Kapitel 4

Biometrische Rechnungsgrundlagen

In diesem Abschnitt geht es um die Erstellung von *Sterbetafeln*, d.h. von Tabellen, in denen dokumentiert ist, mit welcher Wahrscheinlichkeit Personen in welchem Alter sterben.

Präziser gesagt, sollen Schätzer für die Wahrscheinlichkeit q_x, $x \in \mathbb{N}$, dass eine Person, die ihren x-ten Geburtstag erlebt, vor ihrem $(x+1)$-ten Geburtstag stirbt, hergeleitet werden. Die Wahrscheinlichkeiten q_x heißen *theoretische Sterbewahrscheinlichkeiten.*

4.1 Ermittlung der rohen Sterbewahrscheinlichkeiten

Die *rohen Sterbewahrscheinlichkeiten* sind erste (vorläufige) Schätzer für die theoretische Sterbewahrscheinlichkeiten, die (noch) nicht ausnutzen, dass q_x eine gewisse Stetigkeit in x besitzt.

Wir nehmen an, dass wir einen Bestand an Personen über einen Zeitraum B von n Jahren beobachten.

Wir bezeichnen mit G_y (i.d.R. $y < 0$) die Menge derjenigen Personen in Bestand, die zwischen $y-1$ und y Jahren nach Beobachtungsbeginn geboren werden. Diese sind also zu Beobachtungsbeginn $-y$ Jahre alt.

Für eine Personengruppe G, einen Beobachtungszeitraum B und ein Alter x bezeichne $T_x(B,G)$ die Menge der Personen aus G, die innerhalb von B im Alter x sterben, und $L_x(B,G)$ die Menge der Personen aus G, die ihren x-ten Geburtstag erleben.

Nun ist der Schätzer für q_x nach der *Geburtsjahrmethode* definiert durch

$$\hat{q}_x^G := \frac{\#T_x(B, \cup_{i=1}^{n-1} G_{i-x})}{\#L_x(B, \cup_{i=1}^{n-1} G_{i-x})} = \frac{\sum_{i=1}^{n-1} \#T_x(B, G_{i-x})}{\sum_{i=1}^{n-1} \#L_x(B, G_{i-x})}.$$

Man ermittelt also die Geburtsjahrgänge, für die der x-te und der $x+1$-te Geburtstag (und somit alle Todesfälle im Alter x) in den Beobachtungszeitraum fallen. Dann schätzt man die Sterbewahrscheinlichkeit durch die relative Häufigkeit aller Personen, die in Alter x sterben, unter allen beobachteten Personen dieser Geburtsjahrgänge, die ihren x-ten Geburtstag erleben.

Die Sterbewahrscheinlichkeit ist ungefähr

$$\frac{\sum_{i=0}^{n} \#T_x(B, G_{i-x})}{\frac{1}{2}\#L_x(B, G_{-x}) + \sum_{i=1}^{n-1} \#L_x(B, G_{i-x}) + \frac{1}{2}\#L_x(B, G_{n-x})}.$$

Die Zahl der Personen aus $L_x(B, G_{-x})$ und $L_x(B, G_{n-x})$ werden mit $\frac{1}{2}$ multipliziert, da diese Personen im Schnitt nur ein halbes Jahr innerhalb von B das Alter x haben.

Ein Problem ist, dass sich $\#L_x(B, G_{-x})$ nicht aus den Daten ermitteln lässt. Deshalb nennt man $L_x^A(B)$ die Menge der x-Jährigen, die sich zu Beobachtungsbeginn im Bestand befinden, und $L_x^E(B)$ die Menge der x-Jährigen am Beobachtungsende und approximiert

$$L_x^A(B) - L_x(B, G_{-x}) \approx L_x^E(B) - L_x(B, G_{n-x}).$$

Nun ist der Schätzer für q_x nach der *Sterbejahrmethode* definiert durch

$$\hat{q}_x^G := \frac{\sum_{i=0}^n \#T_x(B, G_{i-x})}{\sum_{i=1}^n \#L_x(B, G_{i-x}) + \frac{1}{2}\#L_x^A(B) - \frac{1}{2}\#L_x^E(B)}.$$

Sowohl Geburts- als auch Sterbejahrmethode sind nur für *geschloßene* Personenbestände geeignet, d.h. während des Beobachtungszeitraums gibt es keine Zugänge und keine Abgänge außer durch Tod.

Vorteile Sterbejahr- gegenüber Geburtsjahrmethode:

- keine Informationen werden "weggeworfen"
- für $n = 1$ anwendbar

Vorteil Geburtsjahr- gegenüber Sterbejahrmethode:

- Ändern sich die Sterbewahrscheinlichkeiten im Laufe der Zeit, bieten die Werte der Geburtsjahrmethode mit $n = 2$ die klarste Interpretationsmöglichkeit.

Nun wollen wir auch Zu- und Abgänge mit berücksichtigen.

Wir zerlegen den Beobachtungszeitraum $B = \cup_{i=1}^n B_i$ in einzelne Jahre $B_i,\ i = 1, \dots, n$. Es bezeichne

- $L_x^A(B_i)$ die Menge der x-Jährigen am Anfang von B_i,
- $L_x^E(B_i)$ die Menge der x-Jährigen am Ende von B_i und
- $T_x(B)$ die Menge der Personen, die innerhalb von B im Alter x sterben.

Nun ist die *Sterbeziffer* definiert durch

$$k_x = \frac{\#T_x(B)}{\sum_{i=1}^n \frac{1}{2}(\#L_x^A(B_i) + \#L_x^E(B_i))}.$$

Die Sterbeziffer ist allerdings als Schätzer für die theoretische Sterbewahrscheinlichkeit ungeeignet, weil im Nenner diejenigen Personen unberücksichtigt bleiben, die nach ihrem x-ten Geburtstag, aber noch vor dem darauffolgenden Jahreswechsel sterben. Dies sind ungefähr $\frac{1}{2}\#T_x(B)$ Personen. Deshalb definiert man den Schätzer nach der *Sterbeziffermethode* für q_x durch

$$\begin{aligned} q_x^Z &= \frac{\#T_x(B)}{\frac{1}{2}\#T_x(B) + \sum_{i=1}^n \frac{1}{2}(\#L_x^A(B_i) + \#L_x^E(B_i))} \\ &= \frac{\#T_x(B)}{(\frac{1}{2}k_x + 1)\sum_{i=1}^n \frac{1}{2}(\#L_x^A(B_i) + \#L_x^E(B_i))} \\ &= \frac{k_x}{\frac{1}{2}k_x + 1} \\ &= \frac{2k_x}{k_x + 2} \end{aligned}$$

Zu- und Abgänge werden nun approximativ berücksichtigt.

Nun wollen wir Zu- und Abgänge exakt berücksichtigen.

Hierzu bezeichnen wir für jede Person j, die irgendwann (innerhalb von B) zum beobachteten Bestand gehört, mit $d_{x,j}$ die *Verweildauer*, d.h. die Länge des Zeitraums innerhalb von B, in dem j zum Bestand gehört und Alter x hat. Stirbt j während der Beobachtung im Alter x, so setzt man $d_{x,j} := 1$.

Weiter bezeichne $L_x(B)$ die Menge der Personen, die zu mindestens einem Zeitpunkt in B zum Bestand gehören und Alter x haben, d.h. die Personen j mit $d_{x,j} > 0$.

Wir definieren nun den Schätzer nach der *Verweildauermethode* für q_x durch

$$\hat{q}_x = \frac{\#T_x(B)}{\sum_{j \in L_x(B)} d_{x,j}}.$$

Die Verweildauer $d_{x,j}$ approximiert dabei die Wahrscheinlichkeit, dass der Tod von j in den Daten erfasst wird, falls j im Alter x stirbt.

Anmerkung: Es ist gerechtfertigt, für Personen, die durch Tod aus dem Bestand ausscheiden, eine höhere als die tatsächliche Verweildauer anzusetzen. Aber der Wert 1 ist zu hoch. Vielmehr müsste geschätzt werden, wie hoch die Verweildauer wäre, wenn die Person nicht gestorben wäre.