

Contents

1	Motivation und Aufgabenstellung	2
1.1	Motivation	2
1.2	Ziele - Aufgabenstellung	3
2	Daten	6
2.1	Die Daten der Sterbetafel 2004R	6
2.2	Öffentlich zugängliche Sterblichkeitsdaten	6
3	Schätzung von Generationstafeln	10
3.1	Die Basistafel der Sterbetafel 2004R	10
3.1.1	Auswahl des Basisjahrs	10
3.1.2	Bestimmung roher Sterbewahrscheinlichkeiten	10
3.1.3	Glättung der Sterbewahrscheinlichkeiten	11
3.1.4	Extrapolation von hohen Altern	11
3.2	Trendfunktionen der Sterbetafel 2004R	11
3.3	Intuitives Modell	12
3.4	Stein-Schätzer	12
4	Ergebnisse der Schätzmethoden für die Mortality.org Daten 1876 - 1987	14
4.1	Ergebnisse der 2004R inspirierten Methoden	14
4.2	Ergebnisse meiner Methoden	14
5	Qualitätsvergleich der Modelle mittels Simulation	15
6	Qualitätsvergleich der Modelle mittels Tests und Theorie	16
7	<i>optional</i> Internationalisierung	17

1 Motivation und Aufgabenstellung

Ziel dieses Kapitels ist es diese Arbeit zu motivieren und die Ziele dieser Arbeit aufzuzeigen.

1.1 Motivation

Die Idee für diese Arbeit ist entstanden, als ich im Winter 2018/2019 ein Praktikum bei einer Lebensversicherung gemacht habe. Dort habe ich eine Programmbibliothek geschrieben, die aus einer Sterbetafel Kommutationssymbole bestimmt.

Eine Sterbetafel ist eine Tafel, ist eine Tabelle, die für ein bestimmtes Kollektiv, die Ausscheidewahrscheinlichkeiten tabelliert. Um diese Aussage mit Inhalt zu füllen, definiere ich im folgenden die Begriffe. Bei einem Kollektiv handelt es sich um eine Menge an Personen, die im folgenden mit P_1, \dots, P_n bezeichnet werden. Jede dieser Personen hat einen Zeitpunkt T_i^g zu dem er geboren ist und einen zufälligen Zeitpunkt T_i^t zu dem er stirbt. Weiterhin ist das Alter X_i beim Ausscheiden aus dem Kollektiv interessant. Mit diesen Definitionen können wir nun 1-jährige Überlebenswahrscheinlichkeiten definieren:

$$q_x = \mathbb{P}(X \leq x + 1 | X > x), x \in \mathbb{N}_0 \quad (1)$$

Das heißt, q_x ist die Wahrscheinlichkeit, dass eine Person, die bereits x Jahre alt geworden ist, vor ihrem $x + 1$ -ten Geburtstag verstirbt. Diese Wahrscheinlichkeiten nennt man dann die Ausscheidewahrscheinlichkeiten durch Tod. Es gibt in der Personenversicherungsmathematik auch noch Invalidität als Ausscheidgrund, doch dieser Ausscheidgrund wird in dieser Arbeit nicht weiter untersucht.

Als nächstes werde ich darauf eingehen, warum Sterbetafeln für Lebensversicherer relevant sind und was Kommutationssymbole sind. In dieser Arbeit wird hauptsächlich die Sterbetafel 2004R verwendet. Dabei steht das R für Rente. Das heißt, diese Tafel wird für Versicherungsprodukte mit Erlebnisfallcharakter, wie die Rente, konzipiert. Erlebnisfallprodukte zeichnen sich dadurch aus, dass die versicherte Person nach eine Leistung erhält, wenn sie einen bestimmten Zeitraum n überlebt. In diesen Produkten, wird die Sterblichkeit als Rechnungsgrundlage neben dem Rechnungszins und den Kosten mit einbezogen, da nur bei Erleben geleistet wird. Konkret werden zwei Arten von Sterbetafeln generiert: Zum einen wird versucht die Sterblichkeiten möglichst genau zu schätzen, man spricht von einer Sterbetafel 2. Ordnung, zum anderen werden Risikoabschläge genommen, um sicher zu gehen, dass die Versicherung ihren Verpflichtungen auf alle Fälle nachkommen kann. Hier spricht man von einer Sterbetafel 1. Ordnung. Für die Konstruktion von Verträgen wird die Tafel 1. Ordnung verwendet. In dieser Arbeit interessiert uns der statistische Aspekt der Versicherungsmathematik, weshalb uns nur die 2. Ordnung interessiert.

An dieser Stelle sei noch kurz erwähnt, dass genau dieser Unterschied in den Sicherheitsabschlägen den Unterschied zwischen Tafeln mit Sterbe- und mit Rentencharakter ausmachen. Für Rentenversicherungen ist es sicherer eine etwas zu geringe Sterblichkeit anzunehmen, während es für Risikolebensversicherungen sicherer ist, eine etwas zu hohe

Sterblichkeit anzunehmen. Deshalb findet man auch in der DAV Tafel 2008T über keine Schätzung von Sterbetrends. Es ist sicherer für Versicherungen mit Todesfallcharakter kein Verbesserung der Sterblichkeit an zu nehmen.

Außerdem soll noch kurz auf Kommutationswerte eingegangen werden: Um Produkte mathematisch präzise beschreiben zu können, werden sogenannte Kommutationssymbole eingeführt. Dies wird am folgenden Beispiel anhand der Einmalprämie einer n -jährigen Erlebnisfallversicherung mit Versicherungssumme 1 Geldeinheit (GE) für eine Person mit Alter x erläutert. Sei diese Einmalprämie mit ${}_nE_x$ bezeichnet. Weiterhin bezeichne l_x die Lebenden im Alter x . Das heißt, es gelte $l_{x+1} = l_x \cdot (1 - p_x)$. Um an dieser Stelle Rundungsfehler zu vermindern, wird normalerweise eine recht große Startkollektivgröße angenommen. Zum Beispiel sei $l_0 = 1.000.000$. Dann ist

$${}_nE_x = \frac{l_{x+n} \cdot v^{x+n}}{l_x \cdot v^x}$$

wobei v den Diskontierungsfaktor bezeichnet. Er sei durch die folgende Gleichung definiert:

$$v^m = \left(\frac{1}{1+i} \right)^m$$

Bei dieser Gleichung bezeichne i den Rechnungszins. Insgesamt sieht man, dass die Einmalprämie also genau die diskontierte Anzahl an Personen am Ende der Aufschubzeit dividiert durch die diskontierte Anzahl an Personen am Beginn der Aufschubzeit ist. Deshalb ist spielt die Startgröße des Kollektivs keine Rolle und kann auf eine beliebige, numerisch angenehme Größe gesetzt werden.

Dieses Beispiel soll dieses Kapitel abschließen. Ich hoffe, dass ich eine gewisse Motivation geben konnte, warum Sterbetafeln von Interesse sind. Im nächsten Abschnitt wird auf die Ziele und die Aufgabenstellung dieser Arbeit eingegangen.

1.2 Ziele - Aufgabenstellung

Wir haben in dem obigen Abschnitt 1.1 gesehen, was Sterbetafeln sind und warum sie relevant sind. In diesem Abschnitt werden die Ziele dieser Arbeit vorgestellt.

Ein Problem bei der Verwendung von Sterbetafeln ist, dass sich die Sterblichkeit im Laufe der Zeit verändert. Genauer gesagt, hat sich die Lebenserwartung in den letzten Jahrzehnten erhöht. Diese Veränderung in der Lebenserwartung muss auch in Sterbetafeln berücksichtigt werden. Betrachten wir dazu zwei Säulendiagramme, die die Entwicklung der Sterblichkeit der Jahre 1990 und 2017 in der Bundesrepublik Deutschland zeigen:

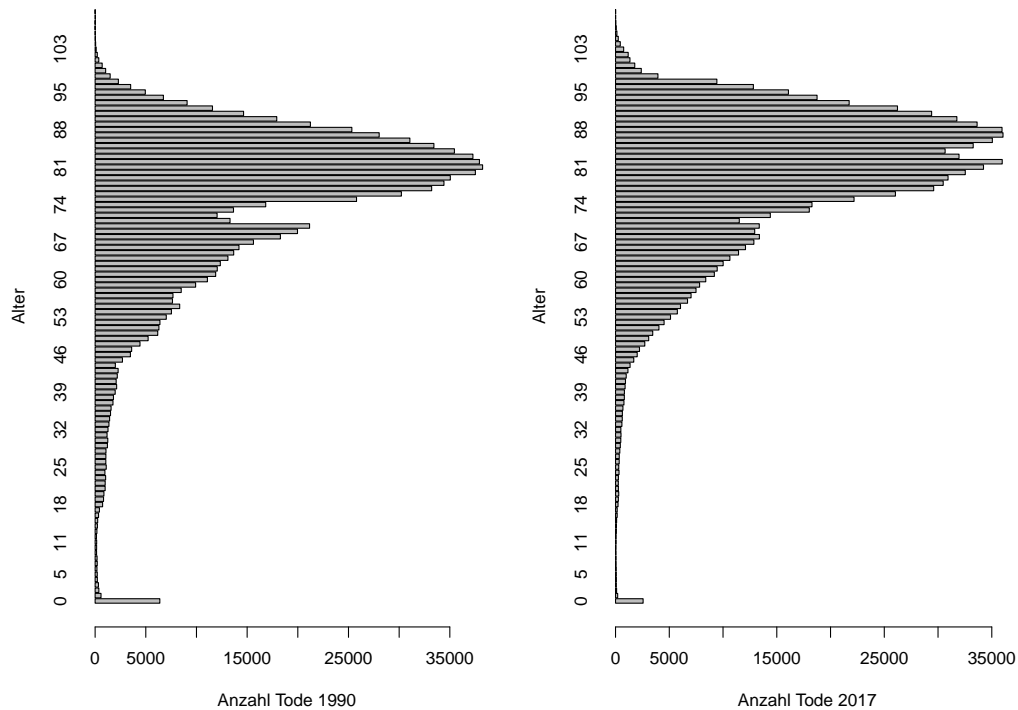


Figure 1: Anzahl an Toden in der Bundesrepublik Deutschland. Auf der linken Seite im Jahr 1990 und auf der rechten Seite im Jahr 2017

Wenn man die beiden Diagramme vergleicht, sieht man, dass im Jahr 2017 im Vergleich zum Jahr 1990 deutlich mehr Personen später versterben. Das Diagramm für das Jahr 2017 ist nach oben, zu den höheren Altern, verschoben. Diese Erhöhung der Lebenserwartung nennt man Sterbetrends. Um diese Trends zu schätzen, geht man in zwei Schritten vor. Im ersten Schritt wird eine Basistafel geschätzt und im zweiten Schritt werden dann Sterblichkeitstrends geschätzt.

Wie man diese Sterblichkeitstrends schätzt ist, wie bereits gesagt, auch in der Praxis sehr relevant. Entsprechend beschäftigt sich auch die Deutschen Aktuarvereinigung (DAV) mit diesem Problem und hat mit der Sterbetafel 2004R eine Richtlinie für deutsche Personenversicherer herausgegeben. Die Dokumentation für die Sterbetafel 2004R, sowie weiterer von der DAV veröffentlichte Sterbetafeln, können unter <https://aktuar.de/unsere-themen/lebensversicherung/Seiten/sterbetafeln.aspx> eingesehen werden.

Das Ziel dieser Arbeit ist es dann, sich Gedanken über die Eigenschaften dieser Schätzer zu machen und welche anderen Möglichkeiten man hat, um diese Trendfunktionen zu schätzen. Außerdem gibt es verschiedene Modelle und die Frage ist, wie man das beste Modell auswählt.

Diese Ziele geben der weitere Aufbau der Arbeit vor. Im nächsten Abschnitt, das heißt

in Abschnitt 2 gehen wir auf die Daten ein, die der Sterbetafel 2004R zugrunde liegen und auf Daten, die öffentlich verfügbar sind. In den nächsten beiden Abschnitten 3 und 5 gehen wir dann auf das Schätzen von Basistafeln, das Schätzen von Sterblichkeitstrends und auf die Auswahl von Modellen ein.

OPTIONAL In dem vorletzten Abschnitt wird die Qualität der Schätzer betrachtet und im letzten Abschnitt schauen wir uns die Situation im internationalen Vergleich an.

Abschließend möchte ich noch kurz auf die Daten und Codes, die zum erstellen der Diagramme (1) verwendet wurden, eingehen. Die Daten stammen von der Onlinedatenbank <https://www.mortality.org/cgi-bin/hmd/country.php?cntr=DEU&level=2>. Weitere Information zu diesen Daten können im Abschnitt 2 gefunden werden. Dieses Histogramm wurde in **R** erstellt. Der Code für dieses Diagramm, sowie alle anderen Codes, die in dieser Arbeit verwendet werden, können auf meiner Github-Seite unter <https://github.com/fake1884/moRtRe> eingesehen werden.

2 Daten

Das Konstruieren von Statistischen Modellen und Schätzern, ohne dass man Daten hat, auf deren Grundlage man arbeiten kann ist nicht sehr sinnvoll. Deshalb werden in diesem Abschnitt Datenbanken vorgestellt, die benutzt werden/man nutzen kann.

2.1 Die Daten der Sterbetafel 2004R

Die Sterbetafel 2004R basiert laut Kapitel 2 der Herleitung auf den folgenden Daten:

- Versichertendaten für die Jahre 1995 bis 2002, die von der Münchner Rück und der GenRe ausgewertet wurden,
- Abgekürzte Bevölkerungssterbetafeln des Statistischen Bundesamtes für das frühere Bundesgebiet ab 1971/73 bis 1998/200 (für 1986/88 stattdessen die Allgemeine Deutsche Sterbetafel 1986/88) und
- Daten der gesetzlichen Rentenversicherung (GRV) für das frühere Bundesgebiet von 1986 bis 2002 für die Alter von 66 bis 98 Jahren

Die Daten der Münchner Rück und der Gen Re sind Privateigentum der jeweiligen Unternehmen und deshalb nicht öffentlich zugänglich. Auch die Daten des Statistischen Bundesamtes beziehungsweise der gesetzlichen Rentenversicherung sind auf ihren jeweiligen Webseiten nicht ohne weiteres öffentlich zugänglich. Allerdings gibt es auf der Internetseite Mortality.org Daten über die Sterblichkeit für viele Staaten. Auf diese Daten wird im nächsten Abschnitt genauer eingegangen.

Allerdings gibt es in dem 2. Kapitel der Herleitung der 2004R noch anderen interessante Punkte, die im folgenden Betrachtet werden sollen: TODO

Wiedervereinigung 1989 problematisch, da die Daten vorher separat erfasst wurden.

Ein weiteres Problem mit den Daten ist, dass es im Jahr 1969/1970 eine Grippewelle in Deutschland gab, die das Datenmaterial verzerrt.

2.2 Öffentlich zugängliche Sterblichkeitsdaten

In diesem Abschnitt werden die Daten von Mortality.org diskutiert. Beginnen wir, indem wir einige Grundlegende Begriffe klären:

- Periodentafel/Periodensterbetafel: Um eine Periodensterbetafel zu erstellen wird, meist über einen relativ kurzen Zeitraum, ein Kollektiv bestehend aus Personen P_1, \dots, P_n betrachtet. Aus diesem Kollektiv werden dann, im einfachsten Fall, die beiden Kennziffern I_x und t_x bestimmt. Dabei bezeichnet I_x die anzahlgewichtete Verweildauer der Lebenden des Alters x . Das heißt, eine Person die ein halbes Jahr nach ihrem x -ten Geburtstag verstirbt, trägt $1/2$ zu I_x bei. Auf der anderen Seite

bezeichnet t_x die Anzahl an Todesfällen von Personen im Alter x . Man bestimmt dann

$$q_x^{\text{roh}} = \frac{t_x}{I_x} \quad (2)$$

als die Rohe Sterbewahrscheinlichkeit. Weiter Details zum Schätzen der q_x erfolgt im Kapitel 3. Hier sei nur noch einmal darauf hingewiesen, dass solche Daten einfach zu erfassen sind. Allerdings werden Sterblichkeiten von gleichzeitig lebenden Personen ermittelt und daraus q_x geschätzt. Das Geburtsjahr geht nicht als Faktor in die Betrachtung mit ein. Deshalb sind solche Tafeln für Versicherungsprodukte von langfristiger Natur ungeeignet. Sie sind ebenfalls ungeeignet, wenn man sich für die Sterblichkeit der Gesamtbevölkerung eines Landes in Zukunft interessiert.

- Kohortensterbetafeln: Im Gegensatz zu Periodentafeln sind Kohortensterbetafeln Tafeln, die das Absterben eines Kollektivs mit Geburtsjahr τ beschreiben. Das heißt, man beginnt mit einem Kollektiv von n Personen $P_1^\tau, \dots, P_n^\tau$ und erfasst jedes Jahr die Anzahl an Todesfällen. Bei einer solchen Tafel ist das Geburtsjahr als Faktor relevant. Leider sind solche Daten nur schwer zu beschaffen. Dies liegt daran, dass es bei der Gesamtbevölkerung eines Landes 2 Eintritts- und Austrittsgründe gibt. Zum einen kommen Personen in das Kollektiv durch Geburt, zum anderen durch Einwanderung. Bei den Austrittsgründen gibt es analog dazu Tod und Auswanderung. Dieses Problem sieht man auch in den Daten: Betrachtet man die Anzahl an Personen in Deutschland, die in einem gewissen Jahr geboren sind und das Risiko haben, zu versterben, so ist diese Anzahl nicht monoton fallend. Das heißt, es kommen in jedem Alter neue Personen in das Kollektiv, das betrachtet wird. Da diese Daten nicht zu beschaffen sind, betrachtet man die nachfolgende Art von Tafeln.
- Generationssterbetafel: Unter einer Generationensterbetafel versteht man eine Sterbetafel, bei der die Sterblichkeit nicht nur vom Alter (und eventuell vom Geschlecht), sondern zusätzlich vom Geburtsjahrgang abhängt. Es handelt sich also im Prinzip um eine Kohortentafel. Der Unterschied ist, dass bei einer Generationstafel nicht am Anfang ein Kollektiv steht, sondern eine Gesamtbevölkerung, bei der auch das Geburtsjahr τ erfasst wird. Somit ist eine Generationstafel eine Mischung aus einer Periodentafel und einer Kohortentafel. Die Daten haben mehr Information, sind allerdings trotzdem erfassbar. Es gibt allerdings nach wie vor das Problem, dass man mit zensierten Daten arbeitet. Das heißt, man keine Beobachtungen darüber, wie Personen mit einem bestimmten Geburtsjahr τ in Zukunft versterben. Dies muss mit statistischen Verfahren ausgeglichen werden.

Werfen wir nun einen Blick auf einen Ausschnitt aus "Brief Summary of the Methods Protocol for the Human Mortality Database", um zu sehen, welche Daten vorliegen. Dieses Dokument kann auf der Webside von Mortality.org gefunden werden. Ich habe die Punkte etwas gekürzt.

- "Births: Annual counts of live births by sex and by month are collected for each population over the longest time period available. At a minimum, a complete series of birth counts by sex is needed for the time period over which mortality rates and period life tables are computed."
- "Deaths: Death counts are collected by sex, completed age, year of birth, and year of death if available. Before making subsequent calculations, deaths of unknown age may be distributed proportionately across the age range."
- "Population size: Below age 80, estimates of population size on January 1st of each year are either obtained from another source or derived using intercensal survival. When necessary, persons of unknown age are distributed proportionately across age. Above age 80, population estimates are used. For non-extinct cohorts aged 80 to 90 at the end of the observation period, population estimates are obtained either from another source or by applying the method of intercensal survival."
- "Exposure-to-risk: Estimates of the population exposed to the risk of death during some age-time interval are based on annual population estimates, with a small correction that reflects the timing of deaths during the interval and variation in cohort's birthdays by month."
- "Death rates: For both periods and cohorts, death rates are simply the ratio of deaths to exposure-to-risk in matched intervals of age and time."

Diese Daten sind von 1990 an für das ganze Bundesgebiet verfügbar und für West-beziehungsweise Ostdeutschland separat für die Jahre 1945 bis 1989.

Weiterhin beachte man, dass manchmal Beobachtungen nicht zuordenbar sind. In diesen Fällen werden die Beobachtungen proportional auf die verschiedenen Kategorien verteilt. Dies kann dann dazu führen, dass man bei Todeszahlen kein ganzen Zahlen bekommt, sondern gebrochene.

Die folgende Tabelle soll einen kurzen Einblick in die Struktur der Daten geben:

Year	Age	Female	Male	Total
1949	0	.	.	.
1949	1	.	.	.
1949	2	.	.	.
1949	3	.	.	.
1949	4	.	.	.
1949	5	.	.	.
1949	6	.	.	.
1949	7	0.000412	0.000714	0.000566
1949	8	0.000363	0.000632	0.000501
1949	9	0.000335	0.000539	0.000440
1949	10	0.000260	0.000484	0.000375
1949	11	0.000298	0.000446	0.000374
1949	12	0.000332	0.000445	0.000390
1949	13	0.000282	0.000432	0.000359
1949	14	0.000315	0.000542	0.000432
1949	15	0.000384	0.000667	0.000529
1949	16	0.000467	0.001051	0.000767
1949	17	0.000500	0.001344	0.000933
1949	18	0.000621	0.001643	0.001145
1949	19	0.000562	0.001676	0.001133
1949	20	0.000571	0.001785	0.001190

Man sieht anhand Tabelle, dass zum Beispiel für den Jahrgang 1949 Daten erst ab einem Alter von 7 Jahren gibt. Es gibt dann Daten bis zu einem Alter von 67 Jahren. Das Endalter bei diesen Tabellen ist 110. Alle Personen, die 110 Jahre oder mehr alt sind, werden in eine Kategorie, genannt 110+, gruppiert.

Man erhält nur einen Ausschnitt, da die Generationstafel aus einer Periodentafel bestimmt wird: Wenn man die Anzahl an Toden von Personen mit Alter und aktueller Jahreszahl hat, kann man zurückrechnen, wie viele Personen mit einem bestimmten Geburtsjahr in welchem Alter verstorben sind. Dies kann man auch dem, oben bereits erwähnten, Dokument "Brief Summary of the Methods Protocol for the Human Mortality Database" entnehmen:

"Raw data are usually obtained in a period format (i.e., by the year of occurrence). In the HMD, most data are presented in a period format, but we also provide death rates and life tables in a cohort format (i.e., by year of birth) if the observation period is sufficiently long to justify such a presentation."

3 Schätzung von Generationstafeln

In diesem Abschnitt wird vorgestellt, wie man Generationstafeln schätzen kann. Dazu wird in den ersten beiden Abschnitten das Vorgehen bei der Sterbetafel 2004R beschrieben. Die Sterbetafel 2004R geht in zwei Schritten vor: Im ersten wird eine Basistafel geschätzt, im zweiten wird auf Grundlage der Basistafel eine Trendfunktion geschätzt. Diese Trendfunktion soll Verbesserungen in der Sterblichkeit modellieren. An diesem Vorgehen finde ich problematisch, dass die Basistafel als Periodentafel erstellt wird. Das heißt bei der Erstellung der Basistafel wird das Geburtsjahr nicht berücksichtigt.

In den folgenden zwei Abschnitten werden zwei alternative Methoden vorgestellt.

3.1 Die Basistafel der Sterbetafel 2004R

Hier wird besprochen, wie die Basistafel der Sterbetafel 2004R geschätzt wird. Bei einer Basistafel handelt es sich um eine Periodensterbetafel bei der ein bestimmtes Kalenderjahr zugrunde gelegt wird. Dazu wird in 4 Schritten vorgegangen: Zuerst muss ein Basisjahr für die Tafel festgelegt werden. Das heißt, es muss ein Jahr bestimmt werden, für die die Tafel gelten soll. Danach werden die rohen Sterbewahrscheinlichkeiten bestimmt und im nächsten Schritt geglättet. Zum Schluss werden Sterbewahrscheinlichkeiten für hohe Alter extrapoliert.

3.1.1 Auswahl des Basisjahrs

-> Zur Auswahl des Basisjahrs siehe Anhang 4

Ich würde einfach den größten Jahrgang nehmen, um die Varianz der Schätzer zu verringern (Trend zum Mittel). Allerdings hat man keine Kohortendaten; was heißt also größten? geburtenstärksten?

3.1.2 Bestimmung roher Sterbewahrscheinlichkeiten

Wie bereits in Abschnitt 2 beschrieben werden zur Bestimmung der rohen Sterbewahrscheinlichkeiten die beiden Kennzahlen I_x und t_x für ein Kollektiv bestimmt. Dabei bezeichnet I_x die anzahlgewichtete Verweildauer der Lebenden des Alters x . Das heißt, eine Person die ein halbes Jahr nach ihrem x -ten Geburtstag verstirbt, trägt $1/2$ zu I_x bei. Auf der anderen Seite bezeichnet t_x die Anzahl an Todesfällen von Personen im Alter x . Man bestimmt dann

$$q_x^{\text{roh}} = \frac{t_x}{I_x} \quad (3)$$

Dieses Verfahren nennt man Intervallmethode. Es findet auch in der Sterbetafel 2008T Anwendung.

3.1.3 Glättung der Sterbewahrscheinlichkeiten

Die rohen Sterbewahrscheinlichkeiten werden mit dem Verfahren von Whittaker-Henderson geglättet:

Beschreibung des Verfahrens: -> TODO

-> Quellen aus 2008T Kapitel 2.2: Kakies et al 1985 "Methodik der Sterblichkeitsuntersuchung" Loebus, 1994 "Bestimmung einer angemessenen Sterbetafel für Lebensversicherungen mit Todesfallcharakter."

-> Quellen aus 2004R Kapitel 3.2.3: KBLOZ S.92 ff = Kakies 1985 "Methodik der Sterblichkeitsuntersuchung" L1 S. 508/509 = Loebus, 1994 "Bestimmung einer angemessenen Sterbetafel für Lebensversicherungen mit Todesfallcharakter."

Danach werden die tatsächlich beobachtete Anzahl von Toten verglichen mit der Anzahl von Toten, die bei Ansatz der ausgeglichenen Sterbewahrscheinlichkeiten zu erwarten wäre.

3.1.4 Extrapolation von hohen Altern

-> Siehe Anhang 9

3.2 Trendfunktionen der Sterbetafel 2004R

Hier wird das Schätzen von Sterblichkeitstrends betrachtet. Im vorherigen Kapitel haben wir gesehen, wie man am besten eine Basistafel schätzt. Ausgehend von dieser Basistafel gibt es verschiedene Möglichkeiten die Sterbetrends zu schätzen. Im folgenden Betrachten wir zuerst das vorgehen bei der Sterbetafel 2004R und danach TODO Dieser Abschnitt basiert auch dem 4. Kapitel der Herleitung zur Basistafel 2004R.

traditionelles Modell:

$$\frac{q_{x,t+1}}{q_{x,t}} = \exp(-F(x)) \quad (4)$$

laut Anhang 11 wird F mittels einer Methode der kleinsten Quadrate durch lineare Regression aus den Werten für $\ln(q_{x,t})$ für $t_0 \leq t \leq t_1$ geschätzt.

Kohortenmodell:

$$\frac{q_{x,t+1}}{q_{x,t}} = \exp(-G(t+1-x)) \quad (5)$$

Synthesenmodell

$$\frac{q_{x,t+1}}{q_{x,t}} = \exp(-F(x) - G(t+1-x)) \quad (6)$$

Die Ergebnisse der Trendschätzung können in DAV 2004R Anhang 1 Teil B gefunden werden.

Alle diese Modelle sind nicht parametrisch. Vielleicht wäre ein linear Parametrisches Modell besser? also $F(x) = ax + b$?

Dann gibt auch noch die Trenddämpfung TODO

3.3 Intuitives Modell

In einem ersten Schritt muss man klären, wie viele Zufallsvariablen man in seinem Modell haben möchte. Zum einen gibt es die Möglichkeit davon aus zu gehen, dass es nur eine Zufallsvariable $X \sim F$ gibt. Wir beobachten dann Realisierungen von $g(X, \tau)$ wobei $\tau \in \{1876, \dots, 1978\}$ ein Geburtsjahr ist. Weiterhin bezeichnet g eine unbekannte Funktion, die die Verbesserung der Sterblichkeit abhängig vom Geburtsjahr τ beschreibt. Dieses Vorgehen führt allerdings dazu, dass eine Funktion g geschätzt werden muss, über man schlechte Annahmen treffen kann.

Deshalb wird im Folgenden von $m = 1978 - 1876 = 102$ Zufallsvariablen ausgegangen. Das heißt für jedes Geburtsjahr τ gibt es eine zugehörige Zufallsvariable $X^\tau \sim F^\tau$. Weiterhin haben wir Beobachtungen der Quantile q_x von X^τ . Allerdings sind unsere Beobachtungen unvollständig. Dies liegt zum einen daran, dass die Daten nicht vorliegen, zum anderen an der Zensur der Daten: Für Geburtsjahr und Alterkombinationen, die in der Zukunft liegen, liegen natürlich noch keine Beobachtungen vor. Deshalb gehen wir von dem folgenden Zusammenhang aus:

$$\frac{q_{x,\tau+1}}{q_{x,\tau}} = f(x, \tau) \quad (7)$$

In diesem Modell beschreibt die Funktion f die Verbesserung der Sterblichkeit. Das heißt wir versuchen in einem ersten Schritt $q_{x,\tau}$ soweit wie möglich zu schätzen. In einem zweiten Schritt schätzen wir f .

Jetzt haben wir ein Modell und können uns Gedanken darüber machen, wie wir die Parameter schätzen. Für die $q_{x,\tau}$ für die Daten vorliegen benutzt kann man direkt die Sterblichkeitsraten von mortality.org nehmen. Die Frage ist jetzt wie man f schätzt.

Ideen zum Schätzen von f :

- $f(x, \tau) = \exp(-F_\tau(x))$ also eine Funktion $F_{\tau_{au}}$ pro Jahrgang τ und dann betrachtet man das ganze als Interpolationsproblem.
- $f(x, \tau) = \exp(-(ax + b) \cdot \tau)$ also ein einfaches lineares Modell. und man schätzt a und b .
- Vielleicht findet man bei Whittaker-Henderson noch eine Inspiration.

Die übrigen Quantile $q_{x,\tau}$ werden dann über die Beziehung (7) geschätzt.

3.4 Stein-Schätzer

Angenommen man hat das Modell $X \sim N_d(\theta, I_d)$ mit unbekanntem Parameter $\theta \in \mathbb{R}^d$. Wir wollen θ schätzen. Dazu haben wir nur eine Beobachtung B von X . James and Stein haben in ihrem Artikel "Estimation with quadratic loss" (1961) gezeigt, dass der Schätzer $\hat{\theta} = B$ nicht zulässig ist. Das heißt, sie haben gezeigt, dass es einen Schätzer mit einem kleineren Risiko gibt. Konkret wird gezeigt, dass der Schätzer

$$T_{JS}(B) = \left(1 - \frac{d-2}{\|B\|^2}\right) B \quad (8)$$

für $d \geq 3$ ein kleineres Risiko hat. Da dieses Ergebnis nicht intuitiv ist, wird es auch als das Stein-Paradoxon bezeichnet.

Wie aber passt der Stein-Schätzer in diese Arbeit? Anstatt davon auszugehen, dass wir zugrundeliegende Zufallsvariablen X^τ haben und ihre Quantile beobachten können wir von folgendem Modell ausgehen: Wir beobachten $Y_x^\tau \sim N(q_x^\tau, 1)$. Das heißt, wir gehen von davon aus, dass wir Realisierungen $Y_x^\tau = q_x^\tau + \epsilon$ beobachten. Hier ist $\epsilon \sim N(0, 1)$. Dann ist der Stein-Schätzer dem intuitiven Ansatz aus dem vorhergehenden Abschnitt überlegen.

Wir können also, alternativ zu dem vorhergehenden Abschnitt auch,

$$\hat{q}_x^\tau = \left(1 - \frac{111-2}{\sqrt{\sum_{i=0}^{110} \hat{q}_i^{\tau_{\text{tau}}}}}\right) \hat{q}_x^\tau \quad (9)$$

schätzen. Dann kann man genau wie im letzten Abschnitt f schätzen und damit die Generationstafeln auffüllen. Man würde vermuten, dass dieser Schätzer einen geringeren Fehler aufweist.

Diese und die vorhergehenden Schätzmethoden sind in dem beiliegenden Paket *moRtRe*, der Name steht für "Mortality Trends", für die Daten aus Kapitel 2 implementiert. Die Ergebnisse werden im nächsten Kapitel präsentiert.

4 Ergebnisse der Schätzmethoden für die Mortality.org Daten 1876 - 1987

Die Idee ist die drei/fünf? Schätzmethoden auf den selben Datensatz an zu wenden. Dann wird in den nächsten beiden Kapiteln verglichen, welche Schätzmethode die beste ist.

4.1 Ergebnisse der 2004R inspirierten Methoden

Wie die Überschrift sagt.

4.2 Ergebnisse meiner Methoden

Wie die Überschrift sagt.

5 Qualitätsvergleich der Modelle mittels Simulation

Qualitätsvergleich über Simulation; dazu hat man mehrere Möglichkeiten:

- Man teilt die Daten in Teildatenbanken D_1 und D_2 auf. Auf Grund von D_1 wird der Schätzer bestimmt. Dann kann man mit dem Schätzer die Werte von D_2 schätzen und mit den wahren Werten vergleichen.
- Man zieht neue Datensätze aus der empirischen Verteilungsfunktion der Daten, die man hat, und benutzt diese neuen Daten, um die Qualität der Schätzer zu bestimmen.
- Nachdem man \hat{q}_x^τ gefunden hat, gibt man $\tilde{q}_x^\tau = \hat{q}_x^\tau + \epsilon, \epsilon \sim N(0, 1)$ für einige x vor und versucht \hat{q}_x^τ wieder zu finden. Dies testet, wie gut die Stein-Methode ist, wenn die Voraussetzungen per Definition erfüllt sind.

6 Qualitätsvergleich der Modelle mittels Tests und Theorie

Eine mögliche Fragestellung ist dann, welche Qualität diese Modelle besitzen:

-> Anhang 11 : Likelihood-Ratio-Test zum Modellvergleich

- Sind sie konsistent?
- Sind sie unverzert und vielleicht UMVU?
- Wie sind die Fehler verteilt?

7 *optional* Internationalisierung

Man könnte sich eventuell auch internationale Sterbetrends anschauen.