Universität Leipzig
Fakultät für Mathematik und Informatik
Mathematisches Institut

Diplomarbeit zum Thema:

Modellierung eines Risikoäquivalentes für isolierte Ereignisse und singuläre Ereignisketten in Kranken- und Lebensversicherungsbiographien - mit Anwendung

vorgelegt von:
Tobias, Riedel
geb. am: 25. 10. 1984
Studiengang Mathematik

Leipzig, im Februar 2014

Betreuer: HD Dr. Walter Warmuth
Betreuender Hochschullehrer: Prof. Dr. Manfred Riedel

Kurzzusammenfassung

PLATZHALTER

Danksagung

PLATZHALTER

PLATZHALTER2

Inhaltsverzeichnis

1. Einführung in das Thema

1.1. Mitnahme von Alterungsrückstellung in der PKV - Auswirkungen auf das Versichertenkollektiv

Im deutschen Gesundheitssystem entstehen jährlich Kosten in Höhe von 294 Mrd. Euro[1]. Um diese Summe abzudecken, existieren in Deutschland zwei Systeme: Die gesetzliche Krankenversicherung (GKV) und die private Krankenversicherung (PKV). Aufgrund der Krankenversicherungspflicht müssen alle Personen, mit Wohnsitz in Deutschland, in einem der beiden Systeme versichert sein. Für die GKV besteht deshalb eine Aufnahmepflicht und aus diesem Grunde ist der Großteil der Bevölkerung(69,86 Mio.[2]) in der GKV versichert. Personen die bestimmte Aufnahmekriterien erfüllen können auch in die PKV wechseln. Seit 2009 zahlt jeder gesetzlich Versicherte monatlich einen Teil seines Arbeitseinkommens (z.B. Gehalt bzw. Rente) in den Gesundheitsfonds ein. Andere Einkommensquellen, wie zum Beispiel Mieteinnahmen, werden dabei nicht berücksichtigt. Dieses Geld wird anschließend an die verschiedenen gesetzlichen Krankenkassen verteilt. Bei der Verteilung werden, neben dem Alter und dem Geschlecht der Versicherten, auch ausgewählte, besonders kostenintensive Krankheiten berücksichtigt. Bei diesem System wird der Beitragsatz jährlich, auf Basis des prognostizierten Behandlungsbedarfs, neu berechnet.

In Deutschland waren im Jahr 2011 fast neun Millionen Menschen (siehe Tabelle 1.1.) privat krankenversichert. Im Gegensatz zur GKV handelt es beim PKV-System um eine Individualversicherung. Beim Versicherungseintritt wird für jeden Versicherten eine individuelle Prämie berechnet, welche die Kosten des Versicherten bis zu seinem Tod abdecken sollen. Diese wird auf Basis des Erwartungswerts der künftigen Leistungsausgaben (unternehmensinterne oder bundesweite Statistiken) und der erwarteten Sterblichkeit bestimmt. Das erwarteten Kündigungsverhalten und der Rechnungszins haben außerdem

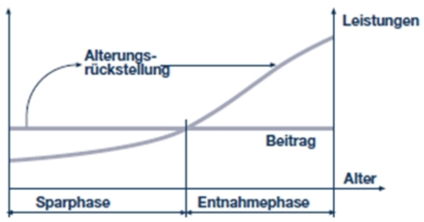
Versicherte mit Krankenvollversicherung	8,98 Mio.
Zusatzversicherungen	22,50 Mio.
Beitragseinnahmen	34,67 Mrd. Euro
Ausgezahlte Versicherungsleistungen	22,77 Mrd. Euro
Alterungsrückstellungen Bestand	169,43 Mrd. Euro

Tabelle 1.1.: Endgültige Werte für das Geschäftsjahr 2011, Stand: November 2012

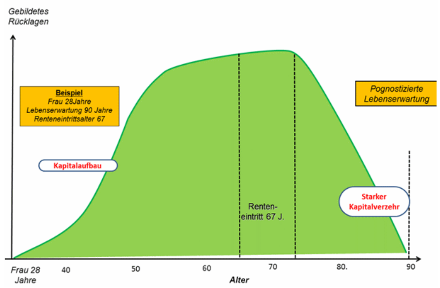
[1] siehe: Statistisches Bundesamt Zahlen in GKV für 2011 (www.destatis.de).
[2] siehe: http://bundesgesundheitsministerium.de Zahlen aus der KM1 für November 2012.

Einfluss auf die Höhe der Prämie. Versicherte mit dem gleichen Tarif[3] und aus derselben Altersgruppe bilden dabei ein Kollektiv, wobei individuelle Gesundheitsmerkmale durch Risikoaufschläge berücksichtigt werden.

Die Prämie ist so kalkuliert das sie, trotz der im Alter erwartungsgemäß steigenden Leistungskosten, über die gesamte Versicherungsdauer weitestgehend konstant bleibt. Deshalb liegen in den ersten Jahren die Prämien voraussichtlich über den in Anspruch genommen Leistungen. Aus dieser Differenz wird eine Rücklage (die Alterungsrückstellung)gebildet die dazu verwendet, die im Alter steigenden Kosten zu decken. Diese beiden Prozesse werden als Anspar- und Entnahmephase bezeichnet. Damit ein weitestgehend konstanter Beitrag sichergestellt werden kann, muss die Versicherungsprämie in einer Weise kalkuliert sein, die zum Einen die dauernde Erfüllbarkeit der vom Versicherer versprochenen Leistungen sicherstellt, und zum Anderen Prämiensteigerungen nur aus solchen Gründen zulässt, die vom Versicherer nicht zu beeinflussen sind. Dazu zählen beispielsweise nicht vorhersehbare Kostensteigerungen im Gesundheitswesen. Dabei ist zu berücksichtigen, dass gewisse Annahmen über die Entwicklung solcher Faktoren bereits bei der Kalkulation eines Tarif getroffen werden und damit eine Erhöhung nur dann zulässig ist, wenn die tatsächlichen Anstiege die prognostizierten Werte noch übertreffen.



(a) Leistungsverlauf



(b) Rücklagen

Abbildung 1.1.: Grafik Alterungsrückstellung

Seit dem 1. Januar 2009 sind die privaten Krankenversicherer gesetzlich dazu verpflichtet, im Fall des Krankenkassenwechsels eines Versicherten innerhalb der PKV, ihm die angesparte Alterungsrückstellung im Umfang des Basistarifs mitzugeben. Für den Versicherten hat das den Vorteil, dass er die Krankenkasse wechseln kann, ohne eine komplett neue Rückstellung aufbauen zu müssen. Da die mitgegebene Rückstellung allerdings nur den Basistarif umfasst, verliert er in der Regel einen Teil seiner Alterungsrückstellung.

Ein Wechsel hat nicht nur Auswirkungen auf ein einzelnes Individuum, sondern betrifft auch das Kollektiv von Versicherten, welches der Einzelne verlässt. In dem Zeitraum vom

[3]Der Eintrittszeitpunkt ist dabei entscheidend, da die einzelnen Tarife regelmäßig neu berechnet werden und dadurch verschiedene Tarifgenerationen entstehen. In einem Kollektiv werden nur Versicherte aus einer Tarifgeneration zusammengefasst.

Tarifabschluss bis zum Wechsel können sich die Risikomerkmale des Versicherten verändern. Er kann z.B. eine chronische Krankheit entwickeln, nach einem Unfall zum Pflegefall werden, oder im Gegenteil überhaupt keine nennenswerten Leistungen verursachen. Das Kollektiv, welches die Versicherten einer Tarifgeneration bilden, ist normalerweise dafür gedacht, diese Schwankungen auszugleichen. Dies erfolgt dadurch, indem die Gesünderen mit ihren Prämien die Kosten der Morbideren abdecken. Falls der Versicherte zum Zeitpunkt des Wechsels „besonders gesund“[4] war, entsteht dem Kollektiv damit ein Schaden, da die Überschüsse aus seiner Prämie nicht mehr dafür verwendet werden können, die Kosten morbiderer Versicherter abzudecken.

Die Höhe des Schadens ist dabei schwer zu ermitteln. Man benötigt ein Modell, das auf Basis des aktuellen Gesundheitszustands eines Versicherten eine Prognose erstellt, wie viele Leistungen er bis zu seinem Tod noch verursachen wird. Ein solches Modell wird im nächsten Kapitel vorgestellt.

1.2. Kohorten Modell

Die Gesundheitsforen Leipzig haben ein Modell zur Bestimmung der Höhe der noch zu erwartenden Leistungskosten eines Versicherten, basierend auf seinen Morbiditätsinformationen, entwickelt. Damit ist es möglich, für einen Versicherten zum Zeitpunkt des Kassenwechsels seine restliche Risikolast[5] zu berechnen und damit zu entscheiden, wie dieser das Kollektiv beeinflusst. Die nachfolgende Beschreibung des Modells stützt sich dabei in erster Linie auf den in der „Zeitschrift für Versicherungswesen“ (Heft 23/2011 und 24/2011) erschienenen Artikel „Auf Leben und Tod- Spezifische Implikationen eines vermeintlich längeren Lebens für die Versicherungswirtschaft“[6].

Das Ziel der Modellierung war es, möglichst wenige und möglichst gleichartige „Bausteine“ zu finden, aus denen sich die vielen Krankenversicherungsbiographien[7] zusammensetzen. Aus Millionen von Leistungsfällen, die über viele Jahre beobachtet wurden, konnten Milliarden von „individuellen“ Leistungspfaden analysiert werden. Dafür wurden Daten verwendet, die überwiegend aus dem Umfeld gesetzlich Versicherter stammen. Diese wurden durch Schätzungen, und vielfach durch einen bilanziellen Abgleich mit öffentlichen Gesamtdaten, vervollständigt.

Dabei ließ sich eine gewisse Homogenität, jeweils innerhalb von drei Lebensphasen, erkennen:

- Lebensabschnitt: „0-jährige“

[4] Das heißt der Versicherte beansprucht weniger Leistungen als der Durchschnitt.

[5] Die restliche Risikolast ist die Summe aller Krankheitskosten, die eine Person im Laufe des restlichen Lebens verursacht. Eine zeitunabhängige Vergleichbarkeit wird durch den Übergang zu Barwerten der zukünftigen (zufälligen) Leistungen erreicht.

[6] Quelle: Verweis auf Literaturverzeichnis.

[7] Gesundheitskosten in Zusammenhang mit den Leistungszeitpunkten im Verlauf des Lebens werden als Krankenversicherungsbiographie bezeichnet.

- Lebensabschnitt: „Phase regen Lebens“
- Lebensabschnitt: „Prämortalitätsphase“

Die Abgrenzung der Lebensabschnitte erfolgt dabei vom Rand der Krankenversicherungsbiographie her (Geburtsdatum, Sterbedatum) jeweils Tag genau. Die Längen dieser Abschnitte sind möglichst kurz gewählt und in ganzen Jahren angegeben. Kommt es zu Überlagerungen durch einen frühen Tod (jünger als 6 Jahre), so wird zuerst versucht den Abschnitt der „0-jährige“ vollständig abzubilden. Die verbleibenden Lebensjahre fallen anschließend in die„Prämortalitätsphase“. Bei dem Tod eines Neugeborenen gibt es demzufolge nur einen Lebensabschnitt der betrachtet werden kann. Die „Phase des regen Lebens“ ist aber in der Regel die längste Phase und wird zusätzlich von zwei Arten von Ereignissen überlagert:

- Ereignis: Isolierte Ereignisse
- Ereignis: Singuläre Ereignisketten

Die restliche Risikolast ergibt sich aus einer Mischung von Kohorten[8] der Restbiographien. Eine Restbiographie bezeichnet dabei den Anteil einer Versichertenbiographie, der noch nicht eingetreten ist, das heißt, noch in der Zukunft liegt. Jede Restbiographie setzt sich aus entsprechenden Anteilen aus der Phase der 0-jährigen, der Phase regen Lebens, aus isolierten Ereignissen, aus singulären Ereignisketten und der Prämortalitätsphase zusammen.

Die folgende Grafik veranschaulicht die Zusammensetzung einer Krankenversicherungsbiographie aus den drei Lebensabschnitten mit den überlagernden Ereignissen:

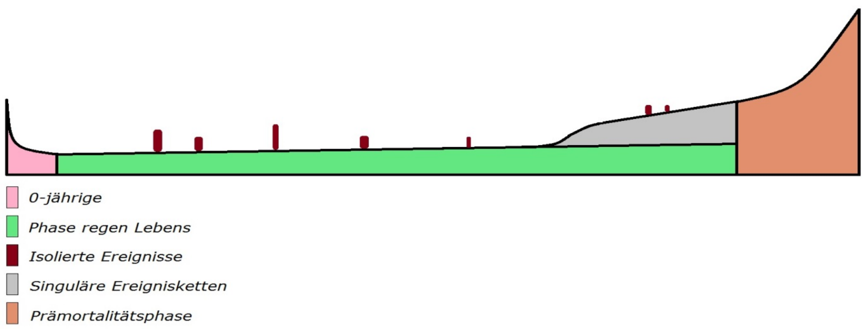


Abbildung 1.2.: Grafik Kohorten-Modell

Im Folgenden sollen die einzelnen Lebensabschnitte und Ergebnisse noch einmal im Detail beleuchtet und die Besonderheiten dargestellt werden.

[8]„Kohorten sind Jahrgänge oder Gruppen von Jahrgängen, die der Abgrenzung von Bevölkerungsgruppen dienen. Sie sind durch ein zeitlich gemeinsames, längerfristig prägendes Startereignis definiert.“ (Quelle: Wikipedia) In diesem Fall ist das Kohorten Merkmal eine gleiche restliche Lebenserwartung.

1.2.1. Lebensabschnitt: „0-jährige“

Dieser Lebensabschnitt beginnt mit dem „Tag der Geburt“ und dauert maximal ein Jahr. Für den Lebensabschnitt der 0-jährigen gilt: Mit jedem Tag des Überlebens steigt die Entfernung vom Tod, zunächst stark und später abgeschwächt. Eine umfangreiche Beschreibung dieser Modellierung findet sich in folgendem Artikel: ...

1.2.2. Lebensabschnitt: „Prämortalitätsphase“

Dieser Lebensabschnitt endet mit dem „Tag des Todes“. Der Abschnitt wird alters- und geschlechtsunabhängig auf maximal fünf Jahre virtuell „rückwärts“ ausgedehnt. Zu Beginn dieser Zeitspanne kann dadurch sehr gut an die Kosten der Vorphase angeknüpft werden. Die eigentlichen Kostenentwicklungen „kurz vor dem Tod“ sind über alle Altersbereiche in diesem Abschnitt abgebildet. Dieser Abschnitt ist auch das zentrale Thema des eingangs erwähnten Artikels.

1.2.3. Lebensabschnitt: „Phase regen Lebens“

Der Zeitraum vom ersten Tag des 2. Lebensjahres bis zum letzten Tag vor der Prämortalitätsphase entspricht dem dritten Lebenszeitabschnitt. Zur Kennzeichnung findet das Wort „rege“ Verwendung. Reges Leben findet in der Regel über eine Länge von vielen Jahren statt. In diesem Lebensabschnitt finden sich die Bereiche einer eher unauffälligen, „gleichbleibenden“ Kostenstruktur. Allerdings sind zwei Ereignisarten mit einem spezifischen Überlagerungscharakter aus dieser Zeitspanne abzugrenzen.

1.2.4. Ereignis: Isolierte Ereignisse

Seltene Ereignisse, welche die Gesundheit eines Versicherten beeinträchtigen, sich nicht ankündigen und bei denen der Eintrittszeitpunkt der eigentliche Auslöser von begrenzten Leistungsabfolgen sind, werden nachfolgend als isolierte Ereignisse bezeichnet. Solche Ereignisse (z. B. Unfälle) treten selten, und in der Regel unabhängig voneinander, auf. Isolierte Ereignisse können auch in der Phase der 0-jährigen und in der Prämortalitätsphase auftreten. Diese werden für die Modellierung der Alters- und Geschlechtsabhängigkeit isolierter Ereignisse zwar hinzugezogen, aber als jeweilige Phasen-Leistungen in der Krankenversicherungsbiographie kumuliert. Auf diese Weise „verschwinden“ keine Leistungen und das Phänomen der isolierten Ereignisse kann innerhalb der Phase regen Lebens separiert beschrieben werden.

1.2.5. Ereignis: Singuläre Ereignisketten

Eine nachhaltige, in der Regel bis zum Tod des Versicherten andauernde, beträchtliche Verschlechterung der Gesundheit wird nachfolgend als singuläre Ereigniskette bezeichnet. Es werden auch singuläre Ereignisketten berücksichtigt, die mit keinen Kosten für die Krankenversicherung verbunden sind (z. B. pflegebedürftig, da dann die Pflegeversicherung greift).

Das Ziel dieser Arbeit soll es sein, eine mögliche Ausgestaltung für den Teil der isolierten Ereignisse und der singulären Ereignisketten zu modellieren.

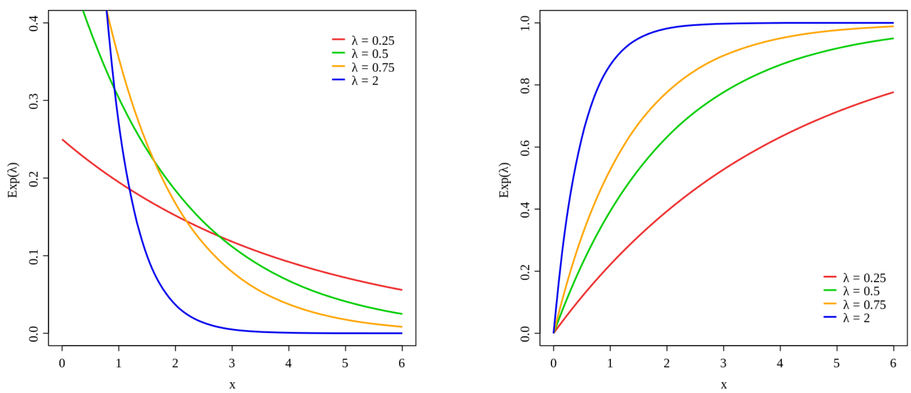
2. Theoretische Grundlagen für die Modellierung

In diesem Kapitel sollen die in der Arbeit verwendeten mathematischen Hilfsmittel vorgestellt werden und die wichtigsten Eigenschaften bewiesen werden. Dabei werden Grundlegende Kenntnisse im Bereich der Wahrscheinlichkeitstheorie vorausgesetzt und dass Begriffe wie Wahrscheinlichkeitsraum, Zufallsvariable, Dichte- und Verteilungsfunktion bekannt sind. Im Folgenden ist der Wahrscheinlichkeitsraum immer definiert durch $(\Omega, \mathfrak{A}, \mathbb{P})$.

Zunächst werden wir die Verteilungen vorstellen, die in dieser Arbeit verwendet werden, und die wichtigsten Eigenschaften ableiten.

Definition 2.1. *Eine Zufallsvariable $X : \Omega \to \mathbb{R}$ heißt* ***exponentialverteilt zum Parameter*** *λ (kurz: $\sim exp(\lambda)$) wenn sie die folgende Dichtefunktion besitzt:*

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{für } x \geq 0 \\ 0 & \text{für } x < 0 \end{cases}$$



(a) Dichtefunktion (b) Verteilungsfunktion

Abbildung 2.1.: Dichte- und Verteilungsfunktion der Exponentialverteilung

Die Exponentialverteilung hat folgende Eigenschaften:

(i) Die Verteilungsfunktion der Exponentialverteilung ist:

$$F_\lambda(x) = \int_{-\infty}^{x} f_\lambda(t)dt = \begin{cases} 1 - e^{-\lambda x} & \text{für } x \geq 0 \\ 0 & \text{für } x < 0 \end{cases}$$

(ii) Der Erwartungswert ist:

$$\mathbb{E}(X) = \int_0^\infty \lambda x e^{-\lambda x} dx = \left[-\frac{e^{-\lambda x}(\lambda x + 1)}{\lambda} \right]_0^\infty = \frac{1}{\lambda}$$

(iii) Die Varianz ist:

$$Var(X) = \int_\infty^0 \left(x - \frac{1}{\lambda} \right)^2 \lambda e^{-\lambda x} dx = \frac{1}{\lambda^2}$$

(iv) Die Exponentialverteilung ist gedächtnislos (auch Nichtalterungseigenschaft genannt), d.h.:

$$\begin{aligned} F(x|t) := \mathbb{P}(x + t \leq X, t < X) &= \frac{\mathbb{P}(t < X < x + t)}{\mathbb{P}(t < X)} \\ &= \frac{\mathbb{P}(X \leq x + t) - \mathbb{P}(t \leq X)}{1 - \mathbb{P}(t \leq X)} \\ &= \frac{1 - e^{-\lambda(x+t)} - (1 - e^{-\lambda t})}{1 - (1 - e^{-\lambda t})} \\ &= \frac{-e^{-\lambda x - \lambda t} + e^{-\lambda t}}{e^{-\lambda t}} \\ &= \frac{e^{-\lambda t}(1 - e^{-\lambda x})}{e^{-\lambda t}} \\ &= 1 - e^{-\lambda x} = \mathbb{P}(x \leq X) = F(x) \end{aligned}$$

Es lässt sich sogar zeigen, dass die Exponentialverteilung die einzige stetige Verteilung[1] mit Nichtalterungseigenschaft ist:

Lemma 2.2. *Wenn die Überlebenswahrscheinlichkeit ($\overline{F}(x) := 1 - F(x)$) einer nicht ausgearteten, nicht negativen Zufallsvariable X der folgenden Beziehung genügt:*

$$\mathbb{P}(x + t \leq X) = \mathbb{P}(x \leq X)\mathbb{P}(t \leq X) \tag{2.1}$$

dann gilt $X \sim exp(\lambda)$ für ein $\lambda > 0$.

Beweis: Mit der Definition $f(x) := \mathbb{P}(x < X)$ lässt sich 2.1 zu folgender Gleichung umformen

$$f(x + t) = f(x)f(t), \text{mit} f(0) = \mathbb{P}(0 \leq X) = 1$$

[1] Im diskreten Fall ist dies die geometrische Verteilung.

die sich wie folgt umformen lässt:

$$ln(f(x+t)) = ln(f(x)) + ln(f(t))$$

Unter der Voraussetzung, das die Dichte $p_X(t) = -f(t+0), t \geq 0$ existiert, ergibt sich durch Differentiation nach t:

$$\frac{f^{'}(x+t)}{f(x+t)} = \frac{f^{'}(t)}{f(t)}$$

Für $t = 0$ folgt damit:

$$\frac{f^{'}(x)}{f(x)} = f^{'}(t)$$

Da $f(0) = 1$ der maximale Wert der Funktion f(x) ist, gilt $0 \geq f^{'}(0) := -\lambda$ und damit erhalten wir die bekannte Differentialgleichung:

$$\begin{aligned}
&f^{'}(x) = -\lambda f(x) \\
\Rightarrow\ & f(x) = e^{-\lambda x} \\
\Rightarrow & F(x) = 1 - f(x) = 1 - e^{-\lambda x} \\
\Rightarrow\quad & X \sim exp(\lambda)
\end{aligned}$$

□

Definition 2.3. *Eine diskrete Zufallsvariable $X : \Omega \to \mathbb{N}$ heißt **Poisson-verteilt zum Parameter** $\lambda \in \mathbb{R}_{>0}$ (kurz $X \sim Poi(\lambda)$), wenn gilt:*

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, k = 0, 1, 2, ...$$

Die Poisson-Verteilung hat folgende Eigenschaften:

(i) Die Verteilungsfunktion der Poisson-Verteilung ist:

$$F_\lambda(n) = \sum_{k=0}^{n} \mathbb{P}_\lambda(k) = e^{-\lambda} \sum_{k=0}^{n} \frac{\lambda^k}{k!}$$

(ii) Der Erwartungswert ist:

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = 0 + \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} \\
&= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \\
&= \lambda e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = \lambda e^{-\lambda} e^{\lambda} = \lambda
\end{aligned}$$

Der Parameter λ der Poisson-Verteilung kann also, als die erwartete Ereignishäufigkeit pro Zeiteinheit interpretiert werden.

(iii) Die Varianz ist:

$$\begin{aligned}
\mathbb{E}(X^2) &= \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} \\
&= e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^k}{(k-1)!} \\
&= e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^k}{(k-2)!} + e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} \\
&= \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} + \lambda e^{-\lambda} \sum_{1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \\
&= \lambda^2 + \lambda \\
\\
Var(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda
\end{aligned}$$

(iv) Seien X_1 und X_2 unabhängige poissonverteilte Zufallsvariablen mit $X_1 \sim Poi(\lambda_1)$ und $X_2 \sim Poi(\lambda_2)$, dann gilt für $X := X_1 + X_2$:

$$\begin{aligned}
\mathbb{P}(X = x) &= \sum_{k=0}^{x} \mathbb{P}(X_1 = k)\mathbb{P}(X_2 = x - k) \\
&= e^{-\lambda_1} e^{-\lambda_2} \sum_{k=0}^{x} \frac{\lambda_1^k}{k!} \frac{\lambda_2^{x-k}}{(x-k)!} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{x!} \sum_{k=0}^{x} \frac{x!}{k!(x-k)!} \lambda_1^k \lambda_2^{x-k} \\
&= e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1 + \lambda_2)^x}{x!} \\
\Rightarrow X &\sim Poi(\lambda_1 + \lambda_2)
\end{aligned}$$

Das heißt die Summe von poissonverteilten Zufallsvariablen ist wieder poissonverteilt.

Die Poisson-Verteilung hat außerdem eine besondere Bedeutung, da sie unter den richtigen Voraussetzungen die Grenzverteilung der Binomialverteilung ist. Dieser Zusammenhang wird in folgendem Lemma verdeutlicht:

Lemma 2.4. *Die Poisson-Verteilung ist die Grenzverteilung der Binomial-Verteilung $B_{n,p}(k)$, wenn für die Erfolgswahrscheinlichkeit p gilt $p \to 0$ und für die Anzahl der Versuche n gilt $n \to \infty$ und außerdem für das Produkt $\lambda := np$ gilt $\lambda \neq 0$ und $\lambda \neq \infty$.*

Beweis: Entsprechend der Nebenbedingung können wir p durch $\frac{\lambda}{n}$ ersetzen und damit ist die Bedingung $p \to 0$ äquivalent zu $n \to \infty$ und wir müssen somit nur einen Grenzwert bestimmen. Wir zeigen, dass der Grenzwert $n \to \infty$ einer binomialverteilten Zufallsvariable an der Stelle k, gegen den Wert einer poissonverteilten Zufallsvariablen an der Stelle k geht.

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}(X=k) &= \lim_{n\to\infty} \frac{n!}{k!(n-k)!}\left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} \\
&= \lim_{n\to\infty} \frac{\lambda^k}{k!} \frac{n(n-1)(n-2)...(n-k+1)}{n^k} \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-k} \\
&= \frac{\lambda^k}{k!} * 1 * e^{-\lambda} * 1 \\
&= \frac{\lambda^k e^{-\lambda}}{k!}
\end{aligned}$$

□

Definition 2.5. *Ein* ***Stochastischer Prozess*** *X ist eine Familie von Zufallsvariablen $\{X_t\}$ mit $X_t : \Omega \to \mathfrak{A}, t \in T$. Für die Indexmenge T gilt in der Regel $T \in \{\mathbb{R}_{\geq 0}, \mathbb{N}_0\}$. Das heißt X ist eine Abbildung*

$$X : \Omega \times T \to Z,\ (\omega, t) \mapsto X_t(\omega)$$

sodass $X_t : \omega \mapsto X_t(\omega)$ für alle $t \in T$ eine messbare Abbildung ist.

(i) Ein stochastische Prozess heißt **zeitdiskret** wenn T abzählbar ist, z.B. $T = \mathbb{N}_0$. Ansonsten heißt er **zeitstetig**. Analog heißt ein Prozess mit diskreten Zustandsraum $\mathfrak{A}$ **wertdisktret** oder auch **Punktprozess**.

(ii) Ein stochastische Prozess heißt **stationär**, wenn für alle $s > 0$, sowie $t_1, t_2, ..., t_k \in T$ und $x_1, x_2, ..., x_k \in \mathfrak{A}$ gilt:

$$\mathbb{P}(X_{t_1+s} = x_1, X_{t_2+s} = x_2, ..., X_{t_k+s} = x_k) = \mathbb{P}(X_{t_1} = x_1, X_{t_2} = x_2, ..., X_{t_k} = x_k)$$

Das heißt, das zufällige Verhalten des Prozesses hängt nicht vom Zeitpunkt der Beobachtung ab.

(iii) Ein stochastischer Prozess besitzt **Unabhängige Zuwächse**, wenn die Zufallsvariablen $X_{t_0}, X_{t_1} - X_{t_0}, ..., , X_{t_n} - X_{t_{n-1}}$ für alle n=1,2,... und $0 \leq t_0 < t_1 < ... < t_n$ unabhängig sind.

Definition 2.6. *Sei $T_1, T_2, ... : \omega \to [0, \infty)$ eine Folge von unabhängig und identisch verteilten Zufallsvariablen, dann ist $N := \{N_t, t \geq 0\}$ mit*

$$N_t = \sum_{k=1}^{\infty} \mathbb{I}(S_k \leq t) \text{ und } S_n = T_1 + ... + T_n$$

ein stochastischer Prozess und wird als Zählprozess bezeichnet.

Prozesse dieser Art werden z.B. in der Zuverlässigkeitstheorie eingesetzt um die Ausfälle einer Komponente in einem bestimmten Zeitraum zu zählen. Deshalb werden die T_n häufig als **Zwischenankunftszeiten** bezeichnet.

Definition 2.7. *Ein Zählprozess $\{N_t, t \geq 0\}$ mit exponentialverteilten Zwischenankunftszeiten $T_n \sim exp(\lambda)$ heißt* ***homogener Poisson-Prozess mit der Intensität*** *λ.*

In folgendem Theorem werden die wichtigsten Eigenschaften des Poisson deutlich:

Theorem 2.8. *Die folgenden Aussagen sind äquivalent*[2]*:*

(i) $\{N_t, t \geq 0\}$ *ist ein Poisson-Prozess mit der Intensität* λ

(ii) Die Zufallsvariablen N_t sind poissonverteilt zum Parameter λt für alle $t \geq 0$. Unter der Bedingung $\{N_t = n\}$, hat für beliebige $n = 1, 2, ...$ der Zufallsvektor $(S_1, S_2, ..., S_n)$, die gleiche Verteilung wie die Ordnungsstatistik von n unabhängigen, in $[0, t]$ gleichverteilten Zufallsvariablen.

(iii) Der stochastische Prozess $\{N_t\}$ hat unabhängige Zuwächse und es gilt $\mathbb{E}(N_1) = \lambda$. Unter der Bedingung $\{N_t = n\}$, hat für beliebige $n = 1, 2, ...$ der Zufallsvektor $(S_1, S_2, ..., S_n)$, die gleiche Verteilung wie die Ordnungsstatistik von n unabhängigen, in $[0, t]$ gleichverteilten Zufallsvariablen.

(iv) Der stochastische Prozess $\{N_t\}$ hat unabhängige Zuwächse und ist stationär und es gilt für $h \downarrow 0$:

$$\begin{aligned} \mathbb{P}(N_h = 0) &= 1 - \lambda h + o(h), \text{ und} \\ \mathbb{P}(N_h = 1) &= \lambda h + o(h) \end{aligned}$$

(v) Der stochastische Prozess $\{N_t\}$ hat unabhängige Zuwächse und ist stationär. Außerdem gilt für jedes $t \geq 0$ das $N_t \sim Poi(\lambda t)$.

Beweis:

- $(i) \Rightarrow (ii)$: Aus (i) folgt, dass $S_n = \sum_{i=1}^n T_i$ eine Summe von n unabhängigen und zum Parameter λ exponentialverteilten Zufallsvariablen ist, d.h., $S_n \sim Erl(n, \lambda)$, wobei $Erl(n, \lambda)$ die Erlang-Verteilung mit den Parametern n und λ bezeichnet. Hieraus folgt $\mathbb{P}(N_t = 0) = \mathbb{P}(S_1 > t) = e^{-\lambda t}$ und damit folgt:

$$\begin{aligned} \mathbb{P}(N_t = n) &= \mathbb{P}(N_t \geq n) - \mathbb{P}(N_t \geq n+1) \\ &= \mathbb{P}(S_n \leq t) - \mathbb{P}(S_{n+1} \leq t) \\ &= \int_0^t \frac{\lambda^n v^{n-1}}{(n-1)!} e^{-\lambda v} \mathrm{d}v - \int_0^t \frac{\lambda^{n+1} v^n}{n!} e^{-\lambda v} \mathrm{d}v \\ &= \int_0^t \frac{\mathrm{d}}{\mathrm{d}v} \left(\frac{(\lambda v)^n}{n!} e^{-\lambda v} \right) \mathrm{d}v \\ &= \frac{(\lambda t)^n}{n!} e^{-\lambda t} \end{aligned}$$

[2]siehe: `http://www.mathematik.uni-ulm.de/stochastik/lehre/ss05/wt/skript/node15.html`

Dies gilt für jedes $n \geq 1$, und damit folgt, dass $N_t \sim Poi(\lambda t)$, womit der erste.

- $(ii) \Rightarrow (iii)$: TODO

$\square$

Die folgende sehr nützliche Eigenschaft ist bereits von der Poisson-Verteilung bekannt:

Lemma 2.9. *Die Überlagerung von zwei unabhängigen Poisson-Prozessen $\{N_t^1, t \geq 0\}$ und $\{N_t^1, t \geq 0\}$ mit der Intensität λ_1 bzw. λ_2 ist wieder ein Poisson-Prozess mit Intensität $\lambda = \lambda_1 + \lambda_2$.*

Beweis: Im vorangegangen Theorem hab wir gezeigt, dass für einen Poisson-Prozess $\{N_t, t \geq 0\}$ gilt $N_t \sim Poi(\lambda t)$. Da $\{N_t^1, t \geq 0\}$ und $\{N_t^1, t \geq 0\}$ unabhängig sind gilt für die Summe $N_t^1 + N_t^2 \sim Poi((\lambda_1 + \lambda_2)t)$. Damit ist die Überlagerung der beiden Poisson-Prozesse wieder ein Poisson-Prozess mit Intensität $\lambda = \lambda_1 + \lambda_2$. $\square$

Als nächstes werden wir eine weitere wichtige Klasse an stochastischen Prozessen vorstellen, doch vorher müssen wir noch den Begriff der stochastischen Matrix einführen:

Definition 2.10. *Eine Matrix $P = (p_{i,j})$ heißt* ***stochastisch****, falls für alle $i, j \in I$ (Indexmenge) gilt $p_{i,j} \in [0, 1]$ und $\sum_{j \in I} p_{i,j} = 1$.*

Definition 2.11. *Ein Stochastischer Prozess $\{X_t, t \geq 0\}$ besitzt die* ***Markoveigenschaft****, wenn für alle $i_0, i_1, ..., i_{n+1} \in \mathfrak{A}$ gilt:*

$$\mathbb{P}(X_{t+1} = i_{t+1} | X_t = i_t, X_{t-1} = i_{t-1}, ..., X_0 = i_0) = \mathbb{P}(X_{t+1} = i_{t+1} | X_t = i_t)$$

Ein solcher Prozess heißt im stetigen Fall ***Markovscher Prozess*** *und* ***Markov-Kette*** *im diskreten Fall. Die* ***Startverteilung*** *der Markov-Kette ist definiert durch $v(i) = \mathbb{P}(X_0 = i)$ und die Wahrscheinlichkeiten $\mathbb{P}(X_{t+1} = i_{t+1} | X_t = i_t) =: p_{i_t, i_{t+1}}$ werden als* ***Übergangswahrscheinlichkeiten*** *bezeichnet. Die Matrix $P = (p_{i,j})_{i,j \in \mathfrak{A}}$ die sich aus den Übergangswahrscheinlichkeiten ergibt, ist eine stochastische Matrix und heißt* ***Übergangsmatrix****.*

Der nächste Zustand einer Markov-Kette hängt also immer nur von dem aktuellen Zustand ab. D.h. die Kette wird durch die Übergangswahrscheinlichkeiten charakterisiert.

TODO:

- Eigenschaften der Markov-Kette charakterisieren
- Pfade durch Potenzen der Übergangsmatrix
- Irreduzibilität
- Rekurrenz und Transienz
- Absorbierende Zustände
- Stationäre Verteilung
- Ergodizität

3. Modellierung Isolierte Ereignisse

In diesem Kapitel soll ein Risikomodell für die Isolierten Ereignisse in der Phase regen Lebens (im Folgenden alternativ auch als Unfälle bezeichnet) erstellt werden. Die Definition der Isolierten Ereignisse wird deshalb genauer analysiert, um die Anforderungen an das spätere Risikomodell abzuleiten. Es werden zwei Ansätze vorgestellt, die diese Anforderungen erfüllen. Zum Test werden mithilfe dieser Ansätze konkrete Modelle zu einem Anwendungsfall erstellt und ausgewertet.

3.1. Grundlegende Betrachtungen

Im Kohorten-Modell werden die Unfälle als **seltene Ereignisse**, welche die Gesundheit eines Versicherten beeinträchtigen, **sich nicht ankündigen** und bei denen der Eintrittszeitpunkt der eigentliche **Auslöser von begrenzten Leistungsabfolgen** ist, beschrieben. Außerdem sollen sie selten, und in der Regel **unabhängig voneinander**, auftreten. Das heißt die Eintrittswahrscheinlichkeit für einen Unfall hängt nicht davon ab, ob der Versicherte bereits einen Unfall hatte. Es gibt also insbesondere **keinen Lerneffekt** und **keine Folgeschäden** die im Modell berücksichtigt werden müssen. Außerdem ist die durch den Unfall hervorgerufene Leistungsfolge begrenzt was bedeutet, dass **keine Chronizität** durch einen Unfall entstehen kann. Dadurch sind die isolierten Ereignisse klar zu den singulären Ereignisketten abgegrenzt, die im nächsten Kapitel behandelt werden.

Für das Kohorten-Modell müssen die isolierten Ereignisse in der Phase „regen Lebens“ beschrieben werden. Deshalb werden für die Modellierung zwar alle Unfälle hinzugezogen, aber die Kosten von Unfällen aus den anderen beiden Phasen werden als jeweilige Phasen-Leistungen in der Krankenversicherungsbiographie kumuliert. Das heißt insbesondere, dass Unfälle mit Todesfolge keine besondere Betrachtung mehr benötigen, da sie bereits in der dritten Phase des Modells erfasst wurden.

Die Erfahrung zeigt, dass die Unfallwahrscheinlichkeit in der Regel sowohl vom Alter- als auch vom Geschlecht des Versicherten abhängt. Einen guten Beleg dafür liefert die Statistik über die „Unfallentwicklung auf deutschen Straßen 2012“ vom Statistischen Bundesamt[1]. Daraus geht hervor das besonders junge Erwachsene im Alter zwischen 18 und 24 Jahren in schwere Verkehrsunfälle mit Personenschäden verwickelt sind. Hinzukommt

[1] Quelle: Statistisches Bundesamt unter `https://www.destatis.de/DE/Publikationen/Thematisch/TransportVerkehr/Verkehrsunfaelle/PK_Unfallentwicklung.html`

dass in 86% aller Motorradunfälle in diesem Altersbereich der Verunglückte männlich war. Neben Alter und Geschlecht hat auch die Saison in bestimmten Fällen Einfluss auf die Unfallwahrscheinlichkeit. Dies spiegelt sich zum Beispiel bei Verkehrsunfällen wieder, denn im Zeitraum von Mai bis September gab es die meisten Unfälle mit Personenschäden, da es aufgrund der höheren Geschwindigkeiten bei schönen Wetter zu schwereren Verkehrsunfällen kam. Dabei ist zu beachten, dass die Einflüsse von Alter, Geschlecht und Saison auf die Unfallwahrscheinlichkeit auch vom Typ des Unfalls abhängen.

Zusammenfassend lässt sich sagen, dass die Unfallwahrscheinlichkeit unabhängig von bereits eingetretenen Unfällen ist, aber je nach Unfalltyp, vom Alter und Geschlecht des Versicherten abhängen und einer Saisonalität unterliegen kann. Für die Modellierung werden wir die Saisonalität allerdings nicht besonders betrachten, da aufgrund des in der Regel sehr langen Beobachtungszeitraums, in der Phase regen Lebens, diese Effekte nur einen sehr geringen Einfluss auf das Modell haben werden. Auf Basis dieser Erkenntnisse werden nun zwei Modelle vorgestellt:

3.2. Modell - Poisson-Prozess

Für den ersten Modellansatz nehmen wir an, dass aufgrund des Charakters der Unfälle als seltene Ereignisse, die Genesungsdauer vernachlässigt werden kann. Das heißt die aus dem Unfall resultierenden Folgeleistung werden auf den Zeitpunkt des Unfalls konzentriert. Außerdem wird die Unfallwahrscheinlichkeit, selbst direkt nach einem Unfall, nicht verändert. Laut Voraussetzung hängt die Unfallwahrscheinlichkeit vom Alter und Geschlecht der Versicherten ab. Doch zunächst werden wir ein Modell entwickeln für den Fall das die Unfallwahrscheinlichkeit konstant ist. *Diese ist intuitiv definiert als die Anzahl der Unfälle pro Zeiteinheit.* Anschließend wird ein erweitertes Modell um die verschiedenen Abhängigkeiten zu berücksichtigen.

Wir führen folgende Bezeichnungen ein:

(i) Die Unfallwahrscheinlichkeit wird mit λ bezeichnet.

(ii) Die Unfallzeitpunkte seien gegeben durch $t_0, t_1, t_2, t_3, ...$, wobei $t_0 = 0$ den Beginn der Beobachtung kennzeichnet.

(iii) Die Zeit zwischen den einzelnen Unfällen sei definiert als $\{T_k\}_{k\geq 0}$ mit $T_k := t_k - t_{k-1}$ und $T_0 := 0$.

(iv) Die Anzahl der Unfälle wird bezeichnet durch $N_t := max\{k : t_k \leq t\}$

Sei nun die Wahrscheinlichkeit für einen Unfall λ konstant. Dann gilt für die Wartezeiten $\{T_k\}_{k\geq 0}$, dass die Zeit bis zum nächsten Unfall unabhängig davon ist, wie viel Zeit bereits ohne Unfall verstrichen ist. Sei $F(x)$ die Verteilung der Wartezeiten, dann gilt also:

$$
\begin{aligned}
F(x|t) &= F(x) \\
=> F(x+t) &= F(x)F(t)
\end{aligned}
$$

Damit folgt gemäß Lemma 2.2 das $F(x) \sim Exp(\lambda)$. Damit ist $N_\lambda = \{N_t\}_{t>0}$ gemäß Definition 2.7 ein Poisson-Prozess mit Parameter λ. Wie bereits am Anfang erwähnt, ist die Unfallwahrscheinlichkeit von verschiedenen Faktoren abhängen und dabei in erster Linie natürlich vom Unfalltyp. Deshalb zerlegen wir den Prozess in mehrere Teilprozesse:

Seien $U_1, U_2, ..., U_n$ mit $n \in \mathbb{N}$ die verschiedenen relevanten Unfalltypen und $\lambda_{U_1}, \lambda_{U_2}, ..., \lambda_{U_n}$ die zugehörigen Unfallwahrscheinlichkeiten. Dann gilt für die daraus resultierenden Prozesse $N_\lambda = \sum_{i=1}^{n} N_{\lambda_{U_i}}$. Das heißt der Prozess lässt sich beliebig in Unterprozesse zerlegen die jeweils einen Typ bzw. auch eine Familie von Typen beschreiben. Dadurch lässt sich das Modell beliebig verfeinern, falls dies in einem Bereich notwendig ist.

Da Unfälle vom Alter und Geschlecht abhängen können, reicht es nicht aus einen Homogenen Poisson-Prozess zu verwenden. Deshalb ist es notwendig das unfallspezifische λ als Funktion dieser drei Einflüsse zu berücksichtigen. Die Stärke der jeweiligen Einflussfaktoren hängt natürlich in erster Linie vom Unfalltyp ab. Als praktikable Lösung zur Bestimmung der einzelnen Funktionen bietet es sich deshalb an, λ als Treppenfunktion zu modellieren, bei der die einzelnen Stufen von der empirischen Verteilung der Unfallwahrscheinlichkeit abhängen. Das detaillierte Vorgehen bei der Bestimmung von λ wird im weiteren Verlauf dieses Kapitels anhand eines Beispiels erläutert.

Ein analoges Vorgehen (empirische Verteilung) bietet sich auch bei der Modellierung der Kostenverteilung für die Unfälle an, insbesondere wenn diese nur schlecht durch eine Verteilungsfunktion beschrieben werden können. Bei Leistungskosten in der Krankenversicherung ist zu beobachten, dass es viele Fälle mit niedrigen Kosten und einige mit sehr hohen Kosten (Hochkostenfällen), aber nur einen sehr geringen Teil mit durchschnittlichen Kosten. Aus diesem Grund ist es wichtig eine Verteilungsfunktion zu wählen, die zu Einen die Standardfälle abdeckt und zum Anderen noch genug Masse in den hohen Kostenbereichen hat, um die Hochkostenfälle nicht zu vernachlässigen. Verteilungsfunktionen mit dieser Eigenschaft werden auch als Heavy-Tail-Verteilung[2] bezeichnet. Das optimale Vorgehen bei der Modellierung der Kostenverteilung hängt letztendlich vom Unfalltyp und den zur Verfügung stehenden Daten ab.
- Zusammenfassung

3.3. Modell - Markov-Kette

- Grundannahme: kein Lerneffekt und Gleichverteilte Eintritts- bzw. Übergangswahrscheinlichkeiten

[2]z.B. Weibull-Verteilung mit Formparameter<1 oder Log-Normalverteilung

- Ergodisch (nur bei Unfällen) => Startverteilung spielt auf längere Sicht keine Rolle
- Übergangswahrscheinlichkeiten werden aus den Daten geschätzt (Mittelwert) => exponentlialverteilte Genesungszeiten plausibel?
- Ansatz der den Prozess eines Unfalls mit den verschiedenen Phasen der Genesung gut abbildet
- Zur Bestimmung der zu Erwartenden Anzahlen evtl. Simulation? Monte-Carlo-Methode
- Verweildauer und Kosten evtl. aus DRG Katalog?

3.4. Anwendung und Test eines Modells

Im folgenden Abschnitt wird zunächst die Datengrundlage vorgestellt, auf der das Modell später getestet werden soll. Anschließend wird die Vorgehensweise zur Identifizierung der Unfälle beschrieben. Aus diesen werden dann die Parameter des Modells geschätzt und die Modellgüte bewertet.

3.4.1. Datengrundlage + Aufbereitung

Die Datengrundlage für die Anwendung des Modells besteht aus Versicherten- und Krankenhausabrechnungsdaten, die über einen Zeitraum von 4 Jahren erfasst wurden.Für die Aufbereitung der Daten wurden Duplikate entfernt, die Datenformate vereinheitlicht und fehlerhafte bzw. unvollständige Datensätze herausgefiltert. Bevor im Folgenden die Datengrundlage näher betrachtet wird, soll erst einmal ein kurzer Einblick in das deutsche Abrechnungssystem im Krankenhaus gegeben werden um die vorhandenen Information zu erläutern.

Seit 2003 wird in Deutschland die Vergütung im Kranhausbereich über ein Fallpauschalensystem realisiert. Dieses basiert auf den sogenannten Diagnosis Related Groups (DRG) oder auf deutsch diagnosebezogene Fallgruppen. Behandlungsfälle, die medizinisch und hinsichtlich des Ressourcenverbrauchs ähnlich sind, sind dort zu Fallgruppen zusammengefasst. Die Zuordnung erfolgt auf Basis der Patienten- und Falldaten die während eines Krankenhausaufenthaltes gesammelt werden. Dazu gehören neben dem Alter und Geschlecht des Patienten auch die gestellten Diagnosen und die vorgenommen Prozeduren. Komplikationen oder erschwerende Begleiterkrankungen (Komorbiditäten) werden über verschiedene Schweregrade berücksichtigt. Für jede DRG gibt es eine Fallpauschale die in Verbindung mit der Verweildauer die Kosten für den Behandlungsfall bestimmt. Das System wird vom „Institut für das Entgeltsystem im Krankenhaus“[3] (kurz InEK) gepflegt und weiterentwickelt.

[3] `www.g-drg.de`

Damit die DRG sauber bestimmt werden kann, müssen die Diagnosen und die Prozeduren nach den medizinischen Klassifikationen ICD-10-GM und OPS kodiert werden. Das ICD-10-GM ist eine für Deutschland angepasste Version der „internationalen statistischen Klassifikation der Krankheiten und verwandter Gesundheitsprobleme“(kurz ICD[4]). Der zugehörige Katalog wird vom „deutschen Institut für Medizinische Dokumentation und Information“[5],kurz DIMDI, jährlich aktualisiert und angepasst. Er ist hierarchisch strukturiert und enthält 22 Krankheitskapitel, die sich in Gruppen, Kategorien und Subkategorien aufsplitten. Das DIMDI schätzt die Anzahl der Schlüsselnummern auf ca. 13.400. Die jährlichen Anpassungen stellen auch eine besondere Herausforderung bei der Arbeit mit ICD-Schlüsseln dar. Es kann vorkommen das in zwei unterschiedlichen Jahren einem ICD-Schlüssel eine völlig neue Diagnose zu geordnet wird. Der Operationen- und Prozedurenschlüssel (OPS) wird ebenfalls vom DIMDI gepflegt und ist die amtliche Klassifikation zum Verschlüsseln von Operationen, Prozeduren und allgemein medizinischen Maßnahmen im stationären Bereich und beim ambulanten Operieren.

Sowohl die DRG- als auch ICD-Informationen finden sich in den Grunddaten wieder und sind für jeden einzelnen Krankenhausfall aufgeschlüsselt. Es ist zu beachten, dass alle während eines Kranhausfalles gestellten Diagnosen erfasst werden. Das heißt es können auch Diagnosen vorkommen die mit der DRG nicht in Zusammenhang stehen[6]. Die Datenstruktur der aufbereiteten Daten wird in folgendem Diagramm veranschaulicht:

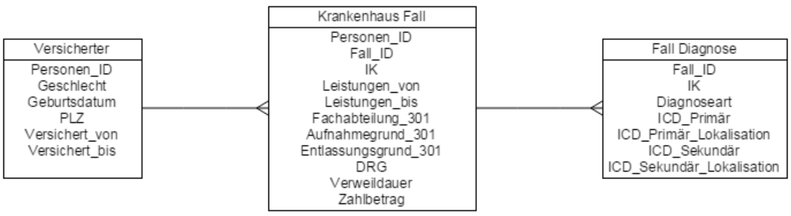


Abbildung 3.1.: ER-Diagramm Datengrundlage

(i) **Versicherter:** In der ersten Tabelle sind Information zu den Versicherten hinterlegt. Neben Geschlecht und Geburtsdatum, gibt es auch Information über den Wohnort(Postleitzahl) und den Zeitraum in dem er versichert war. Letzteres ist wichtig, da logischerweise nur während dieses Zeitraums Leistungsdaten über den entsprechenden Versicherten vorliegen. Als Primärschlüssel wird in dieser Tabelle die Personen_ID verwendet, wodurch dann auch eine Zuordnung zu den Krankenhaus Falldaten möglich ist.

[4]Aufgrund des englischen Namens: „International Statistical Classification of Diseases and Related Health Problems“

[5]`www.dimdi.de`

[6]z.B. ein Herzinfarkt während einer Operation

(ii) **Krankenhaus_Fall:** Ein Krankenhausfall umfasst alle Maßnahmen die von der Einweisung bis zu Entlassung eines Versicherten fällig werden. In dieser Tabelle kommt ein zusammengesetzter Primärschlüssel aus Fall_ID und Institutionskennzeichen des Krankenhauses (IK) zum Einsatz [7]. Für jeden Behandlungsfalls gibt es Daten zum Zeitpunkt und Grund der Aufnahme und Analog für die Entlassung. Außerdem wird die abgerechnete DRG und die damit verbundenen Kosten angegeben.

(iii) **Fall_Diagnose:** Zu jedem Fall sind außerdem die dazugehörigen Diagnosen hinterlegt. Diese können wieder über die Kombination aus Institutionskennzeichen und Fall_ID zugeordnet werden. Die Diagnosen werden nach dem oben beschrieben ICD-Katalog codiert und in Haupt- und Nebendiagnosen unterteilt, durch das Feld Diagnoseart. Jeder Datensatz enthält eine Primäre ICD zu der falls notwendig auch die Lokalisation angegeben wird. Das ist z.B. notwendig um zu spezifizieren, ob der linke oder der rechte Arm gebrochen ist. In einigen wenigen Fällen gibt es auch noch eine Sekundäre ICD.

3.4.2. Identifizierung von Unfällen

Eine Herausforderung dieser Arbeit war es, aus den sehr umfangreichen Abrechnungsdaten, die Fälle zu identifizieren, die als Unfälle im Sinne des Kohorten-Modells interpretiert werden können. In diesem Zusammenhang sollte nochmal erwähnt werden, das alle Fälle die nicht über die in diesem Abschnitt beschrieben Methodik erfasst werden können, trotzdem bilanziell im Kohorten Modell erfasst sind. Diese werden dann der Phase regen Lebens zugeordnet und somit gehen keine Kosten verloren.

Das wichtigste Erkennungsmerkmal der Unfälle ist, wie bereits beschrieben, dass sie sich nicht ankündigen. Deshalb ist der Aufnahmegrund im Krankenhaus ein guter Indikator dafür, ob der Patient mit Überweisung oder Termin aufgenommen wurde oder nicht. In den Daten ist der Aufnahmegrund entsprechend den gesetzlichen Vorgaben codiert[8]. Die für uns relevanten Information sind in der 3. und 4. Stelle codiert. Die folgenden Tabelle liefert eine Übersicht der möglichen Ausprägungen und der relativen Häufigkeiten in den Daten:

[7] Das ist notwendig da jedes Krankenhaus die Fall_ID selbst vergibt und es so vorkommt das zwei Krankenhäuser die selbe ID vergeben.

[8] siehe A.2

Code	Aufnahmegrund	Anteil
01	Normalfall	56,704%
07	Notfall	38,566%
03	Verkehrsunfall / Sportunfall / Sonstiger Unfall	0,105%
02	Arbeitsunfall / Wegeunfall / Berufskrankheit	0,021%
06	Kriegsbeschädigten-Leiden / BVG-Leiden	0,003%
04	Hinweis auf Einwirkung von äußerer Gewalt	0,002%
05	(frei) früher Hinweis auf Selbstmord / Selbstbeschädigung[9]	0,001%
	Keine Angabe (Feld nicht gefüllt)	4,598%

Tabelle 3.1.: Aufnahmegrund und relative Häufigkeiten

Mehr als die Hälfte der Krankenhausfälle werden als Normalfall klassifiziert und kommen somit über eine Überweisung oder mit Termin ins Krankenhaus. Ein sehr kleiner Teil wird aufgrund einer Kriegsbeschädigung(06) behandelt oder aufgrund eines Gewaltverbrechens(04). Für uns interessant sind also die Notfälle und die Fälle die tatsächlich als Unfall aufgenommen werden (02 und 03). Letztere machen leider nur einen sehr kleinen Teil aus, was auf die Natur der Daten zurückzuführen ist. Diese Schlüssel werden nur vergeben wenn es versicherungsrelevant ist. Damit konnten wir den Anteil der möglich Unfälle bereits auf ca. 40% der Krankenhausfälle einschränken.

Den nächsten Anhaltspunkt liefert der Entlassungs- bzw. Verlegungsgrund. Allerdings gibt es 25 verschiedene Ausprägungen und zusätzlich die Information, ob der Versicherte arbeitsfähig entlassen wurde[10]. Deshalb erläutern wir lediglich die interessanten Fälle, welche in der nachfolgenden Tabelle zusammengefasst sind. Die Anteile beziehen sich dabei bereits nur auf Notfälle und Unfälle:

Entlassungsgrund	Anteil
Behandlung beendet	83,210%
Tod oder Entlassung in ein Hospiz	5,796%
Entlassung in eine Pflegeeinrichtung	3,199%
Entlassung in eine REHA-Einrichtung	2,219%
Sonstige	5,575%

Tabelle 3.2.: Entlassungsgrund und relative Häufigkeiten für Not- und Unfälle

Für den Großteil der Not- und Unfälle, endet mit dem Krankenhausaufenthalt auch die Behandlung. Ein Teil der Patient stirbt oder muss in ein Hospiz verlegt werden. Diese Fälle werden in der Prämortalitätsphase des Kohorten-Modells erfasst und sind somit für die Singulären Ereignisse, zumindest kostentechnisch, nicht relevant. Allerdings kann die Tatsache, dass ein entsprechender Unfall passiert ist, bei der Modellierung der Unfall-

[10]siehe A.2

häufigkeiten verwendet werden. Fälle mit anschließender Pflege sind für die Singulären Ereignisketten von Relevanz. Diese werden im nächsten Kapitel behandelt. Unter Sonstige sind insbesondere die nicht trivialen Fälle zusammengefasst. Das bedeutet in erster Linie eine Verlegung oder eine Entlassungen aus abrechnungstechnischen Gründen. In diesen Fällen muss der weitere Fallverlauf von aufwändig rekonstruiert werden, was nicht immer gelingt. Insbesondere bei einen Kassenwechsel des Patient stehen die Information i.d.R. nicht zur Verfügung. Hier muss dann abhängig vom jeweiligen Unfalltyp entschieden werden, ob sich der Aufwand lohnt. In diese Fall müssten die Kosten aller relevanten Anschlussbehandlung zu den Unfallkosten hinzugezählt werden.

Nachdem wir die möglichen Unfälle aufgrund des Aufnahme und Entlassungsgrundes eingegrenzt haben, sind in den Daten nur noch die DRG und ICD Informationen vorhanden. Deshalb an dieser Stelle eine kurze Erläuterung wie diese Informationen zusammenhängen. Zu jedem Fall gibt es eine DRG und mehrere Diagnosen, wobei eine davon als Hauptdiagnose ausgezeichnet wird. Die Hauptdiagnose ist in der Regel der Grund für die Hospitalisierung. Es kommt vor das während der Behandlung andere Erkrankung entdeckt werden und dadurch die Hauptdiagnose am Ende für die DRG und damit für die Kosten, kaum eine Rolle spielt[11].

Aufgrund der Komplexität der mögliche Krankenhausfälle ist es schwierig eine allgemeine Vorgehensweise zu beschreiben um Unfälle zu identifizieren. Um geeignete Kandidaten für ein Unfallmodell zu finden, kann man allerdings die häufigsten oder teuersten Fälle betrachten, die nach den Filter des Aufnahme- und Entlassungsgrundes übrig bleiben. Der Vorteil des Kohorten-Modells ist es, dass unabhängig davon wie vollständig man die Unfalltypen modelliert keine Kosten verloren gehen. Es stets möglich das Phänomen der isolierten Ereignisse durch weitere Modelle genauer zu beschreiben, falls die erforderlich ist. Die restlichen Kosten werden alle in der Phase regen Lebens erfasst.

In den verwendeten Daten waren z.B. die häufigsten Fälle Hypertonie, Herzinsuffizienz, Ohnmacht oder eine Gehirnerschütterung. Die konkreten DRGs und ICDs die man einem Unfalltyp zuordnen möchte lassen sich teilweise nur mit medizinischen Hintergrundwissen identifizieren. Aus diesem Grund wurde für die Erstellung der Modelle ein einfacher Fall ausgewählt: die Gehirnerschütterung. Für die Modellierung im nächsten Kapitel werden dabei nur Versicherte betrachtet, die über den gesamten Zeitraum bei der Kasse versichert waren und somit vollständige Datensätze besitzen.

- Hochkostenfälle ?
- Fälle von Versicherten die nicht die ganze Zeit da sind gehören zwar zum Phänomen Unfall aber die Kosten können nicht berücksichtigt werden

[11] Z.B. wenn ein Patient wegen einer Ohnmacht eingeliefert wird, kommt es vor, dass am Ende das einsetzen eines Herzschrittmachers abgerechnet wird.

3.4.3. Anwendung und Test der Modelle

- Modell ableiten am Beispiel von gehirnerschüterrung
- Saissonalität und Geschlechts-/Altersabhängigkeit testen und Parameter schätzen
- Ergodizität voraussetzen
- Eingrenzung der Fälle
- Analyse der Daten
- Häufigkeitsverteilung
- Kostenverteilung
- Zeitpunktverteilung
- Altersverteilung
- Hochkostenfälle
- Anpassungstest an Poissonverteilung
- Verknüpfung mit Kosten
- Test des Modells

A. Anhang

A.1. Abkürzungsverzeichnis und Zeichenerklärungen

$\mathbb{N}$	Natürliche Zahlen
$\mathbb{N}_0$	Natürliche Zahlen *inklusive* 0
$\mathbb{R}$	Reelle Zahlen
$\mathbb{R}^+$	die Menge reeller, nicht negativer Zahlen
F_X	Verteilungsfunktion der indizierten Zufallsgröße X
$\mathbb{E}(X)$	Erwartungswert der Zufallsgröße X
$Var(X)$	Varianz der Zufallsgröße X
$(\Omega, \mathfrak{A}, \mathbb{P})$	Wahrscheinlichkeitsraum
$X \sim F$	die Zufallsgröße X habe die Verteilung F
p_k	Wahrscheinlichkeitsfunktion diskreter Zufallsgrößen, $\mathbb{P}(X = k)$
$\mathbb{I}$	Identische Abbildung

A.2. Codierung nach § 301 Abs. 3 SGB V

Im Folgenden wird ein Auszug aus der Anlage 2 zur § 301-Vereinbarung wiedergegeben[1]

Schlüssel 1: Aufnahmegrund

1. u. 2. Stelle	01	Krankenhausbehandlung, vollstationär
	02	Krankenhausbehandlung vollstationär mit vorausgegangener vorstationärer Behandlung
	03	Krankenhausbehandlung, teilstationär
	04	Vorstationäre Behandlung ohne anschließende vollstationäre Behandlung
	05	Stationäre Entbindung
	06	Geburt
	07	Wiederaufnahme wegen Komplikationen (Fallpauschale) nach KFPV 2003
	08	Stationäre Aufnahme zur Organentnahme
	09	- frei -
3. u. 4. Stelle	01	Normalfall
	02	Arbeitsunfall / Berufskrankheit (§ 11 Abs. 5 SGB V)
	03	Verkehrsunfall / Sportunfall / Sonstiger Unfall (z. B. § 116 SGB X)
	04	Hinweis auf Einwirkung von äußerer Gewalt
	05	- frei -
	06	Kriegsbeschädigten-Leiden / BVG-Leiden
	07	Notfall

[1] Quelle: GKV - Spitzenverband http://www.gkv-datenaustausch.de/leistungserbringer/krankenhaeuser/krankenhaeuser.jsp

Schlüssel 5: Entlassungs-/Verlegungsgrund

1.u. 2. Stelle	01	Behandlung regulär beendet
	02	Behandlung regulär beendet, nachstationäre Behandlung vorgesehen
	03	Behandlung aus sonstigen Gründen beendet
	04	Behandlung gegen ärztlichen Rat beendet
	05	Zuständigkeitswechsel des Kostenträgers
	06	Verlegung in ein anderes Krankenhaus
	07	Tod
	08	Verlegung in ein anderes Krankenhaus im Rahmen einer Zusammenarbeit (§ 14 Abs. 5 Satz 2 BPflV in der am 31.12.2003 geltenden Fassung)
	09	Entlassung in eine Rehabilitationseinrichtung
	10	Entlassung in eine Pflegeeinrichtung
	11	Entlassung in ein Hospiz
	12	interne Verlegung
	13	externe Verlegung zur psychiatrischen Behandlung
	14	Behandlung aus sonstigen Gründen beendet, nachstationäre Behandlung vorgesehen
	15	Behandlung gegen ärztlichen Rat beendet, nachstationäre Behandlung vorgesehen
	16	externe Verlegung mit Rückverlegung oder Wechsel zwischen den Entgeltbereichen der DRG-Fallpauschalen, nach der BPflV oder für besondere Einrichtungen nach § 17b Abs.1 Satz 15 KHG mit Rückverlegung
	17	interne Verlegung mit Wechsel zwischen den Entgeltbereichen der DRG-Fallpauschalen, nach der BPflV oder für besondere Einrichtungen nach § 17b Abs.1 Satz 15 KHG
	18	Rückverlegung
	19	Entlassung vor Wiederaufnahme mit Neueinstufung
	20	Entlassung vor Wiederaufnahme mit Neueinstufung wegen Komplikation
	21	Entlassung oder Verlegung mit nachfolgender Wiederaufnahme
	22	Fallabschluss (interne Verlegung) bei Wechsel zwischen voll- und teilstationärer Behandlung
	23	Beginn eines externen Aufenthalts mit Abwesenheit über Mitternacht (BPflV-Bereich – für verlegende Fachabteilung)
	24	Beendigung eines externen Aufenthalts mit Abwesenheit über Mitternacht (BPflV-Bereich – für Pseudofachabteilung 0003)
	25	Entlassung zum Jahresende bei Aufnahme im Vorjahr (für Zwecke der Abrechnung – PEPP*)
3. Stelle	1	arbeitsfähig entlassen
	2	arbeitsunfähig entlassen
	9	keine Angabe

Abbildungsverzeichnis

Tabellenverzeichnis