Einführung in die Überlebensanalyse: Teil 1

Einführung

Festlegung: Die Ausfallszeit T sei von nun an eine nicht negative Zufallsvariable aus einer homogenen Population.

Die Verteilung von T wird dann durch vier Funktionen charakterisiert:

- die **Überlebensfunktion** S: Wahrscheinlichkeit, dass ein Individuum nach der Zeit t noch lebt
- die **Risikofunktion** λ: Wahrscheinlichkeit, dass ein Individuum mit Alter t das Ereignis das nächste Mal noch erlebt
- die **Wahrscheinlichkeitsdichte** f: Wahrscheinlichkeit, dass das Ereignis T zum Zeitpunkt t stattfindet
- die **durchschnittliche Restzeit** mrl: durchschnittliche Zeit bis zum Ereignis T, gegeben, dass das Ereignus bei t nicht stattfand

Kennen wir eine dieser vier Funktionen, so können die anderen drei eindeutig bestimmt werden.

Die Überlebensfunktion

Def. *Die Überlebensfunktion ist definiert als*

$$S(t) = \mathbb{P}[T > t]$$

Bem.

- $S(t)$ ist eine monoton fallende Funktion mit $S(0) = 1$ und $lim_{t\to\infty} S(t) = 0$
- Falls T stetig ist mit Dichte f, so ist $S(t)$ streng monoton fallend und es gilt

  $$S(t) = 1 - F(t) = \int_t^\infty f(x)dx$$

  In diesem Fall lässt sich $f(t)$ schreiben als

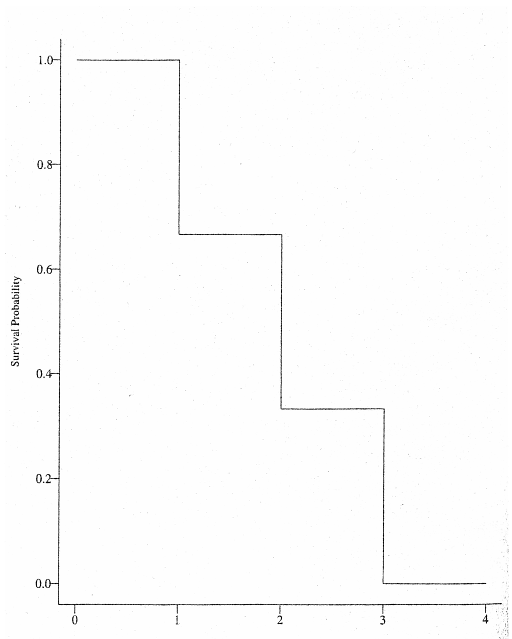
  $$f(t) = -\frac{dS(t)}{dt}$$

- Ist T diskret mit Werten t_j, $j = 1, 2, \ldots$ s.d. $t_1 < t_2 < \ldots$, so ist $S(t)$ eine monoton fallende Treppenfunktion und es gilt:

  $$S(t) = \sum_{t_j > t} \mathbb{P}[T = t_j] =: \sum_{t_j > t} p(t_j)$$

Bsp. Sei T diskret mit Wahrscheinlichkeiten $p(t_j) = \mathbb{P}[T = j] = 1/3,\ j = 1, 2, 3$. Dann hat $S(t)$ die Form

$$S(x) = \mathbb{P}[T > t] = \sum_{t_j > t} p(t_j) = \begin{cases} 1, & \text{wenn } t < 1 \\ 2/3, & \text{wenn } 1 \leq t < 2 \\ 1/3, & \text{wenn } 2 \leq t < 3 \\ 0, & \text{wenn } t \geq 3 \end{cases}$$



Die Risikofunktion

Def. *Die Risikofunktion ist definiert als*

$$\lambda(t) := lim_{\Delta t \to 0} \frac{\mathbb{P}[t \leq T < t + \Delta t | T \geq t]}{\Delta t} \geq 0$$

Bem. Falls T eine stetige Zufallsvariable ist, so gilt:

$$\lambda(t) = f(t)/S(t) = -\frac{d\,ln(S(t))}{dt}$$

Def. *Eine verwandte Grösse ist die* ***kumulierte Risikofunktion:***

$$\Lambda(t) := \int_0^t \lambda(u) du = -ln[S(t)]$$

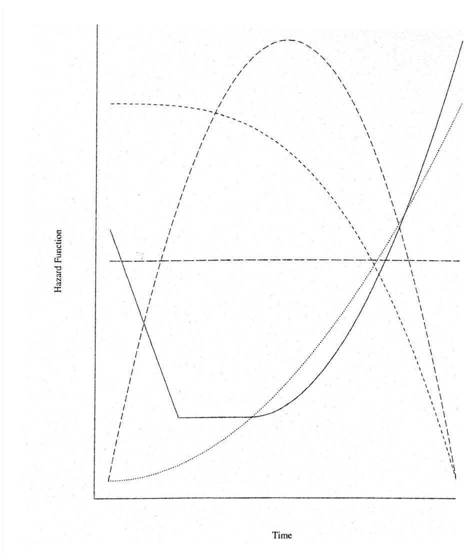
Hierbei muss T wiederum stetig sein und es gilt dann:

$$S(t) = exp[-\Lambda(t)] = exp\left[-\int_0^t \lambda(u)du\right]$$

Bem. Mit der obigen Gleichung sieht man, dass man $\lambda(t)\Delta t$ als eine Näherung nehmen kann für die Wahrscheinlichkeit, dass ein Individuum mit Alter t das nächste Ereignis noch erlebt.

Bsp.

- Monoton wachsende Risikofunktionen sind zum Beispiel Alter und Abnutzung.
- Monoton fallende Risikofunktionen sind viel seltener und fanden vorallem Verwendung bei frühen Ausfalls-Likelihood.
- Eine badewannenförmiges Risiko wird meist bei Populationsmortalität beobachtet. Hier erfolgen während einer frühen Periode Todesfälle, danach stabilisiert sich die Todesrate und steigt schlussendlich aufgrund des natürlichen Alterungsprozesses.
- Schlussendlich gibt es noch die hügelförmige Risikofunktion. Diese beschreibt zum Beispiel die Todesrate nach einer erfolgreichen Operation: Das Risiko steigt zu Beginn aufgrund der Gefahr von Infektionen und anderen Komplikationen an und sinkt dann stetig ab.



Bem. Ist T eine diskrete Zufallsgrösse, so wird die Risikofunktion beschrieben durch

$$\lambda(t_j) = \mathbb{P}[T = t_j | T \geq t_j] = \frac{p(t_j)}{S(t_{j-1})}, \;\; j = 1, 2, \ldots$$

mit $S(t_0) = 1$.
Wegen $p(t_j) = S(t_{j-1}) - S(t_j)$ gilt: $\lambda(t_j) = 1 - S(t_j)/S(t_{j-1}), \;\; j = 1, 2, \ldots$ Die Überlebensfunktion kann also geschrieben werden als

$$S(t) = \prod_{t_j \leq t} S(t_j)/S(t_{j-1}) = \prod_{t_j \leq t} (1 - \lambda(t_j))$$

Bem. Sei $k \leq n$ die Anzahl der distinkten Ausscheidungszeiten $T_1^o < \ldots < T_k^o$.(Die Daten werden häufig in Invervalle gegliedert).Definiere D_j als die Anzahl der bei T_j^o beobachteten Ausscheidungszeiten. Sei $0 = t_0 < t_1 < \cdots < t_m = t$ eine Partition der Intervalls $[0, t]$ und seien d_l bzw. y_l die Anzahl der Todesfälle in $[t_{l-1}, t_l)$ bzw. die Anzahl der Fälle, die nicht vor t_{l-1} scheiterten. y_l wird auch die Anzahl der zur Zeit t_l gefährdeten Fälle genannt.
Für kleine Δt gilt dann:

$$\Lambda(t + \Delta t) - \Lambda(t) \approx \lambda(t)\Delta t \approx \mathbb{P}[t \leq T < t + \Delta t | T \geq t]$$

Somit ist für $y_l > 0$ d_l/y_l ein naiver Schätzer für $\Lambda(t_l) - \Lambda(t_{l-1})$ und

$$\hat{\Lambda}(t) = \sum_{l:t_l \leq t} d_l/y_l$$

Nehme nun an, $m \to \infty$ und $max_{1 \leq l \leq m}|t_l - t_{l-1}| \to 0$. Für grosse m enthält dann jedes Intervall noch höchstens eine distinkte Ausscheidungszeit, für alle anderen Intervalle gilt $d_l = 0$. Im Grenzfall gilt:

$$\hat{\Lambda}(t) = \sum_{k:T_k^o \leq t} D_k/\bar{Y}_k$$

wobei $\bar{Y}_k$ die Anzahl der zur Zeit T_k^o gefährdeten Fälle ist, $k = 1, \ldots, L$.
Da Nelson als erster mit $\hat{\Lambda}$ arbeitete, wird $\hat{\Lambda}$ auch oft der kummulative Nelson-Risiko-Schätzer genannt.
Für $D_k/\bar{Y}_k \approx 0$ gilt:

$$\tilde{S}(t) = exp(-\hat{\Lambda}) = \prod_{k:T_k^o \leq t} exp(-D_k/\bar{Y}_k) \approx \prod_{k:T_k^o \leq t} (1 - D_k/\bar{Y}_k)$$

Bem. Die kumulierte Risikofunktion ist für diskrete Zufallsgrössen definiert als

$$\Lambda(t) = \sum_{t_j \leq t} \lambda(t_j)$$

Die Beziehung $S(t) = exp[-\Lambda(t)]$ gilt in diesem Fall nicht mehr. Aus diesem Grund wird $\Lambda(t)$ auch oft wiefolgt definiert:

$$\Lambda(t) = \sum_{t_j \leq t} ln[1 - \lambda(t_j)]$$

Sind die $\lambda(t_j)$ klein, so ist die zweite Definition eine Näherung der ersten. Hier wird jedoch die erste Definition verwendet werden.

Betrachten wir nochmals unser Beispiel:
Dann gilt für die Risikofunktion:

$$\lambda(t_j) = 1 - S(t_j)/S(t_{j-1}) = \begin{cases} 1/3, & \text{wenn } j = 1 \\ 1/2, & \text{wenn } j = 2 \\ 1, & \text{wenn } j = 3 \\ 0, & \text{sonst} \end{cases}$$

Bem. Für diskrete Zufallsgrössen ist die Risikofunktion überall Null, ausser bei Punkten, bei denen eine Ausscheidung möglich ist.

Die mittlere Restzeit

Def. *Die mittlere Restzeit $mrl(t)$ ist definiert als*

$$mrl(t) = \mathbb{E}[T - t | T > t]$$

Dabei entspricht genau $mrl(0)$ der mittleren Lebenszeit μ.

Bem.

- Für stetige Zufallsvariablen gilt:

$$mrl(t) = \frac{\int_t^\infty (x-t) f(x) dx}{S(t)} = \frac{\int_t^\infty S(x) dx}{S(t)}$$

und

$$\mu = \mathbb{E}T = \int_0^\infty x f(x) dx = \int_0^\infty S(x) dx$$

- Auch die Varianz von T steht in Beziehung zur Überlebensfunktion:

$$Var(T) = 2 \int_0^\infty t S(x) dx - [\int_0^\infty S(x) dx]^2$$