionskurve (der Befehl `scatter.smooth(...)` kann nützlich sein). Definiere eine Transformation dieser Kovariable, $ZZ = f(Z_k)$ (mit einer vernünftigen Funktion f und für den passenden Index k) und bilde ein neues Modell, der auch ZZ enthält. Evaluiere nun dies und vergleiche es mit den anderen.

(9 Punkte)

2. Gegeben sei der Datensatz $D = (T_i^*, \delta_i, \vec{Z}_i)_{i=1}^n$. Es sei $\vec{Z}_i \in \mathbb{R}^4$ ($i = 1, \dots, n$). Formuliere die Nullhypothese um herauszufinden, ob

$$\begin{cases} \beta_1 &= \beta_3 \\ \beta_2 &= 2\beta_4 \end{cases}$$

gilt. Wie ist die Teststatistik unter H_0 verteilt?

(3 Punkte)