

Survival analysis

1. In einer Studie ist die Dauer (in Wochen) der Arbeitslosigkeit nach Beendigung einer Zusatzausbildung von Interesse. Um welche Art von Daten handelt es sich in den folgenden Fällen? Gib T_i^* in Abhängigkeit von T_i und den Zensierungsindikator δ_i sowie den Likelihoodbeitrag der Beobachtung an. Nimm dabei an, dass die Ereigniszeit T absolutstetig mit Survivalfunktion S und Dichte f ist und eventuelle zufällige, dazu unabhängige Zensierungen die Dichte g haben.

- (a) Person 1 beginnt nach 3 Wochen ein Arbeitsverhältnis.
- (b) Person 2 stirbt nach 35 Tagen an einem Herzinfarkt, ohne einer Beschäftigung nachgegangen zu sein.
- (c) Person 3 findet schon während der Ausbildung eine Anstellung.
- (d) Person 4 gibt an, irgendwann zwischen der 15. und der 20. Woche nach Ausbildungsende eine Arbeit begonnen zu haben.
- (e) Person 5 war bei Studienende nach einem Jahr noch immer arbeitslos.

Nimm nun an, dass nur dann Daten vorliegen, wenn die Arbeitssuche nach Ausbildungsende länger als ein Viertel Jahr gedauert hat. Welche Datensituation ist hier also anzunehmen?

2. Gegeben sei eine stetige Überlebenszeit T mit Hazardrate

$$h(t) = 2\lambda^2 t, \quad \lambda > 0, t \geq 0.$$

- a) Bestimme die zugehörige kumulierte Hazardrate $H(t)$, Survivalfunktion $S(t)$ und Dichte $f(t)$.
 - b) Es seien nun Daten (T_i^*, δ_i) , $i = 1, \dots, n$ mit rechtszensierten berlebenszeiten T_i^* und Zensierungsindikatoren δ_i gegeben. Bestimme die Likelihoodfunktion $L(\lambda)$, die Loglikelihoodfunktion $\ell(\lambda)$, die Scorefunktion $U(\lambda)$ und die Informationsmatrix $I(\lambda)$ für den unbekannten Parameter λ .
 - c) Bestimme den Maximum-Likelihood-Schätzer $\hat{\lambda}$.
 - d) Aufgrund des Vorgehens bei der Datengewinnung weiß man, dass die Überlebenszeiten nur dann beobachtbar waren, wenn sie größer als 2 waren. Welche Form hat die Likelihoodfunktion jetzt?
3. In einem (parametrischen) Modell für die beiden Risiken T_1 und T_2 sei die gemeinsame Survivalfunktion gegeben durch:

$$S(t_1, t_2) = (1 - t_1)(1 - t_2) \left(1 + \frac{t_1 t_2}{2} \right), \quad t_i \in (0, 1), \quad i = 1, 2.$$

- (a) Bestimme die marginalen Survivalfunktionen S_{T_1} , S_{T_2} und die marginalen Hazardraten $\tilde{h}_{T_1}$, $\tilde{h}_{T_2}$.

- (b) Bestimme die risiko-spezifische Hazardraten $h_1(t)$ und $h_2(t)$ und die kumulierte Inzidenzfunktion $F_1(t)$ für T_1 .
- (c) Sind die zwei Risiken unabhängig?
- (d) Was kann tatsächlich beobachtet werden?

4. Bestimme zu den folgenden rechtszensierten Daten (+ bedeutet zensiert)

10 7 32+ 23 22 6 16 34+ 32+ 25+ 11+ 20+ 19+ 6 17+ 35+ 6 13 9+ 6+ 10+

- (a) Den Nelson-Aalen-Schätzer $\hat{H}(13)$ sowie den zugehörigen Standardfehler $\hat{\sigma}_{NA}(13)$,
- (b) Den Kaplan-Meier-Schätzer $\hat{S}(13)$, sowie
- (c) Das log-log-Konfidenzintervall zum Niveau 1% für $S(13)$.
- (d) Den Zählprozess $N_{\bullet}(11)$ und eine Schätzung für seinen Kompensator an der Stelle $t = 11$, $\hat{\Delta}_{\bullet}(11)$.

5. Um zwei Patientengruppen auf Gleichheit der Hazardraten zu überprüfen, kann die Nullhypothese

$$\mathbf{H}_0 : \{h^{(1)}(t) = h^{(2)}(t) \text{ für alle } t \leq \tau\}$$

mithilfe des log-Rang-Tests getestet werden. Dieser ist in R als χ^2 -Test implementiert, d.h. die Teststatistik

$$T^2 = \frac{Z_1(\tau)^2}{V_{11}(\tau)} \text{ ist unter } \mathbf{H}_0 \text{ asymptotisch } \chi_1^2\text{-verteilt.}$$

Zur Erinnerung: $Z_1(\tau)$ ist definiert als

$$Z_1(\tau) = \int_0^\tau L(t) d(H_{NA}^{(1)}(t) - H_{NA}^{(2)}(t)),$$

mit $L(t) = \frac{Y_1(t)Y_2(t)}{Y_1(t) + Y_2(t)}.$

Zeige, dass $Z_1(\tau)$ sich darstellen lässt als

$$O - \int_0^\tau \frac{Y_1(t)}{Y_{\bullet}(t)} dN^{(\bullet)}(t),$$

wobei O die Anzahl der beobachteten Ausfälle in der ersten Gruppe (bis zum Zeitpunkt τ) ist.

6. Es seien zwei Populationen mit unterschiedlichen Überlebenswahrscheinlichkeiten gegeben. In der ersten Gruppe sei die Überlebenszeit weibullverteilt und unterliege einer festen Zensierung am Zeitpunkt 1.

In der zweiten Gruppe sei die Überlebenszeit exponentialverteilt mit Intensität λ unter zufälligen exponential(γ)-verteilten (rechts-) Zensierungen.

(zur Erinnerung: wenn $T \sim Wei(\alpha, \theta)$ ist, ist $S(t) = e^{-\theta t^\alpha}$; Wenn $T \sim exp(\lambda)$, dann ist $f(t) = \lambda e^{-\lambda t}$)

- (a) Berechne für die erste Gruppe die Verteilung von T^* und für die zweite Gruppe $P(T^* \leq t, \delta = 1)$

- (b) Welchen Test stellt der folgende Output dar? Interpretiere die Ergebnisse

Call:

```
survdifff(formula = Surv(time, delta) ~ group, rho = 1)
```

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
group=1	100	47.8	52.7	0.447	1.43
group=2	100	49.8	44.9	0.524	1.43

Chisq= 1.4 on 1 degrees of freedom, p= 0.232

Gib auch Nullhypothese, Teststatistik und Testentscheidung ($\alpha = 5\%$) an.

7. Im Rahmen einer Studie soll die Dauer (**time**) bis zur Tumorentwicklung (**delta**: 1=Tumor, 0=kein Tumor bis zum Studienende) bei Ratten mit einem Cox-Modell untersucht werden. Falls sich in der Beobachtungszeit kein Tumor gebildet hat, wird die Beobachtung als rechtszensiert angesehen. Als Kovariablen sollen die Behandlung mit einem bestimmten Medikament (**drug**: 1=ja, 0=nein), das Geschlecht der Ratte (**gender**: 1=männlich, 0=weiblich) sowie das Alter (**age**) in Monaten in das Modell aufgenommen werden.

- (a) Formuliere für die gegebenen Daten die Modellgleichung im Cox-Modell (mit $r(\vec{\beta}, \vec{Z}) = \exp(\vec{\beta}^T \vec{Z})$), wobei neben den Haupteffekten der Kovariablen noch die Interaktion von Geschlecht und Medikament berücksichtigt werden soll.
- (b) Zunächst wurden nur die Kovariablen **drug** und **gender** in das Modell aufgenommen. Interpretiere alle Größen des folgenden R-outputs:

Call:

```
coxph(Surv(time,delta) ~ drug + gender)
```

	coef	exp(coef)	se(coef)	z	Pr(> z)
drug	0.718	2.05	0.215	3.33	0.00086
gender	0.376	1.46	0.210	1.79	0.07300

Likelihood ratio test=14.3 on 2 df, p=0.000796 n= 100.

- (c) Gib anhand des R-outputs in b) das relative Risiko einer Tumorerkrankung einer männlichen Ratte mit Verabreichung des Medikaments bezüglich einer weiblichen Ratte ohne Medikament an.
- (d) Wie kann man den Einfluss der Interaktion von Geschlecht und Medikament im Modell aus a) auf Signifikanz zu einem vorgegebenen Niveau α testen? Formuliere die entsprechende Nullhypothese und gib die Teststatistik sowie die Testentscheidung an.
- (e) Wie könnte man prüfen, ob die Kovariable 'age' tatsächlich einen linearen Effekt hat?