

Contents

1	Motivation und Aufgabenstellung	2
1.1	Motivation	2
1.2	Ziele - Aufgabenstellung	3
2	Grundlegenden Definitionen und historische Daten	5
2.1	Definitionen relativer Sterbehäufigkeiten und Beschreibung von Personenge- samtheiten	5
2.2	Ein Problem bei der Schätzung roher Sterbewahrscheinlichkeiten: Welche Personen sind zu berücksichtigen	7
2.3	Weitere Probleme beim Schätzen roher Sterbewahrscheinlichkeiten	9
2.4	Arten von Sterbetafeln	11
2.5	Öffentlich zugängliche Sterblichkeitsdaten	12
3	Einfache Modelle	16
3.1	Geburtsjahr als erklärende Variable	16
3.2	Alter als erklärende Variable	18
3.3	Eigenschaften des einfachen Modells	22
4	Übliche Verfahren zur Gewinnung von Sterbetafeln und die Sterbetafel 2004R	35
4.1	Auswahl des Basisjahrs	35
4.2	Bestimmung roher Sterbewahrscheinlichkeiten	36
4.3	Glättung der Sterbewahrscheinlichkeiten	37
4.3.1	Überblick über Glättungsverfahren	37
4.3.2	Das Verfahren von Whittaker-Henderson	40
4.4	Übliche Methoden zur Trendschätzung	44
5	Verbesserung des Einfachen Modells	56
5.1	Einfaches Trendmodell	56
6	Qualitätsvergleich der Modelle mittels Simulation	59
6.1	Qualitätsvergleich durch Simulation	59
6.2	Qualitätsvergleich der Modelle mittels Theorie	60
7	Ausblick	61
7.1	Stein-Schätzer	61
7.2	Qualitätsvergleich mittels Tests	61
7.3	Extrapolation von hohen Altern	62
7.4	Modellbildung	62
7.5	Verschiedenes	63

1 Motivation und Aufgabenstellung

In diesem Kapitel wird in Abschnitt 1.1 diese Arbeit und insbesondere der Begriff der Sterbetafel motiviert. Dann werden in Abschnitt 1.2 die Ziele dieser Arbeit genauer diskutiert. Allgemein gesagt, ist das Ziel dieser Arbeit die Bestimmung von Eigenschaften und der Vergleich von Sterbetafeln.

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. Um die Begriffe Sterbetafel und Kommutationssymbole mit Leben zu füllen betrachten wir als Beispiel die Einmalprämie einer n -jährigen Erlebnisfallversicherung mit Versicherungssumme 1 Geldeinheit für eine Person mit Alter x . Sei diese Einmalprämie mit ${}_nE_x$ bezeichnet. Um ${}_nE_x$ mathematisch beschreiben zu können benötigen wir einige Definitionen:

Sei also q_x ist die Wahrscheinlichkeit, dass eine Person, die bereits x Jahre alt geworden ist, vor ihrem $x+1$ -ten Geburtstag verstirbt. Man nennt q_x eine 1-jährige Sterbewahrscheinlichkeiten. Aus diesen q_x bestimmt man dann weiter Werte, die zur Produktkalkulation notwendig sind.

Bezeichne also p_x die 1-jährige Überlebenswahrscheinlichkeit. Sie gibt die Wahrscheinlichkeit an, dass eine Person, die bereits x Jahre alt ist, auch noch das nächste Jahr überlebt. Dann gilt in unserem Fall per Definition $p_x = 1 - q_x$, da wir davon ausgehen, dass eine Person entweder stirbt oder nicht. Man kann auch noch andere Ausscheidegründe betrachten.

Weiterhin bezeichne l_x die Lebenden im Alter x . Das heißt, es gelte $l_{x+1} = l_x \cdot p_x = l_x \cdot (1 - q_x)$.

Weiterhin benötigt man in der Versicherungsmathematik regelmäßig einen Rechnungszins, der hier mit i bezeichnet wird. Weiterhin sei der Diskontierungsfaktor v durch definiert:

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

Dann kann man die Einmalprämie einer n -jährigen Erlebnisfallversicherung mit Versicherungssumme 1 Geldeinheit für eine Person mit Alter x über die Gleichung

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

berechnen. Das heißt, man definiert die Einmalprämie also genau als die diskontierte Anzahl an Personen am Ende der Aufschubzeit dividiert durch die diskontierte Anzahl an Personen am Beginn der Aufschubzeit ist. Deshalb und um bei der Berechnung der l_x Rundungsfehler zu vermindern, wird normalerweise eine recht große Startkollektivgröße angenommen. Zum Beispiel sei $l_0 = 1.000.000$. Diese Größe hat keinen Einfluss auf die Prämie, die sie sowohl im Zähler als auch im Nenner vorkommt und sich somit kürzt.

Kommen wir noch kurz auf die Begriffe am Anfang dieses Abschnitts zurück: Eine Sterbetafel ist eine Tabelle, in der 1-jährige Sterbewahrscheinlichkeiten q_x tabelliert sind und die l_x sind ein Beispiel für Kommutionswerte.

Ich hoffe, dass dieses Beispiel 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 ein Beispiel gesehen, warum Sterbetafeln wichtig 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:

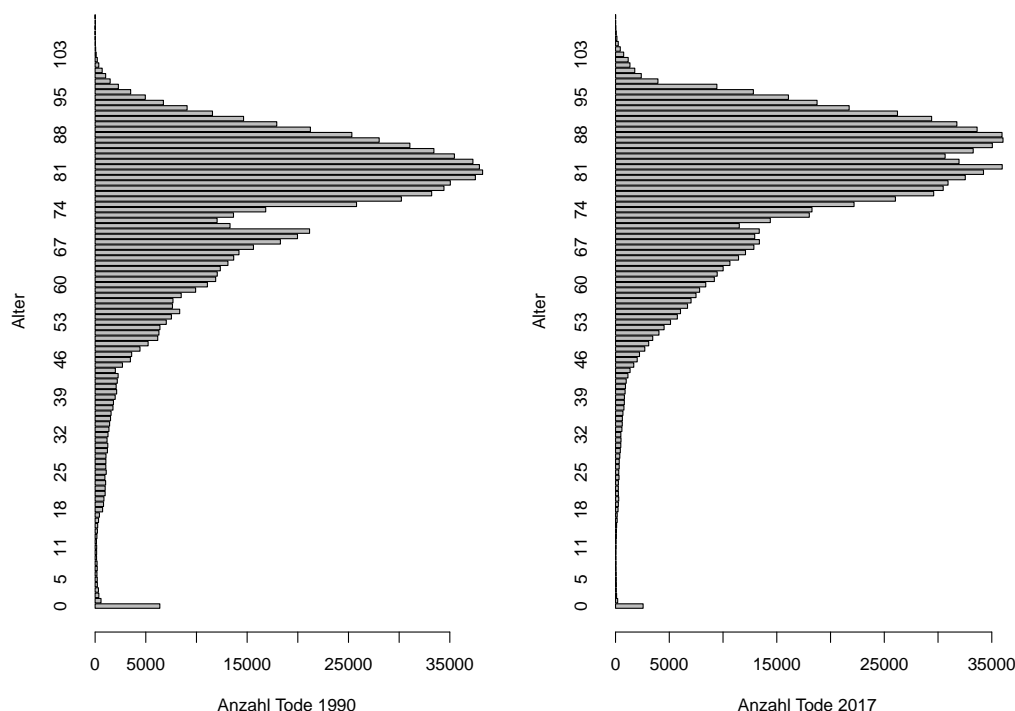


Figure 1: Anzahl an Toden in der Bundesrepublik Deutschland. Auf der linken Seite im Jahr 1990 und auf der rechten Seite im Jahr 2017

Wenn man die beiden Diagramme vergleicht, sieht man, dass im Jahr 2017 im Vergle-

ich zum Jahr 1990 deutlich mehr Personen später sterben. Das Diagramm für das Jahr 2017 ist nach oben, zu den höheren Altern, verschoben. Diese Erhöhung der Lebenserwartung nennt man Sterbetrends. Weil es auch in der Praxis sehr relevant ist, wie man diese Sterblichkeitstrends schätzt, beschäftigt sich auch die Deutschen Aktuarvereinigung (DAV) mit diesem Problem. Mit der Sterbetafel 2004R hat die DAV eine Richtlinie für deutsche Personenversicherer herausgegeben. Die Dokumentation für die Sterbetafel 2004R [DAV04], sowie weiterer von der DAV veröffentlichte Sterbetafeln, können unter <https://aktuar.de/unsere-themen/lebensversicherung/Seiten/sterbetafeln.aspx> eingesehen werden und waren der Ansatzpunkt für diese Arbeit.

Diese Arbeit beginnt im nächsten Kapitel 2 damit grundlegende Definitionen zu geben und historische Daten und ihre Qualität zu diskutieren. Für die Definitionen wurde das Buch "Methodik von Sterblichkeitsuntersuchungen" [Kak85] von Peter Kakies verwendet. Die historischen Daten stammen von der Website <https://www.mortality.org>, [Mor18], wo man auch weitere Informationen zu den Daten finden kann.

Im nächsten Kapitel, 3, werden einfache Modelle betrachtet, die zum Vergleich mit den üblicherweise verwendeten Modellen dienen sollen.

Daraufhin werden in Kapitel 4 bekannte Modelle diskutiert. Insbesondere wird das Whittaker-Henderson Verfahren zur Glättung von Sterbetafeln betrachtet. Dazu wird wieder das Buch "Methodik von Sterblichkeitsuntersuchungen" [Kak85] von Peter Kakies verwendet. Außerdem wurden "An extension of the Whittaker-Henderson method" von N-con und Scott, [NS12], sowie "Bestimmung einer angemessenen Sterbetafel für Lebensversicherungen mit Todesfallcharakter" von Loebus, [Loe94] verwendet. Ein weiteres häufig verwendetes Modell ist das Lee-Carter Modell, das auch in Kapitel 4 diskutiert wird. Für dieses Modell wird die Originalquelle "Modeling and Forecasting U. S. Mortality" von Lee und Carter, [LC92], sowie "Understanding the Lee-Carter Mortality Forecasting Method" von Girosi und King [GK07] verwendet.

In Kapitel 5 wird dann eine Verbesserung für das einfache Modell aus Kapitel 3 vorgeschlagen.

In Kapitel 6 werden dann die Eigenschaften der Modelle verglichen und die Ergebnisse von Simulationen gegeben. Diese Simulationen, sowie fast alle Graphiken in dieser Arbeit wurden 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. Der Name des Pakets "moRtRe" steht für "Mortality Trends" mit großen "R"s für die Programmiersprache **R**.

2 Grundlegenden Definitionen und historische Daten

In diesem Kapitel werden im ersten Abschnitt grundlegende Begriffe wie Kollektiv und relative Sterbehäufigkeiten geklärt. Im nächsten Abschnitt werden Methoden zur Bestimmung von rohen Sterbewahrscheinlichkeiten angegeben. Im daran anschließenden Abschnitt werden methodenunabhängige Einflüsse auf die Gewinnung roher Sterbewahrscheinlichkeiten diskutiert. Diese ersten drei Abschnitte orientieren sich an dem Buch "Methodik von Sterblichkeitsuntersuchungen", siehe [Kak85].

Der nächste Abschnitt behandelt verschiedene Arten von Sterbetafeln. Für diesen Abschnitt gab es keine brauchbare Quelle: In der Literatur werden die verschiedenen Arten von Sterbetafeln nicht präzise definiert.

Im letzten Abschnitt dieses Kapitels werden Daten vorgestellt, auf deren Grundlage Sterbetafeln erstellt wurden und werden. Insbesondere wird auf die öffentlich zugänglichen Sterbetafeln der Website mortality.org eingegangen, siehe [Mor18].

2.1 Definitionen relativer Sterbehäufigkeiten und Beschreibung von Personengesamtheiten

Da das Ziel dieser Arbeit das Schätzen von Sterbewahrscheinlichkeiten ist, müssen wir Notation einführen, um Sterblichkeit zu definieren. Dazu definieren wir zuerst verschiedene Arten von Jahren und was Personen und Personengemeinschaften sind:

- Jahre: Seien $\iota, \tau, \gamma \in \mathbb{N}$ Jahre. Dabei bezeichne ι immer ein Kalenderjahr, τ ein Geburtsjahr und γ immer ein Todesjahr. Daraus werden gleich Personen definiert. Ich verwende γ für das Todesjahr, da wir später eine Regression auf Y dem zufälligen Todesalter machen möchten.
- Personen: Seien P_1, \dots, P_n Personen. Eine Person P_i ist definiert durch ein Geburtsjahr τ_i und durch ein Todesjahr γ_i . Ist $\iota = \gamma_i$ ist die Person P_i im Jahr ι verstorben. Aus diesen Definitionen kann man direkt die folgenden Begriffe ableiten:

$$\begin{aligned}\text{Alter} &= X = \text{Kalenderjahr} - \text{Geburtsjahr} = \iota - \tau, \text{ für } \iota \leq \gamma \\ \text{Todesalter} &= Y = \text{Todesjahr} - \text{Geburtsjahr} = \gamma - \tau \\ \text{Lebenserwartung} &= LE = \mathbb{E}(Y)\end{aligned}$$

Seine weiterhin E und S jeweils End- und Startalter einer Person. Wir setzen in dieser Arbeit $S = 0$ und $E = 110$. Alle Alter größer 110 die beobachtet werden, werden auf 110 gesetzt. Außerdem unterscheiden wir Personen nicht nach Geschlecht.

- Kollektive: Sei weiterhin K_j eine Menge von lebenden Personen P_i . Das heißt, aus $P_i \in K_j$ folgt, dass $\gamma_i < \iota_i$. Alle Personen unter Beobachtung werden mit K bezeichnet. Der Index j dient dazu, Personen in Risikoklassen zu unterteilen. Zum

Beispiel $J = \{1900, \dots, 2000\}$, also eine Unterscheidung nach Geburtsjahren. Sei J immer derart, dass $K = \cup_{j \in J} K_j$. Die Bezeichnung K kommt daher, dass man von einem Kollektiv spricht. Ein Synonym zu Kollektiv ist eine Personengesamtheit. Wir bezeichnen die Kardinalität einer Personengemeinschaft K_j mit k_j also $|K_j| = k_j$ beliebige Personengemeinschaften.

- Austrittsgründe: Es gibt verschiedene Gründe für den Austritt aus einem Kollektiv. Von besonderem Interesse ist der Austritt durch Tod. Andere Austrittsgründe können Invalidität oder Auswanderung sein. Invalidität wird in dieser Arbeit keine Rolle spielen. Wir werden Auswanderung später als Störgröße diskutieren. Man unterscheidet weiterhin offene und geschlossene Personengesamtheit. Falls aus einer Personengesamtheit Personen nur durch Tod ausscheiden, nennen wir diese Personengesamtheit geschlossen.

Führen wir nun Zufallsvariablen (Z.V.) ein: Sei Y als das zufällige Todesalter und y als eine Realisierung von Y . Analog ist das Todesjahr zufällig und wird mit Γ bezeichnet. Eine Realisierung von Γ wird mit γ bezeichnet. Die Notation soll unterstreichen, dass wir Γ_i beziehungsweise Y_i voraussagen wollen, gegeben dass wir für eine Person P_i die Größe τ_i kennen. Außerdem sind wir in einem bestimmten Kalenderjahr ι . Das heißt, wir wollen eine Methode M finden um $Y_i \approx M(\iota_i, x_i)$ sinnvoll zu schätzen. Aus den Größen Γ beziehungsweise Y wird in der Literatur dann q_x abgeleitet:

$$\text{1-jährige-Sterbewahrscheinlichkeit} = q_x = \mathbb{P}(y \leq x + 1 | y > x), x \in [S, E] \subset \mathbb{N}_0 \quad (1)$$

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 1-jährige-Sterbewahrscheinlichkeit

- Teilgesamtheit der Ausgeschiedenen: Um die q_x für $x \in \{S, S+1, \dots, E\}$ zu schätzen, betrachtet man Kollektive und zählt die Personen, die zu einem bestimmten Kalenderjahr ι , das Risiko haben zu sterben und im darauf folgenden Jahr $\iota + 1$, die Anzahl an Todesfällen. Das konkrete vorgehen wird im nächsten Abschnitt definiert. An dieser Stelle sollen nur die beiden Teilmengen T und K' definiert werden: Wir bezeichnen mit T die Teilmenge der Toten aus K und mit K' die Teilmenge der Überlebenden. Seien weiterhin die Kardinalitäten mit t beziehungsweise k' bezeichnet. Es gilt:

$$\text{Kollektiv} = K = T \cup K' = \text{Tote} \cup \text{Überlebende}$$

- Rohe Sterbehäufigkeiten: Mit den Definitionen der Teilgesamtheiten kann man rohe (1-jährige) Sterbewahrscheinlichkeiten schätzen:

$$\hat{q}^{roh} = \frac{t}{k} = \frac{t}{t + k'} \quad (2)$$

Man betrachtet also ein Kollektiv und schätzt daraus die Sterbewahrscheinlichkeit. Das in dieser Arbeit meistens betrachtete Kollektiv ist die west-deutsche Gesamtbevölkerung. Warum genau dieses Kollektiv betrachtet wird, wird in dem Abschnitt 2.5 diskutiert.

- Basistafel und Sterbetafel: Es fehlt noch die Definition einer Sterbetafel beziehungsweise einer Basistafel. Leider werden diese beiden Größen in der Literatur nicht definiert, sondern intuitiv verwendet. Eine Sterbetafel ST ist intuitiv $ST = (q_S, q_1, \dots, q_E)$. Allerdings hat man mit dieser Definition das Problem, dass man gerne Sterbewahrscheinlichkeiten pro Geburtsjahrgang angeben möchte. Das heißt es interessiert ST_1, \dots, ST_m für Indizes $1, \dots, m \in \{\tau_1, \dots, \tau_m\}$. Deshalb sei $B = (q_S, q_1, \dots, q_E)$ eine Basistafel und $ST = (B_1, \dots, B_m)$ eine Sterbetafel.

2.2 Ein Problem bei der Schätzung roher Sterbewahrscheinlichkeiten: Welche Personen sind zu berücksichtigen

Im letzten Kapitel wurde bereits definiert, was man unter einer rohen Sterbewahrscheinlichkeit versteht. In diesem Kapitel werden die verschiedenen Methoden rohe Sterbewahrscheinlichkeiten zu gewinnen diskutiert. Es gibt die Geburtsjahrmethode, die Sterbjahrmethode, das Sterbeziffernverfahren und die Verweildauermethode. Diese vier Methoden existieren, weil nicht ganz klar ist welche Personen zu berücksichtigen sind. Dieser Sachverhalt soll an Graphik 2 veranschaulicht werden. Zu sehen sind die Lebensdauer von drei x -jährigen Personen. Die Lebensdauer der Personen sind durch einen horizontalen Strich dargestellt. Der Beginn zeigt den Eintritt ins Kollektiv durch Erreichen des Alters x und das Ende den Austritt aus dem Kollektiv der x -jährigen. In diesem Fall gehen wir von Tod aus und die Frage ist, welches die mittlere zu erwartenden Lebensdauer einer x -jährigen Person ist. Wir beobachten den Zeitraum 2002 bis 2004, der durch senkrechte Linien eingegrenzt wird. Man sieht, dass Person 1 innerhalb des Beobachtungszeitraums in das Kollektiv eintritt und innerhalb des Beobachtungszeitraums ausscheidet. Das heißt diese Person kann ohne Probleme erfasst werden. Person 2 dagegen ist bereits vor Beginn der Beobachtungsperiode in das Kollektiv eingetreten. Hier stellt sich nun die Frage, ob diese Person erfasst werden sollte. Das Problem ist das es auch Personen wie Person 3 gibt, die gleichzeitig mit Person 2 in das Kollektiv eingetreten sind, allerdings bereits vorher das Kollektiv verlassen haben. Man würde also, wenn man Person 2 mit in die Schätzung aufnimmt und Person 3 nicht, die Sterblichkeit von x -jährigen Personen mit dem selben Geburtsjahr unterschätzen.

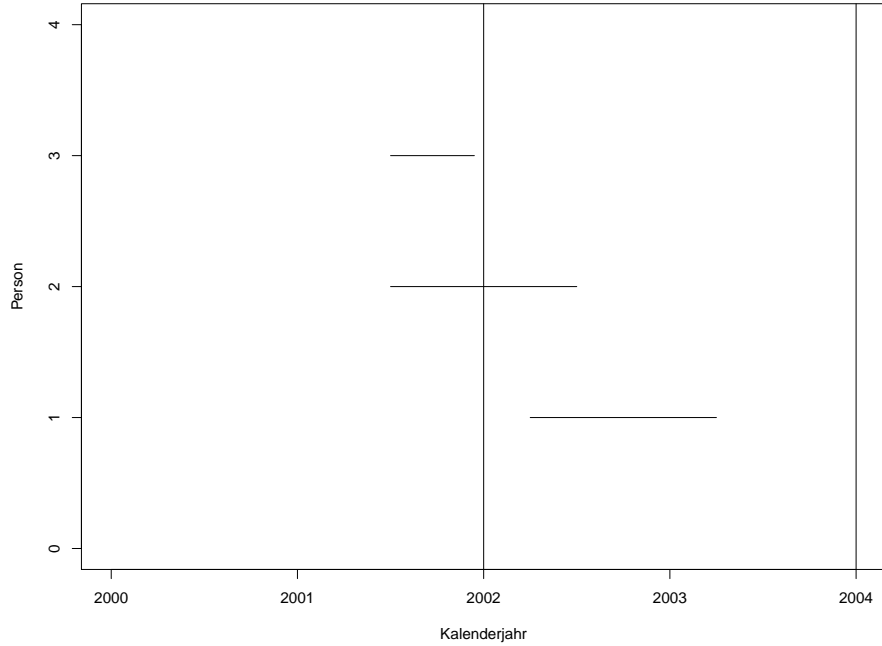


Figure 2: Lebensdauer von verschiedenen x -jährigen Personen. Die Frage ist, welche Personen zu berücksichtigen sind.

Entsprechend gibt es verschiedene Methoden, je nachdem welchen Ansatz man verfolgen möchte. Die Geburtsjahrmethode berücksichtigt nur die Personen, die innerhalb des Beobachtungszeitraums in das Kollektiv eintreten. Das heißt, alle Personen mit dem selben Geburtsjahr werden gleich behandelt und man erhält

$$q_x^G = t_{x,1}(b, g) / l_{x,1}(b, g)$$

dabei bezeichnet der Index G das es sich um die Geburtsjahrmethode handelt und die 1 im Index der Anzahl an Toten und Lebenden, dass die Person ein Jahr zum Kollektiv gehört haben muss. Wir gehen von einer Beobachtungsperiode von mindestens einem Jahr aus. Der Parameter b steht für den Beobachtungszeitraum und g für das Geburtsjahr der Personen, die berücksichtigt werden.

Alternativ gibt es die Sterbejahrmethode, bei der man alle in einem Beobachtungszeitraum gestorbenen berücksichtigt. In Formeln erhält man

$$q_x^S = \frac{t_x(b)}{l_x(b_1/b_2, g_1/g_2) - 1/2 \cdot l_x(b_2, g_2) + 1/2 l_x(b_1, g_0)}$$

wobei eine Reihe an Korrekturen vorgenommen wurden, um bessere Schätzer zu erhalten. Auf diese Korrekturen wird hier nicht weiter eingegangen. Mehr Informationen kann man in [Kak85] finden. In [Kak85] wird als nächstes das Sterbeziffernverfahren beschrieben.

Dieses Verfahren ist technisch einfacher als die Sterbejahrmethode. Es handelt sich um ein Anfangs-Plus-Endbestand-Halbe-Methode. Zuerst muss man die Sterbeziffer definieren:

$$k_x = \frac{t_x(b_1, g_0/g_1)}{1/2 \cdot (l_x(\dot{b}_1) + l_x(\ddot{b}_1))}$$

und erhält damit

$$q_x^Z = \frac{2 \cdot k_x}{2 + k_x}$$

Abschließend gibt es noch das Verweildauerverfahren. Dieses Verfahren erfasst die Teilzeiten, die Personen zu einem Kollektiv gehören und löst damit das Problem welche Personen zu erfassen sind am besten. Allerdings ist es auch das aufwendigste. Personenversicherer sind eventuell in der Lage Sterblichkeiten in ihren Beständen durch diese Methode zu schätzen, aber um Bevölkerungssterblichkeiten zu schätzen ist es zu kompliziert.

$$q_x^D = \frac{t_x(b)}{t_x(b) + \sum_d d \cdot l'_{x,d}(b)}, 0 < d \leq 1$$

Insgesamt lässt sich feststellen, dass die Methoden aufwändiger werden. Dies liegt daran, dass auch Personen erfasst werden, die auch nur einen Teil der Beobachtungsperiode zu dem betrachteten Kollektiv gehören. Das heißt, dass insbesondere die Geburtsjahrmethode geeignet ist, wenn man ein geschlossenes Kollektiv betrachtet, während die Verweildauermethode insbesondere bei offenen Kollektiven geeignet ist.

Als letzten Punkt möchte ich noch erwähnen, dass in Deutschland 1910/1911 und 1932/34 sowie 1949/1951 die Geburtsjahrmethode angewandt wurde. Für die Sterbetafel von 1960/1962 wurde die Sterbejahrmethode angewandt. Das heißt, wenn man Daten aus diesen Sterbetafeln zusammenfasst um zum Beispiel Sterblichkeitstrends zu schätzen, sollte man vermutlich diese Änderung in der Bestimmung der rohen Sterbewahrscheinlichkeit berücksichtigen.

2.3 Weitere Probleme beim Schätzen roher Sterbewahrscheinlichkeiten

In diesem Kapitel werden weitere Einflussgrößen diskutiert, die nicht von der Methode zur Erfassung der rohen Sterbewahrscheinlichkeiten abhängen, aber trotzdem einen Einfluss haben. Es handelt sich um eine Aufzählung interessanter Punkte. Für diese Einflussgrößen werden keine Methoden zur Korrektur entwickelt:

- Jahreszeitliches Schwankungen: Wie man in der folgenden Graphik (3) sieht, schwankt die Sterblichkeit, je nach Jahreszeit. Das heißt, je nachdem zu welchem Monat der Beobachtungszeitraum beginnt und endet hat eventuell einen Einfluss auf die Sterbetafel.

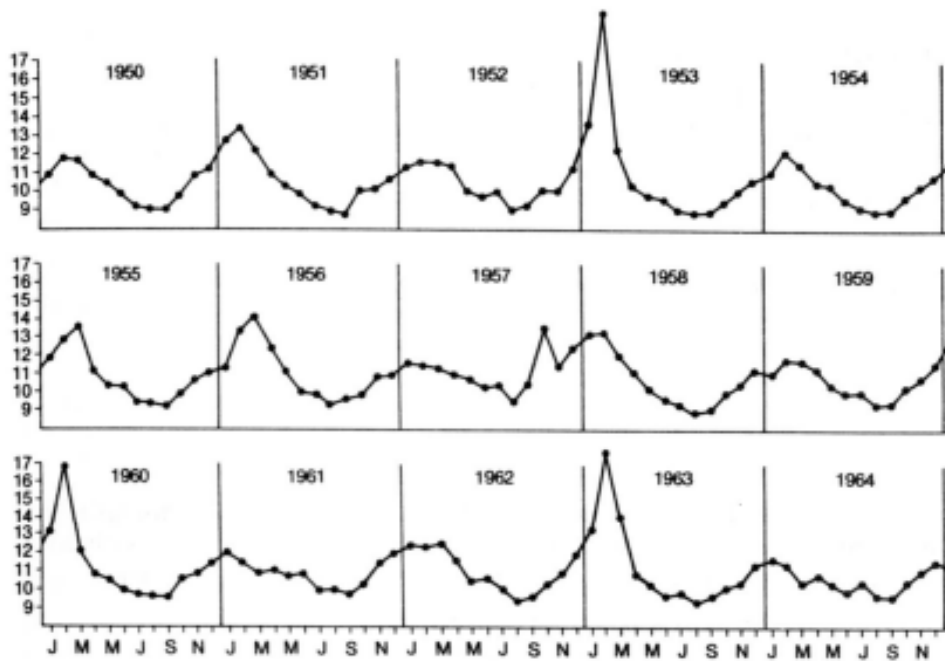


Abb. 10. Monatliche, auf ein Jahr normierte Sterbeziffern 1950–1964 in Prozent.

Figure 3: jahreszeitliche Schwankung der Sterblichkeit; diese Graphik stammt aus "Methodik" [Kak85]

- Extremereignisse: Ein weiteres Problem ist, wie man mit Extremereignissen umgeht. Sowohl die DAV in der Dokumentation zur Sterbetafel 2004R, als auch Lee und Carter ihrer Originalveröffentlichung [LC92] stellen die Frage, wie man mit Extremereignissen umgehen soll. Solche Ereignisse sind zum Beispiel Grippewellen, die durchaus wieder auftreten können.
- Änderung der Datenerfassung: Natürlich hat eine Änderung in der Datenerfassung einen Einfluss auf die Daten. Hier ist insbesondere die Wiedervereinigung Deutschlands 1989/90 problematisch. Zum Beispiel ist in Westdeutschland eine Lebendgeburt durch §31 Absatz 1 der Verordnung zur Ausführung des Personenstandsgesetzes wie folgt geregelt:

„Eine Lebendgeburt, für die die allgemeinen Bestimmungen über die Anzeige und die Eintragung von Geburten gelten, liegt vor, wenn bei einem Kinde nach der Scheidung vom Mutterleib entweder das Herz geschlagen oder die Nabelschnur pulsiert oder die natürliche Lungenatmung eingesetzt hat.“

Dagegen waren in der DDR mindestens zwei Merkmale erforderlich. Dies führt dazu, dass Säuglinge mit nur einem Lebenszeichen und somit schlechter Lebenserwartung

nicht als Lebendgeburt zählen und somit die Säuglingssterblichkeit besser geschätzt wird, als sie ist. Diese Information entstammt [Bel96]

2.4 Arten von Sterbetafeln

In diesem Abschnitt werden Periodentafeln, Kohortentafeln, Generationstafeln und Selektionstafeln definiert. Die Definitionen hier sind recht locker gehalten. Zum Schluss werden noch als Beispiele für Sterbetafeln die 2004R und 2008 diskutiert. In dieser Arbeit werden Periodentafeln und Generationstafeln benutzt

- **Periodensterbetafeln/Aggregattafeln:** 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

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

als die Rohe Sterbewahrscheinlichkeit. Weiter Details zum Schätzen der q_x erfolgte im Kapitel 2.2. 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 sterben, 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.
- Selektionstafel: Da der Bestand von Lebensversicherern oft nicht repräsentativ für die Gesamtbevölkerung ist, gibt es hier eigene Tafeln. Konkret ist die Mortalität im Bestand von Lebensversicherern niedriger als die Mortalität der Gesamtbevölkerung. Da man um solche Tafeln zu erstellen Bestandsdaten von Lebensversicherern braucht, werden solche Tafeln hier nicht erstellt.

Ein Beispiel für eine Sterbetafel ist die Sterbetafel 2004R. 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.

2.5 Öffentlich zugängliche Sterblichkeitsdaten

Die Sterbetafel 2004R basiert laut Kapitel 2 der Herleitung der Sterbetafel 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.

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 Website 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."

Man beachte weiterhin die Spalte "Year". Diese Spalte kann sich entweder auf das Kalenderjahr oder auf das Geburtsjahr beziehen. Auf Mortality.org unter dem Link <https://www.mortality.org/Public/ExplanatoryNotes.php> findet man die folgende Erklärung:

"Period data are indexed by calendar year, whereas cohort data (if available) are indexed by year of birth."

Im *R*-Paket "moRtRe", in dem die Graphiken und Simulationen dieser Arbeit ausgeführt wurden, sind deshalb die Daten im Format "Kalenderjahr Alter Total" beziehungsweise "Geburtsjahr Alter Total" in /data abgelegt.

3 Einfache Modelle

Wir haben uns im letzten Abschnitt grundlegende Definitionen und Probleme bei der Datenerfassung gesehen. In diesem Kapitel werden einfache Modelle eingeführt, um die Sterblichkeitsraten zu schätzen. Diese Modelle sind einfach, da nur eine Variable, entweder das Geburtsjahr oder das Alter verwendet werden. Eine Trendschätzung wird nicht durchgeführt. Diese Modelle dienen als Basisfall und sollten beim Vergleich mit den komplexeren Modellen schlechter abschneiden.

Diese und die vorhergehenden Schätzmethoden sind in dem beiliegenden Paket *moRtRe*, der Name steht für "Mortality Trends", für die Daten aus Kapitel 2.5 implementiert.

3.1 Geburtsjahr als erklärende Variable

Beginnen wir mit dem Modell, dass den Todeszeitpunkt in Abhängigkeit vom Geburtszeitpunkt angibt. Wir benutzen die folgenden Definitionen:

- Bezeichne $\gamma = (\gamma_1, \dots, \gamma_n)$ einen Vektor von Todesjahren.
- Bezeichne $\tau = (\tau_1, \dots, \tau_n)$ einen Vektor von Geburtsjahren.
- Bezeichne $\epsilon = (\epsilon_1, \dots, \epsilon_n)$ einen Vektor von Fehlern. Zu diesem Zeitpunkt machen wir keine Annahmen über die Fehler.

Und gehen von einem funktionalen Zusammenhang

$$\gamma = f(\tau) + \epsilon$$

aus. Dann ist das Todesalter $\gamma - \tau$ und die Lebenserwartung $\mathbb{E}(\gamma - \tau)$. Sei weiterhin die Zufallsvariable X definiert durch $X = f(\tau) + \epsilon$ mit Dichte f_X . Dann kann man eine Sterbetafel aufstellen mittels:

$$\begin{aligned} q_x &= \mathbb{P}(\gamma - \tau < x + 1 | \gamma - \tau > x) \\ &= \frac{\mathbb{P}(\gamma - \tau < x + 1 \cap \gamma - \tau > x)}{\mathbb{P}(\gamma - \tau > x)} \\ &= \frac{\mathbb{P}(f(\tau) + \epsilon - \tau < x + 1 \cap f(\tau) + \epsilon - \tau > x)}{\mathbb{P}(f(\tau) + \epsilon - \tau > x)} \\ &= \frac{\int_x^{x+1} f_X(y) dy}{\int_x^\infty f_X(y) dy} \end{aligned}$$

Offensichtlich hat dieses Modell das Problem, dass das Alter einer zu versichernden Person bekannt ist und diese Info nicht berücksichtigt wird. Ein anderes Problem ist, dass bei Anwendung auf die Daten um so später eine Person geboren ist, um so früher stirbt sie. Siehe dazu Graphik (4)

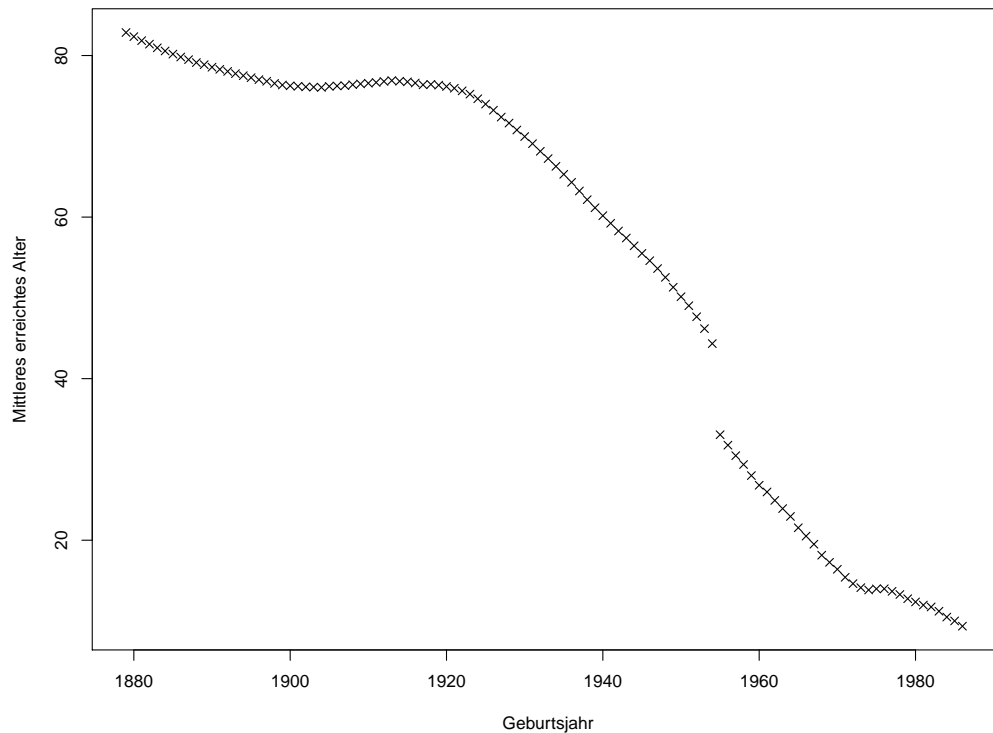


Figure 4: Ein Plot der Geburtsjahr versus mittleres erreichtes Alter bei der Generationstafel 1876 West zeigt.

Dies macht auch Sinn, wenn man einen Moment darüber nachdenkt: Das mittlere Alter, in dem wir Personen sterben sehen, die 1950 geboren sind, ist natürlich höher als das mittlere Alter von Personen, die 2017 geboren sind. Hier sieht man deutlich das Problem mit der Zensur der Daten. Zum Abschluss dieses Kapitels betrachten wir noch eine Heatmap, das die Anzahl an Todesfällen in Abhängigkeit zu Geburtsjahr und Alter beschreibt:

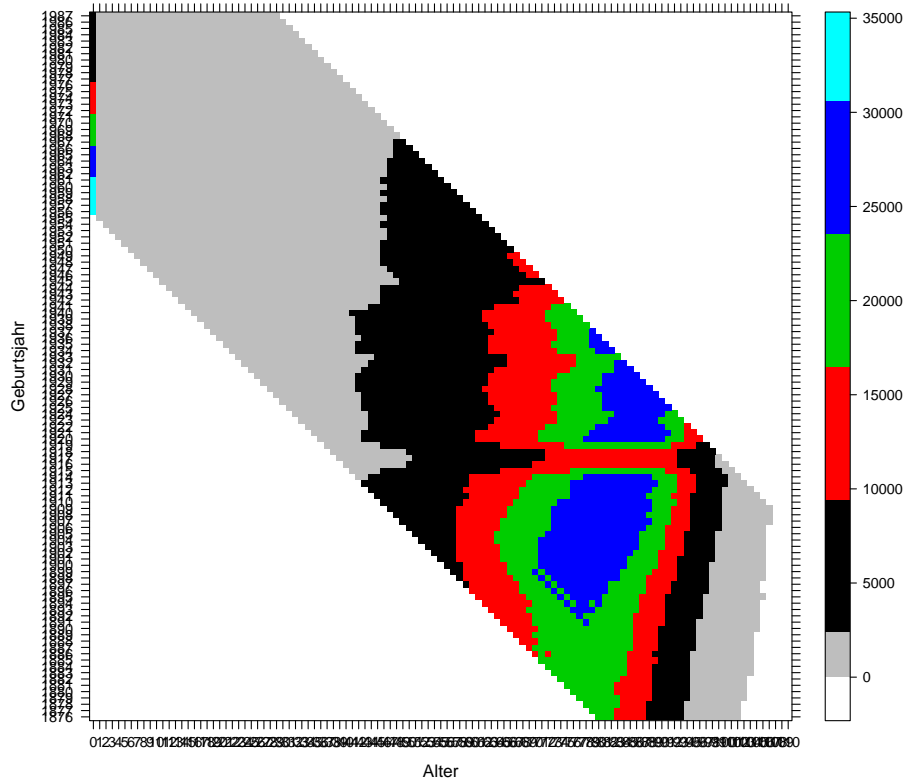


Figure 5: Ein Heatmap der Todesfälle der Generationstafel 1876 West. Auf der x-Achse ist das Geburtsjahr und auf der y-Achse ist das Alter der Personen.

Man sieht, dass man eine Diagonale von links oben, hohe Alter und frühe Geburtsjahre, nach unten links, junge Alter und späte Geburtsjahre hat. Außerhalb der Diagonale gibt es keine Beobachtungen. Deshalb eignet sich nur das Geburtsjahr nicht als einzige erklärende Variable.

3.2 Alter als erklärende Variable

Betrachten wir als nächstes ein weiteres einfaches Modell, doch diesmal mit dem Alter als erklärende Variabel. Beginnen wir wieder mit den Definitionen:

- Bezeichne $Y = (Y_1, \dots, Y_n)$ einen Vektor mit der Anzahl an Todesfällen im Alter i geteilt durch die Gesamtzahl an Todesfällen. Das heißt, sei $b = (\iota_1, \iota_2)$ ein Beobachtungszeitraum und γ, τ das Geburtsjahr beziehungsweise Todesjahr der Person P , dann ist Y_i definiert durch:

$$Y_i = \frac{\#\{P \text{ mit } \gamma \in b \text{ und } \gamma - \tau = i\}}{\#\{P \text{ mit } \gamma \in b\}}$$

- Bezeichne $X = (X_1, \dots, X_n)$ einen Vektor mit Altern. Dabei ist $X_1 = S$ das Startalter und X_n das Endalter
- Bezeichne $\epsilon = (\epsilon_1, \dots, \epsilon_n)$ einen Vektor mit Fehlern.

Wir unterstellen einen funktionalen Zusammenhang zwischen X und Y

$$Y = f(X) + \epsilon \quad (4)$$

Betrachten wir einen Plot der Daten (6):

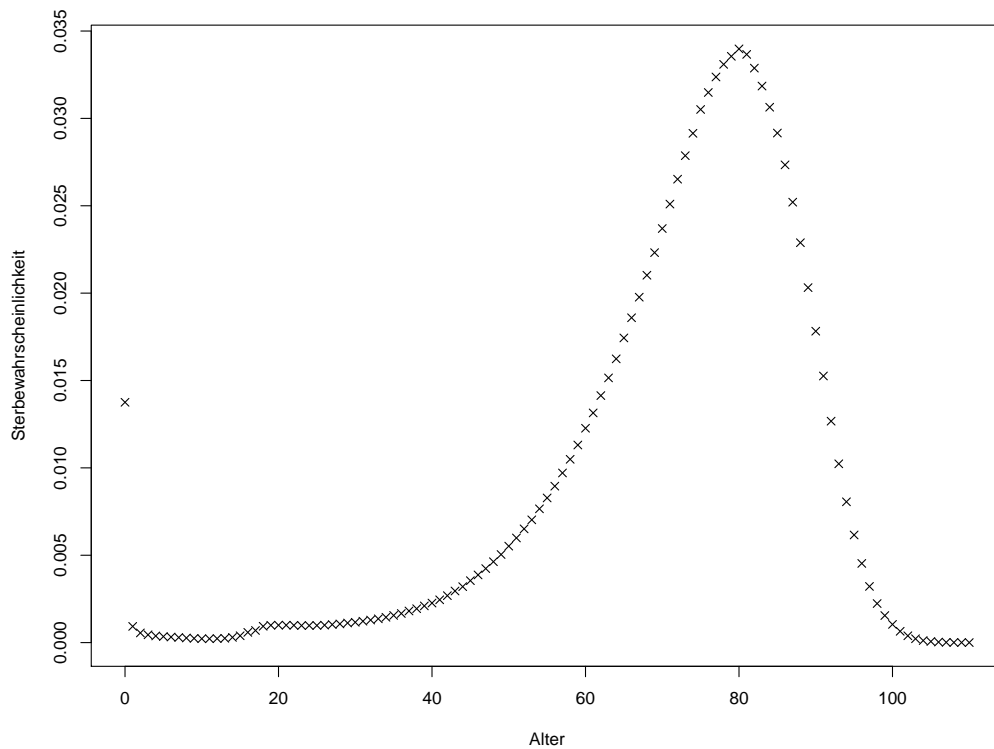


Figure 6: Alter versus relative Anzahl an Sterbefällen pro Alter bei der Generationstafel 1876 West

Man sieht, dass der funktionale Zusammenhang f nicht linear ist. Zum Beispiel in [GK07] wird das folgende Modell vorgeschlagen, dass als Lee-Carter Modell bezeichnet wird: Bezeichne $m_{a,t}$ den Logarithmus der Sterblichkeit der Altersgruppen $a(a = 1, \dots, A)$ und $t(t = 1, \dots, T)$ die Zeit. Dann geht man von

$$m_{a,t} = \alpha_a + \beta_a \gamma_t + \epsilon_{a,t} \quad (5)$$

aus. Hier bezeichnen $\alpha_\alpha, \beta_\alpha$ und γ_t Parameter, die zu schätzen sind und $\epsilon_{a,t}$ zufälligen Schwankungen. Logarithmiert man nun die Sterbewahrscheinlichkeiten und passt dann ein lineares Modell an, so erhält man folgende Graphik (7):

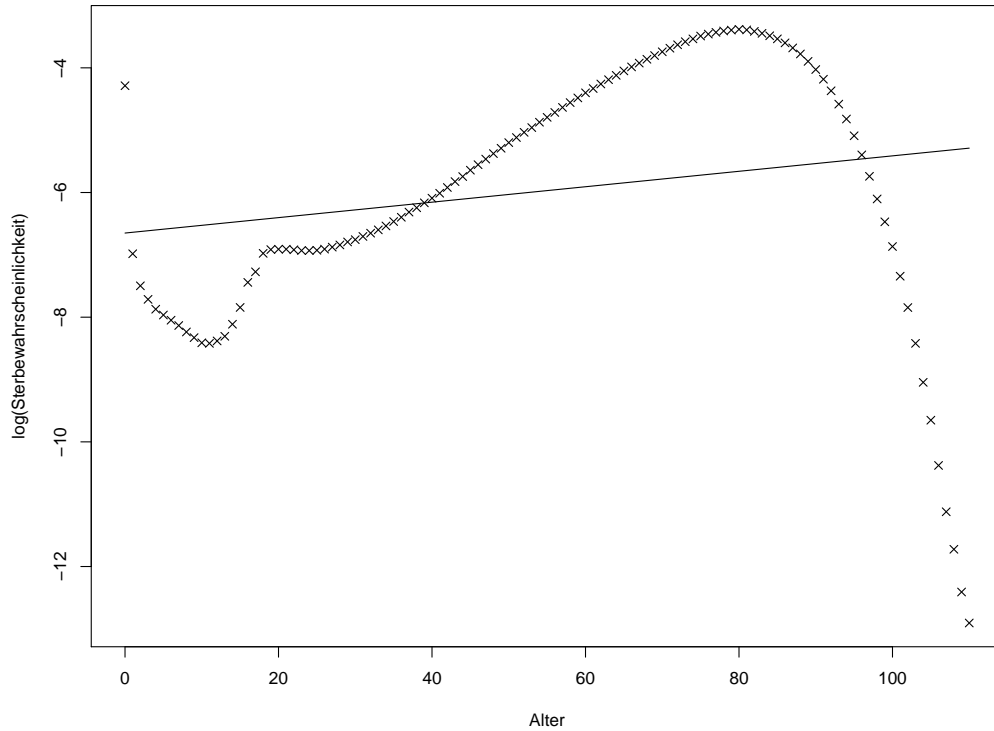


Figure 7: Alter versus logarithmierte relative Anzahl an Sterbefällen pro Alter bei der Generationstafel 1876 West. Außerdem ist eine Regressionsgerade eingezeichnet bei der mit der einfachen kleinsten Quadraten Methode (OLS) $\log(Y)$ auf X regressiert wurde.

Durch die Logarithmierung werden die Daten schon deutlich linearisiert, allerdings noch nicht zu einem Grad, der mir ausreichend scheint.

Betrachten wir noch einmal Graphik (6). Die relativen Sterbehäufigkeiten ergeben in Summe 1, sprich $\sum_{x=S}^E Y_i = 1$. Jede Person stirbt in unserem Modell, spätestens bei Erreichen des Alters E . Diese Eigenschaft haben auch Dichten. Es liegt also nahe, Y_i für alle i als Ausschnitt aus einer Dichte zu interpretieren. Das heißt, es scheint sinnvoll den funktionalen Zusammenhang f als Dichte zu betrachten. Die bekannteste Dichte ist die Normalverteilung, und die Daten scheinen optisch auch recht gut zu einer Normalverteilung zu passen. Wir legen also das folgende Modell an:

$$Y = \frac{1}{\sqrt{2 \cdot \pi \cdot s^2}} \cdot \exp\left(-\frac{(m - X)^2}{2 \cdot s^2}\right) + \epsilon \quad (6)$$

Bei Modell (6) sind alle Operationen vektorwertig zu verstehen. Das heißt, $Y, X \in \mathbb{R}^{E-S+1}$. In folgender Graphik sieht man noch einmal die Daten aus Graphik (6) und zusätzlich der geschätzte, funktionale Zusammenhang f :

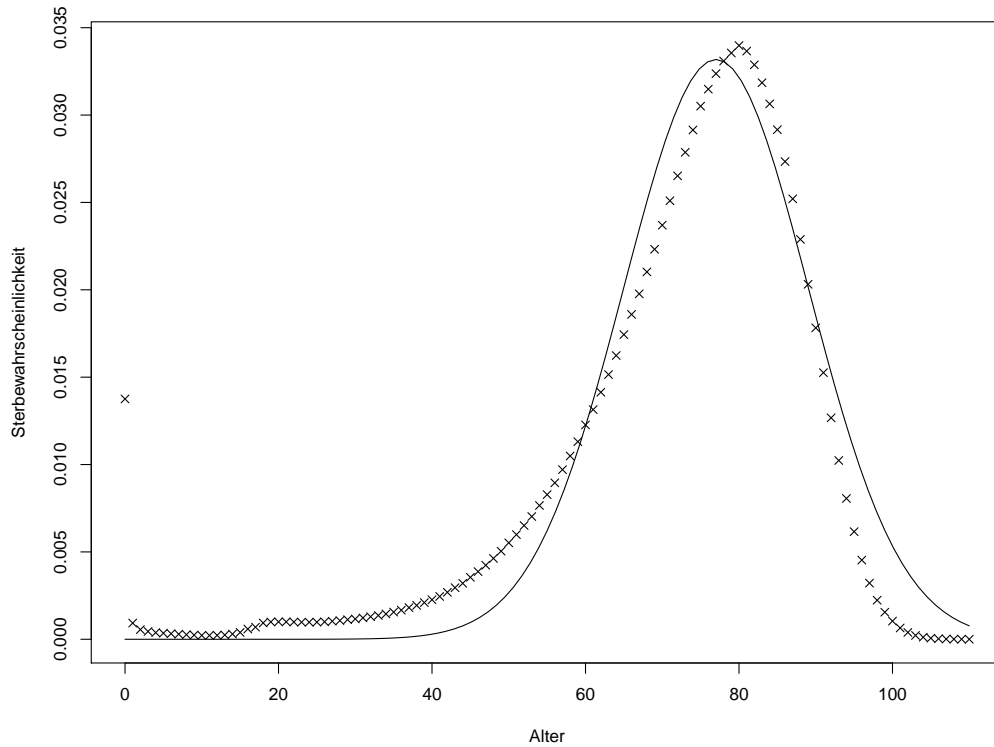


Figure 8: Noch einmal das Alter versus die Sterblichkeit bei der Generationstafel 1876 West. In der durchgezogenen Linie ein Schätzer für das Modell.

Die Interpretation der Parameter ist in diesem Modell recht einfach: Der Parameter m ist das mittlere erreichte Lebensalter und der Parameter s ist die Standardabweichung von dem mittleren erreichten Lebensalter. Hier heißt m der Lageparameter Mitte im Sinne des Mittelwertes zu verstehen.

Dieses Modell ist das Basismodell. Das heißt, die anderen beiden Modelle müssen zeigen, ob sie besser sind. Es hat sicherlich viele Schwächen. So hat man beispielsweise Masse im negativen Bereich. Das heißt, man hat eine von Null verschiedene Wahrscheinlichkeit mit einem negativen Alter zu sterben. Außerdem wird das Geburtsjahr als erklärende Variable vernachlässigt. Allerdings ist es als Vergleichsfall sehr praktisch, da man davon ausgehen kann, dass dieses Modell schlechter ist, als die anderen, die in dieser Arbeit betrachtet werden.

3.3 Eigenschaften des einfachen Modells

In diesem Abschnitt werden Eigenschaften des einfachen Modells besprochen. Das Ziel dieses Abschnittes ist es, das Theorem von Cramér anzuwenden:

Theorem 1 (Cramér). Sei $\mathbf{X} = (X_1, \dots, X_n)$ ein Sample aus einer parametrischen Familie $\mathcal{P} = \{P_\theta : \theta \in \Theta \subset \mathbb{R}^k\}$, wobei P_θ eine Dichte f_θ bezüglich des Lebesguemaßes hat. Sei θ_0 der wahre Parameter. Sei weiterhin

1. Die Menge Θ ist eine offene Teilmenge des \mathbb{R}^n
2. Das Modell ist identifizierbar.
3. Die zweite Ableitung von f_θ existiert und ist stetig für alle $x \in \chi$ und kann unter das Integral $\int f_\theta(x)dx$ geschoben werden.
4. Es existiert eine messbare, integrierbare Funktion $K(x)$, sodass $|l''_\theta(x)| \leq K(x)$ für alle $\theta : |\theta - \theta_0| < \epsilon$.
5. Die Fisher-Informationsmatrix $i(\theta_0)$ ist positiv definit.

Dann existiert eine stark konsistente Folge $\hat{\theta}_n$ von Wurzeln von $n^{-1} \sum_{i=1}^n l'_\theta(X_i) = 0$ mit $\sqrt{n}(\hat{\theta} - \theta) \rightarrow^D \mathcal{N}(0_k, i(\theta_0)^{-1})$.

Proof. Für den vollständigen Beweis sehen Fergusons Buch "A course in large sample theory", [Fer96], Theorem 18. Hier eine kurze Zusammenfassung des Beweises:

- Existenz von konsistenten Wurzeln: Wende Theorem 17 in Fergusons an.
- Asymptotische Normalität: TODO

□

Nachdem wir das Theorem von Cramér für das einfache Modell gezeigt haben, haben wir damit gezeigt, dass das Modell identifizierbar ist und dass ein kritischer Wert der Likelihoodfunktion stark konsistent und asymptotisch normalverteilt ist. Allerdings müssen wir sicherstellen, dass wir auch das richtige Maximum benutzen, beziehungsweise dass es nur ein Maximum gibt.

Außerdem müssen wir für das Cramér Modell noch eine Annahme bezüglich des Fehlers machen. Wir gehen davon aus, dass $\epsilon \sim N(0, \sigma_\epsilon)$ unabhängig und identisch (u.i.v.) verteilt ist und erhalten somit das Modell:

$$Y = \frac{1}{\sqrt{2\pi s^2}} \cdot \exp\left(-\frac{(m - X)^2}{2 \cdot s^2}\right) + \epsilon \sim N\left(\frac{1}{\sqrt{2\pi s^2}} \cdot \exp\left(-\frac{(m - X)^2}{2 \cdot s^2}\right), \sigma_\epsilon\right)$$

Dieses Modell hat die Dichte:

$$f_{(m,s)}(y) = \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp\left(-\frac{\left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x)^2}{2s^2}\right) - y\right)^2}{2\sigma_\epsilon^2}\right)$$

und die log-Likelihoodfunktion

$$\begin{aligned} l_y(m, s) &= \sum_{i=1}^n \ln(f_{y_i}(m, s)) = \sum_{i=1}^n \ln\left(\frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp\left(-\frac{\left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i\right)^2}{2\sigma_\epsilon^2}\right)\right) \\ &= \sum_{i=1}^n \left\{ -\frac{\ln(2\pi\sigma_\epsilon^2)}{2} - \frac{\left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i\right)^2}{2\sigma_\epsilon^2} \right\} \end{aligned}$$

Um das Theorem von Cramér zu zeigen, brauchen wir die zweiten Ableitungen der Dichte und der log-Likelihoodfunktion. Beginnen wir damit, die ersten Ableitungen der log-Likelihoodfunktion zu bestimmen. Berechnen wir zuerst die Ableitung in m Richtung:

$$\begin{aligned} &\frac{d}{dm} l_y(m, s) \\ &= \frac{d}{dm} \sum_{i=1}^n \left\{ -\frac{\ln(2\pi\sigma_\epsilon^2)}{2} - \frac{\left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i\right)^2}{2\sigma_\epsilon^2} \right\} \\ &= -\frac{1}{2\sigma_\epsilon^2} \sum_{i=1}^n \frac{d}{dm} \left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right)^2 \\ &= -\frac{1}{2\sigma_\epsilon^2} \sum_{i=1}^n 2 \cdot \left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right) \cdot \frac{d}{dm} \left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right) \\ &= -\frac{1}{2\sigma_\epsilon^2} \cdot \frac{2}{\sqrt{2\pi s^2}} \sum_{i=1}^n \left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right) \cdot \frac{d}{dm} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \\ &= -\frac{1}{\sigma_\epsilon^2 \sqrt{2\pi s^2}} \sum_{i=1}^n \left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right) \cdot \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot \frac{d}{dm} \left(-\frac{(m-x_i)^2}{2s^2} \right) \\ &= \frac{1}{\sigma_\epsilon^2 2s^2 \sqrt{2\pi s^2}} \sum_{i=1}^n \left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right) \cdot \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot 2(m-x_i) \\ &= \frac{1}{\sigma_\epsilon^2 s^3 \sqrt{2\pi}} \sum_{i=1}^n \left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right) \cdot \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot (m-x_i) \end{aligned}$$

Es ist \hat{m} das Maximum der Likelihoodfunktion. Das heißt, \hat{m} ist ein kritischer Wert der Likelihoodfunktion und somit eine Nullstelle der ersten Ableitung der Log-Likelihoodfunktion in m Richtung. Diese Nullstellen sind analytisch nur schwer zu bestimmen. Weiterhin sind die Nullstellen invariant unter Multiplikation mit einer positiven Konstante. Plotten

wir also die erste Ableitung der Log-Likelihoodfunktion in m Richtung ohne den vorstehenden Faktor $\frac{1}{\sigma_\epsilon^2 s^3 \sqrt{2\pi}}$ und mit $s = 1$, um zu sehen, ob es Nullstellen gibt. Dieser Plot ist (9).

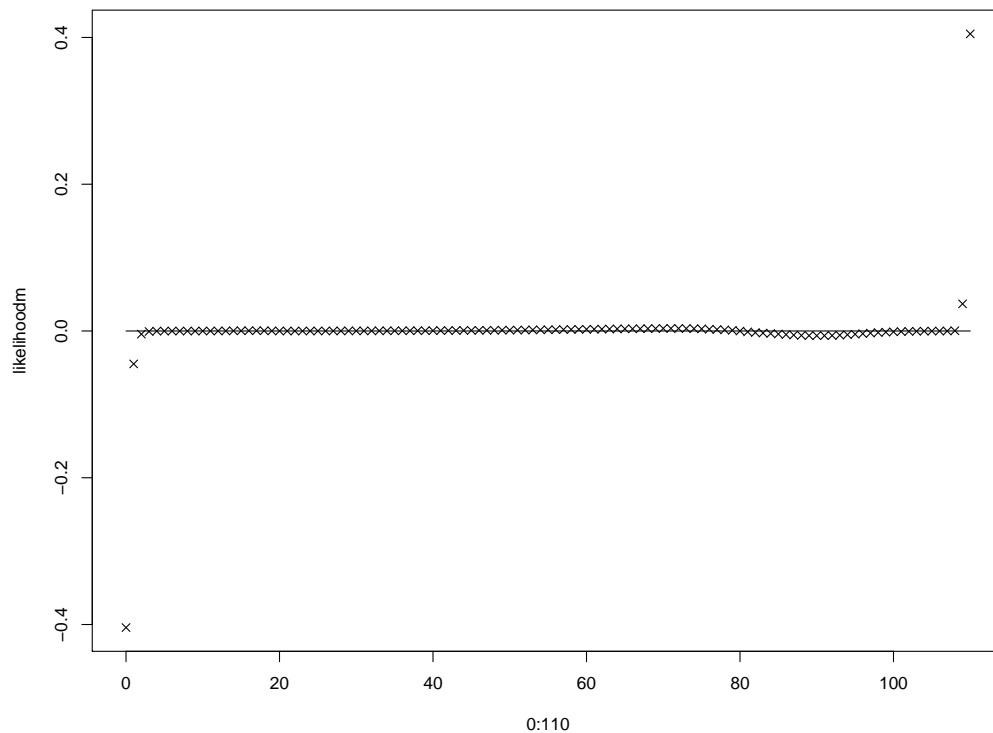


Figure 9: Die erste Ableitung der Log-Likelihoodfunktion in m Richtung. Auf der x -Achse mögliche Werte für \hat{m} und auf der y -Achse der Wert der Log-Likelihoodfunktion an dieser Stelle.

Man sieht, dass die Werte an den Rändern sehr groß sind. Deshalb ist in Graphik (10) dieselbe Funktion noch einmal geplottet, diesmal allerdings ohne die Ränder:

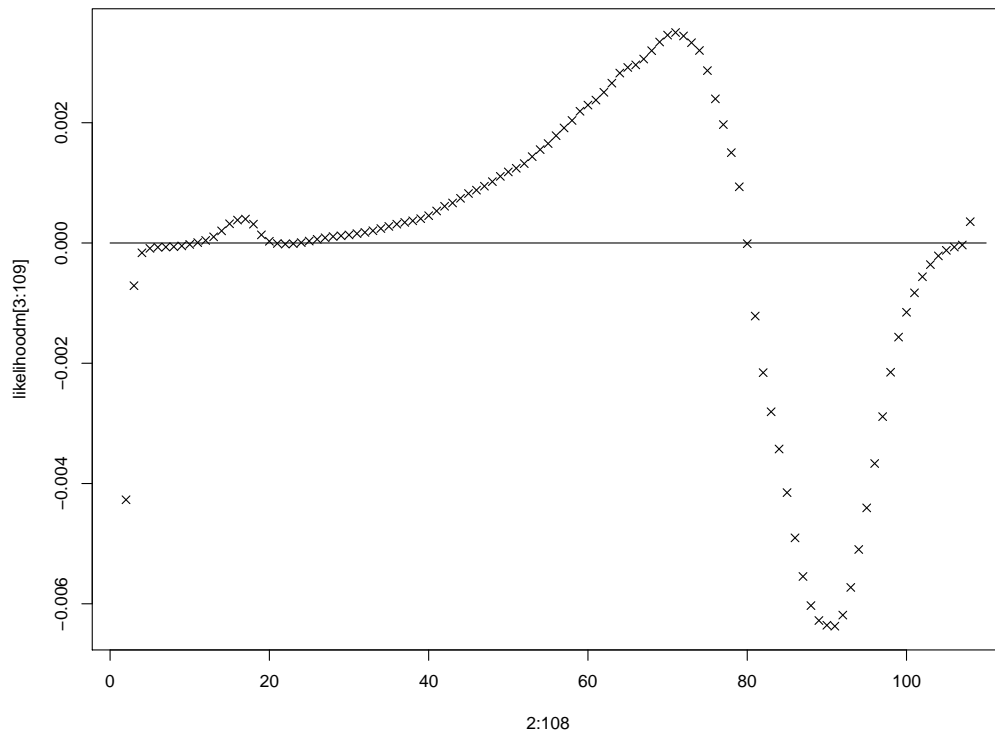


Figure 10: Die erste Ableitung der Log-Likelihoodfunktion in m Richtung ohne die Randwerte. Auf der x -Achse sind wieder mögliche Werte für \hat{m} und auf der y -Achse der Wert der Log-Likelihoodfunktion an dieser Stelle.

Man sieht vier Mögliche Nullstellen. Die Nullstelle bei 80 scheint am sinnvollsten, wenn man sich die Daten (8) anschaut. Berechnen wir jetzt die erste Ableitung in s Richtung:

$$\begin{aligned}
& \frac{d}{ds} l_y(m, s) \\
&= \frac{d}{ds} \sum_{i=1}^n -\frac{\ln(2\pi\sigma_\epsilon^2)}{2} - \frac{\left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i\right)^2}{2\sigma_\epsilon^2} \\
&= -\frac{1}{2\sigma_\epsilon^2} \sum_{i=1}^n \frac{d}{ds} \left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right)^2 \\
&= -\frac{1}{2\sigma_\epsilon^2} \sum_{i=1}^n 2 \left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right) \frac{d}{ds} \left\{ \frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \right\} \\
&= -\frac{1}{\sigma_\epsilon^2 \sqrt{2\pi}} \sum_{i=1}^n \left[\left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right) \cdot \dots \right. \\
&\quad \left. \dots \left\{ \left(\frac{d}{ds} s^{-1} \right) \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) + s^{-1} \frac{d}{ds} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \right\} \right] \\
&= -\frac{1}{\sigma_\epsilon^2 \sqrt{2\pi}} \sum_{i=1}^n \left[\left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right) \cdot \dots \right. \\
&\quad \left. \dots \left\{ \left(-s^{-2} \right) \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) + s^{-1} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \frac{d}{ds} \left(-\frac{(m-x_i)^2}{2s^2} \right) \right\} \right] \\
&= -\frac{1}{\sigma_\epsilon^2 \sqrt{2\pi}} \sum_{i=1}^n \left[\left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right) \cdot \dots \right. \\
&\quad \left. \dots \left\{ \left(-s^{-2} \right) \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) + s^{-1} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot \left(-\frac{(m-x_i)^2}{2} \right) \cdot \left(-2s^{-3} \right) \right\} \right] \\
&= -\frac{1}{\sigma_\epsilon^2 \sqrt{2\pi}} \sum_{i=1}^n \left[\left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right) \cdot \dots \right. \\
&\quad \left. \dots \left\{ \left(-s^{-2} \right) \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) + \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot (m-x_i)^2 \cdot \left(s^{-4} \right) \right\} \right] \\
&= -\frac{1}{\sigma_\epsilon^2 \sqrt{2\pi}} \sum_{i=1}^n \left[\left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right) \cdot \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot \left(\frac{(m-x_i)^2}{s^4} - \frac{1}{s^2} \right) \right]
\end{aligned}$$

Wir wollen wieder die erste Ableitung der Log-Likelihoodfunktion plotten. Diesmal in s Richtung aber wieder ohne störenden Vorfaktor und mit $m = 80$, der Schätzer für m aus dem Plot (10). Man erhält den Plot (12):

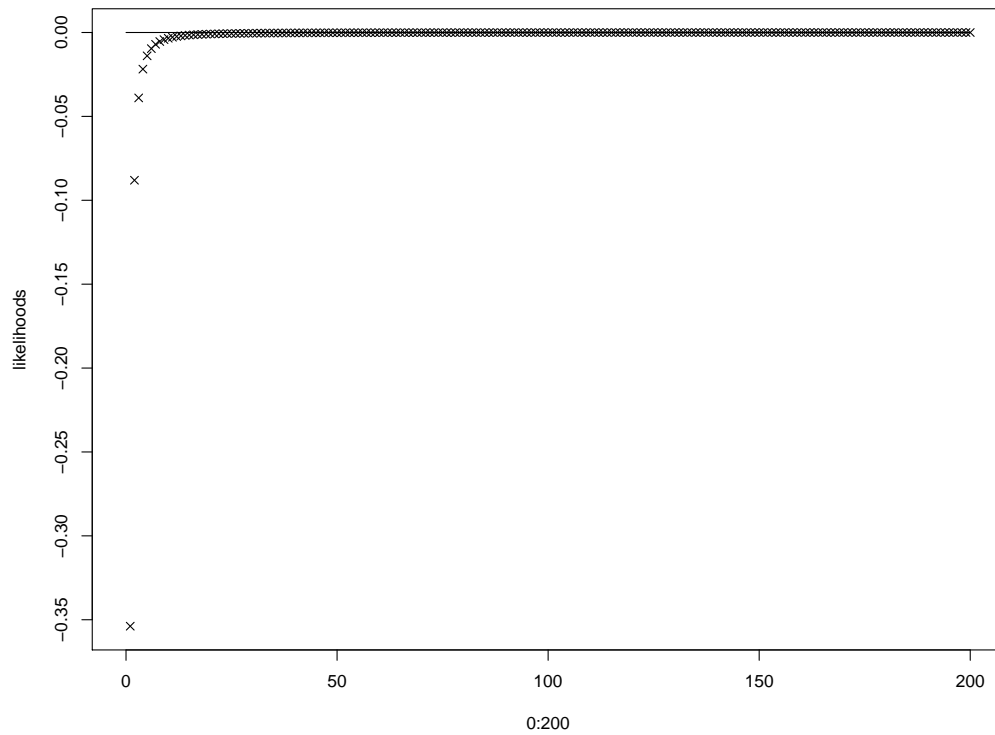


Figure 11: Die erste Ableitung der Log-Likelihoodfunktion in s Richtung. Auf der x -Achse mögliche Werte für \hat{s} und auf der y -Achse der Wert der Log-Likelihoodfunktion an dieser Stelle.

Plot für die Likelihood für s feiner

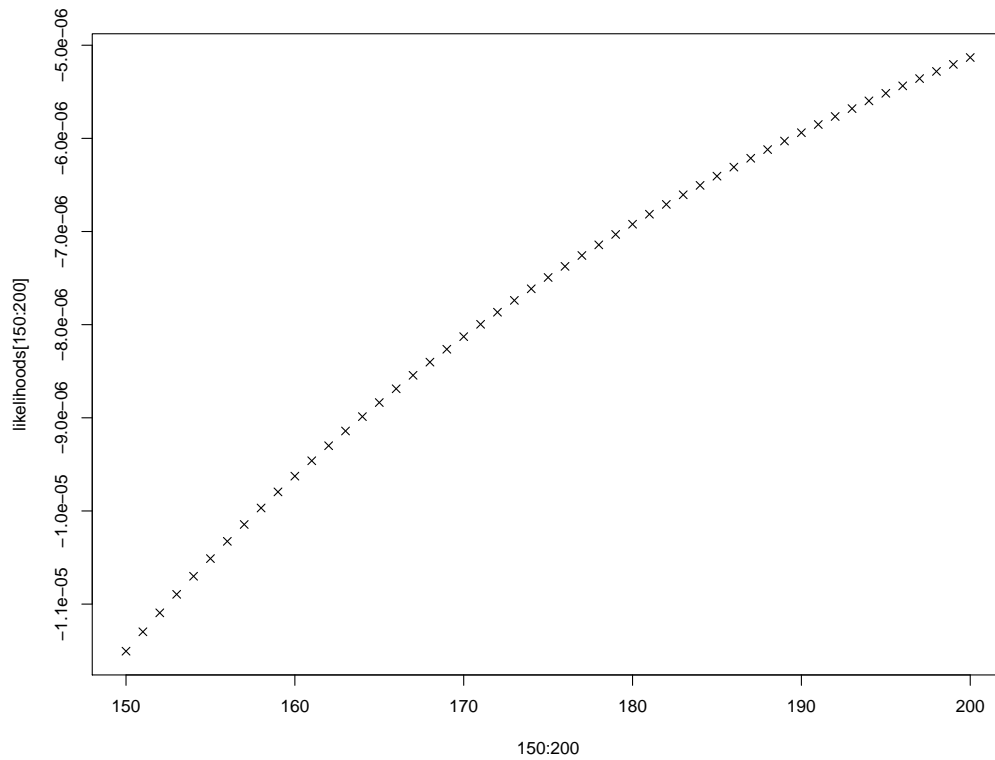


Figure 12: Die erste Ableitung der Log-Likelihoodfunktion in s Richtung im Bereich. Auf der x -Achse sind wieder mögliche Werte für \hat{m} und auf der y -Achse der Wert der Log-Likelihoodfunktion an dieser Stelle.

Berechnen wir als nächstes die zweiten Ableitungen der Log-Likelihood. Beginnen wir mit der zweiten Ableitung in m Richtung:

$$\begin{aligned}
& \frac{d^2}{dm^2} l_{x,y,s,\sigma_\epsilon}(m) \\
&= \frac{d}{dm} \frac{1}{\sqrt{2\pi}s^3\sigma_\epsilon^2} \sum_{i=1}^n \left\{ \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right\} \cdot \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot (m-x_i) \\
&= \frac{1}{\sqrt{2\pi}s^3\sigma_\epsilon^2} \sum_{i=1}^n \frac{d}{dm} \left[\left\{ \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right\} \cdot \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot (m-x_i) \right] \\
&= \frac{1}{\sqrt{2\pi}s^3\sigma_\epsilon^2} \sum_{i=1}^n \left[\exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot \frac{-(m-x_i)}{s^2} \cdot \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot (m-x_i) + \dots \right. \\
&\quad \dots \left\{ \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right\} \cdot \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot \frac{-(m-x_i)}{s^2} \cdot (m-x_i) + \dots \\
&\quad \left. \dots \left\{ \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right\} \cdot \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \right] \\
&= \frac{1}{\sqrt{2\pi}s^3\sigma_\epsilon^2} \sum_{i=1}^n \left[\exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot \dots \right. \\
&\quad \dots \left\langle -\exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot \frac{(m-x_i)^2}{s^2} - \dots \right. \\
&\quad \left. \dots \left\{ \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right\} \cdot \frac{(m-x_i)^2}{s^2} + \left\{ \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right\} \right\rangle \left. \right]
\end{aligned}$$

dann in m und s Richtung

$$\begin{aligned}
& \frac{d}{ds} \frac{d}{dm} l_{x,y,\sigma_\epsilon}(m, s) \\
&= \frac{d}{ds} \frac{1}{\sqrt{2\pi} s^3 \sigma_\epsilon^2} \sum_{i=1}^n \left\{ \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right\} \cdot \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot (m-x_i) \\
&= \frac{1}{\sqrt{2\pi} \sigma_\epsilon^2} \left\{ \frac{d}{ds} s^{-3} \right\} \sum_{i=1}^n \left[\left\{ \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right\} \cdot \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot (m-x_i) \right] + \dots \\
&\quad \dots \frac{1}{\sqrt{2\pi} s^3 \sigma_\epsilon^2} \sum_{i=1}^n \frac{d}{ds} \left[\left\{ \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right\} \cdot \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot (m-x_i) \right] \\
&= -\frac{3}{\sqrt{2\pi} \sigma_\epsilon^2 s^4} \sum_{i=1}^n \left[\left\{ \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right\} \cdot \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot (m-x_i) \right] + \dots \\
&\quad \dots \frac{1}{\sqrt{2\pi} s^3 \sigma_\epsilon^2} \sum_{i=1}^n \left[\frac{d}{ds} \left\{ \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right\} \cdot \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot (m-x_i) + \dots \right. \\
&\quad \left. \dots \left\{ \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right\} \cdot \frac{d}{ds} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot (m-x_i) \right] \\
&= -\frac{3}{\sqrt{2\pi} \sigma_\epsilon^2 s^4} \sum_{i=1}^n \left[\left\{ \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right\} \cdot \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot (m-x_i) \right] + \frac{1}{\sqrt{2\pi} s^3 \sigma_\epsilon^2} \dots \\
&\quad \dots \sum_{i=1}^n \left[\left\{ \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right\} \left(\frac{2(m-x_i)^2}{2s^3} \right) \cdot \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot (m-x_i) + \dots \right. \\
&\quad \left. \dots \left\{ \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right\} \cdot \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \left(\frac{2(m-x_i)^2}{2s^3} \right) \cdot (m-x_i) \right] \\
&= -\frac{3}{\sqrt{2\pi} \sigma_\epsilon^2 s^4} \sum_{i=1}^n \left[\left\{ \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right\} \cdot \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot (m-x_i) \right] - \dots \\
&\quad \dots \frac{1}{\sqrt{2\pi} s^3 \sigma_\epsilon^2} \sum_{i=1}^n \left[\left\{ \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right\} \cdot \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \left(2 \cdot \frac{(m-x_i)^3}{s^3} \right) \right]
\end{aligned}$$

Und zweimal in s Richtung

$$\begin{aligned}
& \frac{d^2}{ds^2} l_{x,y,m,\sigma_\epsilon}(s) \\
&= -\frac{1}{\sigma_\epsilon^2 \sqrt{2\pi}} \sum_{i=1}^n \frac{d}{ds} \left[\left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right) \cdot \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot \left(\frac{(m-x_i)^2}{s^4} - \frac{1}{s^2} \right) \right] \\
&= -\frac{1}{\sigma_\epsilon^2 \sqrt{2\pi}} \sum_{i=1}^n \dots \\
&\quad \dots \left[\frac{d}{ds} \left(\frac{1}{s\sqrt{2\pi}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right) \cdot \left\{ \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot \left(\frac{(m-x_i)^2}{s^4} - \frac{1}{s^2} \right) \right\} + \dots \right. \\
&\quad \dots \left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right) \cdot \frac{d}{ds} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot \left(\frac{(m-x_i)^2}{s^4} - \frac{1}{s^2} \right) + \dots \\
&\quad \left. \dots \left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right) \cdot \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot \frac{d}{ds} \left(\frac{(m-x_i)^2}{s^4} - \frac{1}{s^2} \right) \right] \\
&= -\frac{1}{\sigma_\epsilon^2 \sqrt{2\pi}} \sum_{i=1}^n \dots \\
&\quad \left[\left\{ -\frac{1}{s^2 \sqrt{2\pi}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) + \frac{1}{\sqrt{2\pi s^2}} \cdot \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot \left(\frac{(m-x_i)^2}{s^3} \right) \right\} \cdot \dots \right. \\
&\quad \dots \left\{ \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot \left(\frac{(m-x_i)^2}{s^4} - \frac{1}{s^2} \right) \right\} + \dots \\
&\quad \dots \left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right) \cdot \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot \frac{(m-x_i)^2}{s^3} \cdot \left(\frac{(m-x_i)^2}{s^4} - \frac{1}{s^2} \right) + \dots \\
&\quad \left. \dots \left(\frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) - y_i \right) \cdot \exp\left(-\frac{(m-x_i)^2}{2s^2}\right) \cdot \left(\frac{2}{s^3} - \frac{4(m-x_i)^2}{s^5} \right) \right]
\end{aligned}$$

Zeigen wir jetzt, dass das Theorem von Cramér gilt:

- Die Menge Θ ist eine offene Teilmenge des \mathbb{R}^n : Dies lässt sich leicht über die Definition von Θ steuern. Wir gehen davon aus, dass $m \in (0, 110)$ und $s \in (0, c)$ mit einem großen c . Dies sind beides offene Teilmengen des \mathbb{R}^n ist Θ eine offene Teilmenge des \mathbb{R}^n . Noch ein Wort zur Wahl der Mengen: Wir haben $m \in (0, 110)$ gewählt, weil in dem Beispiel mit den Generationstafel von 1876 genau Start- und Endalter sind. Die Wahl von s liegt daran, dass wir gegebenenfalls s numerisch bestimmen müssen und das ist auf einem endlichen Intervall einfacher.
- Das Modell ist identifizierbar: Zeigen wir jetzt Identifizierbarkeit. Zuerst wird diese Eigenschaft definiert, dann sehen wir einige Beispiele und dann wird gezeigt, dass das einfache Modell identifizierbar ist.

Defintion 1 (Identifizierbarkeit). Sei $\Theta \subset \mathbb{R}^n$. Wir nennen Θ einen Parameterraum. Sei P_θ ein Wahrscheinlichkeitsmaß, dass von θ abhängt. Wir bezeichnen eine Menge

von P_θ mit $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ und nennen \mathcal{P} ein statistisches Modell. Ein statistisches Modell \mathcal{P} heißt identifizierbar, genau dann, wenn \mathcal{P} folgende Eigenschaft besitzt:

$$P_{\theta_1} = P_{\theta_2} \implies \theta_1 = \theta_2 \quad (7)$$

Das heißt, wir können das einfache Modell (6) als ein statistisches Modell auffassen. Dabei ist $\Theta = \mathbb{R} \times \mathbb{R}^+$, da wir von $m \in \mathbb{R}$ und $s \in \mathbb{R}^+$ ausgehen. Die Maße P_θ sind dann Transformationen von ϵ um $f(\theta, x)$ wobei f die Dichte einer Standardnormalverteilung ist. Als nächstes zeigen wir, dass das Modell (6) identifizierbar ist. Hier noch einmal das Modell

$$P_\theta = P_{m,s} : Y = \frac{1}{\sqrt{2 \cdot \pi \cdot s^2}} \cdot \exp\left(-\frac{(m - X)^2}{2 \cdot s^2}\right) + \epsilon$$

Setzen wir jetzt die Definition der Identifizierbarkeit an:

$$\begin{aligned} & P_{\theta_1} = P_{\theta_2} \\ \iff & \frac{1}{\sqrt{2 \cdot \pi \cdot s_1^2}} \cdot \exp\left(-\frac{(m_1 - X)^2}{2 \cdot s_1^2}\right) + \epsilon = \frac{1}{\sqrt{2 \cdot \pi \cdot s_2^2}} \cdot \exp\left(-\frac{(m_2 - X)^2}{2 \cdot s_2^2}\right) + \epsilon \\ \iff & -\frac{\ln(2\pi s_1^2)}{2} - \frac{(m_1 - X)^2}{2 \cdot s_1^2} = -\frac{\ln(2\pi s_2^2)}{2} - \frac{(m_2 - X)^2}{2 \cdot s_2^2} \\ \iff & \ln(s_1^2) + \frac{(m_1 - X)^2}{s_1^2} = \ln(s_2^2) + \frac{(m_2 - X)^2}{s_2^2} \\ \iff & x^2 \left(\frac{1}{s_1^2} - \frac{1}{s_2^2}\right) - 2x \left(\frac{m_1}{s_1^2} - \frac{m_2}{s_2^2}\right) + \left(\frac{m_1^2}{s_1^2} - \frac{m_2^2}{s_2^2} + \ln(s_1) - \ln(s_2)\right) = 0 \end{aligned}$$

Diese letzte Gleichung gilt, genau dann wenn die Koeffizienten Null sind. Das heißt, es muss gelten $m_1 = m_2$ und $|s_1| = |s_2|$. Da wir vorausgesetzt haben, dass $s_1, s_2 \in \mathbb{R}^+$ folgt $\theta_1 = \theta_2$ und somit ist das Modell identifizierbar.

- Die zweite Ableitung von f_θ existiert und ist stetig für alle $x \in \chi$ und kann unter das Integral $\int f_\theta(x) dx$ geschoben werden:

Für diese Bedingung gibt es bei Ferguson noch ein weiteres Lemma

Lemma 1. Wenn $(d/d\theta)g(x, \theta)$ existiert und stetig in θ für alle x und alle θ in einem offenen Intervall S ist, und wenn $|(d/d\theta)g(x, \theta)| \leq K(x)$ auf S wo $\int K(x) dv(x) < \infty$ und wenn $\int g(x, \theta) dv(x)$ auf S existiert dann gilt

$$\frac{d}{d\theta} \int g(x, \theta) dv(x) = \int \frac{d}{d\theta} g(x, \theta) dv(x)$$

Zuerst die zweite Ableitung von f_θ

$$\frac{d^2}{d\theta^2} f_\theta = \text{TODO}$$

Weiterhin erfüllt die folgende Funktion $K(x)$ die Bedingung:

$$K(x) = \text{TODO}$$

TODO: Kann ich das so schreiben?

Die Funktion besteht aus Polynomen, Exponentialfunktionen und Wurzeln. Alle diese Bestandteile sind stetig diffbar und deshalb sollte auch die zweite Ableitung stetig sein. Außerdem handelt es sich um eine Exponentialfamilie und bei Exponentialfamilien kann man immer Integration und Differentiation vertauschen.

- Es existiert eine messbare, integrierbare Funktion $K(x)$, sodass $|l''_\theta(x)| \leq K(x)$ für alle $\theta : |\theta - \theta_0| < \epsilon$:
- Die Fisher-Informationsmatrix $i(\theta_0)$ ist positiv definit: Die Fisher-Information ist definiert als:

$$\begin{aligned} i(\theta) &= -\mathbb{E} \begin{pmatrix} \frac{d^2}{dm^2} l_x(\theta) & \frac{d}{dm} \frac{d}{ds} l_x(\theta) \\ \frac{d}{ds} \frac{d}{dm} l_x(\theta) & \frac{d^2}{ds^2} l_x(\theta) \end{pmatrix} \\ &= \begin{pmatrix} -\mathbb{E} \frac{d^2}{dm^2} l_x(\theta) & -\mathbb{E} \frac{d}{dm} \frac{d}{ds} l_x(\theta) \\ -\mathbb{E} \frac{d}{ds} \frac{d}{dm} l_x(\theta) & -\mathbb{E} \frac{d^2}{ds^2} l_x(\theta) \end{pmatrix} \end{aligned}$$

Die Fisher-Informationsmatrix ist symmetrisch, da zweite Ableitungen in die gleichen Richtungen gleich sind. Für symmetrische Matrizen gilt: Positiv Definit \iff alle Eigenwerte sind positiv. Die Eigenwerte kann man über das Charakteristische Polynom bestimmen:

$$\begin{aligned} \chi_{i_\theta}(\lambda) &= \lambda^2 - (i_{11} + i_{22})\lambda + i_{11}i_{22} - i_{21}i_{12} \\ &= \text{TODO Erwartungswerte ausrechnen} \end{aligned}$$

Um den MSE des Schätzers für das einfachen Modell zu bestimmen, müssen wir zuerst eine explizite Formel für $\hat{\theta}$ finden und dann Bias und Varianz bestimmen. Betrachten wir also nochmal die Definition für $\hat{\theta}$:

$$\hat{\theta} = \operatorname{argmin}_{(m,s)} \sum_{i=1}^n \left(y_i - \frac{1}{\sqrt{2\pi s^2}} e^{-\frac{(x_i - m)^2}{2s^2}} \right)^2 := g(m, s)$$

Das Maximum von g ist der Nullpunkt der Jacobi-Matrix

$$\frac{d}{d(m, s)}g(m, s) = \left(\frac{d}{dm}g(m, s), \frac{d}{ds}g(m, s) \right)$$

Berechnen wir die beiden Teile der Jacobi-Matrix einzeln

$$\begin{aligned} \frac{d}{d(m)} \sum_{i=1}^n \left(y_i - \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-m)^2}{2\sigma^2}} \right)^2 &= \sum_{i=1}^n \frac{d}{d(m)} \left(y_i - \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-m)^2}{2\sigma^2}} \right)^2 \\ &= \text{TODO} \end{aligned}$$

und

$$\begin{aligned} \frac{d}{d(s)} \sum_{i=1}^n \left(y_i - \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-m)^2}{2\sigma^2}} \right)^2 &= \text{TODO} \\ &= \text{TODO} \end{aligned}$$

Als nächstes wird der Mean-Squared-Error (MSE) bestimmt. Dieser ist wie folgt definiert:

Defintion 2 (Mean-Squared-Error (MSE)). Sei $\hat{\theta}$ ein Schätzer für θ . Dann ist

$$\text{MSE}(\hat{\theta}) = \mathbb{E} \left((\hat{\theta} - \theta)^2 \right) \quad (8)$$

Das folgende Resultat wird als bekannt vorausgesetzt

$$\text{MSE}(\hat{\theta}) = \left(\mathbb{E}(\hat{\theta} - \theta) \right)^2 + \text{Var}(\hat{\theta}) \quad (9)$$

TODO : Versucht man das Maximum konkret auszurechnen wird es bald recht schwer; außerdem ist der Plan nachher Cramér anzuwenden -> wahrscheinlich sollte ich $\hat{\theta}$ als den Maximum-Likelihood-Schätzer nehmen.

Im nächsten Kapitel werden wir betrachten welche Modelle es in der Literatur gibt und wie die DAV Sterbetafeln schätzt.

Im übernächsten Kapitel werden wir dann dieses einfache Modell verbessern.

4 Übliche Verfahren zur Gewinnung von Sterbetafeln und die Sterbetafel 2004R

Wir haben bereits in der Motivation in Abschnitt 1.1 gesehen, dass einfache Basistafeln nicht ausreichen. Deshalb werden wir in diesem Abschnitt statistische Modelle angeben, um Sterblichkeiten und unterliegende Trends zu schätzen.

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.

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. Die folgende Beschreibung orientiert sich an Kapitel 3.3.1 der Herleitung der Tafel 2004R. Man beachte, dass die 2004R zum einen eine Tafel für die Rentenbezugszeit und zum anderen für die Aufschubszeit hat. Da es sich bei der Tafel zur Rentenbezugszeit um eine Selektionstafel handelt, wird deren Herleitung hier nicht betrachtet.

4.1 Auswahl des Basisjahrs

Die Auswahl des Basisjahrs wird in Anhang 4 der Herleitung der Sterbetafel 2004R erklärt:

"Der mittlere Zeitpunkt des Beobachtungsmaterials wird definiert durch Gewichtung der Verweildauer der Jahre 1995 bis 2002, also als

$$\frac{\sum_{t=1995}^{2002} t \cdot VWD(t)}{\sum_{t=1995}^{2002} VWD(t)} \quad (10)$$

wobei $VWD(t)$ die gesamte nach Rentenhöhen gewichtete Verweildauer des Beobachtungsmaterials im Jahr t bezeichnet."

Die DAV schlägt also vor als Bezugsjahr der Basistafel das mittlere Jahr des Beobachtungszeitraums zu nehmen. Dies wird nicht weiter motiviert, allerdings spricht dafür, dass im mittleren Jahr der Effekt der Sterblichkeitsverbesserung zu den Randaltern am geringsten sei dürfte.

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?

4.2 Bestimmung roher Sterbewahrscheinlichkeiten

Wie bereits in Abschnitt 2.5 beschrieben werden zur Bestimmung der rohen Sterbewahrscheinlichkeiten die beiden Kennzahlen I_x und t_x für ein Kollektiv bestimmt 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

$$q_x^{\text{roh}} = \frac{t_x}{I_x} \quad (11)$$

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

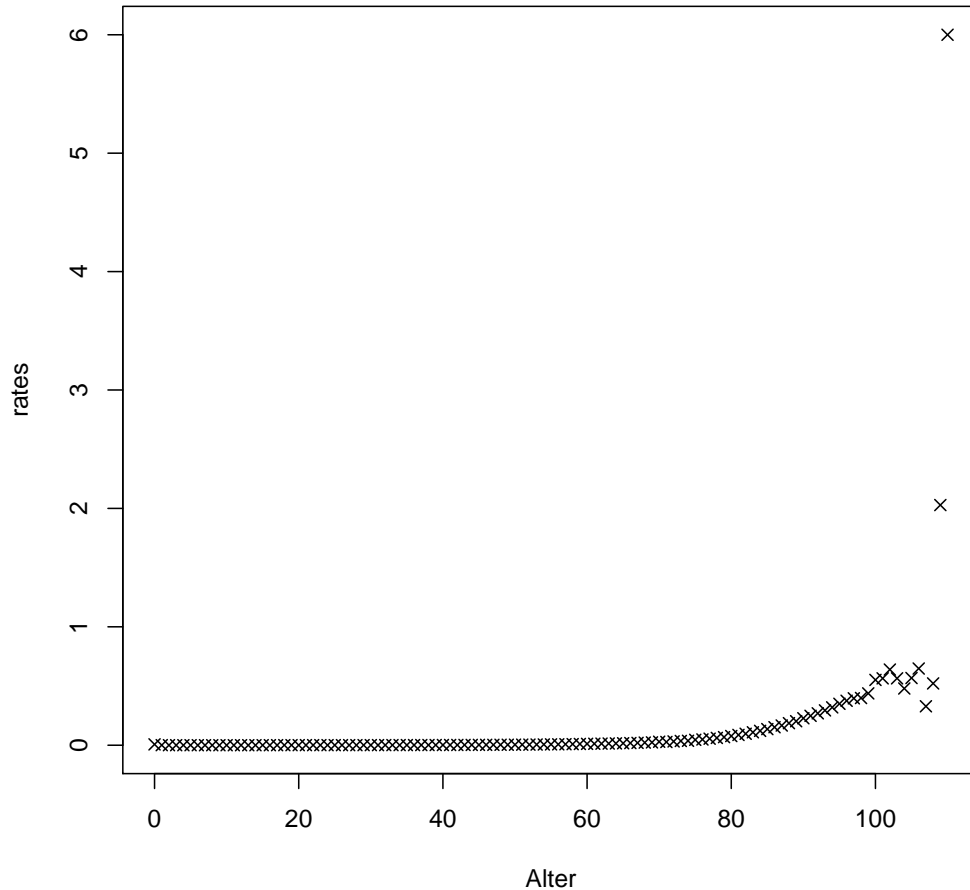


Figure 13: Periodensterblichkeit im Basisjahr 1990. Die Sterbetafel wurde aus Periodensterbetafel erzeugt.

4.3 Glättung der Sterbewahrscheinlichkeiten

Dieses Kapitel basiert auf [Kak85] Kapitel 3. Die DAV hat für die Whittaker-Henderson Verfahren die Quelle [Loe94] empfohlen. Diese fand ich nicht sehr tiefgehend und habe deshalb die Quelle [NS12] verwendet.

4.3.1 Überblick über Glättungsverfahren

Im vorhergehenden Abschnitt haben wir gesehen, wie man rohe Sterbewahrscheinlichkeiten $Q = \{q_S, \dots, q_E\}$ schätzt. Das Ziel von Ausgleichsmethoden ist dann, diese Werte sinnvoll abzuändern. Früher wurde der Ausgleich mittels Millimeterpapier und Bleistift vorgenommen, doch da diese Art Daten auszugleichen sehr subjektiv ist, wurden objektivere Ver-

fahren verwendet. Der theoretische Hintergrund ist die Ausgleichsverfahren als eine Abbildung $q_x \rightarrow q'_x$ zu betrachten. Das heißt, man führt eine Funktion F ein mit:

$$q'_x = F(Q, x, a_k(Q, x)), k = 1, \dots, r \quad (12)$$

Bei der Definition der Funktion F , die auch als Ausgleichsfunktion bezeichnet wird, in (12) bezeichnet x das Alter, Q die Menge der rohen Sterbewahrscheinlichkeiten und a_k sind sogenannte Ausgleichsparameter. Die Ausgleichsparameter dienen dazu, verschiedenen Werte verschieden zu gewichten. Sie müssen entweder im Vorfeld gesetzt oder aus den Daten berechnet werden.

Man unterscheidet zwei große Gruppen an Ausgleichsverfahren. Zum einen gibt es die Mechanische Methoden zum anderen gibt es die Analytische Methoden. Die Zuordnung eines Verfahrens zu einer der beiden Methoden ist nicht immer eindeutig.

- Mechanische Ausgleichsrechnung: F ist eine lineare Funktion der Form

$$q'_x = \sum_{k=-r}^{+s} a_k \cdot q_{x+k}, r, s \in \mathbb{N}$$

und die a_k sind unabhängig von x und von q_x . Man beachte, dass k kleinere und s größere Werte den auszugleichenden Wert beeinflussen. Das heißt, die Werte an den Rändern können nicht voll ausgeglichen werden. Diese Werte an den Rändern nennt man Randwerte. Die zugrunde liegende Idee ist Fehler durch die Zufallsschwankungen auf mehrere Werte zu verteilen und dadurch die Fehler zu verringern. Dadurch dass die Fehler im Allgemeinen verschieden Vorzeichen haben, hofft man, dass sich einige Fehler gegenseitig aufheben. Die mechanische Ausgleichsrechnung hat eine glättende Wirkung.

- Analytische Ausgleichsrechnung: F ist eine nichtlineare Funktion in x und a_k . Die Parameter a_k können von x und q_x abhängig sein, sodass die allgemeine Darstellung

$$q'_x = F(x, a_k(x, Q))$$

lautet. Dies bedeutet, dass wir die Parameter a_k für jeden Altersbereich und jede Wertereihe neu bestimmen müssen. Die Ausgleichsfunktion geben wir explizit vor, zum Beispiel durch eine Exponentialfunktion der Form $\exp(a_2 x^2 + a_x x + a_0)$. Dadurch postulieren wir ein sogenanntes Sterbengesetz, von dem wir annehmen, dass es den Verlauf der tatsächlichen Sterbewahrscheinlichkeit wiedergibt. Die Parameter a_k bestimmt man mit Hilfe zusätzlicher Forderungen, die man an die Ausgleichsfunktion F stellt.

- Zusammenhang zwischen den Methoden: Der Zusammenhang zwischen den mechanischen und den analytischen Ausgleichsverfahren besteht darin, dass man von einem Modell

$$q_x = f(x) + E_x$$

ausgeht. Dabei bezeichnet E_x zufällige Fehler und f ist eine unbekannte Funktion. Man geht also davon aus, dass es ein SterbeGesetz gibt, dass mit f bezeichnet wird. Wendet man nun ein mechanisches Ausgleichsverfahren auf q_x an, erhält man

$$q'_x = \sum_{k=-r}^{+s} a_k \cdot q_{x+k} = \sum_{k=-r}^{+s} a_k \cdot f(x+k) + \sum_{k=-r}^{+s} a_k \cdot E_{x+k}$$

und fordert von einem mechanischem Ausgleichsverfahren, dass es bezüglich f invariant ist und die Fehler verringert:

$$\begin{aligned} \sum_{k=-r}^{+s} a_k \cdot f(x+k) &= f(x) \text{ (Invarianz)} \\ \sum_x \left(\sum_{k=-r}^{+s} a_k \cdot E_{x+k} \right)^2 &< \sum_x E_x \text{ (Anpassung)} \end{aligned}$$

Daß heißt in einem mechanischen Ausgleichsverfahren werden die Fehler minimiert, ohne dass ein SterbeGesetz angenommen werden muss. In einem analytischen Ausgleichsverfahren wird dagegen ein SterbeGesetz angenommen und danach die Fehler minimiert. Man kann also die mechanischen Ausgleichsverfahren als Allgemeiner ansehen.

- Beispiele mechanischer Verfahren: Ein bekanntes Verfahren ist die 9-Punkte-Formel von Schärtlin, bei dem

$$q'_x = \frac{1}{27}(9q_x + 8q_{x+1} + 2q_{x+2} - q_{x+4})$$

gesetzt wird. Ein anderes ist die 15-Punkte-Formel von Spencer, bei der

$$q'_x = \frac{1}{320}(74q_x + 67q_{x+1} + 46q_{x+2} + 21q_{x+3} + 3q_{x+4} - 5q_{x+5} - 6q_{x+6} - 3q_{x+7})$$

gesetzt wird. Ein anderes Verfahren ist das Whittaker-Henderson-Verfahren, dass in dieser Arbeit verwendet wird und deshalb gesondert vorgestellt wird.

- Beispiele analytischer Ausgleichsverfahren: Eine Klasse analytischer Ausgleichsverfahren basiert darauf, dass man die Anzahl an Lebenden als Folge betrachtet und ein Vorschrift angibt, wie man weiter Folgenglieder berechnen kann. Zwei Möglichkeiten sind die SterbeGesetze von de Moivre (1724) und Lambert (1765):

$$\begin{aligned} l'(x) &= l(0) \left(1 - \frac{x}{86} \right) \text{ de Moivre} \\ l'(x) &= 10.000 \left(\frac{96-x}{96} \right)^2 - 6176 \left(e^{\frac{x}{13,68}} - e^{\frac{x}{2,43}} \right) \text{ Lambert} \end{aligned}$$

Historisch gesehen die nächsten SterbeGesetze, die entwickelt wurden, sind die von Gompertz (1824) und Makeham (1860). Diese sind von der folgenden Form, wobei k, s, g, c geschätzt werden müssen:

$$l'(x) = k \cdot s^x \cdot g^{c^x}$$

Nach der Wahl eines SterbeGesetzes müssen als nächstes die Parameter bestimmt werden. Dazu bieten sich verschiedene Möglichkeiten an. Eine ist das Verfahren der kleinsten Quadrate, bei dem folgende Bedingung erfüllt werden muss:

$$\Pi_{i=0}^{m-1} F(x_0 + (v-1)m + i, a_k) = \Pi_{i=0}^{m-1} q_{x_0+(v-1)m+i}$$

Eine andere ist das Verfahren von King-Hardy, bei dem

$$\sum_{x=x_0}^{x_0+n} g(x, a_k) (F(x, a_k) - q_x)^2 = \text{Minimum}$$

erfüllt werden muss. Eine dritte Möglichkeit ist das Verfahren der Spline-Funktionen nach Reinsch:

$$\int_{x_0}^{x_0+n} \left(\frac{d^2}{dx^2} F(x, a_k) \right)^2 dx = \text{Minimum}$$

mit der Nebenbedingung

$$\sum_{i=0}^n \frac{(F(x_0 + i, a_k) - q_{x_0+i})^2}{g_i} \leq A$$

Schlussendlich gibt es noch das Verfahren der Ordinatenmomente:

$$\sum_{x=x_0}^{x_0+n} x^v F(x, a_k) = \sum_{x=x_0}^{x_0+n} x^v q_x$$

4.3.2 Das Verfahren von Whittaker-Henderson

Wie bereits oben erwähnt, werden wir im Folgenden das Verfahren von Whittaker-Henderson verwenden. Die Idee dieser Konstruktionsmethode ist die Bestimmungsgleichungen für die a_k aus einer Optimierungsaufgabe herzuleiten. Sei das Bestimmungsmaß definiert durch

$$\sum_{k=0}^n (q'_{x_0+k} - q_{x_0+k})^2 \quad (13)$$

und das Glättemaß durch

$$\sum_{k=0}^{n-s} \left(\sum_{v=0}^s (-1)^v \binom{s}{v} q'_{x_0+k+v} \right)^2 \quad (14)$$

Das Glättemaß wird in der Literatur auch mit $\Delta^s q'_{x_0+k}$ bezeichnet. Insgesamt erhält man, wenn man Gleichung (13) und (14) kombiniert die Optimierungsaufgabe

$$\sum_{k=0}^n w_k (q'_{x_0+k} - q_{x_0+k})^2 + g \cdot \sum_{k=0}^{n-s} \left(\sum_{v=0}^s (-1)^v \binom{s}{v} q'_{x_0+k+v} \right)^2 = \text{Minimum} \quad (15)$$

dabei sind w_k zusätzliche positive Gewichte und g eine positive Konstante, die das Maß der Glättung vorgibt. Für $g \rightarrow 0$ erhält man die originalen Daten und für $g \rightarrow \infty$ erhält man ein Polynom zweiter Ordnung.

Zur Lösung des Optimierungsproblems (15) bietet es sich an, das Problem in Matrixschreibweise zu formulieren. Diese Reformulierung orientiert sich an [NS12]. Sei K die $(n - p) \times n$ Matrix

$$K_{i,j} = \begin{cases} (-1)^{p+j-i} \binom{p}{j-i} \\ 0 \end{cases}$$

Damit lässt sich das Optimierungsproblem (15) schreiben als

$$\begin{aligned} & \sum_{k=0}^n w_k (q'_{x_0+k} - q_{x_0+k})^2 + g \cdot \sum_{k=0}^{n-s} \left(\sum_{v=0}^s (-1)^v \binom{s}{v} q'_{x_0+k+v} \right)^2 \\ &= (Q' - Q)^t W (Q' - Q) + g \cdot \sum_{k=0}^{n-s} \left(\sum_{v=0}^s K_{k,v} q'_{x_0+k+v} \right)^2 \\ &= (Q' - Q)^t W (Q' - Q) + g (K q')^t (K q') \end{aligned}$$

Weiterhin zeigt Greville ([Gre81], Seite 54), dass für ein Problem von der Form

$$(u - y)^t W (u - y) + (Ku)^t H (Ku) = \text{Minimum}$$

Eine Lösung durch

$$(I + K^t H K) u = y$$

gegeben ist. Setzt man $H = g \cdot I$ und $w_k = 1$ für alle k erhält man eine Lösung für das Verfahren von Whittaker-Henderson, da die Matrix $(I + g K^t K)$ positiv definit und damit invertierbar ist. Die Lösung für Problem (15) ist also gegeben durch

$$Q' = (I + g K^t K)^{-1} Q$$

Es bleibt, sich Gedanken über die Auswahl des Glättungsparameters g zu machen. Wir werden hier das Generalised-Cross-Validation (GCV) Methode benutzen, die in [NS12] Kapitel 4 vorgeschlagen wird. Dabei wird g so gewählt, dass

$$\text{GCV} = \frac{(q - q')^t (q - q')}{(1 - \text{Tr}((I + g K^t K)^{-1}))^2 / n}$$

minimiert.

Betrachten wir jetzt ein Beispiel. Es wird ein Datensatz wie in der folgenden Tabelle generiert:

Jahr	Alter	Anzahl an Toden
1900	0	$f(0, \mu, \sigma, 1900)$
\vdots	\vdots	\vdots
1900	110	$f(110, \mu, \sigma, 1900)$

Wobei f dem bekannten einfachen Modell

$$Y = \frac{1}{\sqrt{2 \cdot \pi \cdot s^2}} \cdot \exp\left(-\frac{(m - X)^2}{2 \cdot s^2}\right) + \epsilon$$

entnommen wird. Hier ist $\epsilon \sim N(0, \sigma_\epsilon)$ Also

$$f(x, \mu, \sigma, \tau) = \frac{1}{\sqrt{2 \cdot \pi \cdot s^2}} \cdot \exp\left(-\frac{(m - X)^2}{2 \cdot s^2}\right)$$

Es werden die folgenden Werte benutzt: $m = 70, s = 15, \sigma_\epsilon = 0.1$. Diese Parameter sind nah an den geschätzten Parametern aus den Daten von [Mor18]. Der Glättungsparameter λ wird um die Berechnung zu Vereinfachen aus $\{100\}$ genommen. Das heißt für $\lambda \in \{100\}$ wird der GCV Wert bestimmt und das λ mit dem kleinsten Wert wird verwendet. Lambda wird dabei aus der Menge $\{2^r : r = 0, \dots, 10\}$ ausgewählt.

Betrachten wir noch die folgenden Graphik 14, die das Verfahren nach Whittaker-Henderson illustriert. Die durchgezogene Linie sind die Originaldaten und die gestrichelte Linie sind die mittels Whittaker geglätteten Daten. Die gepunkteten Linien ist das Modell.

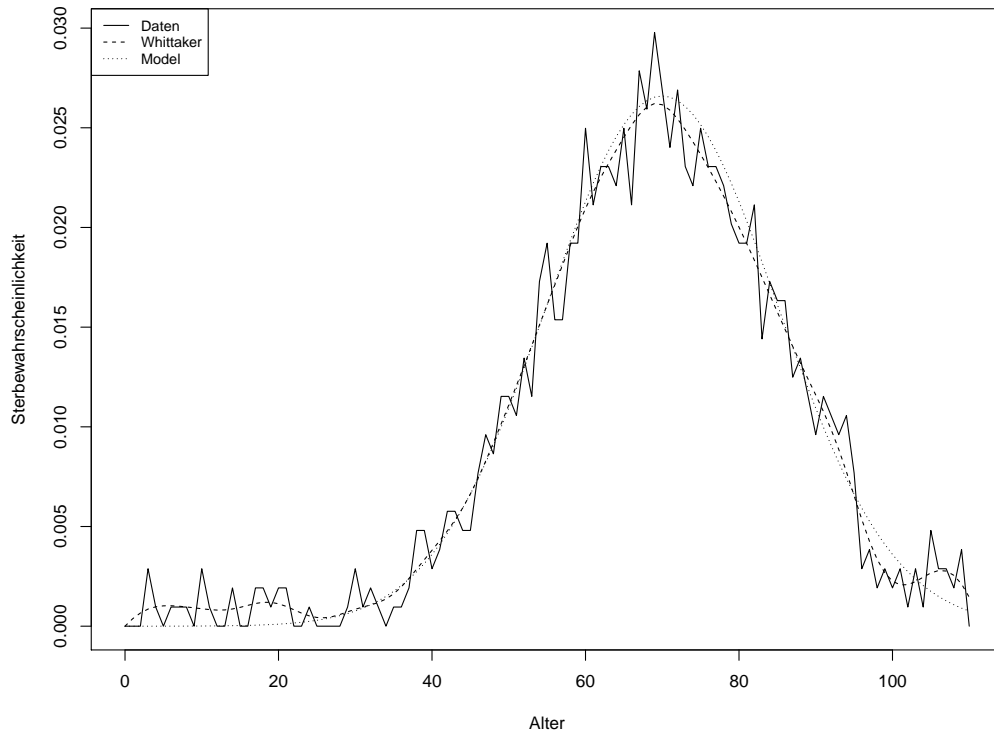


Figure 14: Glättung von Daten mittels Whittaker

Man sieht, dass das geglättete Modell dem wahren Modell stark ähnelt.

Betrachten wir, um den Abschnitt über Ausgleichsmethoden abzuschließen, noch die theoretischen Eigenschaften des Whittaker Modells. Das heißt, es wird wie in Kapitel 6 gezeigt, dass das Modell identifizierbar ist und der MSE bestimmt. Dazu braucht man allerdings zuerst ein statistisches Modell, da wir bisher nur ein numerisches Verfahren haben. Gehen wir zum Beispiel von dem folgenden Modell aus:

$$Q = \tilde{Q} + \epsilon \quad (16)$$

wobei \tilde{Q} die wahren Werte von Q sind und ϵ ein Fehler. Dieses Modell ist dann identifizierbar, da

$$\begin{aligned} P_{\theta_1} = P_{\theta_2} &\iff \tilde{Q}_1 + \epsilon = \tilde{Q}_2 + \epsilon \\ &\iff \tilde{Q}_1 = \tilde{Q}_2 \\ &\iff \theta_1 = \theta_2 \end{aligned}$$

Als nächstes wird der MSE berechnet

$$\begin{aligned}
\text{MSE}(Q') &= \text{Bias}(Q') + \text{Var}(Q') \\
&= \mathbb{E}((I + gK^t K)^{-1}Q - \tilde{Q}) + \text{Var}((I + gK^t K)^{-1}Q) \\
&= (I + gK^t K)^{-1}\mathbb{E}(Q) - \tilde{Q} + (I + gK^t K)^{-2}\text{Var}(Q) \\
&= (I + gK^t K)^{-1}\mathbb{E}(\tilde{Q} + \epsilon) - \tilde{Q} + (I + gK^t K)^{-2}\text{Var}(\tilde{Q} + \epsilon)
\end{aligned}$$

TODO : Jetzt brauchen wir Annahmen über die Verteilung des Fehlers vermute ich.

4.4 Übliche Methoden zur Trendschätzung

In diesem Kapitel werden übliche Methoden zum Schätzen von Sterblichkeitstrends betrachtet. Die DAV schlägt in der Dokumentation zur Sterbetafel 2004R ([DAV04]) im vierten Kapitel die folgenden drei Modelle vor: Das traditionelle Modell, das Kohortenmodell und das Synthesemodell. Hier die drei Modelle im Vergleich:

- traditionelles Modell:

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

- Kohortenmodell:

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

- Synthesenmodell:

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

laut Anhang 11 werden F beziehungsweise G 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. Die Ergebnisse der Trendschätzung können in DAV 2004R Anhang 1 Teil B gefunden werden. Als Sekundärliteratur zu diesen Modellen wird TODO empfohlen. Da diese Quelle nicht zu finden ist, betrachten wir im folgenden ein Modell das besser bekannt ist oder zu dem zumindest einfacher Quellen zu finden sind. Es handelt sich um das Lee-Carter Modell. Die folgenden Beschreibungen zum Lee-Carter Modell orientieren sich an [GK07].

- Das Modell: Seien $m_{a,t}$ die logarithmierten Sterblichkeitsraten in den Altersgruppen $a \in \{S, S + 1, \dots, E\}$ und zu den Zeitpunkten $t = 1, \dots, T$. Die Sterblichkeitsraten $m_{a,t}$ werden modelliert durch

$$m_{a,t} = \alpha_a + \beta_a \gamma_t + \epsilon_{a,t} \tag{17}$$

hier sind α_a, β_a und γ_t Parameter, die geschätzt werden müssen. Bei $\epsilon_{a,t}$ handelt es sich um zufällige Störungen. Die Parametrisierung (17) ist nicht eindeutig, da sie unter den folgenden beiden Transformationen invariant ist:

- $\beta_a \rightarrow c\beta_a$ und $\gamma_t \rightarrow \frac{1}{c}\gamma_t$ für $c \in \mathbb{R}, c \neq 0$
- $\alpha_a \rightarrow \alpha_a - \beta_a c$ und $\gamma_t \rightarrow \gamma_t + c$ für $c \in \mathbb{R}$

Dies führt dazu, dass die Likelihood-Funktion, die zu Modell (17) gehört mehr als ein Maximum hat. Deshalb werden in der Literatur zwei Nebenbedingungen eingeführt. Zum einen fordert man

$$\sum_t \gamma_t = 0 \quad (18)$$

und zum anderen

$$\sum_a \beta_a = 1 \quad (19)$$

Stellt man nun das Modell (17) nach γ_t um erhält man

$$\gamma_t = \frac{m_{a,t} - \alpha_a - \epsilon_{a,t}}{\beta_a}$$

Setzt man diese Gleichung für γ_t nun in die Nebenbedingung (18) ein, erhält man

$$\begin{aligned} 0 &= \sum_t \gamma_t = \sum_t \frac{m_{a,t} - \alpha_a - \epsilon_{a,t}}{\beta_a} \\ \iff 0 &= \sum_t (m_{a,t} - \alpha_a - \epsilon_{a,t}) \\ \iff \sum_t \alpha_a &= \sum_t m_{a,t} - \sum_t \epsilon_{a,t} \\ \iff \bar{m}_a &:= \frac{1}{t} \sum_t m_{a,t} = \alpha_a \end{aligned}$$

Dabei benutzt man, dass die Fehler in den Altersgruppen a über die Zeit t im Erwartungswert Null sind und deshalb in der Rechnung als Null angenommen werden. Dadurch lässt sich das Modell (17) neu schreiben, indem man $\alpha_a = \bar{m}_a$ und $\tilde{m}_{a,t} = m_{a,t} - \bar{m}_a$ setzt:

$$\tilde{m}_{a,t} = \beta_a \gamma_t + \epsilon_{a,t}$$

Weiterhin kann man Nebenbedingung (19) nach γ_t umstellen. Dazu stellt man zuerst das Modell (17) nach β_a um und erhält

$$\beta_a = \frac{m_{a,t} - \alpha_a - \epsilon_{a,t}}{\gamma_t}$$

setzt man diese Charakterisierung für β_a nun in Nebenbedingung (19) ein, erhält man

$$1 = \sum_a \beta_a = \sum_a \frac{m_{a,t} - \alpha_a - \epsilon_{a,t}}{\gamma_t}$$

$$\iff \gamma_t = \sum_a (m_{a,t} - \alpha_a) = \sum_a (m_{a,t} - \bar{m}_a)$$

- Schätzung der Parameter: Wir haben bereits Charakterisierungen für α und für γ gesehen. Es bleibt sich Gedanken über β zu machen. Dazu wird der Ansatz aus [GK07] verwendet. Dort wird behauptet, dass β mittels der Maximum-Likelihood-Methode zu schätzen äquivalent zu folgendem Ansatz ist: Zuerst bestimmt man die Matrix

$$\tilde{M} = \begin{pmatrix} \tilde{m}_{0,1956} & \dots & \tilde{m}_{0,2017} \\ \vdots & \vdots & \vdots \\ \tilde{m}_{110,1956} & \dots & \tilde{m}_{110,2017} \end{pmatrix} \quad (20)$$

falls die Alter $\{0, \dots, 110\}$ und die Kalenderjahre $\{1956, \dots, 2017\}$ sind. Als nächstes macht man eine Singulärwertzerlegung der Matrix $\tilde{M} = BLU'$ und schätzt β als die erste Spalte von B . Hier muss man darauf achten, dass die Spalten der Matrix B Länge eins haben und die Singulärwerte in L absteigend sortiert sind. Ansonsten muss man *beta* normieren und die Spalte aus B nehmen, die dem größten Singulärwert entspricht.

Nachdem man β geschätzt hat, kann man die Charakterisierungen für α und γ , die wir im vorhergehenden Abschnitt zum Modell gesehen haben, benutzen, um Schätzer zu generieren.

- Vorhersagen: Im Lee-Carter Modell geht man davon aus, dass α und β über die Zeit konstant bleiben. Das heißt, es genügt γ_t für t außerhalb der Beobachtungsperiode zu schätzen.

Um aus diesem Modell zukünftige Werte für γ_t zu schätzen, schlagen Lee und Carter ein Random Walk with Drift (RWD) Modell vor:

$$\hat{\gamma}_t = \hat{\gamma}_{t-1} + \nu + \xi_t \quad (21)$$

hier ist ν ein unbekannter Drift Parameter der geschätzt werden muss und $\xi_t \sim \mathcal{N}(0, \sigma_{\text{rw}}^2)$. Der Maximum-Likelihood Schätzer für ν ist

$$\hat{\nu} = (\hat{\gamma}_E - \hat{\gamma}_S) / (E - S)$$

hier sind E, S das Endalter und das Startalter der Sterbetafel.

- Bedeutung der Parameter: In diesem Punkt geht es darum, die Bedeutung der Parameter zu diskutieren. Gehen wir zuerst die Parameter einzeln durch:
 - Alpha: Der Alpha-Parameter ist auf Grund der Nebenbedingungen der Mittelwert der Sterblichkeiten über die Kalenderjahre. Der folgende Plot 15 zeigt die Werte von Alpha für das oben genannte Beispiel. Man kann diesen Parameter als die Sterblichkeit ohne Trends verstehen.

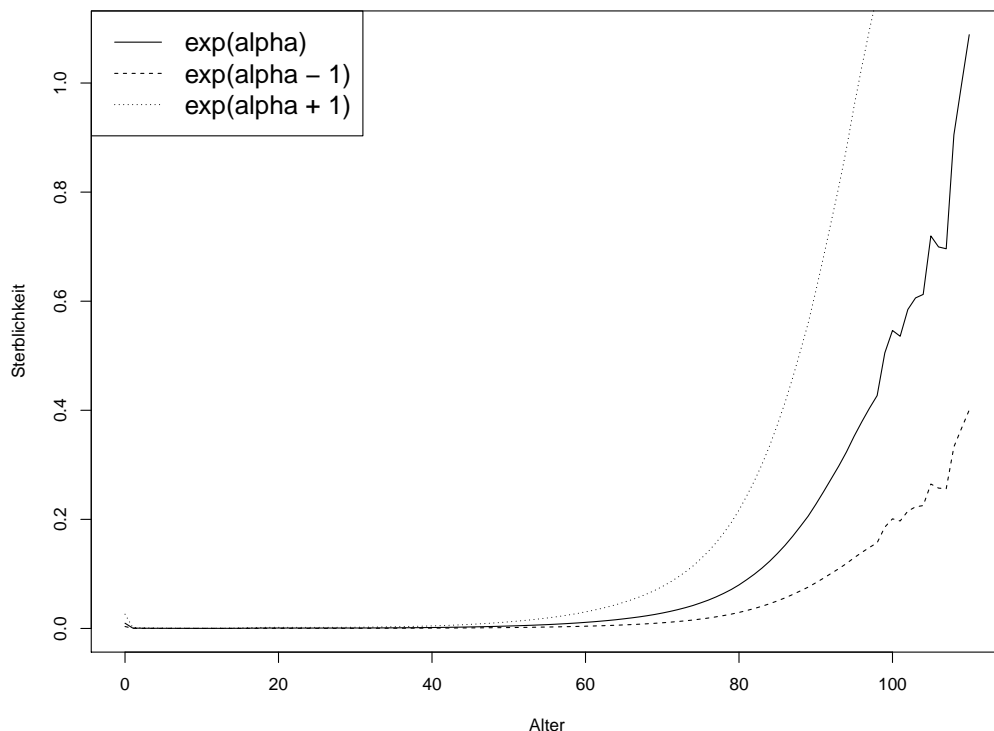


Figure 15: Lee-Carter Schätzung für $\exp(\alpha)$ für Westdeutsche Periodendaten 1956 bis 2017

Betrachtet man die Graphik 15, sieht man, dass die Sterbewahrscheinlichkeit mit den Altern zuerst langsam und dann stark ansteigt. Die Sterbewahrscheinlichkeit für das Endalter beträgt 1.0886442632 und ist damit über 100 %. Insgesamt ist der Verlauf dieser Sterblichkeitsfunktion recht sinnvoll: Wir betrachten Periodendaten, das heißt, wir betrachten nur ein Kalenderjahr und in diesem Kalenderjahr die Verteilung von Toden und wir sehen, dass von den älteren Personen relativ mehr sterben als von den jungen Personen.

Außerdem veranschaulicht die Graphik 15, dass ein höherer Wert von Alpha zu höheren Sterbewahrscheinlichkeiten führt und ein niedrigerer Wert zu niedrigeren. Dies wird bei der Diskussion der Sterblichkeitsverbesserung wichtig.

- Beta: Der Beta-Parameter gibt an wie stark der Einfluss der Sterblichkeitsverbesserung ist, indem er den Gamma-Parameter Gewichtet. Die Gewichtungen, die man bei den Westdeutschen Periodendaten bekommt ist wie in Graphik 16 gezeigt.

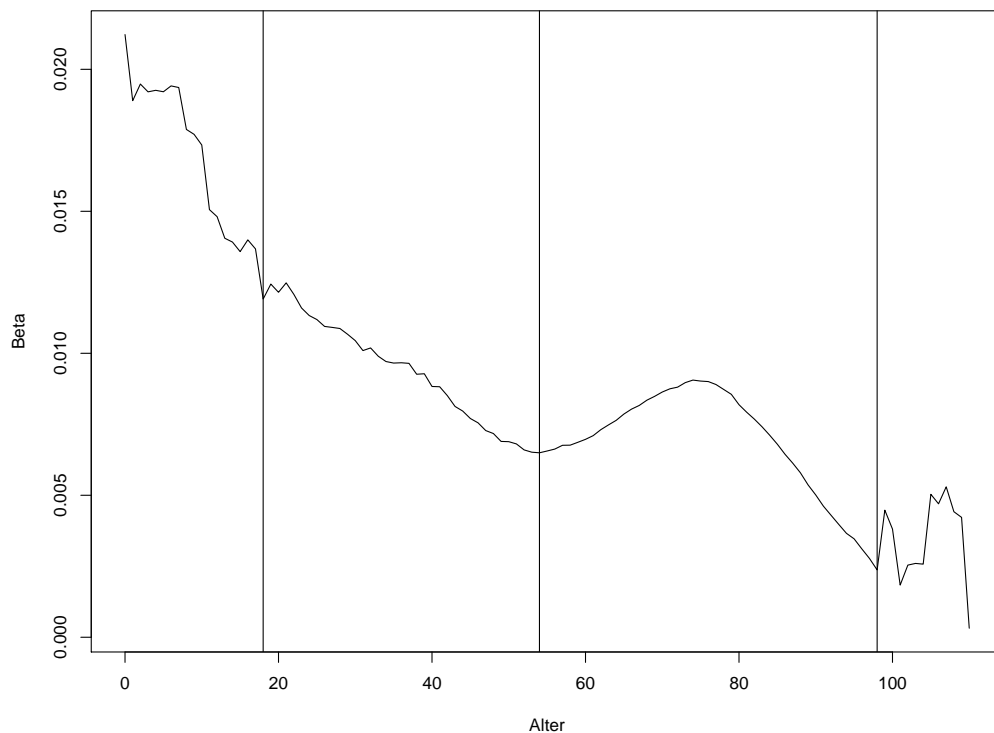


Figure 16: Lee-Carter Schätzung für Beta für Westdeutsche Periodendaten 1956 bis 2017. Als optische Hilfe sind 4 Bereiche verschiedener Sterblichkeitsverbesserung eingetragen.

Man sieht, dass der Einfluss der Sterblichkeitsverbesserung zuerst stark ist sich dann abschwächt, dann wieder steigt und dann wieder absinkt. Um den Einfluss auf die Sterblichkeitsverbesserung zu verstehen, muss man Beta und Gamma gemeinsam betrachten. Dies wird später gemacht.

- Gamma: Der Gamma-Parameter gibt die Verbesserung beziehungsweise die Verschlechterung der Sterblichkeit an. Beides ist grundsätzlich möglich. In der folgenden Graphik 17 sieht man wie sich γ_t im Laufe der Jahre entwickelt hat.



Figure 17: Lee-Carter Schätzung für Gamma für Westdeutsche Periodendaten 1956 bis 2017. Zur optischen Hilfe ist eine Linie auf Höhe der Null geplottet.

Man sieht, dass die Sterblichkeitsverbesserung ungefähr linear verläuft und von einem Wert von plus 60 zu einem Wert von -60 läuft. Da der Beta-Parameter immer positiv ist, gibt der Gamma-Parameter das Vorzeichen vor.

An dieser Stelle kann man versuchen den etwas unintuitiven Verlauf der Beta-Alter-Funktion zu verstehen: Intuitiv würde man denken, dass die Sterblichkeitsverbesserung insbesondere Neugeborenen und alten Personen beeinflusst.

Hier haben wir allerdings eine sehr starke Verbesserung bei unter 19 jährigen dann eine etwas langsermere Verbesserung bis zum Alter 54, dann eine Buckel und ab 98 keine Bestimmte Struktur mehr. Die Sterblichkeit der unter 19 jährigen lässt sich gut durch einen Rückgang der Kindersterblichkeit erklären. So sank laut Wikipedia "die Kindersterblichkeit weltweit [...] von 18,2 % im Jahr 1960 auf 4,3 % im Jahr 2015". Auch der letzte Abschnitt ab 98 lässt sich gut erklären: Für diese Personengruppe sind nicht genug Daten verfügbar, da nur wenige Personen solch ein Alter erreichen. Der Buckel zwischen 54 und 98 lässt sich damit erklären, dass an diesem Altersbereich der Großteil der sichtbaren Sterblichkeitsverbesserung passiert: Das mittlere Alter von Personen ist in diesem Bereich angestiegen und das deckt sich mit der Alltagserfahrung. Perso-

nen die früher nur 70 geworden sind werden heute 80. Für den Bereich zwischen 19 und 54 bietet sich keine Erklärung direkt an. Ich vermute, dass es sich hierbei um Verbesserungen der Arbeitsbedingungen (Arbeitsunfällen) und bessere Verkehrssicherheit (Anschnallpflicht) handelt.

- ν : Der ν -Parameter wird nur für Vorhersagen gebraucht und gibt wieder, wie sich der Gamma-Parameter verändert. Schätzt man ν im Beispiel erhält man -1.908261. Das heißt von einem Jahr zum nächsten ändert sich Gamma und -1.908261. Berechnen wir also beispielsweise des Verhältnis der Sterblichkeit von einem Jahr zum nächsten und benutzen dabei den Mittelwert der Betas von 0.009009, so erhalten wir

$$\frac{e^{\alpha+\beta(\gamma-1.908261)}}{e^{\alpha+\beta\gamma}} = e^{\alpha-\alpha+\beta\gamma-\beta\gamma-\beta 1.908261} = e^{-1.908261\beta} = e^{-1.908261 \cdot 0.009009} \approx 0.98$$

Das heißt, die Sterblichkeit verringert sich pro Jahr im Mittel um zwei Prozent.

Im folgenden Plot 18 sieht man die Werte der Produkte der Beta-Parameter und der Gamma-Parameter. Dass heißt, man sieht die folgende Matrix:

$$\begin{pmatrix} \beta_0 \cdot \gamma_S & \dots & \beta_{110} \cdot \gamma_S \\ \vdots & \vdots & \vdots \\ \beta_0 \cdot \gamma_E & \dots & \beta_{110} \cdot \gamma_E \end{pmatrix} \quad (22)$$

Hier der Plot:

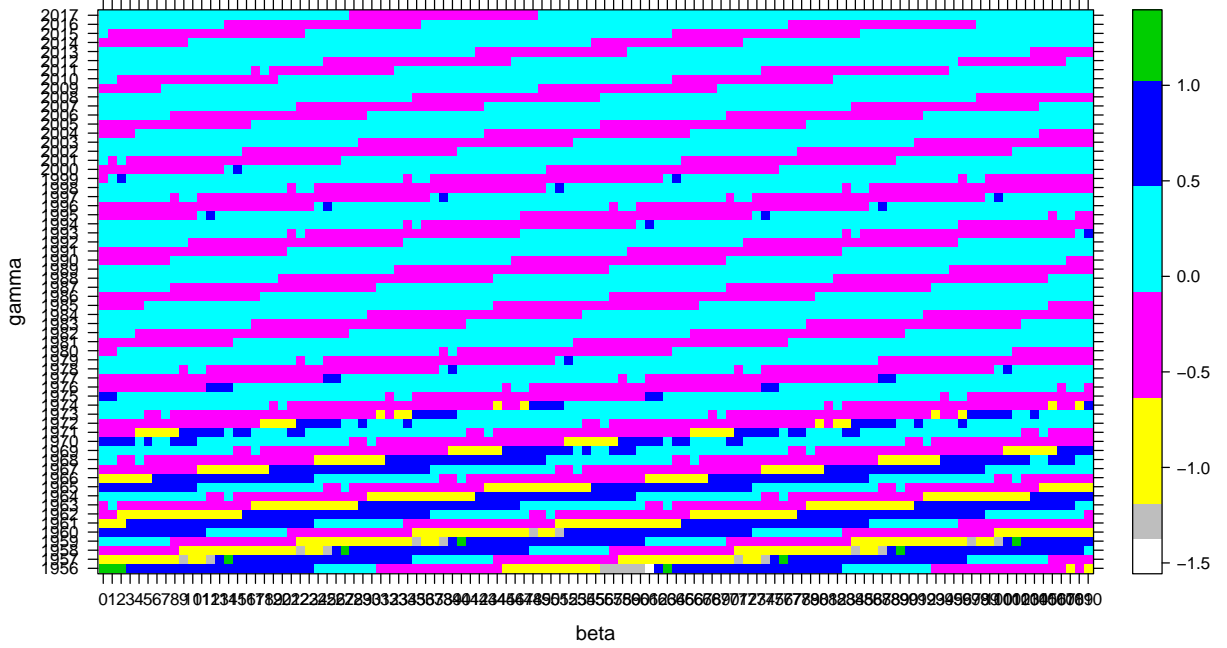


Figure 18: Heatmap für (γ_t, β_a)

Man sieht TODO

Damit wurde das Lee-Carter-Modell angegeben und die Parameter interpretiert. Außerdem wurde das Schätzen der Parameter beschrieben. Dies beendet die Vorstellung des Lee-Carter-Modells. Als nächstes betrachten wir die theoretische Eigenschaften des Lee-Carter-Modells. Das Ziel ist wieder das Theorem von Cramér anzuwenden. Hier noch einmal das Theorem:

Theorem 2 (Cramér). Sei $\mathbf{X} = (X_1, \dots, X_n)$ ein Sample aus einer parametrischen Familie $\mathcal{P} = \{P_\theta : \theta \in \Theta \subset \mathbb{R}^k\}$, wobei P_θ eine Dichte f_θ bezüglich des Lebesguemaßes hat. Sei θ_0 der wahre Parameter. Sei weiterhin

1. Die Menge Θ ist eine offene Teilmenge des \mathbb{R}^n
2. Das Modell ist identifizierbar.

3. Die zweite Ableitung von f_θ existiert und ist stetig für alle $x \in \chi$ und kann unter das Integral $\int f_\theta(x)dx$ geschoben werden.
4. Es existiert eine messbare, integrierbare Funktion $K(x)$, sodass $|l''_\theta(x)| \leq K(x)$ für alle $\theta : |\theta - \theta_0| < \epsilon$.
5. Die Fisher-Informationsmatrix $i(\theta_0)$ ist positiv definit.

Dann existiert eine stark konsistente Folge $\hat{\theta}$ von Wurzeln von $n^{-1} \sum_{i=1}^n l'_\theta(X_i) = 0$ mit $\sqrt{n}(\hat{\theta} - \theta) \rightarrow^D \mathcal{N}(0_k, i(\theta_0)^{-1})$.

Proof. Für den Beweis siehe Fergusons Buch "A course in large sample theory" ,[Fer96], Theorem 18. \square

Wir geben wieder zuerst das Modell noch einmal an und zeigen dann die fünf Eigenschaften. Hierbei untersuchen wir β_a und ν getrennt. Betrachten wir zuerst das Modell für β_a :

$$\tilde{m}_{a,t} \sim N(\beta_a \hat{\gamma}_t, \sigma_{\epsilon_{a,t}})$$

Entsprechend erhalten wir die gemeinsame Dichte

$$\Pi_{a=1}^n \Pi_{t=1}^n f(\tilde{m}_{a,t} | \beta_a, \sigma_{\epsilon_{a,t}}) = \Pi_{a=1}^n \Pi_{t=1}^n \frac{1}{\sqrt{2\pi\sigma_{\epsilon_{a,t}}^2}} \exp\left(-\frac{(\beta_a \hat{\gamma}_t - \tilde{m}_{a,t})^2}{2\sigma_{\epsilon_{a,t}}^2}\right)$$

und die log-likelihood

$$\begin{aligned} & \ln\left(\Pi_{a=1}^n \Pi_{t=1}^n f(\tilde{m}_{a,t} | \beta_a, \sigma_{\epsilon_{a,t}})\right) \\ &= \ln\left(\Pi_{a=1}^n \Pi_{t=1}^n \frac{1}{\sqrt{2\pi\sigma_{\epsilon_{a,t}}^2}} \exp\left(-\frac{(\beta_a \hat{\gamma}_t - \tilde{m}_{a,t})^2}{2\sigma_{\epsilon_{a,t}}^2}\right)\right) \\ &= \sum_{a=1}^n \sum_{t=1}^n \left(\frac{\ln(2\pi\sigma_{\epsilon_{a,t}}^2)}{2} - \frac{(\beta_a \hat{\gamma}_t - \tilde{m}_{a,t})^2}{2\sigma_{\epsilon_{a,t}}^2}\right) \end{aligned}$$

Wir erhalten als Ableitung in β_a Richtung

$$\begin{aligned}
& \frac{d}{d\beta_a} l_{\tilde{m}_{a,t}}(\beta_a) \\
&= \frac{d}{d\beta_a} \sum_{a=1}^n \sum_{t=1}^n \left(\frac{\ln(2\pi\sigma_{\epsilon_{a,t}}^2)}{2} - \frac{(\beta_a \hat{\gamma}_t - \tilde{m}_{a,t})^2}{2\sigma_{\epsilon_{a,t}}^2} \right) \\
&= \frac{1}{2\sigma_{\epsilon_{a,t}}^2} \sum_{a=1}^n \sum_{t=1}^n \frac{d}{d\beta_a} \left(-(\beta_a \hat{\gamma}_t - \tilde{m}_{a,t})^2 \right) \\
&= \frac{1}{2\sigma_{\epsilon_{a,t}}^2} \sum_{a=1}^n \sum_{t=1}^n (-2(\beta_a \hat{\gamma}_t - \tilde{m}_{a,t}) \hat{\gamma}_t) \\
&= \frac{1}{\sigma_{\epsilon_{a,t}}^2} \left(\sum_{a=1}^n \sum_{t=1}^n \tilde{m}_{a,t} \hat{\gamma}_t - \sum_{a=1}^n \sum_{t=1}^n \beta_a \hat{\gamma}_t^2 \right) \\
&= \frac{1}{\sigma_{\epsilon_{a,t}}^2} \left(\sum_{a=1}^n \sum_{t=1}^n \tilde{m}_{a,t} \hat{\gamma}_t - \left(\sum_{a=1}^n \beta_a \right) \cdot \left(\sum_{t=1}^n \hat{\gamma}_t^2 \right) \right)
\end{aligned}$$

und erhalten als Nullstelle:

$$\begin{aligned}
0 &= \frac{1}{\sigma_{\epsilon_{a,t}}^2} \left(\sum_{a=1}^n \sum_{t=1}^n \tilde{m}_{a,t} \hat{\gamma}_t - \left(\sum_{a=1}^n \beta_a \right) \cdot \left(\sum_{t=1}^n \hat{\gamma}_t^2 \right) \right) \\
\iff 0 &= \sum_{a=1}^n \sum_{t=1}^n \tilde{m}_{a,t} \hat{\gamma}_t - \left(\sum_{a=1}^n \beta_a \right) \cdot \left(\sum_{t=1}^n \hat{\gamma}_t^2 \right) \\
\iff \sum_{a=1}^n \sum_{t=1}^n \tilde{m}_{a,t} \hat{\gamma}_t &= \left(\sum_{a=1}^n \beta_a \right) \cdot \left(\sum_{t=1}^n \hat{\gamma}_t^2 \right) \\
\iff \left(\sum_{a=1}^n \sum_{t=1}^n \tilde{m}_{a,t} \hat{\gamma}_t \right) \cdot \left(\sum_{t=1}^n \hat{\gamma}_t^2 \right)^{-1} &= \left(\sum_{a=1}^n \beta_a \right)
\end{aligned}$$

Das heit, wir erhalten als Maximum-Likelihood-Schtzer:

$$\hat{\beta}_a = \left(\sum_{t=1}^n \tilde{m}_{a,t} \hat{\gamma}_t \right) \cdot \left(\sum_{t=1}^n \hat{\gamma}_t^2 \right)^{-1}$$

Die zweite Ableitung ist

$$\begin{aligned}
& \frac{d^2}{d\beta_a^2} l_{\tilde{m}_{a,t}}(\beta_a) \\
&= \frac{d}{d\beta_a} \frac{1}{\sigma_{\epsilon_{a,t}}^2} \left(\sum_{a=1}^n \sum_{t=1}^n \tilde{m}_{a,t} \hat{\gamma}_t - \left(\sum_{a=1}^n \beta_a \right) \cdot \left(\sum_{t=1}^n \hat{\gamma}_t^2 \right) \right) \\
&= -\frac{n}{\sigma_{\epsilon_{a,t}}^2} \left(\sum_{t=1}^n \hat{\gamma}_t^2 \right)
\end{aligned}$$

Die zweite Ableitung ist immer negativ. Also haben wir in der Tat ein Maximum gefunden. Jetzt werden die fünf Eigenschaften gezeigt:

- Die Menge Θ ist eine offene Teilmenge des \mathbb{R}^n : Wir zeigen diese Eigenschaft für alle Parameter gleichzeitig. Das heißt wir setzen $\theta = (\alpha_a, \beta_a, \gamma_t, \nu)$. Dies ist wieder kein Problem, da man Θ passend wählen kann. Zum Beispiel $\alpha_a, \beta_a, \gamma_t \in (c_1, c_2)$ und $\nu \in (c_3, c_4)$ mit c_1, c_2, c_3, c_4 passend.
- Das Modell ist identifizierbar: Wir zeigen Identifizierbarkeit für alle Parameter gleichzeitig. Das heißt wir zeigen Identifizierbarkeit für $\theta = (\alpha_a, \beta_a, \gamma_t, \nu)$. Wir setzen wieder die Definition an:

$$P_{\theta_1} = P_{\theta_2} \iff \alpha_a + \beta_a \gamma_t + \epsilon = \alpha'_a + \beta'_a \gamma'_t + \epsilon \text{ und } \gamma_{t-1} + \nu + \xi = \gamma'_{t-1} + \nu' + \xi$$

betrachten wir zuerst den ersten Summanden und setzen die Nebenbedingungen für $\alpha_a = \bar{m}_a = \alpha'_a$ und für $\gamma_t = \sum_a (m_{a,t} - \bar{m}_a)$ ein:

$$\begin{aligned} \alpha_a + \beta_a \gamma_t + \epsilon = \alpha'_a + \beta'_a \gamma'_t + \epsilon &\iff \alpha_a + \beta_a \gamma_t = \alpha'_a + \beta'_a \gamma'_t \\ &\iff \bar{m}_a + \beta_a \sum_a (m_{a,t} - \bar{m}_a) = \bar{m}_a + \beta'_a \sum_a (m_{a,t} - \bar{m}_a) \\ &\iff \beta_a = \beta'_a \end{aligned}$$

betrachten wir nun den zweiten Summanden

$$\begin{aligned} \gamma_{t-1} + \nu + \xi = \gamma'_{t-1} + \nu' + \xi &\iff \gamma_{t-1} + \nu = \gamma'_{t-1} + \nu' \\ &\iff \sum_a (m_{a,t-1} - \bar{m}_a) + \nu = \sum_a (m_{a,t-1} - \bar{m}_a) + \nu' \\ &\iff \nu = \nu' \end{aligned}$$

Das heißt insgesamt haben wir $\theta = (\alpha_a, \beta_a, \gamma_t, \nu) = (\alpha'_a, \beta'_a, \gamma'_t, \nu') = \theta'$ gezeigt und das Modell ist somit identifizierbar.

- Die zweite Ableitung von f_θ existiert und ist stetig für alle $x \in \chi$ und kann unter das Integral $\int f_\theta(x) dx$ geschoben werden:

?Wir haben wieder eine Verkettung glatter Funktionen und eine Exponentialfamilie; deshalb ist nichts zu zeigen?

- Es existiert eine messbare, integrierbare Funktion $K(x)$, sodass $|l''_\theta(x)| \leq K(x)$ für alle $\theta : |\theta - \theta_0| < \epsilon$:

?Die zweite Ableitung ist unabhängig von θ und deshalb beschränkt?

- Die Fisher-Informationsmatrix $i(\theta_0)$ ist positiv definit:

Die zweite Ableitung ist

$$\frac{d^2}{d\beta_a^2} l_{\tilde{m}_{a,t}}(\beta_a) = -\frac{n}{\sigma_{\epsilon_{a,t}}^2} \left(\sum_{t=1}^n \hat{\gamma}_t^2 \right)$$

damit ist die Fisher-Informationsmatrix

$$\begin{aligned} i(\theta_0) &= -\mathbb{E} \left(\frac{d^2}{d\beta_a^2} l_{\tilde{m}_{a,t}}(\beta_a) \right) \\ &= -\mathbb{E} \left(-\frac{n}{\sigma_{\epsilon_{a,t}}^2} \left(\sum_{t=1}^n \hat{\gamma}_t^2 \right) \right) \\ &= \frac{n}{\sigma_{\epsilon_{a,t}}^2} \left(\sum_{t=1}^n \hat{\gamma}_t^2 \right) \end{aligned}$$

und da dieser letzte Ausdruck immer positiv ist und in \mathbb{R} alle Matritzen symmetrisch und gleich ihrem Eigenwert sind, ist die Fisher-Informationsmatrix positiv definit.

Betrachten wir als nächstes ν . Zuerst das Modell:

$$\sum_a (m_{a,t} - \bar{m}_a) \sim N(\sum_a (m_{a,t-1} - \bar{m}_a + \nu, \sigma_{\xi_t}))$$

Dann die gemeinsame Dichte:

TODO

und die Log-Likelihood:

TODO

Es bleibt Eigenschaften drei, vier und fünf zu zeigen: *TODO*.

Als nächstes wird der MSE berechnet

$$\text{MSE}(\hat{\theta}) = \text{MSE}(\hat{\alpha}_a, \hat{\beta}_a, \hat{\gamma}_t, \hat{\nu}) = \text{Bias}(\hat{\alpha}_a, \hat{\beta}_a, \hat{\gamma}_t, \hat{\nu}) + \text{Var}(\hat{\alpha}_a, \hat{\beta}_a, \hat{\gamma}_t, \hat{\nu})$$

TODO: Varianz von einem Vektor -> Kovarianzmatrix; Bias von einem Vektor -> Summe der Elemente?

5 Verbesserung des Einfachen Modells

Die folgende Graphik zeigt die Sterblichkeit pro Alter und pro Geburtsjahrgang:

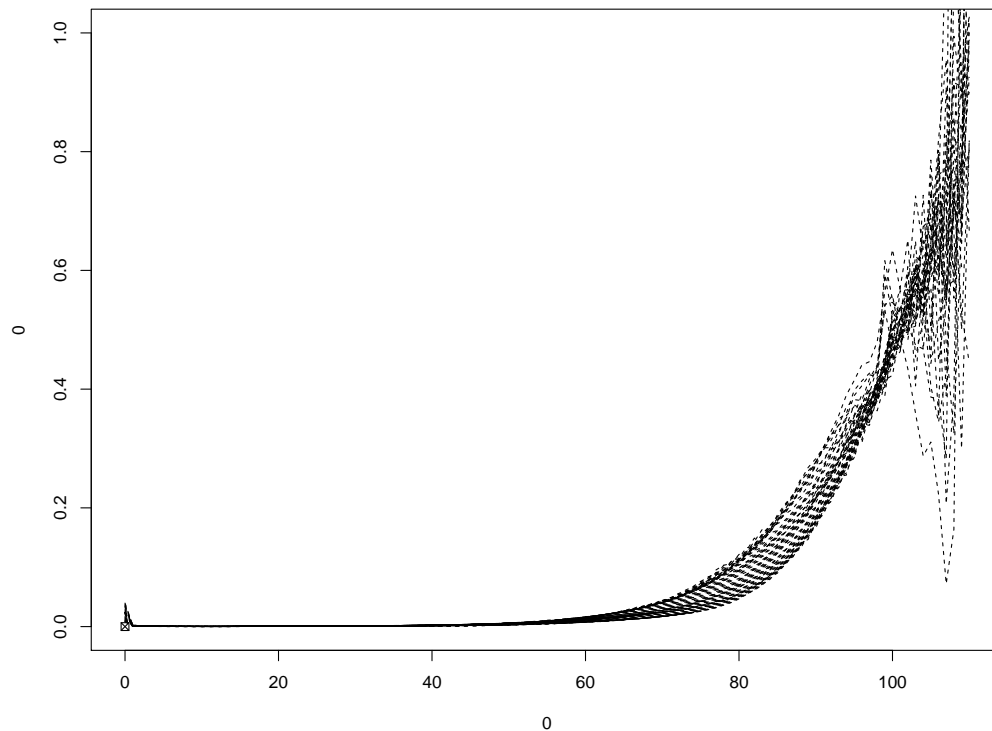


Figure 19: abschnittsweises-Dichteschätzen

In diesem Abschnitt werden Ansätze vorgestellt, dass einfache Modell zu verbessern.

5.1 Einfaches Trendmodell

Bisher sind wir von dem Modell

$$Y = \frac{1}{\sqrt{2 \cdot \pi \cdot s^2}} \cdot \exp\left(-\frac{(m - X)^2}{2 \cdot s^2}\right) + \epsilon$$

ausgegangen. Dieses Modell beachtet in keiner Weise das Geburtsjahr. Das heißt, Änderungen in der Sterblichkeit werden nicht berücksichtigt. Wir haben bereits den Random Walk with Drift (RWD) bei dem Lee-Carter Modell gesehen. Fügen wir also auch diesem Modell eine RWD Komponente hinzu, indem wir von dem Modell

$$Y = \frac{1}{\sqrt{2 \cdot \pi \cdot s^2}} \cdot \exp\left(-\frac{(m_t - X)^2}{2 \cdot s^2}\right) + \epsilon$$

mit

$$m_t = m_{t-1} + \nu + \xi_t \quad (23)$$

ausgehen. Dabei ist ν der unbekannte Driftparameter und ξ_t ein Fehler. Schätzt man die μ_t für den Datensatz der Generationssterbetafel erhält man den folgenden Plot:



Figure 20: Ein Plot der $\hat{\mu}_t$ gegen die Geburtsjahre

-> TODO: Problem : Die Zensur führt dazu, dass die Mittelwerte stark schwanken/ der Trend überschätzt wird. Eine Lösung wäre, μ in (a,b) zu verlangen. Eine andere direkt nur das Quantil zu schätzen (man müsste aus dem Geburtsjahr bestimmen können welchen Abschnitt/Quantilteil man braucht. Vielleicht kann man es dann so einrichten, dass man nur dieses Teil schätzt...)

Als nächste Verbesserung könnten wir uns von dem parametrischen Modell verabschieden und zu Kerndichteschätzern übergehen. Allerdings ist dies mit Zensur? wahrscheinlich? nicht möglich.

TODO: Schätzer in diesem Modell einfügen

Zeigen wir wieder, dass dieses Modell identifizierbar ist, berechnen den MSE und zeigen Konsistenz und eine asymptotische Verteilung.

Beginnen wir mit der Identifizierbarkeit. Wir haben bereits in dem Kapitel über das einfache Modell gezeigt, dass

$$P_\theta = P_{\theta'} \iff (m, s) = (m', s')$$

und daran ändert sich auch durch das Hinzufügen des Random Walk with Drift nichts. Es bleibt zu zeigen, dass $\nu = \nu'$:

$$\begin{aligned} P_\theta = P_{\theta'} &\iff \frac{1}{\sqrt{2 \cdot \pi \cdot s^2}} \cdot \exp\left(-\frac{(m_t - X)^2}{2 \cdot s^2}\right) + \epsilon = \frac{1}{\sqrt{2 \cdot \pi \cdot s'^2}} \cdot \exp\left(-\frac{(m'_t - X)^2}{2 \cdot s'^2}\right) + \epsilon \\ &\quad \text{und } m_{t-1} + \nu + \xi_t = m'_{t-1} + \nu' + \xi_t \\ &\iff m_{t-1} + \nu = m'_{t-1} + \nu' \\ &\iff \nu = \nu' \end{aligned}$$

Also folgt insgesamt

$$P_\theta = P_{\theta'} \iff (m, s, \nu) = (m', s', \nu') \iff \theta = \theta'$$

und somit Identifizierbarkeit.

TODO : MSE berechnen

TODO : Konsistenz und asymptotische Verteilung zeigen

6 Qualitätsvergleich der Modelle mittels Simulation

In diesem Kapitel werden Vergleiche der drei Methoden vorgestellt.

6.1 Qualitätsvergleich durch Simulation

In diesem Unterkapitel werden Simulationen vorgestellt, um die Qualität der verwendeten Methoden zu vergleichen. In einem ersten Schritt wird das einfachste Modell und das Whittaker-Henderson-Verfahren verglichen. Dazu werden zwei Arten von Simulationen durchgeführt: In der ersten Simulation wird aus dem Modell des einfachsten Schätzers, dass heißt

$$Y = f(x) + \epsilon$$

beziehungsweise

$$Y \sim N(70, 15)$$

Daten generiert. Dabei werden $\mu = 70$ und $\sigma = 15$ verwendet. Diese sind genau die Werte, die aus dem Datensatz geschätzt wurden. Es werden $n = 1000$ Personen simuliert. Insgesamt werden $m = 100$ Datensätze erzeugt.

Jahr	Alter	Anzahl an Toden
1900	0	$f(0, \mu, \sigma, 1900)$
\vdots	\vdots	\vdots
1900	110	$f(110, \mu, \sigma, 1900)$
1901	0	$f(0, \mu, \sigma, 1901)$
\vdots	\vdots	\vdots
2000	110	$f(110, \mu, \sigma, 2000)$

TODO : Fehlerfunktion beschreiben

In einem zweiten Schritt wird der Datensatz halbiert. Auf die eine Hälfte werden die Verfahren angewandt und mit der anderen Hälfte wird die Qualität bestimmt. Wegen der Zensur werden die Datensätze nach Geburtsjahr Modus zwei aufgeteilt.

Aus diesen beiden Simulationen erhält man die folgenden Ergebnisse:

Methode	Einfaches Modell	Whittaker-Henderson
Simulation	0.0005537667	0.007324122
Datensatz	0.02319633	0.02302209

Man sieht, dass das parametrische Modell, wenn es stimmt, besser als das nicht-parametrische ist. Dies ist genau, was man erwarten würde.

TODO einen ähnlichen Vergleich für Lee-Carter und Abschnittsweises-Dichteschätzen.

Methode	Verbessertes Modell	Lee-Carter
Simulation	TODO	TODO
Datensatz	TODO	TODO

6.2 Qualitätsvergleich der Modelle mittels Theorie

Wir haben bereits in den Kapiteln zu den verschiedenen Schätzmethoden die folgenden Fragen beantwortet:

- Ist das Modell identifizierbar?
- Was ist der MSE des Schätzers?
- Ist der Schätzer konsistent?
- Was ist die asymptotische Verteilung?

In diesem Kapitel sollen diese Ergebnisse verglichen werden:

TODO : Ergebnisse ausrechnen und hier einfügen.

	einfaches Modell	Whittaker	Lee-Carter	Verbessertes Modell
Identifizierbarkeit	✓	✓	✓	✓
MSE	TODO	TODO	TODO	TODO
Kosistenz	TODO	TODO	TODO	TODO
Asymptotische Verteilung	TODO	TODO	TODO	TODO

7 Ausblick

In diesen Abschnitt kommen Ideen, die es zwischendurch gab, die aber nicht weiter verfolgt wurden.

7.1 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

$$T_{JS}(B) = \left(1 - \frac{d-2}{\|B\|^2}\right) B \quad (24)$$

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,

$$\hat{q}_x^\tau = \left(1 - \frac{111-2}{\sqrt{\sum_{i=0}^{110} \hat{q}_i^{\text{tau}}}}\right) \hat{q}_x^\tau \quad (25)$$

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.

7.2 Qualitätsvergleich mittels Tests

TODO : Testverfahren in der Literatur/Überblick Quelle: Kapitel 2 aus Methodik von Sterblichkeitsuntersuchungen

Insbesondere sollen die folgenden beiden Fragen mittels statistischen Tests beantwortet werden:

- Ist es notwendig Trends zu bestimmen? Dazu wird geschaut, ob eine Sterbetafel aus z.B. 1950 und eine von 2000 signifikant gleich sind.
- Unterscheiden sich die drei Sterbetafeln, die mittels der drei Verfahren bestimmt wurden signifikant?

TODO : Der Kolmogorov–Smirnov Anpassungstest sieht gut aus; vorstellen
 TODO : Das einfache Modell und Whittaker testen
 TODO : Das finale verbesserte Verfahren und Lee-Carter testen

7.3 Extrapolation von hohen Altern

Dieser Abschnitt orientiert sich an der Herleitung der DAV Tafel 2004R Anhang 9. Die DAV Tafel 2004R hat in der Basistafel nur Alter bis 99. Die Alter 100 bis 120 werden mittels der Methodik aus TKV = Thatcher "The force of mortality at Ages 80 to 120" (1998) geschätzt:

sechs Modelle:

- Gompertz-Modell: $q_x = 1 - \exp(-\exp(a + bx))$
- Quadratisches Modell: $q_x = 1 - \exp(-\exp(a + bx + cx^2))$
- Heligmann-Pollard-Modell: $q_x = (a \exp(bx)) / (1 + a \exp(bx))$
- Weibull-Modell: $q_x = 1 - \exp(-a(x + (1/2))^b)$
- Kannisto-Modell: $q_x = 1 - \exp(-((a \exp(bx)) / (1 + a \exp(bx))))$
- Logistische Modell $q_x = 1 - \exp(-(\beta \exp(bx)) / (1 + \alpha \exp(bx)))$

TODO TKV lesen; TKV kann auf <http://www.demogr.mpg.de/Papers/Books/Monograph5/start.htm> eingesehen werden.

Die DAV schätzt dann alle Parameter für die Basistafel und dann wird auf Grund von Gütekriterien das beste Modell ausgewählt. Konkret werden drei herangezogen:

- Vergleich von gemäß extrapolierten Sterbewahrscheinlichkeiten erwarteten Toten mit tatsächlichen Toten $\sum_{x=96}^{99} (\hat{q}_x L_x) / \sum_{x=96}^{99} \hat{q}_x L_x$
- Wert der Loglikelihood (ohne konstanten Summanden): $\sum_{x=96}^{99} T_x \ln(\hat{q}_x) + \sum_{x=96}^{99} (L_x - T_x) \ln(1 - \hat{q}_x)$
- Wert der χ^2 -Statistik: $\sum_{x=96}^{99} (T_x - L_x \hat{q}_x)^2 / (L_x \hat{q}_x (1 - \hat{q}_x))$

7.4 Modellbildung

Hier ein anderer Ansatz zur Modellbildung:

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 schlecht 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:

$$\frac{q_{x,\tau+1}}{q_{x,\tau}} = f(x, \tau) \quad (26)$$

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 (26) geschätzt.

7.5 Verschiedenes

- Internationalisierung: Eventuell ist die Frage nach dem West-Ost-Deutschland Vergleich interessant und der Internationale Vergleich.
- Interpretation als Markovkette: Man kann eventuell einen Sterbeprozess als Markovkette auffassen und damit Wahrscheinlichkeiten schätzen.
-> TODO Graphik mittels Inkscape erstellen
- Weitere Qualitätsvergleiche: Die DAV schlägt außerdem noch einen
-> Anhang 11 : Likelihood-Ratio-Test zum Modellvergleich
- Trenddämpfung: Dann gibt auch noch die Trenddämpfung. Trenddämpfung bedeutet, dass sich der Trend der Sterblichkeitsverbesserung im Laufe der Zeit abschwächt. Diese Abschwächung sollte man auch schätzen und in das Modell integrieren. TODO

References

- [Bel96] E. Beleites, *Grenzen der medizin: Weil die medizin in europa so gut ist ...* <https://www.aerzteblatt.de/archiv/357/Grenzen-der-Medizin-Weil-die-Medizin-in-Europa-so-gut-ist>, Accessed: 26-07-2019, 1996.
- [DAV04] DAV, *Dav 2004r*, <https://aktuar.de/unsere-themen/lebensversicherung/Seiten/sterbetafeln.a> Accessed: 22-05-2019, 2004.
- [Fer96] T. S. A. Ferguson, *Course in Large Sample Theory*. 1996.
- [GK07] F. Girosi and G. King, “Understanding the lee-carter mortality forecasting method,” 2007.
- [Gre81] T. Greville, “Moving-weighted-average smoothing extended to the extremities of the data iii. stability and optimal properties,” 1981. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/0021904581900885>.
- [Kak85] P. Kakies, *Methodik von Sterblichkeitsuntersuchungen*. Verl. Versicherungswirtschaft, 1985.
- [LC92] R. D. Lee and L. R. Carter, “Modeling and forecasting u. s. mortality,” 1992.
- [Loe94] Loebus, *Bestimmung einer angemessenen sterbetafel für lebensversicherungen mit todesfallcharakter*, 1994.
- [Mor18] Mortality, *Mortality*, <https://www.mortality.org/>, Accessed: 22-05-2019, 2018.
- [NS12] A. S. Nocon and W. F. Scott, *An extension of the whittaker-henderson method of graduation*, 2012.