Prof. Dr. U. Stadtmüller
M. Zampiceni

Winter 2012/2013
19.12.2012
Blatt 10

Survival analysis

(Besprechung: Mo., 06.01.2013)

1. Zeige, dass

$$Z_1(t_0) = \int_0^{t_0} L(t) d\left(H_{NA}^{(1)}(t) - H_{NA}^{(2)}(t)\right) = \sum_{j:t_j \le t_0} L_j \left(\frac{D_j^{(1)}}{Y_j^{(1)}} - \frac{D_j^{(2)}}{Y_j^{(2)}}\right).$$

(1 Punkt)

2. Man möchte testen, ob eine Überlebenszeit eine gewisse (parametrische) Verteilung hat. Die Nullhypothese ist dann:

$$H_0 : S(t) = S_0(t) = e^{-\int_0^t h_0(v)dv}, \;\; t \in [0, t_0]$$

Überlege Dir, wie man eine Teststatistik analog zur Vorgehensweise beim Zweistichprobenproblem definieren könnte, und wie sie unter der Nullhypothese verteilt ist.

(3 Punkte)

3. Betrachtet werden zwei Gruppen der Größe $n_1 = n_2 = 150$ einer Gesamtpopulation vom Umfang $n = n_1 + n_2$.
Die Überlebenszeiten in der ersten Gruppe seien Weibullverteilt mit Parametern $\alpha_1 = 3$ und $\lambda_1 = 0,8$ und in der zweite Gruppe besitzen die Überlebenszeiten die Hazardrate $h(t) = 3t^2$. Der für beide Gruppen identische und unabhängige Zensierungsprozess C sei $Exp(9/8)$-verteilt.
Erzeuge in `R` Realisierungen der oben beschriebenen Überlebenszeiten zunächst getrennt für beide Gruppen. Fasse anschließend die erzeugten Survivaldaten (beobachtete Überlebenszeit, Zensierungsindikator und Gruppe) als Dataframe zusammen.

   (a) Schätze die Hazardraten für die erste Gruppe: Benutze hierbei die Funktion `muhaz()` aus dem Paket `muhaz`. Die Hilfsdatei zur Funktion könnte nützlich sein. Definiere und plotte:

      i. Die Schätzer für die Hazardrate, die man bekommt, wenn man als Kernfunktion einmal Epanechinkov-, einmal Biquadratic- und einmal Rechteck-Kernfunktionen nimmt. Benutze die gleiche Bandbreite für alle die Messpunkte (setze hierfür `bw.method='g'`): Vergleiche die drei Schätzungen graphisch miteinander und mit der theoretischen Hazardrate.
      Um mehr Informationen über die Schätzungen zu bekommen, kann man ihren `summary` aufrufen.

      ii. Die Schätzer für die Hazardrate (Epanechinkov- Kernfunktion) mit 3 unterschiedlichen Bandbreiten: Einmal die optimale, die man automatisch bekommt; einmal eine kleinere und einmal eine größere Bandbreite (um sie per Hand zu bestimmen kann man das Argument `bw.grid` der Funktion `muhaz` benutzen). Vergleiche graphisch die drei Schätzungen miteinander und mit der theoretischen Hazardrate.

iii. Variiere diesmal das Argument `bw.method` der Funktion `muhaz`. Man kann dabei zwischen global, lokal und k-nearest neighbours wählen. Vergleiche graphisch die drei Schätzungen miteinander und mit der theoretischen Hazardrate. (benutze wieder Epanechinkov- Kernfunktionen)

(b) Führe mithilfe von `survdiff()` einen Test zum Niveau $\alpha = 5\%$ durch, um die beiden Gruppen auf Gleichheit der Hazardraten zu überprüfen.

(c) Mit `survdiff()` kann man auch testen, ob eine gegebene Population aus einer theoretischen parametrischen Verteilung stammt, indem man mithilfe von `offset()` auf der rechten Seite der Formel die theoretischen Überlebenswahrscheinlichkeiten übergibt. Führe einen Test zum Niveau $\alpha = 5\%$ durch, um zu prüfen, ob eine Weibullverteilung mit Parametern α_1 und λ_1 vorliegt. (Verwende eine Gewichtsfunktion Deiner Wahl)

(5.5+1+1.5 Punkte)

Wir wünschen Euch frohe Weihnachten und einen guten Rutsch ins neue Jahr!!!