Westfälische Wilhelms-Universität
Münster
Institut für Mathematische Statistik

Credibility Theory

Diplomarbeit

vorgelegt von
Michael Schwab

Betreuer: Prof. Dr. Gerold Alsmeyer

29.Dezember 2011

Westfälische Wilhelms-Universität
Münster
Institut für Mathematische Statistik

Credibility Theory

Diplomarbeit

vorgelegt von
Michael Schwab

Betreuer: Prof. Dr. Gerold Alsmeyer

29.Dezember 2011

Zusammenfassung

Dir vorliegende Arbeit soll einen Einblick in die Grundlagen der Credibility-Theory geben. Sie wurde entwickelt, um „glaubwürdige " Schätzer für die Schadenneigung von Versicherten zu haben. Um dieses Anwendungsfeld der mathematischen Statistik besser einschätzen zu können wird zu Beginn auf das Umfeld der Credibility-Theory, die Versicherungswirtschaft eingegangen (1.1). Mit dem Bayes-Ansatz wird dann ein erstes Prämienproblem gelöst (2.4).

Die Stärke der Credibility-Theory ist, dass sie weniger Vorwissen über die auftretenden Wahrscheinlichkeitsverteilungen verlangt als für einen Bayes-Schätzer benötigt werden. So verzichten wir in einem ersten einfachen Credibility-Modell (Kap. 3.1) bereits auf die Kenntnis der genauen Form von Strukturfunktion des Versicherungskollektivs und Schadenverteilung der Individuen. Beim Modell von Bühlmann (3.2) modellieren wir das erste Mal ein ganzes Kollektiv von Versicherten und führen daher den Unterschied zwischen homogenem und inhomogenem Credibility-Schätzer ein.

Fassen wir die Credibility-Schätzer als Orthogonalprojektionen auf Unterräume des $\mathbf{L^2}$ auf (4.1) gewinnen wir einen intuitiven Zugang zu dieser Sichtweise auf das Problem der Prämienberechnung. Wir nutzen die Hilbertraumeigenschaften des $\mathbf{L^2}$ um mit den Normalengleichungen (4.2) einfach nachzurechnende Kriterien für Credibility-Schätzer bereitzustellen. Wir benutzen sie, um den homogenen und inhomogenen Credibility-Schätzer im Bühlmann-Straub-Modell zu finden (4.3). Abschließend (4.5) wenden wir die Ergebnisse und zuvor gefundene Schätzer für Strukturparameter Bühlmann-Straub-Modell auf einen Datensatz aus der KFZ-Haftpflichtversicherung an.

Bei der Darstellung des Themas erfolgt weitgehend wie in [BG05].

Inhaltsverzeichnis

Tabellenverzeichnis

Abbildungsverzeichnis

Kapitel 1

Einleitung

1.1 Was ist eine Versicherung?

Das Eingehen von Risiken ist nahezu unvermeidlich. Jeder menschliche Tätigkeit – auch dem Unterlassen – ist es zu eigen, dass uns daraus Nachteile entstehen können. Von einem wirtschaftswissenschaftlichen Standpunkt aus betrachtet sind solche Nachteile dann Kosten, die bezahlt werden müssen. Dies beginnt bei unvermeidlichen Gesundheitsrisiken die im Krankheitsfall hohe Behandlungskosten hervorrufen. Es setzt sich fort etwa im Straßenverkehr, wo durch Unfälle enorme Schadenssummen entstehen können. Auch beruflich oder in der Freizeit, z.B. beim Bergsteigen, werden Risiken eingegangen, die enorme finanzielle Folgen in sich bergen.

Dementgegen steht ein Bedürfniss nach Sicherheit und Planbarkeit. Eine mögliche Strategie, um die eingegangenen Risiken beherrschbar zu machen, wäre es, zu betrachten, gegen welche Schadenfälle eine Absicherung erwünscht ist. Dann könnte eine entsprechende Summe Geld an die Seite gelegt werden um sich auf das Eintreten vorzubereiten. Diese Möglichkeit ist jedoch für die meisten Menschen nicht gangbar, zum einen wegen fehlender finanzieller Mittel oder Sicherheiten, zum anderen, weil ihnen der Überblick zu einer akuraten Bewertung der eingegangenen Risiken fehlt.

Aus dieser Unsicherheit entstand das Konzept der Versicherung. Der Träger eines Risikos bietet einem Versicherer[1] Geld dafür, ihm das Risiko (oder

[1]Ob es sich dabei um ein profitorientiertes Unternehmen, einen "Verein auf Gegenseitigkeit", ein anderes genossenschaftliches Konstrukt oder eine informelle Vereinigung handelt, ist unerheblich.

zumindest einen Teil davon) abzunehmen. Ein entscheidender Faktor für das Zustandekommen eines solchen Vertrages ist natürlich der Preis, den der Risikoinhaber für die Übernahme des Risikos bereit ist zu zahlen bzw. den der Versicherer für das Absichern des Risikos verlangt. Diesen Preis nennen wir (Versicherungs-)Prämie.

1.2 Die mathematische Formulierung des Problems der Prämienberechnung

1.2.1 Erster Ansatz

Das Problem der Prämienberechnung bezieht sich meist nicht auf einzelne zu versichernde Risiken, sondern auf ein ganzes Portfeuille von Risiken, die wir mit $i = 1, 2, \ldots, N$ indizieren. Jedes Risiko produziert dann in einer bestimmten Periode Kosten von X_i. Wir können die X_i als Zufallsvariablen betrachten, die Werte in den positiven reellen Zahlen annehmen. Eine unmittelbar einsichtige Anforderung an die Versicherungsprämie ist, dass die gesammten Prämieneinnahmen die Schadenskosten decken sollten. Gehen wir davon aus, dass wir die Prämien allein auf Grund dieses Deckungskriteriums berechnen. Ferner wollen wir annehmen, dass wir als Versicherer keinerlei Informationen über Unterschiede zwischen den Individuen haben, die wir in unserem Portfeuille versichern. So setzen wir ersteinmal für alle die gleiche Prämie an. In diesem Fall sprechen wir auch von einem homogenen Portfeuille. Dann ergibt sich aus dem Deckungskriterium und der Homogenität der Risiken

$$P^* = \frac{1}{N} \sum_{i=1}^{N} X_i \tag{1.1}$$

als mögliche Prämie für die aktuelle Periode. P^* ist also einfach ein Schadensmittel. Eine Prämie, die genau die erwarteten Schadenkosten abdeckt, nennen wir reine Risikoprämie. Neben dem reinen Risikoprinzip gibt es noch weitere Überlegungen, die in die Prämienberechnung mit einfließen können. Sie werden im Abschnitt 1.2.4 kurz angesprochen.

Homogenität war lange ein Leitbild der Versicherungsmathematik. Durch das Zuordnen von Risiken in bestimmte Kategorien erhoffte man, Homogenität in erzeugen zu können. Für die KFZ-Haftpflichtversicherung beschreibt Tabelle 1.1 mögliche Kategorisierungen.

Faktor	Merkmal
Auto	Art, Hubraum, Leistung, Nutzlast, Abmessungen, Gewicht, Alter
Gebiet	Bundesland, Kreis, Stadt, geographische Lage, Verkehrsdichte, Fahrzeugdichte, Straßenzustand, Einwohnerzahl
Fahrintensität	Gefahrene Kilometer, Verwendungszweck, berufliche/private Nutzung
Fahrer (objektiv)	Geschlecht, Alter, Familienstand, Führerscheindauer, Beruf
Fahrer (subjektiv)	Intelligenz, Charakter, körperliche Veranlagung, Gesundheitszustand, „Verkehrsgesinnung “

Tabelle 1.1: Risikofaktoren in der KFZ-Haftpflicht, vgl. [Meh62]S.23

Der der Tabelle zugrundeliegende Text [Meh62] kann auch als Beispiel dienen, wie Versicherungsmathematik betrieben wurde, bevor die stochastische Modellierung einzu hielt.

In unserer anfänglichen Betrachtung haben wir von vielem abstrahiert, u.a. den Unterschieden zwischen den versicherten Individuuen. Wie eine Versicherungsprämie aus ihrer Perspektive gestaltet werden könnte, soll nun überlegt werden.

1.2.2 Die individuelle Prämie

In der Regel liegt dem Versicherungsunternehmen nicht nur ein ganzes Portfeuille von Risiken vor, sondern die Individuen und ihre Risiken werden auch über mehrere Perioden hinweg versichert. Die Perioden indizieren wir mit $j = 1, \ldots, J$, so dass also X_{ij} die Kosten des i-ten Risikos (Versicherungsvertrages) in der j-ten Periode darstellt. Das heißt ein Individuum i produziert in der Zeit $j = 1, \cdots, J$ die Forderungen $X_{i,1}, \ldots, X_{i,J}$. Der Index J deutet bereits an, dass es sich in der Praxis bei den Perioden oft um Jahre handelt.

Deshalb werden die beiden Begriffe von nun an gleichbedeutend vewendet.

Um die Daten aus verschiedenen Perioden mathematisch verwertbar zu machen:

Annahme 1.1 (Stationarität) *Für ein Individuum i sind die* $X_{ij},\ j = 1, \ldots, J$ *identisch verteilt.*

Durch diese Annahme können wir nun Datenreihen etwa von individuellen Forderungen sinnvoll analysieren. Die Stationaritätsannahme ermöglicht uns den Vergliech von Daten aus unterschiedlichen Perioden. Identische Verteilung ist natürlich eine rigide Vorraussetzung. Eine Aussage darüber, wie sich die Verteilung der Risiken eines Individuums mit der Zeit entwickelt, wird aber immer nötig sein. Annahme 1.1 stellt auch insofern eine Abstraktion dar, als dass durch unterschiedliche Einflüsse wie etwa Inflation, sich ändernde Vertragsbedingungen u.s.w. die Verteilungen der Risiken meistens nicht direkt, sondern bereits in verarbeiteter Form in die Berechnungen einfließen müssen. Auch diesen Fragen werden wir uns kurz im Abschnitt 1.2.4 widmen.

Annahme 1.2 (bedingte Unabhängigkeit) *Für ein Individuum i sind die* $X_{ij},\ j = 1, \ldots, J$ *bedingt unabhängig.*

Die Formulierung der **bedingten** Unabhängigkeit rührt daher, dass wir nunmehr, anders als in Abschnitt 1.2.1, davon ausgehen, dass es Unterschiede zwischen den Individuen $j = 1, \ldots, N$ gibt. Genauer gesagt können sich die zugrundeliegenden Verteilungen der X_{ij} von einem Individuum zum anderen unterscheiden. Die Unabhängigkeit der jährlichen Schadenforderungen X_{ij}, $j = 1, \ldots, J$ eines Individuums i kann also nur bedingt auf der für dieses Individuum vorliegenden Verteilung $F(x)$ sinnvoll definiert werden.

Kern der versicherungsmathematischen Betrachtung ist nun eben diese Verteilung der Risiken $F(x)$. A priori ist über sie nichts bekannt, nur, dass die Forderungen unterschiedlicher Individuen oder Risiken unterschiedlich verteilt sind. In der praktischen Anwendung wird davon ausgegangen, dass die Individuen gewisse Ähnlichkeiten aufweisen, dass sie zu einer gemeinsamen „Klasse" von Risiken gehören. Das wird etwa in der KFZ-Versicherung dadurch erreicht, dass die Risiken nach Fahrzeugtyp, Motorleistung etc. kategorisiert werden (siehe Tabelle 1.1). Doch erfahrungsgemäß sind auch innerhalb einer Risikoklasse die Unterschiede zwischen den Individuen erheblich. So hängt gerade im KFZ-Bereich viel von der individuellen Verfassung und Risikoneigung des jeweiligen Fahrers ab. Dies sind

Bedingungen, die sich weder gut quantifizieren lassen noch im Vorhinein bekannt sein dürften[2].

Um das eben beschriebene mathematisch greifbar zu machen, wird hier folgende Interpretation angeboten: Die *Ähnlichkeit* der Risiken in einer Klasse wollen wir so auffassen, dass die in diesem Kollektiv vorkommenden Verteilungen zu einer gemeinsamen Familie von Verteilungen gehören. Die Unterschiedlichkeit der individuellen Risiken im Kollektiv sollen durch einen Parameter θ dargestellt werden. Dabei ist der hier gewählte Begriff „Familie“ vorerst nicht in einem engen mathematischen Sinne aufzufassen (etwa Exponentialfamilie). Auch als Parameter kann bei dieser abstrakten Betrachtung noch die Verteilung selbst oder das zu versichernde Individuum herhalten. Bei einigen späteren Modellbetrachtungen werden die Begriffe allerdings genau in dieser Richtung verwand. Θ beinhalte die möglichen „Risikoprofile“, die ein Individuum beim Realisieren von Forderungen entsprechend des Versicherungsvertrages annehmen kann. Wir indizieren also eine Verteilungsfunktion F mit dem Parameter $\theta \in \Theta$.

Aus Sicht des Versicherungsnehmers i währe nun eine „faire“ Versicherungsprämie eine solche, die auf den zu erwartenden Schadensforderungen aufgrund seines individuellen Risiko-Profils beruht. Um die Notation übersichtlich zu halten, verzichten wir auf den Index i oder j, falls Individuum bzw. Periode im betreffenden Zusammenhang erkennbar die selben bleiben.

Wir kommen daher zur

Definition 1.3 *Die korrekte individuelle Prämie zu einem Risikoprofil θ ist gegeben durch:*

$$P^{ind}(\theta) = E\left[X_{J+1}|\theta\right] =: \mu(\theta) \tag{1.2}$$

Der Index $J+1$ soll hier andeuten, dass die Frage der Prämienberechnung sich immer auf zukünftig zu erwartende Schäden bezieht. Das heißt, dass die (unbekannte) Größe $\mu(\theta)$ bestimmt werden muss. Liegen bereits Daten über den bisherigen Forderungsverlauf das Individuums i vor, können wir einen Schätzer $\widehat{\mu(\theta)}$ anwenden.

Die Verteilungseigenschaften der Forderungen des Individuums i sind vollständig im Parameter θ enthalten. Daher könnten wir Gleichung (1.2) auch schreiben als: $P_i^{ind} = E\left[X_{J+1}|\theta_i\right] =: \mu(\theta_i)$.

[2]Auf das Problem der Informationsasymmetrie zwischen Versicherer und Kunden werden wir noch im Abschnitt 1.3.2 zu sprechen kommen.

1.2.3 Die kollektive Prämie

Über den genauen Wert von $\theta_i \in \Theta$ ist, wie gesagt, im Vorraus nichts bekannt, allenfalls, dass er dem Kollektiv Θ entstammt. An dieser Stelle unserer Betrachtung tritt nun das zweite Mal der Zufall auf. Das Risikoprofil θ_i eines Individuum i kann als das Ergebniss einer zufälligen Ziehung aus dem Kollektiv Θ aufgefasst werden. Die Häufigkeiten der verschiedenen möglichen Risikoprofile in Θ, also etwa die Zahl guter und schlechter Fahrer in einer Gruppe bei der KFZ -Versicherung, sind wieder unbekannt. Informationen über die Struktur des Kollektivs Θ könnten wir zum Beispiel aus Informationen über die bisherigen Forderungsverläufe der darin versicherten Individuen erhalten[3]. Dieses – wie auch immer gewonnene – Vorabwissen über die Verteilung der θ_i in Θ führt uns zu einer Verteilungsfunktion $U(\theta)$.

Definition 1.4 *Die Verteilung $U(\theta)$ auf Θ nennen wir die Strukturfunktion des Kollektivs.*

Mit der Strukturfunktion $U(\theta)$ können wir nun zu einem neuen Verständniss eines Schadensmittels kommen, das im Unterschied zu Gleichung (1.1) die Heterogenität des zu versichernden Kollektivs explizit darstellt.

Definition 1.5 *Die kollektive Prämie P^{koll} ist gegeben durch:*

$$P^{koll} = \int_{\Theta} \mu(\theta)\, dU(\theta) =: \mu_0 \tag{1.3}$$

1.2.4 Datenqualität und temporäre als-ob-Statistiken

Wie in Abschnitt 1.2.1 und 1.2.2 angekündigt, wollen wir uns kurz zwei Fragen widmen, die sich im näheren Umfeld der Prämienberechnung befinden. Dabei geht es zum einen um die zugrundeliegenden Daten, mit denen die Prämien berechnet werden sollen. Zum anderen wird andiskutiert, welche weiteren Faktoren bei einem praktischen Problem der Prämienberechnung berücksichtigt werden müssen.

Die Forderungen, die der Versicherungsnehmer dem Versicherer gegenüber geltend machen kann, setzen sich aus einer Reihe von Daten zusammen.

[3] Das wäre dann die sogenannte frequentistische Interpretation. Die Vorstellung, dass ein erfahrener Versicherungsmathematiker oder Statistiker hier Spezialwissen aufwendet (bayesianischer Ansatz), muss vorerst nicht bemüht werden.

Das wird besonders deutlich, wenn wir für den Moment die Prämienberechnung bei Rückversicherungen betrachten. Dabei folgen wir der Darstellung in [BS70]. Das „Schadentotal“ X_{ij}, also die Forderung eines Erstversicherers i im Jahr j gegenüber dem Rückversicherer ergibt sich aus folgenden Basisdaten:

$$\left.\begin{aligned} r_{ij} &= \text{Selbstbehalt je Schadenfall} \\ R_{ij} &= \text{max. Deckungsbeitrag je Schadensfall} \\ A_{ij} &= \text{Anzahl Schadenfälle} \\ Y_{ij}^{m} &= \text{Betrag des m-ten Originalschadens} \end{aligned}\right\} \quad \text{für Versicherer } i \text{ im Jahr } j$$

Wenn wir nun δ_{ij} als die Zusammenfassung dieser Vertragsbedingungen auffassen, d.h. als eine Abbildung, die die beim Erstversicherer angefallen Schäden auf das Schadentotal des entsprechenden Rückversicherungsvertrags i für das Jahr j abbildet, kommen wir zu folgender Darstellung:

$$\left(Y_{ij}^{1}, Y_{ij}^{2}, \ldots, Y_{ij}^{A_{ij}}\right) \xmapsto{\delta_{ij}} \sum_{m=1}^{A_{ij}} \min\left\{\left[Y_{ij}^{m} - r_{ij}\right]^{+},\ R_{ij}\right\}$$

Diese Gleichung vermittelt einen Eindruck von der Komplexität der in einem Versicherungsvertrag erfassten und verarbeiteten Informationen. Zu den Vertragsbedingungen δ_{ij}, die sich jahresweise verändern können[4], kommen weitere Veränderungen wie Inflation, allgemeine Trends (mehr Verkehr, längerer Winter …) hinzu. Doch geht es dem (Rück-)Versicherungsunternehmen darum, die dem jeweiligen Versicherungsnehmer eigene Risikodisposition herauszufinden. Daher wird versucht, allgemeine Trends aus diesen Berechnungen herauszuhalten. In der Praxis bedient man sich dabei „vertragsweiser as-if-Statistiken“. Das heißt, Beirtragsänderungen werden herausgerechnet, Trends eliminiert und Sondereffekte geglättet, um zu einer Datengrundlage zu kommen, die so aussieht **als ob** diese Enflüsse nicht vorlägen. Für unsere weitern Betrachtungen gehen wir im Sinne einer auf das Kernproblem bezogenen Darstellung vom Vorliegen solcher Als-ob-Statistiken aus.

In Abschnitt 1.2.1 haben wir erläutert, was unter einer reinen Risokoprämie verstanden werden soll. Während die zu erwartenden Schadenkosten natürlich ein zentrales Element sind, kann in die Prämienberechnung eines Versicherungsvertags auch noch anderes einbezogen werden. Allgemein verstehen wir dabei unter einem Prinzip zur Prämienberechnung ei-

[4] Das kann auch bei Erstversicherungsverträgen der Fall sein, wenn z.B. der Selbstbehalt oder der Leistungskatalog verändert werden.

ne Abbildung $\mathcal{H}$, die eine Zufallsvariable X (oder ihre zugehörigen Verteilungsfunktion F^X) in die reellen Zahlen abbildet. Für ein $\lambda \in \mathbb{R}^+$ heißt die Abbildung

$$
\begin{array}{ll}
X \longmapsto \mathcal{H}(X) & \\
\text{Erwartungswertprinzip, wenn} & \mathcal{H}(X) = (1+\lambda)\, E(X) \\
\text{Standardabweichungsprinzip, wenn} & \mathcal{H}(X) = E(X) + \lambda \cdot \sigma(X) \\
\text{Varianzprinzip, wenn} & \mathcal{H}(X) = E(X) + \lambda \cdot \sigma^2(X) \\
\text{Exponentialprinzip, wenn} & \mathcal{H}(X) = 1/\lambda \ln E(e^{\lambda X})
\end{array}
$$

Dabei besteht ein Prämienprinzip immer aus der reinen Risikoprämie $E(X)$[5] und einem nichtnegativen Sicherheitsaufschlag. Der Sicherheitsaufschlag (engl. risk loading) deckt zum einen die Kosten, die durch extreme Schadenshäufungen entstehen können. Zum anderen kann er auch als Preis des eingesetzten Risikokapitals interpretiert werden. In der Praxis müssen ferner Verwaltungskosten des Versicherungsunternehmens und ein geplanter Gewinn aus den Prämien finanziert werden. Eine fundierte Darstellung zu Prämienprinzipien und ihrer Eigenschaften findet sich u.a. bei [Kre99].

In der vorliegenden Betrachtung soll es weder um die Absicherung gegen außergewöhnliche Schadenereignisse noch um finanzmathematische Aspekte beim Anlegen von Risikokapital bei oder durch Versicherungsunternehmen gehen. Fokus der Untersuchung ist die Prämienberechnung für Versicherungsnehmer als Teil eines Kollektivs ähnlicher Risiken. Daher beschäftigen wir uns im Folgenden ausschließlich mit der reinen Risikoprämie, die für $\lambda = 0$ ein Erwartungswertprinzip ist.

1.3 Warum Credibility Theory?

1.3.1 Individuelle vs. kollektive Prämie

Mit den Definitionen (1.2) und (1.5) haben wir bisher zwei unterschiedliche Perspektiven auf das Versicherungsproblem eingenommen. Zum einen die des Versicherers, der mindestens so viele Einnahmen durch Prämien erzielen muss, wie er Schadenkosten in der kommenden Periode erwartet. Zum anderen die des Versicherungsnehmers, der ungerne mehr zahlen

[5] Wobei E(X) hier als erwarteter Schaden im Sinne eines der später erläuterten Schätzverfahren zu verstehen ist.

möchte, als er für sich selbst als Schaden erwartet. Über die kollektive Prämie hat der Versicherer aus der Erfahrung mit dem Kollektiv heraus genaue Informationen. Über die individuell anzusetzende Prämie hingegen weiß er wenig, bei einem Neukunden nichts. Durch Kategorisierung kann er erreichen, dass er einen neuen Versicherungsnehmer i einem Kollektiv mit bekannter Strukturfunktion $U(\theta)$ oder zumindest bekanntem kollektiven Mittelwert etc. zuordnet. Über das θ_i, die Risikodisposition des Individuums, weiß er allerdings immer noch nichts.

Wie zu Beginn von Abschnitt 1.2.3 geschildert, können wir diese Situation als ein zweistufiges Auftreten des Zufalls interpretieren. In einer ersten Stufe wird aus dem Kollektiv Θ entsprechend seiner Verteilungs-/ Strukturfunktion $U(\theta)$ wie aus einer Urne ein θ gezogen. Die so erhaltene Risikodisposition des betrachteten Individuums bestimmt nun seine individuellen Schadenforderungen $X_{i,1}, X_{i,2}, \ldots$. Auf Seite 21 ist in Abbildung 2.1 dargestellt, wie das sogenannte Zwei-Urnen-Modell für einen gamma-verteilte Strukturfunktion aussieht. Sie erzeugt Parameter λ_i, die dann als individuelle Risikodisposition die Poisson-verteilten Schadenhäufigkeiten eines Individuums i beeinflussen.

Wenn wir Θ nicht nur als den Raum der möglichen Riskikodisposition auffassen, sondern damit auch die Zufallsvariable identifizieren, mit der wir die individuellen θ aus der ersten Urne ziehen, können wir die korrekte individuelle Prämie nun schreiben als

$$P^{ind} = \mu(\Theta) = E\left[X_{J+1}|\Theta\right]. \tag{1.4}$$

Im Unterschied zu P^{koll} (Definition 1.5) ist P^{ind} hier selbst eine Zufallsvariable, nämlich ein auf Θ bedingter Erwartungswert. Nun lässt sich auch mathematisch ausschreiben, was in Annahme (1.2) mit **bedingter** Unabhängigkeit der X_j gemeint war:

Für ein Individuum i seien X_j, X_k die Schadenforderungen aus zwei Jahren j und k. Dann gilt unter den Voraussetzungen $Cov(X_j, X_k) < \infty$ und $\mathrm{Var}(\Theta) > 0$ mit dem Gesetz von der totalen Kovarianz

$$\begin{aligned}\mathrm{Cov}(X_j, X_k) &= \mathrm{E}\big[\underbrace{\mathrm{Cov}\left(X_j, X_k|\Theta\right)}_{=\,0\text{ n.V.}}\big] + \mathrm{Cov}\big(\underbrace{\mathrm{E}[X_j|\Theta]}_{=\,\mu(\Theta)},\ \underbrace{\mathrm{E}[X_k|\Theta]}_{=\,\mu(\Theta)}\big)\\ &= \mathrm{Cov}(\mu(\Theta), \mu(\Theta))\\ &= \mathrm{Var}(\mu(\Theta)) > 0.\end{aligned}$$

Also ist zwar $\mathrm{Cov}(X_j, X_k) > 0$ und die X_j sind K_k sind weder unabhängig noch unkorelliert. Das liegt an der durch θ_i gegeben individuellen Risikodisposition. Doch betrachten wir die Schadenforderungen eines Individuums isoliert unter gegebenem θ_i verhalten sich die Schadenforderungen $X_{ij}, j = 1, \ldots, J$ wie unabhängige Zufallsvariablen. Näheres zur bedingten Unabhängigkeit und ihrer (teils problematischen) Verwendung in der Versicherungsmathematik findet sich etwa in [HS87].

Ziel unserer Untersuchung ist es, einen möglichst guten Schätzer $\widehat{\mu(\Theta)}$ für die korrekte individuelle Prämie zu finden (später $\widehat{\mu(\Theta_i)}$ die Prämie für eines der I Individuen). Zwar haben wir mit P^{koll} eine Prämie, die die Kosten des Versicherers deckt, doch ist sie wegen des ökonomischen Problems der adversen Selektion praktisch nur anwendbar, wenn uns noch keine Informationen über den Versicherungsnehmer vorliegen. Das dahinterstehende Konzept soll im letzten Abschnitt dieses Kapitels knapp erläutert werden.

Wir stehen auf jeden Fall vor der Herausforderung, die individuelle Prämie schon zu einem frühen Zeitpunkt gut zu schätzen. Dazu ziehen wir die bereits über den Versicherungsnehmer i gewonnen Daten $X_{ij},\ j = 1, \ldots, J$ aus den ersten J Perioden heran und fassen sie als Forderungsverlauf oder Erfahrung im Vektor X zusammen. Den besten Schätzer für die individuelle Prämie unter Heranziehung des Erfahrungsvektors nennen wir Bayes-Schätzer und kommen formal zur

Definition 1.6 *Bayes-Prämie*

$$P^{Bayes} = \widetilde{\mu(\Theta)} := E\left[\mu(\Theta)|X\right] \tag{1.5}$$

Zum Konzept der Bayes-Statistik und wie wir durch sie eine Ordnung auf einer (Teil-)Menge aller möglichen Schätzer für $\mu(\Theta)$ erhalten, kommen wir in Kapitel 2. Hier sei noch festgehalten, dass P^{Bayes} nach seiner Definition als auf X bedingter Erwartungswert selbst wieder eine Zufallsvariable ist. Zum Zeitpunkt der Berechnung einer Prämie für die nächste Periode ist X jedoch bereits bekannt und kann in die Formel eingesetzt werden. Im Unterschied zu P^{ind} ist P^{Bayes} also eine richtige Prämie, mit der sich ein Wert aus den reellen Zahlen berechnen lässt.

1.3.2 Adverse Selektion und Versicherungsmärkte

An dieser Stelle soll noch eine Frage der Motivation für die Credibility-Theorie geklärt werden. Wir haben im reinen Risikoprinzip und mit der kollektiven Prämie aus Definition 1.5 bereits einen Berechnungsansatz gefunden, der eine sehr bedeutende Eigenschaft erüllt: Die kollektive Prämie ist eine gute Vorraussage für die Schadenkosten der nächsten Periode. Wird sie als Durchschnittsprämie von allen Versicherungsnehmern gefordert, sollten mit den Einnahmen alle Schadenkosten beglichen werden können.

Entscheiden sich bei einer Einheitsprämie die Individuen jedoch aufgrund eines rationalen Kalküls, ob sie den Versicherungsvertrag eingehen, kommt u.U. kein effizienter Versicherungsmarkt zustande. Dies wollen wir am folgende Beispiel nach [Gür62] belegen.

Am Versicherungsmarkt gebe es $N = 1000$ potentielle Versicherungsnehmer. Das Versicherungsunternehmen weiß nichts über ihre individuellen Risikodispositionen. Ihm ist lediglich bekannt, dass in den letzten Jahren durchschnittlich jeweils $X_\bullet = 200.000$ EUR Schaden angefallen sind. Die kollektive Durchschnittsprämie beträgt dann $P^{koll} = X_\bullet/N = 200$ EUR.

Im Gegensatz zum Versicherungsunternehmen wissen die Individuen genau über ihre individuelle Risikodisposition bescheid. Das Individuum X_i erwarte den Schaden $\mathrm{E}[X_i] = 2i/5$. Es ist im Rahmen eines rationalen Kalküls nur bereit, einen Versicherungsvertrag einzugehen, wenn die Prämie unterhalb des erwarteten Schaden liegt, also $P < \mathrm{E}[X_i]$.

Der Versicherungsmarkt funktioniere nun folgendermaßen:

1. Der Versicherer bietet eine Versicherung zum Preis P_0 an, Grundlage dafür ist die kollektive Durchschnittsprämie der Vorperiode.
2. Die Individuen entscheiden sich entsprechend ihres individuell rationalen Kalküls für oder gegen das Eingehen eines Versicherungsvertrages.
3. Nun realisieren die versicherten Individuen ihre Schäden entsprechend ihres Erwartungswertes, der Versicherer zahlt die Schadensummen aus.

In unserem Fall beginnt die erste Periode, indem der Versicherer ein Angebot macht, sich für 200 EUR bei ihm versichern zu lassen. Die Individuen $1 \dots 500$ entscheiden sich gegen eine Versicherung, die restlichen dafür. Der Versicherer macht Einnahmen i.H.v. 100000 EUR. Die anfallenden Schäden ergeben sich (in dieser statischen Betrachtung) zu $X_\bullet = \sum_{i=501}^{1000} i * 2/5 = 150100$ (EUR).

Tabelle 1.2 zeigt die weitere Entwicklung in diesem Model des Versicherungsmarktes. Der Markt kommt nach neun Perioden zum Erliegen, das Versicherungsunternehmen macht insgesamt 66760,40 EUR Verlust.

Dieses Modell veranschaulicht das Prinzip der adversen Selektion, das

Periode	1	2	3	4
Prämie (EUR)	200	300, 20	350, 20	375, 20
Anzahl Versicherte	500	250	125	62
Prämieneinnahmen (EUR)	100000	75050	43775	23262, 40
Anfallender Schaden (EUR)	150100	87550	46900	24043, 60
Verlust (EUR)	50100	1250	3125	781, 20

Tabelle 1.2: Versicherungsmarkt mit Durchschnittsprämien und asymmetrischer Information

auf Versicherungsmärkten bei asymmetrischer Information und Durchschnittsprämien auftreten kann. Die „guten Risiken“ wollen sich nichtmehr versichern lassen. Dadurch verschlechtert sich das Portefeuille insgesamt und nach und nach ist die Versicherung nurnoch für „schlechte Risiken“ attraktiv. Kennen die potentielle Versicherte ihre Risikodispositionen, ist es notwendig, die Prämien möglichst nah an diesen zu setzen. Bühlmann und Gisler folgern daraus:

Satz 1.7 (Die korrekte individielle Prämie) *Die korrekte individuelle Prämie ist die wettbewerbsfähigste Prämie.*

Kapitel 2

Statistische Grundlagen

2.1 Entscheidungstheorie

Im mathematischen Sinn ist ein Schätzer wie der im vorherigen Kapitel beschrieben Bayes-Schätzer $\widetilde{\mu(\Theta)}$ eine Abbildung einer Menge beobachteter Daten in eine Menge der möglichen Entscheidungen. Wie auch immer die Daten gewonnen werden, bezeichnen wir diese Beobachtungen als statistisches Experiment.

Definition 2.1 *Ein statistisches Experiment besteht aus drei Elementen*

$$E = (\mathfrak{X}, \mathscr{A}, (W_\theta)_{\theta \in \Theta}) \tag{2.1}$$

wobei

- $\mathfrak{X}$ *Stichprobenraum heißt und die möglichen Beobachtungen beinhaltet.*
- $\mathscr{A}$ *als* σ*-Algebra über* $\mathfrak{X}$ *die beobachtbaren Ereignisse repräsentiert.*
- $(W_\theta)_{\theta \in \Theta}$ *eine Familie von Wahrscheinlichkeitsmaßen ist, die mit* $\theta \in \Theta$ *parametrisiert ist.*

In diesem Rahmen beobachten wir eine Zufallsvariable $X : (\Omega, \mathfrak{A}) \longrightarrow (\mathfrak{X}, \mathscr{A})$, die mit $(P_\theta)_{\theta \in \Theta}$ auf $(\Omega, \mathfrak{A})$ verteilt ist. Wir identifizieren $P_\theta^X = W_\theta$ für alle θ.

In unserem Falle geht es darum, den Paramter θ einer Verteilungsfunktion $F_\theta(x) = W_\theta\{X \leq x\}$, oder, allgemeiner, ein Funktional $g(\theta)$ (z.B. Erwartungswert, Varianz der Verteilung) des Parameters zu schätzen.

$D = \{g(\theta) : \theta \in \Theta\}$ sei die Menge aller möglichen Schätzwerte (Entscheidungen), die wir realistischerweise für θ erwarten können, $\mathfrak{D}$ sei eine σ-Algebra darauf.

Definition 2.2 *$(D, \mathfrak{D})$ heißt Entscheidungsraum. Eine Abbildung*

$$\delta : (\mathfrak{X}, \mathscr{A}) \longrightarrow (D, \mathfrak{D})$$

heißt Entscheidungsfunktion.

In unserem Fall, in dem δ eine Entscheidung darüber trifft, welchen Parameterwert θ oder welchen Wert des entsprechenden Funktionals $g(\theta)$ wir aufgrund der Daten für am wahrscheinlichsten halten, nennen wir δ eine Schätzfunktion oder schlicht Schätzer. Wir suchen nun nach einer Möglichkeit, die Güte einer Entscheidung zu bewerten. Ziel ist es, dadurch eine Ordnung unter Entscheidungsfunktionen zu erzeugen, so dass wir zwei Entscheidungsfunktionen miteinander vergleichen können. Als ersten Schritt dazu formalisieren wir den Verlust, der beim Ergreifen einer Entscheidung (beim Schätzen eines falschen Parameterwertes) entstehen kann.

Definition 2.3 *Die Abbildung $L(\theta, \delta(x)) : \Theta \times D \longrightarrow \mathbb{R}$ heißt Verlustfunktion.*

$L(\theta, \delta(x))$ gibt den Verlust an, der bei Vorliegen des wahren Parameterwerts θ und getroffener Entscheidung δ (aufgrund der Datenlage x) auftritt. Ein typisches Beispiel ist die Verlustfunktion von Gauß-Markov, die wegen ihrer Struktur auch quadratische Verlustfunktion genannt wird:

$$L(\theta, \delta) = (g(\theta) - \delta(x))^2 \tag{2.2}$$

Definition 2.4 *Ein Tripel aus einem statistischen Experiment $\mathscr{E}$, einem Entscheidungsraum $(D, \mathfrak{D})$ und einer Verlustfunktion $L(\theta, \delta(x))$ nennen wir statistisches Modell.*

Im Rahmen eines solchen Modells S können wir nun ein Kriterium zur Bewertung möglicher Entscheidungsfunktionen δ angeben.

Definition 2.5 *Die Risikofunktion eines Schätzers δ ist gegeben durch*

$$\begin{aligned} R_\delta(\theta) &:= \mathrm{E}_\theta\left[L(\theta, \delta)\right] \\ &\quad \textit{und falls, wie in unserem Fall } \mathfrak{X} = \mathbb{R}^n \\ &= \int_{\mathbb{R}^n} L(\theta, \delta(x))\, dF_\theta(x) \end{aligned}$$

Mit der Risikofunktion können wir nun Elemente $\delta_1, \delta_1 \in V$ vergleichen, wobei V die Menge der möglichen Entscheidungsverfahren zum Modell S ist. Wenn $R_{\delta_1}(\theta) \leq R_{\delta_2}(\theta)$ für alle $\theta \in \Theta$ sagen wir, dass δ_1 mindestens so gut ist wie δ_2. Gilt die Ungleichung sogar strikt für wenigstens ein θ heißt δ_1 besser oder weniger riskant als δ_2. Eine Entscheidungsfunktion die mindestens so gut wie alle anderen $\delta \in V$ ist nennen wir gleichmäßig beste Entscheidungsfunktion.

Dies liefert uns allerdings allenfalls eine Teilordnung auf der Menge der Entscheidungsfunktionen. So können unterschiedliche Verfahren sozusagen in verschiedenen Bereichen von Θ ihre Stärken haben. Also etwa $R_{\delta_1}(\theta_1) < R_{\delta_2}(\theta_1), R_{\delta_1}(\theta_2) > R_{\delta_2}(\theta_2)$ für $\theta_1, \theta_2 \in \Theta, \theta_1 \neq \theta_2$ Das schließt einen direkten Vergleich vermöge des eben beschriebenen Kriteriums aus. Im Allgemeinen wird auch aus einem ähnlichen Grund kein in diesem Sinne bestes Verfahren zu finden sein, wie wir es aber für die Prämienkalkulation anstreben.

2.2 Der Bayes-Ansatz

Der Ansatz von Bayes dient dem Zweck, dennoch zu einer vollständigen Ordnung auf der Menge der Entscheidungsfunktionen zu kommen. Dazu wird ein gewichteter Mittelwert des Risikos eines Entscheidungsverfahrens δ über ganz Θ gebildet. Mit Verweis auf Definition 1.4 wollen wir die die Wahrscheinlichkeitsverteilung auf Θ, die wir zur Gewichtung verwenden, mit $U(\theta)$ bezeichnen. In ihr kommt ein Vorwissen über das Auftreten verschiedener Parameterwerte zum Ausdruck, weshalb $U(\theta)$ auch als a-priori Verteilung bezeichnet wird. Die A-priori-Verteilung beinhaltet u.a. Aussagen über erwartete Durchschnittswerte wie $\mathrm{E}[g(\Theta)] = \int_\Theta \delta(\theta)\ dU(\theta)$. Den so gebildeten gewichteten Mittelwert bezeichnen wir als

Definition 2.6 *Das Bayes-Risiko einer Entscheidungsfunktion δ zur A-priori-Verteilung $U(\theta)$*

$$R(\delta) = \int\limits_\Theta R_\delta(\theta)\ dU(\theta) \tag{2.3}$$

Darüber hinaus heißt eine Entscheidungsfunktion δ^* Bayes-Verfahren, wenn es das Bayes-Riskio über alle Entscheidungsfunktionen minimiert, d.h. wenn gilt

$$R(\delta^*) \leq R(\delta)\ \forall\ \delta \in V \tag{2.4}$$

Damit haben wir also ein Optimalitätskriterium, das uns bei Kenntnissen über die vorliegenden Verteilungen erlaubt, einen besten Schätzer zu bestimmen. Mit einigen weiteren Details gelangen wir sogar zu einer Konstruktionsvorschrift für das Bayes-Verfahren. P sei die gemeinsame Verteilung von Θ und X. Als Pendant zur A-priori-Verteilung $U(\theta)$ erhalten wir nach Tätigen einer Beobachtung $X = x$ die a-posteriore Verteilung $U_x(\theta) = P^{\Theta|X=x}$. Ferner bezeichnet F_θ die Verteilung von X im Fall von $\Theta = \theta$. Die Randverteilung X, also die unbedingte Verteilung $F = F_\Theta = P^X$ ist gegeben durch

$$F(\cdot) = \int_\Theta F_\theta(\cdot)\; U(d\theta) = \mathrm{E}\left[F_\Theta(\cdot)\right] \tag{2.5}$$

Damit und mit dem Satz von Fubini können wir das Bayes-Risiko eines Schätzers δ in einer Weise umschreiben, die uns Informationen über das Aussehen des Bayes-Schätzers bereitstellt. Wir beschränken uns von nun an auf den bereits oben erwähnten Fall $\mathfrak{X} = \mathbb{R}^n$, da er für die weitere Betrachtung der maßgebliche ist.

$$\begin{aligned} R(\delta) &= \int_\Theta R_\delta(\theta)\; U(d\theta) \\ &= \int_\Theta \int_{\mathbb{R}^n} L(\theta, \delta(x))\; F_\theta(dx)\; U(d\theta) \\ &= \int_{\mathbb{R}^n} \int_\Theta L(\theta, \delta(x))\; U_x(d\theta)\; F_\Theta(dx) \\ &= \int_{\mathbb{R}^n} \mathrm{E}\left[L(\Theta, \delta(x))|X = x\right]\; F_\Theta(dx) \end{aligned}$$

Dabei bezeichnet $\int_\Theta L(\theta, \delta(x)) \;\; U_x(d\theta) = \mathrm{E}\left[L(\Theta, \delta(x))|X = x\right]$ das A-posteriori Risiko eines Schätzers, das wir erst nach Erhalt der Beobachtung $X = x$ berechnen können. Aus der Gleichungskette folgt unmittelbar, welche Eigenschaft der Bayes-Schätzer in Bezug auf dieses Risiko haben muss:

Satz 2.7 *Bayes-Schätzer ist dasjenige δ^*, welches das A-posteriori-Risiko, also das Bayes-Risiko bezüglich der A-posteriori-Verteilung U_x, für jede Beobachtung x minimiert.*

Als Beispiel zur Definition 2.3 hatten wir bereits die quadratische Verlustfunktion $L(\theta, \delta) = (g(\theta) - \delta(x))^2$. Setzen wir sie in die Formel für das Bayes-Risiko ein können wir die oben eingeführte Ordnung auf den Schätzern präzisieren.

Definition 2.8 *Das quadratische Risiko oder auch der quadratische Verlust eines Schätzers δ ist gegeben durch $\mathrm{E}\left[(\delta(\Theta) - g(\Theta))^2\right]$. Ein Schätzer ist mindestens so gut wie ein anderer, wenn sein quadratisches Risiko nicht größer ist als das des anderen.*

Genau wie oben verwenden wir auch die Formulierungen, dass ein Schätzer besser ist als andere. Ein Schätzer, der mindestens so gut ist wie alle anderen, heißt bester Schätzer.

Das bringt uns auch zu einer besseren Aussage über den Bayes-Schätzer:

Satz 2.9 *Bei quadratischer Verlustfunktion ist der Bayes-Schätzer gegeben durch*

$$\widetilde{\delta} = \widetilde{\delta(\Theta)} = \mathrm{E}[\, g(\Theta)\,|X] \tag{2.6}$$

Beweis Satz 2.9: Definitionsgemäß müssen wir zeigen, dass der Bayes-Schätzer $\widetilde{\delta(\Theta)}$ mindestens so gut ist wie alle anderen. Dafür betrachten wir das Bayes-Risiko eines beliebigen Schätzers $\widehat{\delta(\Theta)}$, der ein θ-Funktional $g(\theta)$ schätzen soll.

$$\begin{aligned}
\mathrm{E}\left[\left(\widehat{\delta(\Theta)} - g(\Theta)\right)^2\right] &= \mathrm{E}\left[\left(\widehat{\delta(\Theta)} - \widetilde{\delta(\Theta)} + \widetilde{\delta(\Theta)} - g(\Theta)\right)^2\right] \\
&= \mathrm{E}\left[\mathrm{E}\left[\left(\widehat{\delta(\Theta)} - \widetilde{\delta(\Theta)} + \widetilde{\delta(\Theta)} - g(\Theta)\right)^2 \middle| X\right]\right] \\
&= \mathrm{E}\left[\mathrm{E}\left[\left(\widehat{\delta(\Theta)} - \widetilde{\delta(\Theta)}\right)^2 \middle| X\right]\right] \\
&\quad + \underbrace{\mathrm{E}\left[2\mathrm{E}\left[\left(\widehat{\delta(\Theta)} - \widetilde{\delta(\Theta)}\right)\left(\widetilde{\delta(\Theta)} - g(\Theta)\right) \middle| X\right]\right]}_{=2\mathrm{E}\left[\mathrm{E}\left[\widehat{\delta(\Theta)}\mathrm{E}[g(\Theta)|X] - \widehat{\delta(\Theta)}g(\Theta) - \mathrm{E}[g(\Theta)|X]^2 + \mathrm{E}[g(\Theta)|X]g(\Theta) \middle| X\right]\right] = 0} \\
&\quad + \mathrm{E}\left[\mathrm{E}\left[\left(\widetilde{\delta(\Theta)} - g(\Theta)\right)^2 \middle| X\right]\right] \\
&= \underbrace{\mathrm{E}\left[\left(\widehat{\delta(\Theta)} - \widetilde{\delta(\Theta)}\right)^2\right]}_{\geq 0} + \underbrace{\mathrm{E}\left[\left(\widetilde{\delta(\Theta)} - g(\Theta)\right)^2\right]}_{\text{quadr. Verl. des Bayes-Schätzers}}
\end{aligned}$$

□

Wir haben nun die Optimalität des Bayes-Schätzers im Sinne von Definition 2.8 gezeigt. Es bleibt noch, die Güte des Bayes-Schätzers anhand seines quadratischen Verlustes anzugeben.

Satz 2.10 *Der quadratische Verlust eines Bayes-Schätzers beträgt*

$$\mathrm{E}\left[\left(\widetilde{\delta(\Theta)} - g(\Theta)\right)^2\right] = \mathrm{E}\left[\mathrm{Var}\left[\widetilde{\delta(\Theta)}\middle| X\right]\right]. \tag{2.7}$$

Beweis Satz 2.10: Die Aussage folgt aus $g(\Theta) = \mathrm{E}\left[\mathrm{E}\left[\widetilde{\delta(\Theta)}\middle| X\right]\right]$ und der Definition der Varianz.

□

2.3 Prämienberechnung mit dem Bayes-Ansatz

Mit Satz 2.9 haben wir gezeigt, dass die in Definition 1.5 beschriebene Bayes-Prämie tatsächlich ein im Sinne der Bayes-Statistik optimaler Schätzer für die individuelle Prämie unter Heranziehung des Erfahrungsvektors X ist. In Tabelle 2.1 werden eine Reihe von Begriffen aus Kapitel 1 mit ihren jeweiligen Pendants aus der bayesianischen Statistik identifiziert.

Versicherungswesen (Kap.1)	Bayes-Statistik (Kap.2)
Θ Menge der möglichen Risikoprofile	Θ Parametermenge
θ Risikodisposition eines Versicherungsnehmers	θ gezogener Parameter einer Verteilungsfamilie F
$U(\theta)$ Strukturfunktion	$U(\theta)$ A-priori-Verteilung
$P^{ind}(\theta) = \mu(\theta)$ korrekte individuelle Prämie	$g(\theta)$ zu schätzendes Funktional
$P^{Bayes} = \widetilde{\mu(\Theta)}$ die Bayes-Prämie	$\widetilde{\delta} = \widetilde{\delta(\Theta)}$ Bayes-Schätzer
$P^{koll} = \int_\Theta \mu(\theta)\ dU(\theta) =: \mu_0$ kollektive (Durchschnitts-) Prämie	$\mathrm{E}[g(\Theta)]$ im Mittel erwarteter Wert des θ-Funktionals $g(\theta)$

Tabelle 2.1: Begriffe aus dem Versicherungswesen in der Sprache der Bayes-Statistik

Die Einbettung der versicherungsmathematischen Terme in die bayesianische Statistik können wir nun zum Beispiel nutzen, um die Qualität verschiedener Prämien miteinander zu vergleichen. So ergibt sich durch Varianzzerlegung der quadratische Verlust der kollektiven Prämie mit

$$\begin{aligned}\mathrm{E}\left[(\mu_0 - \mu(\Theta))^2\right] &= \mathrm{E}\left[(-(\mu(\Theta) - \mu_0))^2\right] = \mathrm{Var}[\mu(\Theta)] \\ &= \mathrm{E}[\mathrm{Var}[\mu(\Theta)|X]] + \mathrm{Var}[\mathrm{E}[\mu(\Theta)|X]]\end{aligned}$$

Man beachte, dass der erste Teil der Zerlegung dem quadratischen Verlust der Bayes-Prämie entspricht, d.h. dem Verlust, der auch beim Anwenden des besten Schätzers auftritt. Die zweite Komponente beinhaltet den Teil des Verlustes, der durch die a priori vermutete Varianz der Strukturfunktion entsteht, also dadurch, dass wir mit der konstanten P^{koll} eine zu grobe Näherung eingehen.

Zum Abschluss sei noch eine weitere Eigenschaft des Bayes-Schätzers bei quadratischer Verlustfunktion erwähnt, nämlich, dass er im Mittel kostendeckend ist. Es sei $C \in \mathscr{A} = \sigma(X_1, X_2, \dots X_J)$ ein beobachteter Schadenverlauf oder allgemeiner, eine messbare Menge aus der σ-Algebra der beobachtbaren Ereignisse. Dann gilt:

$$\begin{aligned} \int\limits_{\Theta}\int\limits_{C} \widetilde{\mu(\theta)}\; F_\theta(dx)\; U(d\theta) &= \int\limits_{\Theta}\int\limits_{C} \mathrm{E}\left[\, \mu(\theta) \,|X = x\right]\; F_\theta(dx)\; U(d\theta) \\ &= \int\limits_{\Theta}\int\limits_{C} \mu(\theta)\; F_\theta(dx)\; U(d\theta) \end{aligned} \tag{2.8}$$

Wenn wir uns das Ereignis C als die Schadenverläufe einer Menge von Individuen vorstellen, repräsentiert die linke Seite von Gleichung (2.8) die Prämiensumme, die wir aufgrund dieser Erfahrungen erheben würden. Die rechte Seite hingegen stellt den zu erwartenden Gesamtschaden einer Menge von Individuen mit der in C beschriebenen Schadenhistorie da. Das bedeutet, die korrekte individuelle Prämie ist nicht nur optimal im Sinne der Bayes-Statistik, sie ist auch erwartungstreu in der Weise, dass sie genau in Höhe des erwarteten Schadens anfällt.

2.4 Die Bayes-Prämie im Poisson-Gamma-Modell

In diesem Abschnitt wollen wir die Bayes-Prämie (d.h. den Bayes-Schätzer) in einem konkreten Beispiel berechnen. Wir betrachten dazu das von Bichsel [Bic64] eingeführte Modell aus der KFZ-Haftpflicht-Versicherung. Modelliert werden soll die Häufigkeit von Schadenforderungen durch Individuen. Vereinfachend wird davon ausgegangen, dass die zu erwartenden Schadenkosten X_j eines beliebigen Individuums im Jahr j von zwei Faktoren abhängen:

- Von der Anzahl N_j der Schadenfälle im Jahr j

- einer Konstanten C, die nur von der Leistung des PKW abhängt

so, dass sich die Schadensumme im Jahr j ergibt durch $X_j = CN_j$. In der Häufigkeit N_j spiegelt sich also die Risikodisposition eines Individuums wieder. Da die Schadenkosten und Prämien linear in der Häufigkeit der Schadenfälle sind, stehen diese im Zentrum unserer Betrachtung. Wir wollen N_j durch eine poisson-verteilte Zufallsvariable mit Parameter θ modellieren. Die individuellen Risikoprofile θ erhalten wir durch die gammaverteilte Zufallsvariable Θ.

Damit kommen wir zu folgenden Modellannahmen:

Annahme 2.11 (Poisson-verteilte Schadenhäufigkeiten) *Bedingt auf $\theta = \Theta$ sind die N_j unabhängig und identisch verteilt. Ihre Verteilung $P^{N_j} = F_\theta$ ist eine Poisson-Verteilung zum Parameter θ, das heißt*

$$P\left(N_j = k \middle| \theta = \Theta\right) = e^{-\theta} \frac{\theta^k}{k!} \tag{2.9}$$

Annahme 2.12 (Gamma-Verteilung als Strukturfunktion) *Die Strukturfunktion $U(\theta)$ des Kollektivs, also die Verteilung der Zufallsgröße Θ hat für $\theta > 0$ die Dichte*

$$u(\theta) = \frac{\beta^\gamma}{\Gamma(\gamma)} \theta^{\gamma-1} e^{-\beta\theta} \tag{2.10}$$

ist also Gamma-verteilt mit Skalenparameter β und Formparameter γ.

In Abbildung 2.1 wird die Zweistufigkeit des Modells noch einmal erläutert. Die erste Zeile stellt die Ziehungen aus der ersten Urne dar. Es werden also entsprechend der Gammaverteilung ($\alpha = 2$, $\beta = 0,5$) die Risikodispositionen $L1, L2, L3$ der Individuen 1, 2 und 3 gezogen. In der zweiten Zeile sind dann die Verteilungen der Schadenhäufigkeiten der einzelnen Individuen durch ihre Dichtefunktionen dargestellt. Für jedes Individuum wurden Zufallszahlen für drei Jahre gezogen.

Erwartungswert und Varianz der betrachteten Zufallsgrößen N_j (Poisson-verteilt) und Θ (Gamma-verteilt) sind

$$\begin{aligned} \mathrm{E}\left[N_j \middle| \theta = \Theta\right] &= \theta & \qquad \mathrm{Var}\left[N_j \middle| \theta = \Theta\right] &= \theta \\ \mathrm{E}\left[\Theta\right] &= \frac{\gamma}{\beta} & \qquad \mathrm{Var}\left[\Theta\right] &= \frac{\gamma}{\beta^2} \end{aligned}$$

Daraus erhalten wir unmittelbar:

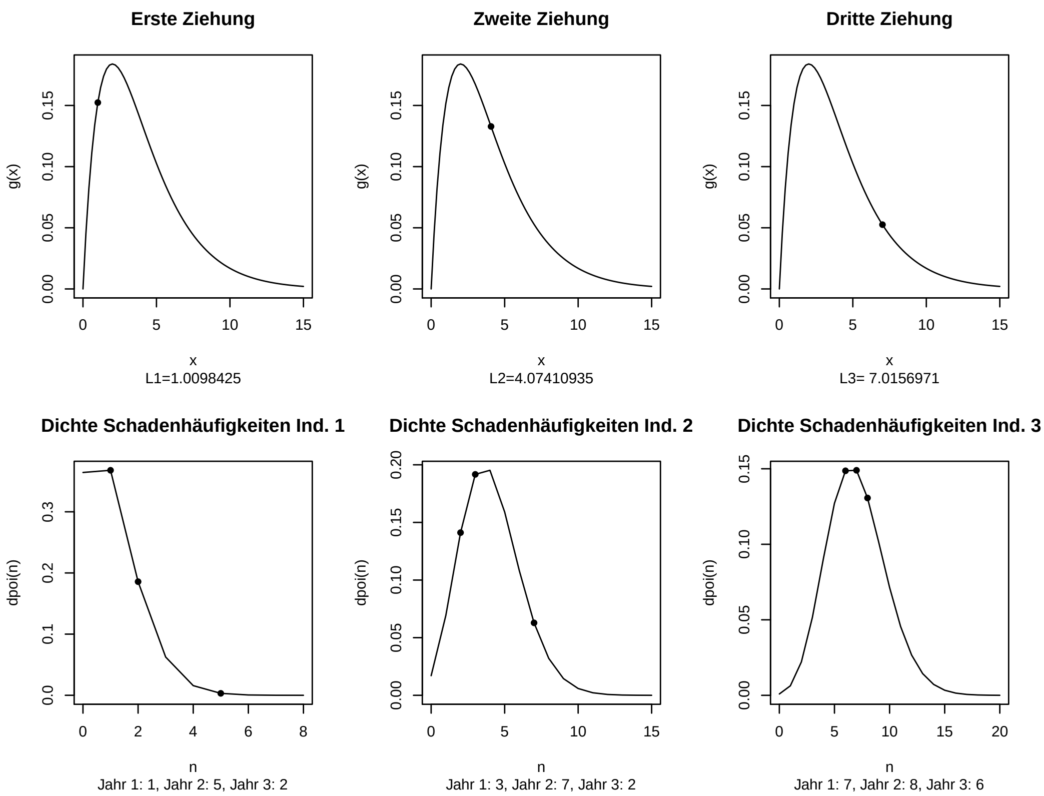

Abbildung 2.1: Das Poisson-Gamma Modell

Satz 2.13 *Mit den Annahmen 2.11 und 2.12 ergeben sich für das Modell von Bichsel die individuelle und kollektive Prämie folgendermaßen:*

$$P^{ind} = \mathrm{E}\left[N_j \middle| \Theta\right] = \Theta \tag{2.11}$$

$$P^{koll} = \mathrm{E}\left[\Theta\right] = \frac{\gamma}{\beta} \tag{2.12}$$

Um die Bayes-Prämie berechnen zu können benötigen wir die A-posteriori-Verteilung von Θ. Ihre Dichte erhalten wir unter Verwendung des folgenden Satzes, der der Wahrscheinlichkeitstheorie entstammt:

Satz 2.14 *Sei Θ eine Zufallsvariable auf $\mathbb{R}$ deren Verteilung $U(\theta)$ die Lebesque-Dichte $u(\theta)$ besitzt. Sei ferner $(W_\theta)_{\theta\in\Theta}$ eine Familie von Zufallsvariablen auf $(\mathbb{R}^n, \mathfrak{B}^n)$ mit dominierendem, σ-endlichen Maß ν und ν-Dichten $f_\theta = \frac{dW_\theta}{d\nu}$. Darüber hinaus ist f_u die Dichte der unbedingten Verteilung W_Θ, wie* *sie in Gleichung (2.5) beschrieben wurde. Dann gilt:*

$$u(\theta|X=x) = \frac{u(\theta)f_\theta(x)}{f_u(x)}\mathbf{1}_{\{f_u>0\}} + u(\theta)\mathbf{1}_{\{f_u=0\}} \tag{2.13}$$

Zur Notation sei noch angemerkt, dass wir mit $N_\bullet$ die Summation von N über den durch einen Punkt ersetzten Index bezeichnen. Nach einer Beobachtung $N = (N_1, N_2, \ldots, N_J)$ ergibt sich also die Dichte der A-posteriori-Verteilung von Θ (die nur auf $\mathbb{R}^+$ positive Werte annimmt) folgendermaßen:

$$u(\theta|N) = \frac{\frac{\beta^\gamma}{\Gamma(\gamma)}\theta^{\gamma-1}\prod_{j=1}^{J} e^{-\theta}\frac{\theta^{N_k}}{N_k!}}{\int_\Theta \prod_{j=1}^{J} e^{-\theta}\frac{\theta^{N_k}}{N_k!}u(\theta)d\theta} = c\;\theta^{\gamma+N_\bullet-1}e^{-(\beta+J)\theta} \tag{2.14}$$

wobei c eine Konstatnte ist, die nicht von θ abhängt. Mit Satz 2.14 wissen wir nun sofort, dass es sich bei $u(\theta|N)$ wieder um die Dichte einer Gamma-Verteilung handelt, nun aber zu den Parametern

$$\gamma' = \gamma + N_\bullet = \gamma + \sum_{j=1}^{J} N_j$$
$$\beta' = \beta + J$$

Durch Ausrechnen von

$$\mathrm{E}[\Theta|N] = \int_\Theta \theta u(\theta|N)\,d\theta \tag{2.15}$$

also dem Erwartungswert einer Gamma-verteilten Zufallsvariable mit Skalenparameter β' und Formparameter γ' erhalten wir

Satz 2.15 *Im beschriebenen Modell von Bichsel ist die Bayes-Prämie nach Tätigen der Beobachtung N gegeben durch*

$$F^{Bayes} = \frac{\gamma + N_\bullet}{\beta + J} \tag{2.16}$$

Ihr quadratischer Verlust beträgt

$$\begin{aligned}\mathrm{E}\left[\left(F^{Bayes}-\Theta\right)^2\right] &= (1-\alpha)\,\mathrm{E}\left[\left(F^{koll}-\Theta\right)^2\right] \\ &= \alpha\,\mathrm{E}\left[\left(\overline{N}-\Theta\right)^2\right]\end{aligned} \tag{2.17}$$

Beweis Satz 2.15: Die Herleitung von (2.16) haben wir oben bereits durchgeführt, es bleibt noch die Aussagen über den quadratischen Verlust zu beweisen. Die ersten beiden Schritte der folgenden Gleichungskette folgen aus der Definition der Bayes-Prämie und ihrer Verteilungseigenschaft.

$$\begin{aligned}
\mathrm{E}\left[\left(F^{Bayes}-\Theta\right)^2\right] &= \mathrm{E}\left[\mathrm{Var}\left[\Theta|N\right]\right] \\
&= \mathrm{E}\left[\frac{\gamma+N_\bullet}{(\beta+J)^2}\right] \\
&= \frac{\gamma}{(\beta+J)^2}\left(1+\frac{J}{\beta}\right) \quad \text{, mit} \quad \frac{J}{\beta}=\gamma\,\mathrm{E}N_\bullet \\
&= \frac{\beta\gamma}{\beta\,(\beta+J)^2}+\frac{\gamma J}{\beta\,(\beta+J)^2} \\
&= \frac{1}{J}\frac{\gamma}{\beta}\alpha \quad \text{, da} \quad \frac{1}{J}\alpha=\frac{1}{J+\beta} \\
&\text{und mit } \alpha=\frac{J}{\beta}(1-\alpha)\text{schließlich} \\
&= \frac{\gamma}{\beta^2}(1-\alpha)
\end{aligned}$$

Für (2.17) wissen wir bereits

$$\mathrm{E}\left[\left(F^{koll}-\Theta\right)^2\right]=\mathrm{Var}\left[\Theta\right]=\frac{\gamma}{\beta^2}$$

und für $\overline{N}$ folgt es aus

$$\begin{aligned}
\mathrm{E}\left[\left(F^{koll}-\overline{N}\right)^2\right] &= \mathrm{E}\left[\mathrm{Var}\left[\overline{N}\middle|\Theta\right]\right]=\frac{1}{n^2}\mathrm{E}\left[\mathrm{Var}\left[\sum_{j=1}^{J}N_j\middle|\Theta\right]\right] \\
&= \frac{1}{n^2}\mathrm{E}\left[n\,\mathrm{Var}\left[N_1\,|\,\Theta\right]\right]=\frac{1}{n}\frac{\gamma}{\beta}
\end{aligned}$$

□

Wir können den Bayes-Schätzer aus (2.16) noch wie folgt zerlegen:

$$\frac{\gamma+N_\bullet}{\beta+J}=\frac{J}{\beta+J}\frac{N_\bullet}{J}+\frac{\beta}{\beta+J}\frac{\gamma}{\beta}=\alpha\overline{N}+(1-\alpha)\frac{\gamma}{\beta} \tag{2.18}$$

also $\alpha=\frac{J}{\beta+J}$. Diese Darstellung gibt uns einen Eindruck über die Funktionsweise des Bayes-Schätzers als gewichtetes Mittel zwischen individuellem Schadenverlauf und dem kollektiven Mittelwert.

Desto länger das Individuum bereits bekannt ist (gemessen in Jahren J), desto stärker wird der individuelle Schadenverlauf (oder, wie hier davon abgeleitet: die durchschnittliche Schadenhäufigkeit $\overline{N}$) gewichtet. Ist nur wenig über den einzelnen Versicherungsnehmer bekannt bleibt α vorerst klein und das Kollektivmittel beeinflußt die Prämie maßgeblich. Ebenso lässt sich der Einfluss von $\beta = \mathrm{E}[\Theta]/\operatorname{Var}[\Theta]$ veranschaulichen. Sind die individuellen Risikoprofile θ weniger breit gestreut, ist also $Var[\Theta]$ klein, können wir uns eher auf die durchschnittlich zu erwartende Schadenhäufigkeit verlassen. Unterscheiden sich die Risikoprofile hingegen deutlich voneinander, müssen wir den beobachteten Schadenverläufen mehr Gewicht verleihen. Da $\operatorname{Var}(N_1|\Theta) = \Theta$ zeigt uns die Gewichtung mit α drittens, dass mit steigender Varianz innerhalb der individuellen Schadenhäufigkeiten bei Berechnung der Bayes-Prämie die individuelle Erfahrungskomponente an Bedeutung verliert. Bichsel macht in seinem Artikel[1] am Rande darauf aufmerksam, dass die hier gefundene Bayes-Prämie im Poisson-Gamma-Modell seiner Einschätzung nach die einfachste Möglichkeit darstellt, die eben beschriebenen plausiblen Zusammenhänge in einem Term zusammenzufassen.

Wie die Bayes-Prämie als gewichtetes Mittel funktioniert lässt sich gut anhand der Werte aus Tabelle 2.2 verstehen. Die Bayes-Prämie dämpft sozusagen den Einfluss der Daten auf den Schätzwert. Daher liegt sie auch näher am Kollektivmittel $\frac{\gamma}{\beta} = 4$. Im vorliegenden Fall sind die Werte aus Ziehung 1 und 3 aus dem unteren 10-%-Quantil (1) bzw. aus dem oberen 15-%-Quantil der Strukturfunktion. Sie bilden also gewissermaßen Randwerte, deren Auftreten im Sinn unwahrscheinlich ist. Daher weicht der Bayes-Schätzer auch deutlich von $\overline{N}$ in Richtung des Erwartungswertes der Strukturfunktion ab.

Individuum	1	2	3
$P^{ind} = \theta$	$1,0$	$4,1$	$7,0$
$\overline{N}$	$2\frac{2}{3}$	4	7
P^{Bayes}	$2\frac{6}{7}$	4	$6\frac{4}{7}$

Tabelle 2.2: Schätzwerte im Poisson-Gamma Model

Man beachte, dass $P^{ind} = \theta$ aus der ersten Zeile der aus den Daten zu schätzende Wert ist. Dass die Bayes-Prämie deutlich $\overline{N}$ Abweicht bei Individuum 1 und 3 liegt auch daran, dass die Werte im unteren 10%-Quantil (1) bzw. im oberen 15%-Quantil der Verteilung der Strukturfunktion liegen.

[1] [Bic64], S.122

Die Voraussetzung, dass die Strukturfunktion bekannt ist, stellt eine restriktive Annahme dar. Sie ist noch praxisferner, wenn wir, wie oben geschehen, die Parameter dieser Verteilung als bekannt annehmen. Um diesem Problem zu entgehen, soll kurz dargestellt werden, wie das Problem von Bichsel bei unbekannten Parametern gelöst werden kann. Dafür wollen wir die Parameter der Strukturfunktion $U(\theta)$, die weiterhin als Gamma-verteilt vorrausgesetzt sein soll, mithilfe der Momenten-Methode schätzen. Dafür benötigen wir natürlich Informationen über das gesamte Kollektiv.

Es seien $(N_i)_{i=1,\dots,I}$ Zufallsvariablen, die die Forderungshäufigkeiten der Individuen $i = 1,\dots,I$ in einer Periode beschreiben. Da hier nur eine Periode betrachtet wird verzichten wir auf die obige Indizierung durch $j, j = 1,\dots,J$ für die Perioden. Entsprechend Annahme 2.11 können wir die Paare von Zufallsvariablen $\{(\Theta_i, N_i), i = 1,\dots,I\}$ als unabhängig und identisch verteilt betrachten. Daraus können wir die ersten beiden Momente der N_i wie folgt berechnen:

$$\mu_N = \mathrm{E}[N_1] = \mathrm{E}[\mathrm{E}[N_1|\Theta]] = \mathrm{E}[\Theta] = \frac{\gamma}{\beta} \tag{2.19}$$

$$\begin{aligned}\sigma_N^2 &= \mathrm{Var}[N_1] = \mathrm{E}[\mathrm{Var}[N_1|\Theta]] + \mathrm{Var}[\mathrm{E}[N|\Theta]] \\ &= \mathrm{E}[\Theta] + \mathrm{Var}[\Theta] = \frac{\gamma}{\beta} + \frac{\gamma}{\beta^2}\end{aligned} \tag{2.20}$$

Als erwartungstreue Schätzer für die ersten beiden Momente μ_N und σ_N^2 haben wir den empirischen Mittelwert und die empirische Varianz

$$\widehat{\mu}_N = \overline{N}_I = \frac{1}{n}\sum_{i=1}^{I} N_i$$

$$\widehat{\sigma}_N^2 = \frac{1}{I-1}\sum_{i=1}^{I}\left(N_i - \overline{N}_N\right)^2$$

Durch das Lösen des Gleichungssystems

$$\begin{aligned}\widehat{\mu}_N &= \mu_N \\ \widehat{\sigma}_N^2 &= \sigma_N^2\end{aligned} \tag{2.21}$$

kommen wir zu empirischen Parameterwerten mit denen wir die Auswertung fortsetzen können. Das Schätzen der Strukturparameter für einen allgemeineren Fall (Bühlmann-Straub-Modell) betrachten wir näher in Abschnitt 4.4.

Kapitel 3

Credibility I: Grundlegende Modelle

Im Abschnitt 2.4 des vorigen Kapitels haben wir uns einen Eindruck davon verschafft, wie unter gewissen Voraussetzungen optimale Schätzer für die individuelle Prämie eines Versicherungsnehmers gewonnen werden können. Mit dem Ziel, sich der Praxis weiter anzunähern, werden wir die Voraussetzungen nun schrittweise vereinfachen. Indem wir uns auf empirisch einfach zu erfassende Größen wie Erwartungswert und Varianz konzentrieren und genaue Verteilungseigenschaften wie etwa die Dichtefunktion in den Hintergrund, stellen können wir zu einer sogenannten Credibility-Prämie kommen. Sie zeichnet sich durch eine gewisse Einfachheit der Formel aus. Dafür geht uns bisweilen ein globales Optimalitätskriterium verloren, wie es bei der Bayes-Prämie vorliegt. Um dennoch zu bewertbaren Ergebnissen zu kommen legen wir uns, im Gegenzug zur Verallgemeinerung bei den Verteilungsannahmen, bei der Auswahl der Schätzer eine Beschränkung auf. Es sollen nur noch solche Schätzer betrachtet werden, die linear (genauer, affin linear) in den Daten sind. Das bedeutet ein Schätzer $\widehat{\mu(X)}$, der Daten über ein Individuum aus den Jahren $J = 1, \ldots, J$ verarbeitet hat die Form

$$\widehat{\mu(X)} = a_0 + \sum_{j=1}^{J} a_j X_j, \quad a_0, \ldots, a_J \in \mathbb{R} \tag{3.1}$$

3.1 Ein einfaches Credibility-Modell

Für das nun zu betrachtende Credibility-Modell sollen die folgenden beiden Annahmen gelten:

Annahme 3.1 (Bedingte Schadenverteilungen) *Bedingt auf $\Theta = \theta$ sind X_j unabhängig und identisch verteilte Zufallsvariablen. Sie beschreiben den von einem Individuum in der Periode j produzierten Schaden. Sie haben die Verteilung F_θ mit den bedingten ersten und zweiten Momenten*

$$\mu(\theta) = \mathrm{E}\left[X_j|\Theta = \theta\right]$$
$$\sigma(\theta) = \mathrm{Var}\left[X_j|\Theta = \theta\right]$$

Annahme 3.2 (Strukturfunktion) *Θ ist eine Zufallsvariable die gemäß $U(\theta)$ verteilt ist.*

Verglichen mit den Annahmen 2.11 und 2.12 aus dem vorherigen Kapitel wird die Verallgemeinerung deutlich. Während dort noch genaue Angaben zum Beispiel zu Dichtefunktionen von F_θ und $U(\theta)$ gemacht wurden wird sich hier auf die Existenz von Erwartungswert und Varianz beschränkt. Es ist klar, dass die versicherungsmathematische Betrachtung nur Sinn ergibt, wenn diese endlich sind. Wie gehabt setzen wir

$$P^{\text{ind}} = \mu(\theta) = \mathrm{E}\left[N_{J+1}|\Theta\right]$$
$$P^{\text{koll}} = \mu_0 = \int_\Theta \mu(\theta)\, U(d\theta)$$

und suchen nach einem Schätzer für P^{ind} der aber von vorneherein festgelegt die Form (3.1) haben soll. Den besten linearen Schätzer $\widehat{\widehat{\mu(X)}}$ nennen wir Credibility-Schätzer (auch P^{cred}). Eine förmliche Definition erfolgspäter (Def 4.1). Für den folgenden Satz genügt aber das eben dargestellte Verständnis.

Satz 3.3 *Im von den Annahmen 3.1 und 3.2 bestimmten Modell hat der Credibility-Schätzer folgende Form:*

$$\widehat{\widehat{\mu(X)}} = \alpha\overline{X} + (1-\alpha)\mu_0 \tag{3.2}$$

dabei ist $\mu_0 = \mathrm{E}\left[\mu(\Theta)\right]$ das Kollektivmittel und wir nennen

$$\alpha = \frac{J}{J + \frac{\sigma^2}{\tau^2}} \tag{3.3}$$

Credibility-Gewicht.

Beweis Satz 3.3: Den Credibility-Schätzer als im Bayes´schen Sinne besten linearen Schätzer bei quadratischer Verlustfunktion erhalten wir durch Lösen des folgenden Minimierungsproblems für die Schätzkoeffizienten $\widehat{a}_0, \widehat{a}_1, \ldots, \widehat{a}_J$:

$$\mathrm{E}\left[\left(\mu(\theta) - \widehat{a}_0 - \sum_{j=1}^{J} \widehat{a}_j X_j\right)^2\right] = \min_{a_0,\ldots,a_J \in \mathbb{R}} \mathrm{E}\left[\left(\mu(\theta) - a_0 - \sum_{j=1}^{J} a_j X_j\right)^2\right]$$

Da die X_j identisch verteilt sind, ist ihre Reihenfolge ohne Bedeutung. Daher muss auch gelten $\widehat{a}_1 = \widehat{a}_2 = \ldots = \widehat{a}_J$. Der Credibility-Schätzer ist also von der Form

$$\widehat{\mu(\Theta)} = \widehat{a} + \widehat{b}\overline{X}, \text{ wobei wie gehabt } \quad \overline{X} = \frac{1}{J}\sum_{j=1}^{J} X_j$$

Der Erfahrungsvektor X geht hier nur als durchschnittliche individuelle Schadenhistorie $\overline{X}$ ein. Das oben genannte Minimierungsproblem vereinfacht sich also zu

$$\mathrm{E}\left[\left(\mu(\theta) - \widehat{a} - \widehat{b}\overline{X}\right)^2\right] = \min_{a,b \in \mathbb{R}} \mathrm{E}\left[\left(\mu(\theta) - a - b\overline{X}\right)^2\right] \tag{3.4}$$

Zur Lösung dieses Problems bilden wir die partiellen Ableitungen nach a und b und setzen diese gleich 0.

$$\frac{\partial}{\partial a}\left(\mathrm{E}\left[\left(\mu(\theta) - a - b\overline{X}\right)^2\right]\right) = 0 \tag{3.5}$$

$$\frac{\partial}{\partial b}\left(\mathrm{E}\left[\left(\mu(\theta) - a - b\overline{X}\right)^2\right]\right) = 0 \tag{3.6}$$

Aus 3.5 erhalten wir sofort

$$\begin{aligned} \mathrm{E}\left[\mu(\theta) - a - b\overline{X}\right] &= 0 \\ \Leftrightarrow \quad \mathrm{E}[\mu(\theta)]\,\mathrm{E}\left[\overline{X}\right] - a\mathrm{E}\left[\overline{X}\right] - b\mathrm{E}\left[\overline{X}\right]^2 &= 0 \end{aligned} \tag{3.7}$$

durch Multiplikation mit $\mathrm{E}[\overline{X}]$.

Subtrahieren wir nun 3.7 von 3.6:

$$\begin{aligned} \mathrm{E}\left[\overline{X}\left(\mu(\theta) - a - b\overline{X}\right)\right] - \left(\mathrm{E}[\mu(\theta)]\,\mathrm{E}\left[\overline{X}\right] - a\mathrm{E}\left[\overline{X}\right] - b\mathrm{E}\left[\overline{X}\right]^2\right) &= 0 \\ \Leftrightarrow \quad \mathrm{E}\left[\overline{X}\mu(\theta)\right] - \mathrm{E}[\mu(\theta)]\,\mathrm{E}\left[\overline{X}\right] - b\left(\mathrm{E}\left[\overline{X}^2\right] - \mathrm{E}\left[\overline{X}\right]^2\right) &= 0 \\ \Leftrightarrow \quad \mathrm{Cov}\left(\overline{X}, \mu(\Theta)\right) - b\,\mathrm{Var}\left(\overline{X}\right) &= 0 \end{aligned} \tag{3.8}$$

Die Modellannahmen 3.1 und 3.2 erlauben uns folgende Darstellungen für $\mathrm{Var}(\overline{X})$ und $\mathrm{Cov}(\overline{X}, \mu(\Theta))$:

$$\mathrm{Cov}\left(\overline{X}, \mu\left(\Theta\right)\right) = \mathrm{Var}\left(\mu\left(\Theta\right)\right) =: \tau^2$$
$$\mathrm{Var}\left(\overline{X}\right) = \frac{\mathrm{E}\left[\sigma^2\left(\Theta\right)\right]}{J} + \mathrm{Var}\left(\mu\left(\Theta\right)\right) =: \frac{\sigma^2}{J} + \tau^2$$

Das Gleichungssystem bestehend aus (3.7) und (3.8) lässt sich nun ohne weitere Umstände lösen und wir erhalten:

$$b = \frac{\tau^2}{\tau^2 + frac\sigma^2 J} = \frac{J}{n + \frac{\sigma^2}{\tau^2}}$$
$$a = (1-b)\mu_0$$

□

Der mit Satz 3.3 erhaltene Credibility-Schätzer weißt eine analoge Struktur auf zum Bayes-Schätzer aus dem Modell von Bichsel aus Satz 2.15. Steigt die Zahl der beobachteten Perioden (Jahre) J, verlagert sich das Credibility-Gewicht weg vom allgemeinen Durchschnittswert μ_0 zum Erfahrungswert $\overline{X}$. Bichsels $\beta = \mathrm{E}[\Theta]/\mathrm{Var}[\Theta]$ hatte den gleichen Einfluss wie hier $\kappa = \frac{\sigma^2}{\tau^2} = (\sigma/\mu_0)^2(\tau/\mu_0)2$. Die Wirkrichtung von τ/μ_0, bei Größerwerden (wegen der wachsenden Heterogenität des Kollektivs) auch das Credibility-Gewicht zu stärken, entspricht der von $\mathrm{Var}(Theta)$ bei Bichsel. Bei poisson-verteilten Schäden (Schadenhäufigkeiten) war $\mathrm{E}[\Theta]$ auch ein Maß für die Streuung der Schadenhistorie des einzelnen Individuums. Im einfachen Credibility-Modell ist dies nun $\sigma/\mu_0 = \sqrt{\mathrm{E}[\mathrm{Var}(X_j|\Theta = \theta)/]}/\mu_0$ und wirkt ebenfalls in gleicher Weise.

Wenn, wie in Satz 2.15, der Bayes-Schätzer linear ist, also bereits den Credibility-Schätzer darstellt, sprechen wir von exakter Credibility.

Haben wir A-priori-Wissen über Erwartungswert und Varianz der Strukturfunktion und kennen wir zudem die Verteilungsfamilie **F**, zu der die Verteilung der X_i gehört, können wir oft exakte Credibility erreichen. Zum Beispiel ist dies immer der Fall, wenn **F** eine einparametrige Exponentialfamilie ist[1] . In der praktischen Anwendung der Credibility-Theory treten diese Fälle allerdings selten auf.

[1]siehe [BG05] S.38ff.

Wir wollen nun noch die quadratischen Fehler des Credibility-Schätzers P^{cred}, der Kollektivprämie $P^{\text{koll}} = \mu_0$ und der durchschnittlichen individuellen Schadenhistorie $\overline{X}$ berechnen. Für P^{koll} gilt

$$\mathrm{E}\left[(\underbrace{\mu_0}_{=\mathrm{E}[\mu(\Theta)]} - \mu(\Theta))^2\right] = \mathrm{Var}(\mu(\Theta)) = \tau^2$$

Als besten linearen Schätzer $\overline{X}$ für die individuelle Prämie, der gegeben $\Theta = \theta$ auch erwartungstreu ist, haben wir $\overline{X}$, das beobachtete Schadenmittel des Individuums. Sein quadratischer Fehler beträgt

$$\begin{aligned}
\mathrm{E}\left[\left(\overline{X} - \mu(\Theta)\right)^2\right] &= \mathrm{E}\left[\mathrm{E}\left[\left(\frac{1}{J}\sum_{j=1}^{J} X_j - \mu(\Theta)\right)^2 \middle| \Theta\right]\right] \\
&= \mathrm{E}\left[\frac{1}{J^2}\underbrace{\mathrm{E}\left[\left(\sum_{j=1}^{J}\left(X_j - \mu(\Theta)\right)\right)^2 \middle| \Theta\right]}_{=\mathrm{Var}(\sum_{j=1}^{J} X_j)} |\Theta\right] \\
&= \mathrm{E}\left[\frac{1}{J}\mathrm{Var}\left(X_j \middle| \Theta\right)\right] = \mathrm{E}\left[\frac{\sigma^2(\Theta)}{J}\right] = \frac{\sigma^2}{J}
\end{aligned}$$

wobei der letzte Schritt aus der auf Θ bedingten Unabhängigkeit der X_j und ihrer ebenso bedingten identischen Verteilung beruht.

Wir wollen nun anhand der quadratischen Fehler, die den beiden Schätzern zu eigen sind, eine Gewichtung der beiden Verfahren durchführen um zu einem gemischten Schätzer für die individuelle Prämie P^{ind} zu kommen. Es bietet sich an, in der Summe das jeweilige Verfahren antiproportional zu seinem quadratischen Verlust zu gewichten. Für $\overline{X}$ wäre dementsprechend das Gewicht

$$\alpha = \frac{\frac{J}{\sigma^2}}{\frac{J}{\sigma^2} + \frac{1}{\tau^2}} = \frac{J}{J + \frac{\sigma^2}{\tau^2}} \tag{3.9}$$

Dies ist uns aus Gleichung (3.3) bekannt. Gewichten wir entsprechend μ_0 mit $(1-\alpha)$ erhalten wir den Credibility-Schätzer wie in Satz 3.3. Im betrachteten Modell führt uns die eben dargelegte Plausibilitätsüberlegung also direkt zum Credibility-Schätzer.

Das Vorgehen, einem gemischten Schätzer zu konstruieren, indem wir die Komponenten proportional zum Kehrwert ihres quadratischen Verlustes zu gewichten, nennen wir **intuitives Prinzip**.

Kommen wir nun zur Güte des Credibility-Schätzers selbst, berechnet anhand seines quadratischen Fehlers:

$$\begin{aligned}
\mathrm{E}\left[\left(\mu(\Theta)-\widehat{\mu(\Theta)}\right)^2\right] &= \mathrm{E}\left[\left(\alpha\mu(\Theta)+(1-\alpha)\mu(\Theta)-\alpha\overline{X}-(1-\alpha)\mu_0\right)^2\right] \\
&= \mathrm{E}\left[\left(\alpha\left(\mu(\Theta)-\overline{X}\right)+(1-\alpha)\left(\mu(\Theta)-\mu_0\right)\right)^2\right] \\
&= \alpha^2\,\mathrm{E}\left[\left(\mu(\Theta)-\overline{X}\right)^2\right]+(1-\alpha)^2\,\mathrm{E}\left[\left(\mu(\Theta)-\mu_0\right)^2\right]
\end{aligned}$$

wobei wir im letzten Schritt verwendet haben, dass $\mu(\Theta)-\overline{X}$ und $\mu(\Theta)$ unabhängig sind und damit $\mathrm{E}[\mu(\Theta)(\mu(\Theta)-\overline{X})]=0$. Mit den vorangegangenen Rechnungen für die quadratischen Fehler folg dann weiter:

$$\begin{aligned}
&= \alpha^2\frac{\sigma^2}{J}+(1-\alpha)^2\tau^2 = \frac{J\sigma^2+\sigma^2\frac{\sigma^2}{\tau^2}}{(J+\frac{\sigma^2}{\tau^2})} = \frac{\sigma^2}{J+\frac{\sigma^2}{\tau^2}} \\
&= \alpha\frac{\sigma^2}{J} \\
&= (1-\alpha)\tau^2
\end{aligned}$$

Die letzten Ergebnisse lassen sich zusammenfassen zu folgender Aussage:

Satz 3.4 *Unter den gegebenen Modellannahmen beträgt der quadratische Verlust des Credibility-Schätzers* $\widehat{\mu(X)}$

$$\mathrm{E}\left[\left(\mu(\Theta)-\widehat{\mu(X)}\right)^2\right] = \alpha\frac{\sigma^2}{J} = (1-\alpha)\tau^2 \tag{3.10}$$

Insbesondere ist er keinesfalls größer als der quadratische Verlust des kollektiven Mittelwerts

$$\mathrm{E}\left[\left(\mu_0-\mu(\Theta)\right)^2\right] = \tau^2 \tag{3.11}$$

und des beobachteten Schadenmittels

$$\mathrm{E}\left[\left(\overline{X}-\mu(\Theta)\right)^2\right] = \frac{\sigma^2}{J} \tag{3.12}$$

3.2 Das Modell von Bühlmann

Im Modell von Bühlmann wollen wir nun erstmals das gesamte Kollektiv modellieren und nicht nur ein einzelnes versichertes Individuum betrachten. Das Kollektiv besteht aus einer Menge von I Individuen, die alle dem Versicherer seit J Jahren bekannt sind. Für jedes Individuum i gelte die Annahme 3.1. Das heißt, bedingt auf seiner Risikodisposition, die durch die Zufallsvariable Θ_i gegeben ist, sind die jährlichen Schadenforderungen $X_{i1}, X_{i2}, \ldots, X_{iJ}$ unabhängig und identisch verteilt. Die Schadenhistorie eines Individuums i, bestehend aus den (jährlichen) Schadenforderungen $X_{i1}, X_{i2}, \ldots, X_{iJ}$ fassen wir in dem Vektor $X_i = (X_{i1}, X_{i2}, \ldots, X_{iJ})'$ zusammen.

Annahme 3.5 (Bedingte Schadenverteilungen) *Bedingt auf $\Theta_i = \theta$ sind X_{ij} unabhängig und identisch verteilte Zufallsvariablen für $j = 1, \ldots, J$. Sie beschreiben den von einem Individuum i in der Periode j produzierten Schaden. Sie haben die Verteilung F_θ mit den bedingten ersten und zweiten Momenten*

$$\mu(\theta) = \mathrm{E}\left[X_{ij}|\Theta_i = \theta\right]$$
$$\sigma(\theta) = \mathrm{Var}\left[X_{ij}|\Theta_i = \theta\right]$$

Annahme 3.6 (Unabhängigkeit der Individuen im Kollektiv) *Die Zufallsvariablen die jeweils zu einem Individum gehören, also $(\Theta_i, (X_{i1}, \ldots, X_{iJ})') = (\Theta_i, X_i)$ sind für $i = 1, \ldots, I$ unabhängig und identsich verteilt. Die Θ_i haben die Strukturfunktion des Kollektivs $U(\theta)$ als Verteilungsfunktion.*

3.2.1 Der inhomogene Credibility-Schätzer

Wir sind nun auf der Suche nach dem Credibility-Schätzer $\widehat{\widehat{\mu(\Theta_i)}}$, der die Risikodisposition eines Individuums i im Kollektiv schätzt. Er soll wieder der beste lineare Schätzer sein. Doch stehen uns hier neben der Schadenhistorie von i, also X_i', auch noch die Informationen über die anderen Individuen zur Verfügung. Sie liegen vor in Form der Schadenhistorien $\{X_k', 1 \leq k \leq I, k \neq i\}$ vor. Der Credibility-Schätzer im Modell von Bühlmann hat also die Form

$$\widehat{\widehat{\mu(\Theta_i)}} = \widehat{a}^{(i)} + \sum_{k=1}^{I} \widehat{b}_k^{(i)} \overline{X}_k$$

Die Verwendung des individuellen Schadenmittel $\overline{X}_k$ statt der Schadenhistorie X_k' ergibt sich aus dem gleichen Invarianz-Argument wie im Bewies von Satz 3.3. Wir stehen wieder vor einem Minimierungsproblem, nämlich:

$$\begin{aligned}
&\mathrm{E}\left[\left(\mu\left(\Theta_i\right)-\widehat{a}^{(i)}-\sum_{k=1}^{I}\widehat{b}_k^{(i)}\overline{X}_k\right)^2\right] \\
=\min_{a^{(i)},b_1^{(i)},\ldots,b_I^{(i)}\in\mathbb{R}}&\mathrm{E}\left[\left(\mu\left(\Theta_i\right)-a^{(i)}-\sum_{k=1}^{I}b_k^{(i)}\overline{X}_k\right)^2\right]
\end{aligned}\tag{3.13}$$

Analog zum Minimierungsproblem 3.4 bilden wir die partiellen Ableitungen nach $a^{(i)}$ und $b_j^{(i)}$ und erhalten

$$\frac{\partial}{\partial a^{(i)}}\mathrm{E}\left[\left(\mu\left(\Theta_i\right)-a^{(i)}-\sum_{k=1}^{I}b_k^{(i)}\overline{X}_k\right)^2\right]=\mathrm{E}\left[-2\left(\mu\left(\Theta_i\right)-a^{(i)}-\sum_{k=1}^{I}b_k^{(i)}\overline{X}_k\right)\right]$$

$$\frac{\partial}{\partial b_j^{(i)}}\mathrm{E}\left[\left(\mu\left(\Theta_i\right)-a^{(i)}-\sum_{k=1}^{I}b_k^{(i)}\overline{X}_k\right)^2\right]=\mathrm{E}\left[-\overline{X}_j\left(\mu\left(\Theta_i\right)-a^{(i)}-\sum_{k=1}^{I}b_k^{(i)}\overline{X}_k\right)\right]$$

Setzen wir diese Terme gleich null und subtrahieren wir (genau wie bei der Lösung von 3.4) nach Multiplikation von $\mathrm{E}[X_k]$ die erste von der zweiten Gleichung führt das zu folgender Aussage:

$$\begin{aligned}
&\mathrm{E}\left[\overline{X}_j\mu\left(\Theta_i\right)\right]-a^{(i)}\mathrm{E}\left[\overline{X}_j\right]-\mathrm{E}\left[\overline{X}_j\sum_{k=1}^{I}b_k^{(i)}\overline{X}_k\right] \\
-&\left(\mathrm{E}\left[\overline{X}_j\right]\mathrm{E}\left[\mu\left(\Theta_i\right)\right]-a^{(i)}\mathrm{E}\left[\overline{X}_j\right]-\mathrm{E}\left[\overline{X}_j\right]\mathrm{E}\left[\sum_{k=1}^{I}b_k^{(i)}\overline{X}_k\right]\right)=0
\end{aligned}$$

Äquivalent dazu ist nach einfachen Termumformungen:

$$\begin{aligned}
&\underbrace{\mathrm{Cov}\left(\overline{X}_j,\mu\left(\Theta_i\right)\right)}_{=\mathrm{E}\left[\overline{X}_j\mu(\Theta_i)\right]-\mathrm{E}\left[\overline{X}_j\right]\mathrm{E}[\mu(\Theta_i)]}-\underbrace{b_j^{(i)}Var\left(\overline{X}_j\right)}_{=b_j^{(i)}\left(\mathrm{E}\left[\overline{X}_j^2\right]-b_j^{(i)}\mathrm{E}\left[\overline{X}_j\right]^2\right)} \\
-&\left(\mathrm{E}\left[\overline{X}_j\sum_{k\neq j}b_k^{(i)}\overline{X}_k\right]-\mathrm{E}\left[\overline{X}_k\right]\mathrm{E}\left[\sum_{k\neq j}b_k^{(i)}\overline{X}_k\right]\right)=0
\end{aligned}$$

Da nach Voraussetzung $\overline{X}_k, \overline{X}_j$ unabhängig sind für $k \neq j$ verschwindet die letzte Klammer. Als Minimierungsbedingung erhalten wir somit:

$$\mathrm{Cov}\left(\overline{X}_j, \mu\left(\Theta_i\right)\right) = b_j^{(i)} Var\left(\overline{X}_j\right)$$

Mit Annahme 3.6 folgt dann $b_j^{(i)} = 0 \; \forall \; i \neq j$, somit reduziert sich die Minimierungsaufgabe zu dem schon bekannten Problem 3.4 und wir erhalten die Lösung aus Gleichung 3.2

Satz 3.7 *Der inhomogene Credibility-Schätzer im Modell von Bühlmann ist gegeben durch:*

$$\widehat{\widehat{\mu(\Theta_i)}} = \alpha \overline{X_i} + (1-\alpha)\mu_0 \tag{3.14}$$

wobei $\alpha = \frac{J}{J + \frac{sigma^2}{\tau^2}}$.

Das heißt, der Credibility-Schätzer im Modell von Bühlmann verwendet nur die allgemeinen Daten, die über das Kollektiv vorhanden sind und das Schadenmittel des Individuums, dessen Prämie zu berechnen ist. Da wir davon ausgehen, mit μ_0 einen eine best-mögliche Auskunft über den Kollektiben Mittelwert zu haben, können die Informationen über die übrigen Individuen keinen zusätzlichen Nutzen bringen.

3.2.2 Der homogene Credibility-Schätzer

Obwohl der gefundene Credibility-Schätzer wie eben gezeigt der beste lineare Schätzer in den Daten, d.h. auf Grundlage der Schadenhistorien aller Mitglieder des Kollektivs ist, kann es nützlich sein, noch einen weiteren linearen Schätzer zu betrachten. Der Schätzer $\widehat{\widehat{\mu(\Theta_i)}}$ verwendet mit $(1-\alpha)\mu_0$ einen konstanten Term in dem eine Annahme über die Struktur des Kollektivs enthalten ist. Wollen wir das durch diese Annahmen gegebene Modellrisiko ausschießen, bietet es sich an, die reine Linearität des Schätzers in den Daten zu fordern (in Abgrenzung zu linear affinen Schätzfunktionen, die einen konstanten Term, der nicht von den Daten abhängt, enthalten dürfen). Ein solcher Schätzer hätte dann die Form

$$\widehat{\widehat{\mu\left(\Theta_i\right)}}^{\text{hom}} = \sum_{k=1}^{I} \sum_{j=1}^{J} b_{kj} X_{kj} \tag{3.15}$$

Wir fordern ferner, dass er erwartungstreu sein soll, also $\mathrm{E}\left[\widehat{\widehat{\mu(\Theta_i)}}^{\mathrm{hom}}\right] = \mathrm{E}[\mu(\Theta_i)]$. Dies lässt sich mathematisch fassen durch

$$\sum_{k=1}^{I} \sum_{j=1}^{J} b_{kj} = 1 \tag{3.16}$$

Denn

$$\mathrm{E}\left[\widehat{\widehat{\mu(\Theta_i)}}^{\mathrm{hom}}\right] = \mathrm{E}\left[\sum_{k=1}^{I} \sum_{j=1}^{J} b_{kj} X_{kj}\right] = \sum_{k=1}^{I} \sum_{j=1}^{J} b_{kj} \mathrm{E}\left[X_{kj}\right] = \mu_0 \sum_{k=1}^{I} \sum_{j=1}^{J} b_{kj}$$

da $\mathrm{E}[\mu(\Theta_i)] = \mathrm{E}[X_{ij}] = \mu_0$.

In Abschnitt 4.3.2 werden wir einen Beweis für folgende Aussage liefern:

Satz 3.8 *Im Modell von Bühlmann ist der homogene Credibility-Schätzer gegeben durch*

$$\widehat{\widehat{\mu(\Theta_i)}} = \alpha \overline{X_i} + (1-\alpha)\overline{\overline{X}}$$

mit $\alpha = \frac{J}{J+\frac{\sigma^2}{\tau^2}}$ *(wie in (3.3)) und*

$$\overline{\overline{X}} = \frac{1}{I \cdot J} \sum_{k=1}^{I} \sum_{j=1}^{J} X_{kj}$$

Entsprechend der allgemeinen Aussagen für das Modell von Bühlmann-Straub (Satz 4.19) kann auch seinen quadratischen Verlust berechnet werden.

Kapitel 4

Credibility II: Verallgemeinerung und das Modell von Bühlmann-Straub

Wir haben nun am einfachen Credibility-Modell 3.1 und am Modell von Bühlmann 3.2 die grundlegenden Techniken der Credibility-Theorie kennengelernt. Dabei hat das Wissen über die Wahrscheinlichkeitsstruktur der vorliegenden Schätzprobleme kontinuierlich abgenommen (Tabelle 4.1).

Modell	$\mathrm{Var}(\mu(\Theta)), \mathrm{E}[\sigma^2(\Theta)]$	Kollektivmittel $\mathrm{E}[\Theta]$	Dichtefkt.Θ
Poisson-Gamma	×	×	×
Einfaches Cred.-Modell	×	×	
Modell v. Bühlmann	×	*	

*im Fall des homogenen Credibility-Schätzers

Tabelle 4.1: A-priori Wissen bei den bisher behandelten Modellen.

Um uns der Praxis weiter zu nähern wollen wir die Reduktion der Voraussetzungen vorantreiben. Bisher sind wir vom Vorliegen von Datensätzen ausgegangen, die sehr einfach strukturiert waren. Im Modell von Bühlmann etwa lagen uns vollständige Daten von I Versicherungsnehmern aus J Jahren vor. Bedingt aud Θ_i sollten die $X_{ij}, j = 1, \ldots, J$ unabhängig und identisch verteilt sein.

Im anstehenden Modell von Bühlmann-Straub (Abschnitt 4.3) wird diese strenge Staionaritätsannahme gelockert. Um die unregelmßigeren Datensätze besser behandeln zu können brauchen wir weitere Schritte an Formalisierung. Wir beginnen bei den beiden Arten von Schätzern, die wir in Abschnitt 3.2 betrachtet haben.

Definition 4.1 *Für ein Schätzproblem, bei dem wir die Zufallsgröße $\mu(\Theta)$ bei Vorliegen eines Vektors von beobachteten Zufallsvariablen $X = (X_1, \ldots, X_n)'$ bestimmen wollen, heißt*

$$\mathbf{L}(\mathbf{X},\mathbf{1}) := \left\{ \widehat{\mu(\Theta)} : \widehat{\mu(\Theta)} = a_0 + \sum_{j=1}^{n} a_j X_j, \quad a_0, \ldots, a_n \in \mathbb{R} \right\} \tag{4.1}$$

Menge der linearen oder inhomogenen Schätzfunktionen. Das Element aus $\mathbf{L}(\mathbf{X},\mathbf{1})$, welches $\mathrm{E}[(\widehat{\mu(\Theta)} - \mu(\Theta))^2]$ minimiert, heißt (inhomogener) Credibility-Schätzer und wird mit P^{cred} oder $\widehat{\widehat{\mu(\Theta)}}$ bezeichnet.

$$\mathbf{L_e}(\mathbf{X}) := \left\{ \widehat{\mu(\Theta)} : \widehat{\mu(\Theta)} = \sum_{j=1}^{n} a_j X_j,\ a_1, \ldots, a_n \in \mathbb{R}, \mathrm{E}\left[\widehat{\mu(\Theta)}\right] = \mathrm{E}\left[\mu(\Theta)\right] \right\} \tag{4.2}$$

ist die Menge der rein-linearen erwartungstreuen oder homogenen Schätzer. Das Element aus $\mathbf{L_e}(\mathbf{X})$), welches $\mathrm{E}[(\widehat{\mu(\Theta)} - \mu(\Theta))^2]$ minimiert heißt homogener Credibility-Schätzer. Wir schreiben $P^{cred_{hom}}$ oder $\widehat{\widehat{\mu(\Theta)}}^{hom}$.

Man beachte bei der Definition des homogenen Credibility-Schätzers, dass die Menge $\mathbf{L_e}(\mathbf{X})$ nur erwartungstreue Schätzer enthalten darf. In diesem Fall bedeutet das, der homogene Credibility-Schätzer muss die Minimalitätsbedingung unter allen möglichen Verteilungen von Θ und X erfüllen (wobei mit „möglich“ hier zulässig im Sinne des Modells gemeint ist). Insbesondere bedeutet das, dass die Koeffizienten $a_j,\ j = 1, \ldots, n$ nicht vom Erwartungswert von Θ abhängen dürfen.

Ferner sei noch ein Hinweis zum besseren Verständniss gegeben: Wir hatten bereits gesehen, dass der inhomogene Credibility-Schätzer im Modell von Bühlmann nur von $\overline{X_i}$ und μ_0, abhängt. Das heißt, er verwendet keine Daten, die wir über andere Mitglieder des Kollektivs haben. Theoretisch wäre dies möglich, da die X_{kj} für $k \neq i$ ebenfalls in $\mathbf{L}(\mathbf{X},\mathbf{1})$ enthalten sind. Auch wenn wir die koleteralen Informationen über die anderen Individuen nicht verwenden, ist $\widehat{\widehat{\mu(\Theta_i)}}$ der beste inhomogene Schätzer, der auch auf ihnen basiert. Dies ist wichtig, wenn wir im späteren Verlauf ausnutzen, dass $\mathbf{L_e}(\mathbf{X}) \subset \mathbf{L}(\mathbf{X},\mathbf{1})$ ein affiner Unterraum ist.

4.1 Der Hilbertraum $\mathbf{L^2}$

Der Weg, der uns zu den bisherigen Credibility-Schätzern geführt hat, war die Lösung von Minimierungsproblemen der Form 3.4

$$\mathrm{E}\left[\left(\mu(\theta)-\widehat{a}-\widehat{b}\overline{X}\right)^2\right]=\min_{a,b\in\mathbb{R}}\mathrm{E}\left[\left(\mu(\theta)-a-b\overline{X}\right)^2\right]$$

wobei, wenn wir $\widehat{a}=0$ fordern, der homogene Credibility-Schätzer gesucht wird, sonst der imhomogene.
Das bedeutet, dass die Berechnung von P^{cred} und $P^{cred_{hom}}$ als ein Problem der Minimierung der Fehlerquadrate betrachtet werden kann, wie es etwa aus der Regeressionsananlyse in der Statistik bekannt ist. Da in der Versicherungsmathematik aber häufig große Datensätze betrachtet werden (bereits aus dem unternehmerischen Kalkül der Versicherung heraus sind große Datensätze ja erstrebenswert) kann die numerische Berechnung an ihre Grenzen stoßen. Wir wollen uns daher von nun an eine Eigenschaft der betrachteten Zufallsprozesse zunutze machen, die wir bisher meist implizit vorausgesetzt haben (etwa Ann.3.1). Für die Berechenbarkeit der Credibility-Schätzer wurde angenommen, dass die X_i und Θ endliche Erwartungswerte und Varianzen haben.

Die Zufallsvariablen, die wir zum Beispiel als jährliche Schadenforderungen der versicherten Individuuen beobachten können und Θ, welches die Risikodisposition eines einzelnen Individuums beschreibt, liegen also in der Menge

$$\mathbf{L^2}:=\left\{X:X\text{ ist }\mathbb{R}\text{-wertige Zufallsvariable mit }\mathrm{E}\left[X^2\right]<\infty\right\}\tag{4.3}$$

Dass diese Menge ein $\mathbb{R}$-Vektorraum ist, ist leicht einzusehen.

Weiter können $\mathbf{L^2}$ mit einem Skalarprodukt, also einer Abbildung $<\cdot,\cdot>$: $\mathbf{L^2}\times\mathbf{L^2}\to\mathbb{R}$ die symmetrisch und in beiden Komponenten linear ist, ausstatten:

$$<X,Y>:=\mathrm{E}[XY]\ \text{ für } X,Y\in\mathbf{L^2}\tag{4.4}$$

Die notwendigen Eigenschaften von $<\cdot,\cdot>$ folgen aus der Linearität des Erwartungswertes. Aus dem Skalarprodukt erhalten wir folgende Norm:

$$||X||:=\sqrt{<X,X>}=\sqrt{\mathrm{E}[X^2]}\tag{4.5}$$

Dieser Norm induziert die Metrik ($d(X,Y):=||X-Y||\forall X,Y\in\mathbf{L^2}$) Das heißt, zwei Zufallsvariablen $X,X'\in\mathbf{L^2}$ sind identisch, wenn ihre Abstandsnorm $\sqrt{\mathrm{E}\left[(X-X')^2\right]} = 0$. Sie dürfen sich also nur auf Nullmengen bezüglich eines Maßes unterscheiden, welches ein dominierendes Maß für die Wahrscheinlichkeitsmaße ist, die modellverträglich zur Definition von $\mathbf{L^2}$ herangezogen werden können.

Genauer ist der $\mathbf{L^2}$ also ein Vektorraum von Äquivalenzklassen quadratintegrierbarer Zufallsvariablen, wobei zwei ZVen zur selben äquivalenzklasse gehören, wenn sie sich nur auf einer Nullmenge unterscheiden.

Der so definierte $\mathbf{L^2}$ mit dem Skalarprodukt $< \cdot, \cdot >$ ist bekanntlich ein Hilbertraum, da er bezüglich der Metrik $d(\cdot, \cdot)$ volsständig ist.

Mit dem Skalarprodukt kommen wir zu folgender Definition:

Definition 4.2 *Gilt für $X, Y \in \mathbf{L^2}$, dass $< X, Y >= 0$, so sagen wir, die Elemente sind orthogonal zueinander und schreiben $X \perp Y$.*

Ferner benötigen wir die Begriffe Unterraum und affiner Unterraum, die üblicherweise so definiert werden:

Definition 4.3 *Eine Teilmenge $\mathbf{U} \subset \mathbf{L^2}$ heißt Unterraum, wenn sie bezüglich der Bildung endlicher Linearkombinationen von Elementen aus $\mathbf{U}$ abgeschlossen ist.*

Eine Teilmenge $\mathbf{U} \subset \mathbf{L^2}$ heißt affiner Unterraum, wenn ein $Z \in \mathbf{L^2}$ und ein Unterraum $\mathbf{U}$ existieren, derart dass

$$\forall \ X \in \mathbf{V} \ \exists \ Y \in \mathbf{U} \ mit \ X = Z + Y \tag{4.6}$$

Äquivalent zu Gleichung 4.6 ist die Forderung, dass $\mathbf{U}$ abgeschlossen unter endlichen normierten Linearkombinationen sein soll. Eine Linearkombination $\sum_{i=1}^n a_i X_i$ heißt normiert, wenn $\sum_{i=1}^n a_i = 1$

Wenn nun $\mathbf{U} \subset \mathbf{L^2}$ ein (affiner) Unterraum ist, können wir bestimmen, welche Komponenten eines Elements aus $\mathbf{L^2}$ in (U) liegen und nennen diesen Anteil orthogonale Projektion.

Definition 4.4 *Sei $Y \in \mathbf{L^2}$, $\mathbf{U} \subset \mathbf{L^2}$ ein abgeschlossener (affiner) Unterraum. Für Y^*, die orthogonale Projektion von Y auf $\mathbf{U}$ gilt*

$$Y^* - Y \perp \mathbf{U} \quad ,d.h. \quad Y^* - Y \perp X_1 - X_2 \quad \forall X_1, X_2 \in \mathbf{U}$$

Es bleibt zu zeigen, dass die so gegebene orthogonale Projektion existiert und eindeutig ist. Das ist allerdings charakteristisch für alle Hilberträume, deshalb unterlassen wir den Beweis an dieser Stelle.

Ebenso charakteristisch für alle Hilberträume sind die folgenden beiden Eigenschaften der orthogonalen Projektion, die auch als alternative Definition verwendet werden können:

Für $Y \in \mathbf{L^2}$ ist die orthogonale Projektion auf einen (affinen) Unterraum $\mathbf{U}$ genau das Element in $\mathbf{U}$, welches $d(Y,X)$ für alle $X \in \mathbf{U}$ minimiert. Im Sinne der durch das Skalarprodukt induzierten Metrik liegt Y^* unter allen Elementen von $\mathbf{U}$ am nächsten an Y, also

$$\|Y - Y^*\| \leq \|Y - X\| \quad \forall \quad X \in \mathbf{U} \tag{4.7}$$

Schreiben wir die Orthogonalitätsbedingung aus der Definition mit Hilfe des Skalarprodukts, ist $Y^* = \mathrm{Pro}(Y|\mathbf{U})$ genau dann wenn $Y^* \in \mathbf{U}$ und

$$< Y^* - Y, X - Y^* >= 0 \quad \forall \quad X \in \mathbf{U} \tag{4.8}$$

Wenn $\mathbf{U}$ nicht nur ein affiner Unterraum ist, genügt es, $< Y^* - Y, X > \forall X \in \mathbf{U}$ zu fordern.

Wir erinnern daran, dass die Metrik $d(X,X) := ||X - Y|| \, \forall X, Y \in \mathbf{L^2}$ durch die Norm aus 4.5 gegeben ist. Wenn $\widehat{\mu(\Theta)} = \mathrm{Pro}\,(\mu(\Theta)|\mathbf{L})$ die orthogonale Projektion eine Zufallsvariablen $\mu(\Theta) \in \mathbf{L^2}$ auf einen (affinen) Unterraum $\mathbf{L}$ ist, minimiert $\widehat{\mu(\Theta)}$ also folgenden Term über alle $X \in \mathbf{L}$:

$$d(\mu(\Theta), X) = ||\mu(\Theta) - X|| = \sqrt{\mathrm{E}\left[(\mu(\Theta) - X)^2\right]} \tag{4.9}$$

Um diese Eigenschaft für unsere Betrachtung nutzbar zu machen, brauchen wir einige (affine) Unterräume.

Für Zufallsvektoren $X = (X_1, X_2, \ldots, X_n)'$, $X_1, X_2, \ldots, X_n \in \mathbf{L^2}$ und reelwertige Funktionen $g : \mathbb{R}^n \longrightarrow \mathbb{R}$ derart, dass $\mathrm{E}\left[g(X)^2\right] < \infty$ bildet

$$\mathbf{G}(\mathbf{X}) := \left\{ Z \in \mathbf{L^2} : Z = g(X) \right\} \tag{4.10}$$

einen Unterraum von L^2.

Dass die Menge der linearen oder inhomogenen Schätzfunktionen $\mathbf{L}(\mathbf{1}, \mathbf{X})$, wie in (4.1) definiert, ein linearer Unterraum von $\mathbf{L^2}$ ist, ist sofort einsichtig. Ferner gilt $\mathbf{L}(\mathbf{1}, \mathbf{X}) \subset \mathbf{G}(\mathbf{X})$ da die linearen Funktionen als $g(X)$ in (4.10) zulässig sind.

Die mit Gleichung (4.2) definierte Menge der rein-linearen oder homogenen Schätzfunktionen $\mathbf{L_e}(\mathbf{X})$ bildet einen affinen Unterraum. Dies rührt von der Bedingung der Erwartungstreue her, die wir in Gleichung (3.16) aufgeschrieben hatten. Sie korrespondiert mit der alternativen Definition eines affinen Unterraums (siehe 4.6 und nachstehenden Text).

In Gleichung (4.9) hatten wir beschrieben, dass die orthogonale Projektion von Y der Punkt des (affinen) Unterraums ist, der den geringsten Abstand zu Y hat (bezüglich der durch das Skalarprodukt induzierten Metrik). Das führt uns zu folgendem

Satz 4.5 *Wollen wir die Zufallsgröße $\mu(\Theta) \in \mathbf{L^2}$ schätzen, sind die folgenden Definitionen für den (in-)homogenen Credibility-Schätzer und den Bayes-Schätzer als orthogonale Projektion auf die jeweiligen Unterräume äquvalent zu denen aus 4.1 bzw. 2.6.*

$$P^{cred} = \widehat{\widehat{\mu(\Theta)}} := \mathrm{Pro}\left(\mu\left(\Theta\right)|\mathbf{L}(\mathbf{1},\mathbf{X})\right) \tag{4.11}$$

$$P^{cred_{hom}} = \widehat{\widehat{\mu(\Theta)}}^{hom} := \mathrm{Pro}\left(\mu\left(\Theta\right)|\mathbf{L_e}(\mathbf{X})\right) \tag{4.12}$$

$$P^{Bayes} = \widetilde{\mu(\Theta)} := \mathrm{Pro}\left(\mu\left(\Theta\right)|\mathbf{G}(\mathbf{X})\right) \tag{4.13}$$

Diese Sichtweise können wir nutzen, um folgenden Satz zu beweisen:

Satz 4.6 **(i)** *Für die Schätzer wie in Satz 4.5 gilt*

$$\mathrm{E}\left[\left(\widehat{\widehat{\mu(\Theta)}} - \mu(\Theta)\right)^2\right] = \mathrm{E}\left[\left(\widehat{\widehat{\mu(\Theta)}} - \widetilde{\mu(\Theta)}\right)^2\right] + \mathrm{E}\left[\left(\widetilde{\mu(\Theta)} - \mu(\Theta)\right)^2\right]$$

(ii) *Die Credibility-Prämie $\widehat{\widehat{\mu(\Theta)}}$ ist unter allen linearen Schätzern $\mathbf{L}(\mathbf{X},\mathbf{1})$ am nächsten an der Bayes-Prämie $\widetilde{\mu(\Theta)}$ in dem Sinne, dass die erwartete quadratische Abweichung minimal ist.*

(iii) *In diesem Sinne ist auch $\widehat{\widehat{\mu(\Theta)}}^{hom}$ unter den homogenen Schätzfunktionen $\mathbf{L_e}(\mathbf{X})$ die beste Näherung sowohl für $\widehat{\widehat{\mu(\Theta)}}$ als auch für $\widetilde{\mu(\Theta)}$.*

Zum Beweis verwenden wir folgenden allgemeinen Satz über Unterräume von Hilberträumen:

Satz 4.7 *Sei $\mathbf{H}$ ein Hilbertraum, $\mathbf{U}$, $\mathbf{V}$ seien abgeschlossene (affine) Unterräume mit $\mathbf{U} \subset \mathbf{V}$. Dann gilt für jedes $Y \in \mathbf{H}$*

(i)

$$\operatorname{Pro}(Y|\mathbf{U}) = \operatorname{Pro}(\operatorname{Pro}(Y|\mathbf{V})|\mathbf{U}) \tag{4.14}$$

(ii)

$$\begin{aligned}\|Y - \operatorname{Pro}(Y|\mathbf{U})\|^2 &= \|Y - \operatorname{Pro}(Y|\mathbf{V})\|^2 \\ &+ \|\operatorname{Pro}(Y|\mathbf{V}) - \operatorname{Pro}(Y|\mathbf{U})\|^2\end{aligned} \tag{4.15}$$

Die orthogonale Projektion ist also iterativ, das bedeutet, sie kann pfadunabhängig über verschachtelte Ketten von Unterräumen durchgeführt werden. Zur Veranschaulichung der letzten Gleichung sei $\mathbf{H}$ der Hilbertraum $\mathbb{R}^3$ mit dem üblichen Skalarprodukt. $\mathbf{V}$ sei eine Ebene im $\mathbb{R}^3$, $\mathbf{U}$ eine Grade, die durch $\mathbf{V}$ verläuft. Dann entspricht (4.15) dem Satz von Pythagoras.

Beweis Satz 4.6:

Die Gleichung in (i) folgt aus Satz 4.7 (ii). Für den Weg dahin verwenden wir, dass $\mathbf{L}(\mathbf{X},\mathbf{1})$ und $\mathbf{G}(\mathbf{X})$ $\mathbf{L^2}$-Unterräume sind mit $\mathbf{L}(\mathbf{X},\mathbf{1}) \subset \mathbf{G}(\mathbf{X})$. Aus $\widehat{\widehat{\mu(\Theta)}} = \operatorname{Pro}(\mu(\Theta)|\mathbf{L}(\mathbf{X},\mathbf{1}))$ und $\widetilde{\mu(\Theta)} = \operatorname{Pro}(\mu(\Theta)|\mathbf{G}(\mathbf{X}))$ (Satz 4.5) und der Anwendung von Satz 4.7 (i) folgt

$$\widehat{\widehat{\mu(\Theta)}} = Pro\left(\widetilde{\mu(\Theta)}\middle|\mathbf{L}(\mathbf{1},\mathbf{X})\right) \tag{4.16}$$

Das zeigt bereits Aussage (ii), da die orthogonale Projektion das Element aus $\mathbf{L}(\mathbf{1},\mathbf{X})$ ist mit dem geringsten Abstand zu $\widetilde{\mu(\Theta)}$. Satz 4.7 (ii) besagt nun

$$\left\|\widehat{\widehat{\mu(\Theta)}} - \mu(\Theta)\right\|^2 = \left\|\widehat{\widehat{\mu(\Theta)}} - \widetilde{\mu(\Theta)}\right\|^2 + \left\|\widetilde{\mu(\Theta)} - \mu(\Theta)\right\|^2$$

Die Definition der Norm (Gleichung (4.5)) bildet den letzten Schritt zum Beweis von (i).

Genau wie für (ii) können wir aus

$$\widehat{\widehat{\mu(\Theta)}}^{hom} = Pro\left(\widetilde{\mu(\Theta)}\middle|\mathbf{L_e}(\mathbf{X})\right)$$

auf die Approximationsaussage in (iii) schließen. Zur Veranschaulichung der benutzten Inklusionszusammenhänge kann Tabelle 4.2 dienen.

□

Zielgröße		Raum/Unterraum
$\mu(\Theta)$	zu schätzendes Funktional in	$\mathbf{L^2}$
$\downarrow$		$\cup$
$\widetilde{\mu(\Theta)} = \text{Pro}(\mu(\Theta)\vert\mathbf{G}(\mathbf{X}))$	bester Schätzer in	$\mathbf{G}(\mathbf{X})$
$\downarrow$		$\cup$
$\widehat{\widehat{\mu(\Theta)}} = \text{Pro}(\widetilde{\mu(\Theta)}\vert\mathbf{L}(\mathbf{X},\mathbf{1}))$	bester Schätzer in	$\mathbf{L}(\mathbf{X},\mathbf{1})$
$\downarrow$		$\cup$
$\widehat{\widehat{\mu(\Theta)}}^{hom} = \text{Pro}(\widehat{\widehat{\mu(\Theta)}}\vert\mathbf{L_e}(\mathbf{X}))$	bester Schätzer in	$\mathbf{L_e}(\mathbf{X})$

Tabelle 4.2: Zusammenhang von Schätzern und entsprechenden Unterräumen

Satz 4.6 beinhaltet also eine Aussage darüber, wie die Güte der betrachteten Schätzer, gemessen durch ihren quadratischen Verlust, abnimmt, wenn die Klasse der zulässigen Schätzfunktionen immer weiter eingeschränkt wird. Die absteigende Kette von Unterräumen aus Tabelle 4.2 geht also ebenfalls mit einem Verlust der Güte der Schätzer einher. Doch die Erfahrung zeigt, dass das Anwachsen des quadratischen Fehlers vom Bayes- zum homogenen Credibility-Schätzer gering ist. Verglichen mit der naiven Schätzung $P^{koll} = \mu_0$ aus Definition 1.5 sind die quadratischen Verluste der hier betrachteten Schätzer alle klein.

Den Genauigkeitsverlust vom Bayes-Schätzer etwa zum homogenen Credibility-Schätzer nehmen wir aber nicht nur aus Gründen der Einfachheit des Schätzers in Kauf. Für den Bayes-Schätzer würden detailierte Informationen z.B. über die Dichte der Strukturfunktion $U(\theta)$ benötigt. Auch für $\widehat{\widehat{\mu(\Theta)}}$ brauchen wir einen A-priori.Wert für μ_0. Doch in der Realität liegen diese Informationen nicht vor. Um zum Bayes-Schätzer zu gelangen blieben uns zwei Möglichkeiten:

- Das Schätzen der relevanten Modellparameter aus den Daten, der sogenannte empirische Bayes-Ansatz

- Wir verlassen uns auf die Einschätzung eines erfahrenen Statistikers oder etwa auf eine theoretische Herleitung bestimmter Größen (reiner Bayes-Ansatz)

Doch durch dieses Vorgehen werden zusätzliche Annahmen über die vorliegende Verteilung gemacht, z.B. zu welcher Familie von Wahrscheinlichkeitsverteilungen die Strukturfunktion und die Verteilung der konkreten

Schadenereignisse X_i gehöhren. Mit dem Ziel, den quadratischen Verlust des Schätzers zu minimieren, würden wir einen Teil des Risikos in die Modellannahmen verlagern.

Anders als der quadratische Verlust von $\widehat{\widehat{\mu(\Theta)}}^{hom}$ gegenüber $\widetilde{\mu(\Theta)}$ lässt sich dieses Risiko allerdings kaum quantifizieren. Der Zielkonflikt zwischen hoher theoretischer Genauigkeit von Schätzer und der ungewissen Qualität der dafür angenommenen Voraussetzungen ist die Motivation der Credibility-Theorie. Im folgenden Abschnitt werden wir zu einem Optimalitätskriterium für Schätzer kommen, das mit gewissermaßen minimalen Vorraussetzungen auskommt.

4.2 Normalengleichungen für Credibility-Schätzer

Im vorangegangenen Kapitel haben wir gesehen, dass wir die Credibility-Schätzer als Orthogonalprojektionen auf (affine) Unterräume des $\mathbf{L^2}$ betrachten können. Die Hoffnung jedoch, daraus eine Konstruktionsvorschrift ableiten zu können, trügt. Wie beim Kleinsten-Quadrate-Schätzer aus der Regressionsanalyse (etwa [Als02] S.70ff) führt uns auch hier die Invertierung einer Matrix sofort zum Ziel[1]. Doch ist es in der Versicherungsmathematik neben den statistischen nicht zuletzt aus betriebswirtschaftlichen Gründen erstrebenswert, möglichst große Datensätze zu verarbeiten. So können Berechnungsverfahren schnell an ihre Grenzen stoßen. Die Hilbertraum-Betrachtung verschafft uns aber neben der (theoretischen) Möglichkeit der Konstruktionsregel noch ein zweites Werkzeug: Ein Kriterium, dass die Optimalität eines vorhandenen Schätzers zeigen kann. Beweisen wir für ein Schätzverfahren, dass es die Orthogonalprojektion auf den Unterraum $\mathbf{L}(\mathbf{X}, \mathbf{1})$ oder $L(X)_e$ ist, wissen wir bereits, dass es sich um den inhomogenen beziehungsweise homogenen Credibility-Schätzer handelt.

Wir beginnen damit, die entsprechenden Kriterien für den inhomogenen Credibility-Schätzer bereitzustellen.

[1] siehe [BG05] S.73, Korollar 3.18.

4.2.1 Der inhomogene Fall

Wie wir bereits gesehen haben verwendet dieser, da μ_0 als bekannt vorausgesetzt wird, nur Daten über das betrachtete Individuum. Bis wir zum homogenen Credibility-Schätzer kommen gilt also $X = (X_1, \ldots, X_J)'$ wobei J die Anzahl der Jahre ist, aus denen wir Daten über das betrachtete Individuum haben.

Satz 4.8 *Sei $\widehat{\widehat{\mu(\Theta)}} \in \mathbf{L}(\mathbf{X}, \mathbf{1})$ und es gelten die folgenden beiden Gleichungen:*

(i)

$$< \mu(\Theta) - \widehat{\widehat{\mu(\Theta)}}, 1 >= 0 \tag{4.17}$$

(ii)

$$< \mu(\Theta) - \widehat{\widehat{\mu(\Theta)}}, X_j >= 0, \quad j = 1, \ldots, J \tag{4.18}$$

Genau dann ist $\widehat{\widehat{\mu(\Theta)}}$ der (inhomogene) Credibility-Schätzer basierend auf X.

Beweis Satz 4.8: Satz 4.5 besagt, dass $\widehat{\widehat{\mu(\Theta)}}$ genau dann der inhomogene Credibility-Schätzer ist, wenn er X orthogonal auf $\mathbf{L}(\mathbf{X}, \mathbf{1})$ projeziert. In Gleichung (4.8) und der nachfolgenden Spezifizierung für den Fall, dass wir die Projektion auf einen echten (nicht affinen) Unterraum betrachten, wird folgendes Kriterium dafür gegeben, dass $\widehat{\widehat{\mu(\Theta)}} \in \mathbf{L}(\mathbf{X}, \mathbf{1})$ die Orthogonalprojektion ist:

$$< \widehat{\widehat{\mu(\Theta)}} - \mu(\Theta), X >= 0 \quad \forall \quad X \in \mathbf{L}(\mathbf{X}, \mathbf{1}) \tag{4.19}$$

Eben dies fordern (4.17) und (4.18) für $(1, X_1, \ldots, X_J)$. Die bilden eine Basis des endlichdimensionalen $\mathbf{L^2}$-Untervektorraums $\mathbf{L}(\mathbf{X}, \mathbf{1})$. D.h. (4.19) gilt bereits, wenn die Gleichung für $(1, X_1, \ldots, X_J)$ erfüllt ist..

□

Wollen wir den stochastischen Aspekt gegenüber dem geometrischen betonen können wir Satz 4.8 auch folgendermaßen schreiben:

Korollar 4.9 *Genau dann ist $\widehat{\widehat{\mu(\Theta)}} \in \mathbf{L}(\mathbf{X}, \mathbf{1})$ der (inhomogene) Credibility-Schätzer basierend auf X wenn die folgenden beiden Gleichungen erfüllt sind:*

(i)

$$\mathrm{E}\left[\mu\left(\Theta\right)-\widehat{\widehat{\mu(\Theta)}}\right]=0 \tag{4.20}$$

(ii) *für* $j=1,\ldots,J$ *gilt*

$$\mathrm{Cov}\left(\mu\left(\Theta\right),X_j\right)=\mathrm{Cov}\left(\widehat{\widehat{\mu(\Theta)}},X_j\right) \tag{4.21}$$

Beweis Korollar 4.9: Gleichung (4.20) folgt sofort aus (4.17) durch Einsetzen der Definition des Skalarprodukts. Im Hilbertraum $\mathbf{L^2}$ ist $\mathrm{Cov}(X,Y)$ gegeben durch $< X-\mathrm{E}X, Y-\mathrm{E}Y >$. Um (4.21) zu zeigen addieren wir zu (4.18) auf beiden Seiten $\mu_{X_j} < \mu\left(\Theta\right)-\widehat{\widehat{\mu(\Theta)}},1 >$ was nach Voraussetzung null ist, also gilt für $j=1,\ldots,J$:

$$\begin{aligned}
& < \mu\left(\Theta\right)-\widehat{\widehat{\mu(\Theta)}},X_j >= 0\\
\Leftrightarrow\quad & < \mu\left(\Theta\right)-\widehat{\widehat{\mu(\Theta)}},X_j > -\mu_{X_j} < \mu\left(\Theta\right)-\widehat{\widehat{\mu(\Theta)}},1 >= 0\\
\Leftrightarrow\quad & < \mu\left(\Theta\right)-\widehat{\widehat{\mu(\Theta)}},X_j-\mu_{X_j} >= 0\\
\Leftrightarrow\quad & < \mu\left(\Theta\right)-\mu_{X_j}-\widehat{\widehat{\mu(\Theta)}}+\mu_{X_j},X_j-\mu_{X_j} >= 0\\
\Leftrightarrow\quad & < \mu\left(\Theta\right)-\mu_{X_j},X_j-\mu_{X_j} > - < \widehat{\widehat{\mu(\Theta)}}-\mu_{X_j},X_j-\mu_{X_j} >= 0\\
\Leftrightarrow\quad & \mathrm{Cov}\left(\mu\left(\Theta\right),X_j\right)=\mathrm{Cov}\left(\widehat{\widehat{\mu(\Theta)}},X_j\right)
\end{aligned}$$

□

Von den Normalengleichungen zu den etwas anschaulicheren wahrscheinlichkeitstheoretischen Ausdrückengelangen wir durch die bekannte Darstellung des Credibility-Schätzers als Linearkombination der Schadenforderungen eines Individuums:

$$\widehat{\widehat{\mu(\Theta)}}=\widehat{a}_0+\sum_{j=1}^{J}\widehat{a}_j X_j$$

So wird (4.20) zu

$$\mu_0-\widehat{a}_0-\sum_{j=1}^{J}\widehat{a}_j\underbrace{\mathrm{E}\left[X_j\middle|\Theta_i\right]}_{=\mu_{X_j}}=0 \tag{4.22}$$

Ebenso erhalten wir durch Einsetzen für $j = 1,\ldots,J$

$$\begin{aligned} \operatorname{Cov}\left(\widehat{\widehat{\mu(\Theta)}}, X_j\right) &= \operatorname{Cov}\left(\widehat{a}_0 + \sum_{k=1}^{J} \widehat{a}_k X_k, X_j\right) \\ &= \sum_{k=1}^{J} \widehat{a}_k \operatorname{Cov}\left(X_k, X_j\right) \underset{(4.21)}{=} \operatorname{Cov}\left(\mu\left(\Theta\right), X_j\right) \end{aligned} \tag{4.23}$$

Darübner hinaus erhalten wir noch eine Aussage über die Varianz von $\widehat{\widehat{\mu(\Theta)}}$.

$$\begin{aligned} \operatorname{Var}\left[\widehat{\widehat{\mu(\Theta)}}\right] &= \operatorname{Cov}\left(\widehat{\widehat{\mu(\Theta)}}, \sum_{j=1}^{J} \widehat{a}_j X_j\right) \\ &= \sum_{j=1}^{J} \widehat{a}_j \operatorname{Cov}\left(\widehat{\widehat{\mu(\Theta)}}, X_j\right) \\ &\underset{(4.21)}{=} \sum_{j=1}^{J} \widehat{a}_j \operatorname{Cov}\left(\mu(\Theta), X_j\right) \\ &= \operatorname{Cov}\left(\mu(\Theta), \widehat{\widehat{\mu(\Theta)}}\right) \end{aligned}$$

4.2.2 Der homogene Fall

Ein Aussage wie in Satz 4.8 lässt sich auch für den homogenen Credibility-Schätzer aufstellen.

Satz 4.10 *Sei* $\widehat{\widehat{\mu(\Theta)}}^{hom} \in \mathbf{L_e}(\mathbf{X})$ *und es gelten die folgenden beiden Bedingungen:*

(i)

$$< \mu\left(\Theta\right) - \widehat{\widehat{\mu(\Theta)}}^{hom}, 1 >= 0 \tag{4.24}$$

(ii)

$$< \mu\left(\Theta\right) - \widehat{\widehat{\mu(\Theta)}}^{hom}, \widehat{\mu(\Theta)} - \widehat{\widehat{\mu(\Theta)}}^{hom} >= 0, \quad \forall \widehat{\mu(\Theta)} \in \mathbf{L_e}(\mathbf{X}) \tag{4.25}$$

Genau dann ist $\widehat{\widehat{\mu(\Theta)}}^{hom}$ *der homogene Credibility-Schätzer basierend auf X.*

Man beachte, dass es sich bei $\mathbf{L_e}(\mathbf{X})$ nur um einen affinen Unterraum handelt und dieser daher im allgemeinen keine Basis hat. Wir können also im homogenen Fall nicht direkt zu einer Aussage gelangen, bei der (4.25) nur für eine endliche Auswahl von Elementen aus $\mathbf{L_e}(\mathbf{X})$ gefordert wird. Der Beweis der Aussagen erfolgt vermöge Gleichung (4.8) genau wie im inhomogenen Fall.

Ebenso gibt es Normalengleichungen für den homogenen Credibility-Schätzer. Der Beweis erfolgt durch analoge Rechnung wie für Satz 4.11:

Korollar 4.11 *Genau dann ist* $\widehat{\widehat{\mu(\Theta)}} \in \mathbf{L_e}(\mathbf{X})$ *der homogene Credibility-Schätzer basierend auf X wenn die folgenden beiden Gleichungen erfüllt sind:*

(i)

$$\mathrm{E}\left[\mu(\Theta) - \widehat{\widehat{\mu(\Theta)}}^{hom}\right] = 0 \tag{4.26}$$

(ii) *für alle* $\widehat{\mu(\Theta)} \in \mathbf{L_e}(\mathbf{X})$ *gilt*

$$\mathrm{Cov}\left(\widehat{\widehat{\mu(\Theta)}}^{hom} - \mu(\Theta), \widehat{\widehat{\mu(\Theta)}}^{hom}\right) = \mathrm{Cov}\left(\widehat{\widehat{\mu(\Theta)}}^{hom} - \mu(\Theta), \widehat{\mu(\Theta)}\right) \tag{4.27}$$

Wie die Ausnutzung der Orthogonalitätsbeziehungen nahelegt, haben wir für den Beweis der Normalengleichungen intensiv von den Hilbertraum-Eigenschaften des $\mathbf{L^2}$ Gebrauch gemacht. Bevor wir zur Anwendung am konkreten Modell übergehen sei allerdings angemerkt, dass sich diese Beziehungen auch auf direktem Wege beweisen lassen [2].

[2] So etwa in [Sun99] S.38 f für den inhomogenen Credibility-Schätzer

4.3 Das Modell von Bühlmann-Straub

Anhand das Modells von Bühlmann-Straub wollen wir die zu Beginn des Kapitels gewonnen Erkenntnisse anwenden. In [BS70] wurde das Modell zum ersten Mal beschrieben. Gegenüber den bisher behandelten Credibility-Modellen weißt es eine zusätzliche Verallgemeinerung auf. Bisher haben sich die Risiken nur anhand ihrer individuellen Risikodisposition unterschieden. Hinzu tritt nun, dass wir die Risiken unterschiedlich gewichten wollen. Dies kommt einer Lockerung der Stationaritätsannahme (dass bedingt auf Θ_i die $X_{ij}, j = 1, \ldots, J$ unabhängig und identisch verteilt sind) gleich. Ein Versicherungsvertrag i im Jahr j soll also ein Gewicht w_{ij} bekommen. Die Bedeutung dieser Gewichtung in der Praxis können wir uns auf zwei verschiedene Weisen vorstellen:

1. Im Versicherungsvertrag/Risiko i sind eine Klasse gleichartiger Risiken zusammengefasst, die bezüglich ihres Risikoverhaltens (bedingt auf den Verteilungseigenschaften der Klasse i) unabhänige und identisch verteilte Zufallsgrößen darstellen. Das Gewicht w_{ij} können wir dann als die Anzahl der Mitglieder der Versicherung in Klasse i im Jahr j auffassen. Ein Beispiel hierfür wäre die Anzahl von Fahrzeugen, die in einer bestimmten Region und einem bestimmten Jahr versichert sind.

2. Die Schadenforderungen des Individuums i sind abhängig von einer a-priori bekannten Größe. Ein Beispiel könnte die versicherte Deckungssumme in einem Rückversicherungsvertrag sein. Dies entspricht etwa der Vorstellung aus dem Modell von Bichsel (Abschnitt 2.4), allerdings sind die Unterschiede hier Teil der Modellbetrachtung und können sich zudem von Jahr zu Jahr ändern.

Im ersten Fall, wenn die Gewichtung durch eine tatsächliche Anzahl von identischen Unterrisiken zustande kommt schreiben wir auch V_{ij} statt w_{ij} um hervorzuheben, dass eine Anzahl Individuen zugrunde liegt. Betrachten wir nur Individuen, die ganze Jahre versichert sind, ist V_{ij} eine natürliche Zahl, sonst allgemeiner wie w_{ij} eine positive reelle Zahl.

Beiden Fällen gemein ist, dass es sich wieder um ein Zwei-Urnen-Modell handelt. Aus der ersten Urne wird das θ_i gezogen, die Risikodisposition der i-ten Klasse. Aus einer zweiten Urne, deren Inhalt von θ_i bestimmt wird erhalten wir nun die jährlichen Schadenforderungen des Individuums i. Neu kommt nun hinzu, dass der Inhalt der zweiten Urne zusätzlich durch die (bekannten) Gewichte w_{ij} beeinflusst wird.

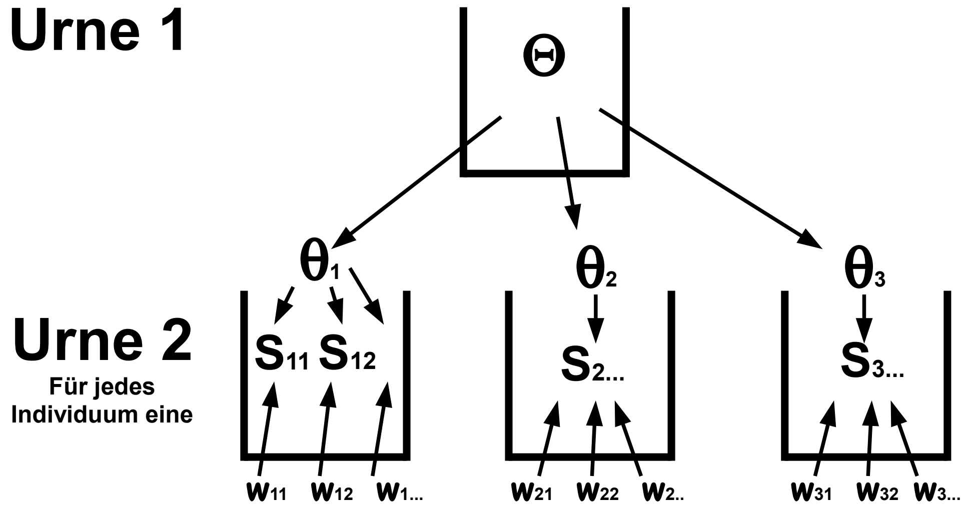

Abbildung 4.1: Der Einfluss von Θ und den w_{ij} im Bühlmann-Straub Modell

Für den ersten Fall stellt sich das folgendermaßen da:

Es sei S_{ij} die Schadenforderung aus einem Versicherungsvertrag i im Jahr j. Abbildung 4.1[3] verdeutlicht die Abhängigkeit der S_{ij} von Θ und den Gewichten w_{ij}.

Das Gewicht $V_{ij} \in \mathbb{N}$ ist dann die Anzahl der Unterrisiken, $S_{ij}^{(k)}, k = 1, \ldots, V_{ij}$ sind die Elementarschäden im Jahr j für Versicherungsvertrag i und es gilt:

$$S_{ij} = \sum_{k=1}^{V_{ij}} S_{ij}^{(k)} \qquad \forall \quad i = 1, \ldots, I \text{ und } j = 1, \ldots, J$$

Wollen wir unterschiedliche Risiken vergleichen, bietet es sich an, die Schadenforderungen mithilfe der Gewichte zu normieren. Dann ist der durchschnittliche Jahresschaden pro versicherter Einheit $X_{ij} := S_{ij}/V_{ij}$ eine Größe, die uns eine gute Auskunft über die Risikodisposition der Individuen

[3]vgl. [BG05] Fig. 1.6 S. 12, Fig. 4.2 S.90

aus Klasse i im Vergleich zu anderen gibt. Es gilt

$$\mathrm{E}[X_{ij}] = \mathrm{E}\left[1/V_{ij}\sum_{k=1}^{V_{ij}} S_{ij}^{(k)}\right] = \mathrm{E}\left[S_{ij}^{(k)}\right] \tag{4.28}$$

für beliebiges $j \in \{1,\ldots,J\}$ und $k \in \{1,\ldots,V_{ij}\}$. Machen wir X_{ij} zum Zentrum unserer Betrachtung stoßen wir allerdings auf das Problem, dass eine Voraussetzung aus den bisher betrachteten Modellen nicht mehr erfüllt ist: Da der Inhalt der zweiten Urne auch von V_{ij} beeinflusst wird, hängt die Varianz von S_{ij} nicht mehr nur von θ_i ab und ist insbesondere nicht konstant über die Perioden hinweg. Für unseren ersten Interpretationsansatz erhalten wir konkret:

$$\begin{aligned}\mathrm{Var}\left(X_{ij}\right) &= \mathrm{Var}\left(\frac{S_{ij}}{V_{ij}}\right)\\ &= \frac{1}{V_{ij}^2}\sum_{k=1}^{V_{ij}} \mathrm{Var}\left(S_{ij}^{(k)}\right)\\ &= \frac{1}{V_{ij}}\mathrm{Var}\left(S_{ij}^{(1)}\right)\end{aligned}$$

Wobei der letzte Schritt daraus folgt, dass die Elementarschäden (bedingt auf Θ_i) unabhängig und identisch verteilt sind. Das bedeutet für festes i $\mathrm{Var}\left(S_{ij}^{(k)}\right) = \mathrm{Var}\left(S_{ij'}^{(k')}\right)$ für alle $j, j' \in \{1,\ldots,J\}, k \in \{1,\ldots,V_{ij}\}, k' \in \{1,\ldots,V_{ij'}\}$.

Erwartungswert und Varianz der S_{ij} hängen damit also sowohl von der Risikodisposition der Klasse i als auch von der Gewichtung V_{ij} ab. Für die normierten Zufallsvariablen X_{ij} hatten wir in (4.28) gezeigt, dass der Erwartungswert nicht von der Gewichtung abhängt. Die Varianz der X_{ij} hängt neben Θ_i sehr wohl von V_{ij} ab.

Das Ergebnis der vorangegangenen Rechnung können wir für den zweiten Interpretationsansatz übernehmen. Dann lautet es $Var(X_{ij}) = \sigma^2(\Theta_i)/w_{ij}$. Die so gegebene Zufallsstruktur ist charakteristisch für das Bühlmann-Straub-Modell:

Annahme 4.12 (Erwartungswert und Varianz der X_{ij}) *Die Risikodisposition θ_i der i-ten Risikoklasse wird bestimmt von der Zufallsvariablen Θ_i. Darauf bedingt sind die X_{ij}, $j = 1,\ldots,J$ unabhängig mit dem gleichen Erwartungswert*

$$\mathrm{E}\left[X_{ij}\middle|\Theta_i\right] = \mu\left(\Theta_i\right) \tag{4.29}$$

Die Varianz hängt zusätzlich vom Gewicht w_{ij} ab:

$$\operatorname{Var}\left(X_{ij}\middle|\Theta_i\right) = \frac{\sigma^2\left(\Theta_i\right)}{w_{ij}} \tag{4.30}$$

Annahme 4.13 (Unabhängigkeit der Θ_i) *Für $X_i = (X_{i1}, \ldots, X_{iJ})'$ sind die Tupel $(\Theta_1, X_1), (\Theta_2, X_2), \ldots, (\Theta_I, X_I)$ unabhängig. Die $\Theta_i, i = 1, \ldots, I$, die die Risikodisposition der Klasse i bestimmen, sind unabhängig und identisch verteilt.*

4.3.1 Der inhomogene Credibility-Schätzer im Bühlmann-Straub-Modell

Bisher haben wir unter dem homogenen Credibility-Schätzern (für ein Θ-Funktional $\mu(\Theta)$ aufgrund von vorliegenden Daten X Entscheidungsfunktionen mit der folgenden Struktur verstanden:

$$\widehat{\mu(\Theta)} = \alpha\overline{X} + (1-\alpha)\mu_0$$

wobei $\overline{X}$ als Mittelwert die Daten, die wir bisher über das betrachtete Individuum gesammelt hatten, zusammenfasste. μ_0 gab Auskunft über das durchschnittliche Verhalten des Kollektivs, zu dem das Individuum gehörte. Das Credibility-Gewicht α machte eine Aussage darüber, wie die beiden zu gewichten seien mit folgenddem Aufbau:

$$\alpha = \frac{N}{N+\kappa}$$

und κ hieß Credibility-Koeffizient. Für das Bühlmann-Staub-Modell müssen wir offensichtlich einige Anpassungen vornehmen.

Im Modell von Bühlmann hatten wir in Abschnitt 3.2.1 den durchschnittlichen Schaden des i-ten Individuums, $\overline{X}_i = \frac{1}{J}\sum_{j=1}^{J} X_{ij}$ als suffiziente Statistik und erwartungstreuen Schätzer für $\mu(\Theta_i)$ in den Credibility-Schätzer eingehen lassen. Schon rein intuitiv vermuten wir hier, dass die unterschiedliche Gewichtung der einzelnen Perioden hier Einfluss haben soll. Ferner ist Erwartungstreue eine wünschenswerte Eigenschaft. Wir definieren daher den gewichteten normierten Mittelwert

$$\breve{X}_i := \sum_{j=1}^{J} \frac{w_{ij}}{w_{i\bullet}} X_{ij} \tag{4.31}$$

Dabei bedeute $\bullet$, dass über den ersetzten Index summiert wird, also $w_{i\bullet} := \sum_{j=1}^{J} w_{ij}$.

Ebenso wie beim Schätzer für $\mu(\Theta_i)$ müssen wir für das Credibility-Gewicht κ Anpassungen vornehmen. Bisher galt, je länger wir das Individuum kennen (je größer J, die Anzahl der Jahre aus denen wir Daten über es haben) desto stärker gewichten wir die Daten, die wir über das Individuum haben. Da die einzelnen Jahre nun mit unterschiedlichen Gewichten w_{ij} versehen sind, liegt es nahe, die Erfahrungsdaten abhängig von ihrem kummulierten Gewicht $w_{i\bullet}$ einzubeziehen.

Es sei wie gehabt

$$\sigma^2 := \mathrm{E}[\sigma^2(\Theta_i)], \qquad \mu_0 = \mathrm{E}\left[\mu(\Theta_i)\right], \qquad \tau^2 := \mathrm{Var}\left(\mu\left(\Theta_i\right)\right)$$

wobei, wie in Annahme 4.12 $\sigma^2(\Theta_i) = w_{ij} Var(X_{ij}|\Theta_i)$ die durch das Gewicht normierte Varianz der bedingten Verteilung der X_{ij} gegeben Θ_i ist.

Diese Überlegungen führen uns zu einem einfachen Kandidaten für den (inhomogenen) Credibility-Schätzer im Bühlmann-Straub-Modell:

$$\widehat{\widehat{\mu(\Theta_i)}} = \alpha_i \check{X}_i + (1-\alpha_i)\,\mu_0 \quad \text{mit } \alpha_i = \frac{w_{i\bullet}}{w_{i\bullet} + \frac{\sigma^2}{\tau^2}} \tag{4.32}$$

Satz 4.14 *Der Schätzer aus Gleichung 4.32 ist der inhomogene Credibility-Schätzer in Bühlmann-Straub-Modell.*

Beweis Satz 4.14: Entsprechend Korollar 4.9 genügt es zu zeigen, dass $\widehat{\widehat{\mu(\Theta_i)}}$ die Gleichungen (4.20) und (4.21) erfüllt. Die erste Gleichung, die im Wesentlichen die Erwartungstreue einfordert, bereitet keine großen Probleme:

$$\begin{aligned}
\mathrm{E}\left[\mu\left(\Theta\right) - \widehat{\widehat{\mu(\Theta)}}\right] &= \mathrm{E}\left[\mu\left(\Theta\right) - \alpha_i \check{X}_i + (1-\alpha_i)\,\mu_0\right] \\
&= \mu_0 - \alpha_i \underbrace{\mathrm{E}\left[\check{X}_i\right]}_{=\mu_0} + (1-\alpha_i)\,\mu_0 \\
&= 0
\end{aligned}$$

Für den zweiten Teil müssen wir zeigen

$$\mathrm{Cov}\left(\mu\left(\Theta\right), X_{kj}\right) = \mathrm{Cov}\left(\widehat{\widehat{\mu(\Theta)}}, X_{kj}\right) \tag{4.33}$$

und zwar für alle $X_{kj}, k = 1, \ldots, I, j = 1, \ldots, J$. Dabei können wir die beiden Fälle

1) $(k,j) \in \{i\} \times \{1,\ldots,J\}$ und

2) $(k,j) \in (\{i\} \times \{1,\ldots,J\})^c = (\{1,\ldots,i-1,i+1,\ldots,I\} \times \{1,\ldots,J\}$

unterscheiden.

Zu 2: Nach Voraussetzung 4.13 sind die X_{kj} für $(k,j) \in (\{1,\ldots,i-1,i+1,\ldots,I\} \times \{1,\ldots,J\}$ unabhängig von $\Theta_i, X_{i1},\ldots,X_{iJ}$. Daraus folgt das Verschwinden von beiden Seiten von Gleichung (4.33), womit der zweite Fall abgehandelt ist.

Zu 1: Auf die linke Seite von Gleichung (4.33) wenden wir den Satz von der totalen Kovarianz an:

$$\begin{aligned}
\operatorname{Cov}\left(\mu(\Theta_i), X_{ij}\right) &= \mathrm{E}\left[\operatorname{Cov}\left(\mu(\Theta_i), X_{ij} \middle| \Theta_i\right)\right] + \operatorname{Cov}\left(\mu(\Theta_i), \mathrm{E}\left[X_{ij} \middle| \Theta_i\right]\right) \\
&= \operatorname{Cov}(\mu(\Theta_i), \mu(\Theta_i)) \\
&= Var(\mu(\Theta_i)) = \tau^2
\end{aligned}$$

wobei $\operatorname{Cov}\left(\mu(\Theta_i), X_{ij} \middle| \Theta_i\right) = 0$ wegen der bedingten Unabhänigkeit der X_{kj} für festes k. Mit der gleichen Rechnung erhalten wir übrigens

$$\operatorname{Cov}\left(X_{il}, X_{ij}\right) = \tau^2 \text{ für } l \neq j \tag{$*$}$$

Diese (unbedingten) Kovarianzen der Schadenforderungen benötigen wir, um die rechte Seite der Gleichung auszurechnen:

$$\operatorname{Cov}\left(X_{il}, X_{ij}\right) = \begin{cases} \dfrac{\sigma^2}{w_{il}} + \tau^2 \text{ falls } l = k \\ \tau^2 \text{ falls } l \neq k \end{cases} \tag{4.34}$$

Für $l \neq k$ hatten wir das das Ergebnis in $(*)$ gesehen. Im Fall $l = k$ betrachten wir die (unbedingte) Varianz von X_{il}:

$$\begin{aligned}
\operatorname{Var}(X_{il}) &= \mathrm{E}\,\underbrace{\left[\operatorname{Var}(X_{il} | \Theta_i)\right]}_{=\frac{\sigma(\Theta_i^2)}{w_{il}} \text{ n.V.}} + \operatorname{Var}\left(\mathrm{E}\left[X_{ij} \middle| \Theta_i\right]\right) \\
&= \frac{\sigma^2}{w_{il}} + \tau^2
\end{aligned} \tag{4.35}$$

Die Werte für die (unbedingten) Kovarianzen können wir in der folgenden Rechnung im zweiten Schritt einsetzen. Zusätzlich zum τ^2 kommt im Fall

$l=k$ noch ein $\frac{\sigma^2}{w_{il}}$ hinzu.

$$\begin{aligned}
\operatorname{Cov}\left(\widehat{\widehat{\mu(\Theta)}}, X_{kj}\right) &= \operatorname{Cov}\left(\alpha_i \sum_{l=1}^{J} \frac{w_{il}}{w_{i\bullet}} X_{il} + (1-\alpha_i)\,\mu_0, X_{ij}\right) \\
&= \alpha_i \sum_{l=1}^{J} \frac{w_{il}}{w_{i\bullet}} \operatorname{Cov}\left(X_{il}, X_{ij}\right) \\
&= \alpha_i \left(\sum_{l=1}^{J} \frac{w_{il}}{w_{i\bullet}} \tau^2 + \frac{w_{ij}}{w_{i\bullet}} \frac{\sigma^2}{w_{ij}}\right) \\
&= \frac{w_{i\bullet}}{w_{i\bullet} + \frac{\sigma^2}{\tau^2}} \left(\tau^2 + \frac{\sigma^2}{w_{i\bullet}}\right) \\
&= \frac{\tau^2\left(w_{i\bullet} + \frac{\sigma^2}{\tau^2}\right)}{w_{i\bullet} + \frac{\sigma^2}{\tau^2}} = \tau^2
\end{aligned}$$

Also gilt Gleichung (4.33) auch für den ersten Fall ($(k,j) \in \{i\} \times \{1,\ldots,J\}$), da linke und rechte Seite übereinstimmen. Die Normalengleichungen sind somit erfüllt, $\widehat{\widehat{\mu(\Theta_i)}}$ ist der gesuchte Credibility-Schätzer.

□

Satz 4.15 *Der quadratische Verlust des (inhomogenen) Credibility-Schätzers im Modell von Bühlmann-Straub beträgt*

$$\mathrm{E}\left[\left(\mu(\Theta) - \widehat{\widehat{\mu(\Theta_i)}}\right)^2\right] = (1-\alpha_i)\tau^2 = \alpha_i \frac{\sigma^2}{w_{i\bullet}} \tag{4.36}$$

Würden wir nur μ_0 als Schätzer für $\mu(\Theta_i)$ verwenden betrüge der quadratische Verlust τ^2. Der quadratische Verlust von $\check{X}_i$ ist $\sigma^2/w_{i\bullet}$.

Beweis Satz 4.15:

$$\begin{aligned}
\mathrm{E}\left[\left(\mu(\Theta_i) - \widehat{\widehat{\mu(\Theta_i)}}\right)^2\right] &= \mathrm{E}\left[\left(\alpha_i \mu(\Theta_i) + (1-\alpha_i)\mu(\Theta_i) - \alpha_i \check{X}_i - (1-\alpha_i)\mu_0\right)^2\right] \\
&= \alpha_i^2\, \mathrm{E}\left[\left(\mu(\Theta_i) - \check{X}_i\right)^2\right] + (1-\alpha_i)^2\, \mathrm{E}\left[\left(\mu(\Theta_i) - \mu_0\right)^2\right]
\end{aligned}$$

Der Term $\alpha_i(1-\alpha_i)\,\mathrm{E}\left[\left(\mu(\Theta_i) - \check{X}_i\right)\left(\mu(\Theta_i) - \mu_0\right)\right]$, der beim Ausmultiplizieren entsteht, verschwindet, da $\mu(\Theta_i) - \check{X}_i$ und $\mu(\Theta_i)$ unabhängig sind. Denn dann erscheint $\mathrm{E}[\mu(\Theta_i) - \check{X}_i]$ als Faktor, doch dieser Term ist null.

Die verbleibenden Erwartungswerte sind die quadratischen Verluste von μ_0 und $\breve{X}_i$. Wenn wir die berechnet haben, sind wir fertig, denn der Rest der Rechnung verläuft genau wie in der Herleitung von Satz 3.4 (der quadratische Verlust des Credibility-Schätzers im Modell von Bühlmann). Für den quadratischen Verlust von $\breve{X}_i$ beginnen wir, indem wir die Definition von $\breve{X}_i$ einsetzen:

$$\mathrm{E}\left[\left(\mu\left(\Theta_i\right)-\breve{X}_i\right)^2\right]=\mathrm{E}\left[\mathrm{E}\left[\left(\sum_{j=1}^{J}\frac{w_{ij}}{w_{i\bullet}}X_{ij}-\mu(\Theta_i)\right)^2\middle|\Theta_i\right]\right]$$

bedingt auf Θ_i sind die X_{ij} für $j=1,\ldots,J$ unabhängig und identisch verteilt, das gilt auch für $X_{ij}-\mu(\Theta_i)$, also

$$\begin{aligned}
&=\mathrm{E}\left[\sum_{j=1}^{J}\frac{w_{ij}^2}{w_{i\bullet}^2}\mathrm{E}\left[\left(X_{ij}-\mu(\Theta_i)\right)^2\middle|\Theta_i\right]\right]\\
&=\mathrm{E}\left[\sum_{j=1}^{J}\frac{w_{ij}^2}{w_{i\bullet}^2}\mathrm{Var}\left(X_{ij}\middle|\Theta_i\right)\right]\\
&=\mathrm{E}\left[\sum_{j=1}^{J}\frac{w_{ij}^2}{w_{i\bullet}^2}\frac{\sigma^2\left(\Theta_i\right)}{w_{ij}}\right]\\
&=\frac{\sigma^2}{w_{i\bullet}}
\end{aligned}$$

Bleibt noch der quadratische Verlust von μ_0:

$$\mathrm{E}\left[\left(\mu\left(\Theta_i\right)-\mu_0\right)^2\right]=\mathrm{Var}\left(\mu\left(\Theta_i\right)\right)=\tau^2$$

□

Mithilfe der Normalengleichungen konnten wir also zeigen, dass ein plausibler Schätzer tatsächlich der Credibility-Schätzer war. Er ist ein gewichtetes Mittel aus dem kollektiven Durchschnittswert μ_0 und der durchschnittlichen Schadenhistorie des i-ten Individduums, $\breve{X}_i$. Die genaue Form der Gewichtung in diesem Fall hatten wir zu Beginn des Abschnitts damit motiviert, dass sie dem Credibility-Gewicht des Modells von Bühlmann ähneln soll.

Eine andere Plausibilitätsüberlegung, die zur gleichen Gewichtung führt, ist als Allgemeines Intuitives Prinzip bekannt. Dabei werden μ_0 und $\breve{X}_i$ anhand der Kehrwerte ihres quadratischen Verlustes gewichtet. Dies folgt der einfachen Überlegung, dass ein Schätzer umso schwächer gewichtet werden sollte, je höher sein erwarteter Verlust (im Vergleich zu anderen verfügbaren Schätzern) ist.

Das Gewicht für μ_0 ist dann:

$$\frac{\frac{1}{\tau^2}}{\frac{w_{i\bullet}}{\sigma^2}+\frac{1}{\tau^2}} = \frac{\frac{\sigma^2}{\tau^2}}{w_{i\bullet}+\frac{\sigma^2}{\tau^2}} = (1-\alpha_i)$$

Für $\check{X}_i$ erhalten wir entsprechend:

$$\frac{\frac{w_{i\bullet}}{\sigma^2}}{\frac{w_{i\bullet}}{\sigma^2}+\frac{1}{\tau^2}} = \frac{w_{i\bullet}}{w_{i\bullet}+\frac{\sigma^2}{\tau^2}} = \alpha_i$$

Im Übrigen verhält sich der Credibility-Koeffizient $\kappa = \frac{\sigma^2}{\tau^2}$ wie man es auch intuitiv erwartet: Für relativ[4] große τ^2, also wenn Θ breit streut, wird mehr Gewicht auf auf die Erfahrungen mit Individuum i gelegt. Dominiert hingegen σ^2 (d.h. ist der quadratische Fehler des Schätzers $\check{X}_i$ groß, insbesondere die Streuung von $\check{X}_i$ groß im Vergleich zu der von Θ) wird das Kollektivmittel stärker betont.

4.3.2 Der homogene Credibility-Schätzer im Bühlmann-Straub-Modell

In Tabelle 4.2 hatten wir die vorangegangen Ergebnisse über Credibility-Schätzer als Projektionen auf Unterräume des $\mathbf{L^2}$ zusammengefasst. Insbesondere folgte aus dem Projektions-Satz 4.7, dass

$$\widehat{\widehat{\mu(\Theta_i)}}^{hom} = \mathrm{Pro}\left(\widehat{\widehat{\mu(\Theta_i)}} | \mathbf{L_e}(\mathbf{X})\right) \tag{4.37}$$

wobei $\mathbf{L_e}(\mathbf{X})$ wie gehabt aus den Linearkombination $\sum_{i=1}^{I}\sum_{j=1}^{J} a_{ij}X_{ij}$ besteht, deren Erwartungswert gleich $\mathrm{E}[\mu(\Theta_i)]$ ist.

Um diese Beziehung direkt ausnutzen zu können benötigen wir folgende allgemeine Aussage über die orthogonale Projektion:

[4] Gemeint ist hier, relativ zu σ^2

Satz 4.16 *Es sei* $\mathbf{H}$ *ein Hilbertraum,* $\mathbf{U}$ *ein affiner Unterraum. Für* $X,Y \in \mathbf{H}$ *und Skalare* a,b *gilt dann:*

$$\mathrm{Pro}\left(\frac{aX+bY}{a+b}\middle|\mathbf{U}\right) = \frac{a\,\mathrm{Pro}\,(X|\mathbf{U}) + b\,\mathrm{Pro}\,(Y|\mathbf{U})}{a+b} \tag{4.38}$$

Für den homogenen Credibility-Schätzer liefert das sofort:

$$\begin{aligned}
\widehat{\mu(\Theta)}^{hom} &= \mathrm{Pro}(\widehat{\mu(\Theta_i)}|\mathbf{L_e}(\mathbf{X})) \\
&= \mathrm{Pro}(\alpha_i\breve{X}_i + (1-\alpha_i)\,\mu_0|\mathbf{L_e}(\mathbf{X})) \\
&= \alpha_i\,\mathrm{Pro}\left(\breve{X}_i\middle|\mathbf{L_e}(\mathbf{X})\right) + (1-\alpha_i)\,\mathrm{Pro}\,(\mu_0|\mathbf{L_e}(\mathbf{X}))
\end{aligned}$$

Da $\breve{X}_i$ ein erwartungstreuer Schätzer für $\mu(\Theta_i)$ ist und eine Linearkombination aus den $X_{ij}, j = 1,\ldots,J$ ist $\breve{X}_i \in \mathbf{L_e}(\mathbf{X})$.

Um den homogenen Credibility-Schätzer zu finden brauchen wir also noch den richtigen Schätzer für μ_0.

Satz 4.17 *Im Bühlmann-Straub-Modell gilt*

$$\widehat{\widehat{\mu_0}} := \mathrm{Pro}\,(\mu_0|\mathbf{L_e}(\mathbf{X})) = \sum_{k=1}^{I}\frac{\alpha_k}{\alpha_\bullet}\breve{X}_k \tag{4.39}$$

wobei $\alpha_k = w_{k\bullet}/(w_{k\bullet} + \sigma^2/\tau^2)$ *das bekannte Credibility-Gewicht im Credibility-Schätzer für* $\mu(\Theta_k)$ *ist.*

Zur Herleitung verfahren wir wie im vorangegangenen Kapitel und suchen einen plausibelen Schätzer von dem wir dann zeigen, dass er der richtige ist. Der Satz sagt bereits, dass es nicht das gewichtete Mittel der Schadenverläufe $\sum_{I,J}(w_{ij}/w_{\bullet\bullet})X_{ij}$ ist.

Eine weitere Statistik, die die Daten etwas weniger reduziert und doch wünschenswerte Eigenschaften bereithält ist $\breve{X}_i$, der normierte gewichtete Mittelwert der Schadenerfahrung des Individuums i. Es sei

$$\widehat{\mu_0} = \mathrm{Pro}\left(\mu_0\middle|\mathbf{L_e}(\breve{\mathbf{X}}_\mathbf{1},\ldots,\breve{\mathbf{X}}_\mathbf{I})\right) = \sum_{i=1}^{I} a_i\breve{X}_i \tag{4.40}$$

Entsprechend Definition 4.4 muss für $\widehat{\mu_0}$ mindestens gelten:

$$\mu_0 - \widehat{\mu_0} \perp \breve{X}_k - \breve{X}_r \Leftrightarrow \mathrm{Cov}\left(\widehat{\mu_0},\breve{X}_k\right) = \mathrm{Cov}\left(\widehat{\mu_0},\breve{X}_r\right) \tag{4.41}$$

wobei $r,k \in \{1,\ldots,I\}$.

Aus $\mathrm{Cov}(\widehat{\mu_0},\breve{X}_k) = \sum_{i=1}^{I} a_i \,\mathrm{Cov}(\breve{X}_i,\breve{X}_k) = a_r \,\mathrm{Var}(\breve{X}_k)$ folgt

$$a_r \,\mathrm{Var}(\breve{X}_r) = a_k \,\mathrm{Var}(\breve{X}_k) \quad \forall r,k \in \{1,\ldots,I\} \tag{4.42}$$

Da wir hier nicht $\mu(\Theta_i)$ sondern μ_0 schätzen wollen, ist die Varianz gleich dem quadratischen Verlust des Schätzers. Bereits in Gleichung (4.35) hatten wir die Varianz der X_{ij} mit $\frac{\sigma^2}{w_{ij}} + \tau^2$ angegeben. Wir wissen auch $\mathrm{Cov}(X_{ij},X_{il}) = \tau^2$ für $j \neq l$.

$$\begin{aligned}
\mathrm{Var}\left(\breve{X}_i\right) &= \sum_{j=1}^{J}\sum_{l=1}^{J} \frac{w_{ij}w_{il}}{w_{i\bullet}^2} \mathrm{Cov}\left(X_{ij},X_{il}\right) \\
&= \sum_{j=1}^{J}\sum_{l=1}^{J} \frac{w_{ij}w_{il}}{w_{i\bullet}^2} \tau^2 + \sum_{j=1}^{J} \frac{w_{ij}^2}{w_{i\bullet}^2} \frac{\sigma^2}{w_{ij}} \\
&= \tau^2 + \frac{\sigma^2}{w_{i\bullet}} \quad \text{und weiter} \\
&= \frac{\tau^2 w_{i\bullet} + \sigma^2}{w_{i\bullet}} \\
&= \tau^2 \alpha_i^{-1}
\end{aligned}$$

Aus Gleichung (4.42) wissen wir, dass $\mathrm{Var}(\breve{X}_r) \sim 1/a_r$ gelten muss. Da $\widehat{\mu_o} \in \mathbf{L_e}(\breve{\mathbf{X}}_\mathbf{1},\ldots,\breve{\mathbf{X}}_\mathbf{I})$ muss es auch erwartungstreu sein, also tritt zur Bedingung (4.42) noch $\sum_{i=1}^{I} a_i = 1$ hinzu. Daraus ergibt sich

$$\widehat{\mu_0} = \sum_{i=1}^{I} \frac{\alpha_i}{\alpha_\bullet} \breve{X}_i \tag{4.43}$$

Beweis Satz4.17: Dazu müssen wir zeigen, dass unser Kandidat $\widehat{\mu_0}$ tatsächlich der homogene Credibility-Schätzer für μ basierend auf dem Informationsvektor X ist. Dabei enthält X nun alle $X_{ij}, i = 1,\ldots,I, j = 1,\ldots,J$. Laut Korollar 4.11 ist äquivalent dazu

$$\mathrm{E}\left[\mu_0 - \widehat{\mu_0}\right] = 0 \tag{4.44}$$

sowie für alle $\widehat{\mu} \in \mathbf{L_e}(\mathbf{X})$

$$\mathrm{Cov}\left(\widehat{\mu_0} - \mu_0, \widehat{\mu_0}\right) = \mathrm{Cov}\left(\widehat{\mu_0} - \mu_0, \widehat{\mu}\right). \tag{4.45}$$

Gleichung (4.44) war bereits Voraussetzung, als wir den besten Schätzer aufgrund der $\check{X}_1, \ldots, \check{X}_I$ gesucht haben und ist daher erfüllt.

Die linke Seite von (4.45) ist gleich der Varianz von $\widehat{\mu_0}$.

$$\begin{aligned} \operatorname{Var}\left(\widehat{\mu_0}\right) &= \sum_{i=1}^{I} \frac{\alpha_i^2}{\alpha_\bullet^2} \operatorname{Var}\left(\check{X}_i\right) = \sum_{i=1}^{I} \frac{\alpha_i^2}{\alpha_\bullet^2} \frac{\tau}{\alpha_i} \\ &= \sum_{i=1}^{I} \frac{\alpha_i}{\alpha_\bullet^2} \tau = \frac{\tau}{\alpha_\bullet} \end{aligned} \tag{4.46}$$

Da $\widehat{\mu_0}$ ein Schätzer für μ_0 ist gibt dieser Wert übrigens auch den quadratischen Fehler des Schätzers an.

Anstatt getrennt zu fordern, dass die Elemente aus $\widehat{\mu} \in \mathbf{L_e}(\mathbf{X})$ Linearkombinationen der X_{ij} sein müssen, die Erwartungstreu sind, können wir auch direkt fordern, dass sie normierte Linearkombinationen sein sollen. Das heißt für $a_{ij} \geq 0$ ist

$$\widehat{\mu} = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{a_{ij}}{a_{\bullet\bullet}} X_{ij} \tag{4.47}$$

Dies können wir in die rechte Seite von (4.45) einsetzen. Unter Ausnutzung der uns bekannten Kovarianzen der X_{ij} Gleichung(4.34) untereinander können wir folgende Rechnung durchführen:

$$\operatorname{Cov}\left(\widehat{\mu_0} - \mu_0, \widehat{\mu}\right) = \sum_{i=1}^{I} \frac{\alpha_i}{\alpha_\bullet} \sum_{j=1}^{J} \frac{w_{ij}}{w_{i\bullet}} \operatorname{Cov}\left(X_{ij}, \sum_{k=1}^{I} \sum_{l=1}^{J} \frac{a_{kl}}{a_{\bullet\bullet}} X_{kl}\right)$$

es gilt $\operatorname{Cov}(X_{ij}, X_{kl}) = 0$ wenn $i \neq k$ nach Modelannahme

$$
\begin{aligned}
&= \sum_{i=1}^{I} \frac{\alpha_i}{\alpha_\bullet} \sum_{j=1}^{J} \frac{w_{ij}}{w_{i\bullet}} \sum_{l=1}^{J} \frac{a_{il}}{a_{\bullet\bullet}} \operatorname{Cov}\left(X_{ij}, X_{il}\right) \\
&= \sum_{i=1}^{I} \frac{\alpha_i}{\alpha_\bullet} \sum_{j=1}^{J} \frac{w_{ij}}{w_{i\bullet}} \left(\sum_{l=1}^{J} \frac{a_{il}}{a_{\bullet\bullet}} \tau^2 + \frac{a_{ij}}{a_{\bullet\bullet}} \frac{\sigma^2}{w_{ij}} \right) \\
&= \tau^2 \sum_{i=1}^{I} \frac{\alpha_i}{\alpha_\bullet} \sum_{j=1}^{J} \frac{w_{ij}}{w_{i\bullet}} \frac{a_{i\bullet}}{a_{\bullet\bullet}} + \sum_{i=1}^{I} \frac{\alpha_i}{\alpha_\bullet} \frac{\sigma^2}{w_{i\bullet}} \frac{a_{i\bullet}}{a_{\bullet\bullet}} \\
&= \tau^2 \sum_{i=1}^{I} \frac{\alpha_i a_{i\bullet}}{\alpha_\bullet a_{\bullet\bullet}} + \sum_{i=1}^{I} \frac{\frac{\sigma^2}{w_{i\bullet} + \frac{\sigma^2}{\tau^2}}}{\alpha_\bullet} \frac{a_{i\bullet}}{a_{\bullet\bullet}} \\
&= \tau^2 \sum_{i=1}^{I} \frac{\alpha_i a_{i\bullet}}{\alpha_\bullet a_{\bullet\bullet}} + \tau^2 \sum_{i=1}^{I} \frac{1-\alpha_i}{\alpha_\bullet} \frac{a_{i\bullet}}{a_{\bullet\bullet}} \\
&= \tau^2 \sum_{i=1}^{I} \frac{\alpha_i a_{i\bullet}}{\alpha_\bullet a_{\bullet\bullet}} + \frac{\tau^2}{\alpha_\bullet} - \tau^2 \sum_{i=1}^{I} \frac{\alpha_i a_{i\bullet}}{\alpha_\bullet a_{\bullet\bullet}} = \frac{\tau^2}{\alpha_\bullet}
\end{aligned}
$$

Damit ist gezeigt, dass $\widehat{\mu_0}$ der Credibility-Schätzer für μ_0 basierend auf allen Daten $X_{ij}, i = 1, \ldots, I, j = 1, \ldots, J$ ist.

$\square$

Als direkte Konsequenz erhalten wir:

Korollar 4.18 *Im Bühlmann-Straub-Modell ist der homogene Credibility-Schätzer für $\mu(\Theta_i)$ gegeben durch*

$$
\widehat{\widehat{\mu(\Theta ß)}}^{hom} = \alpha_i \breve{X}_i + (1-\alpha_i)\widehat{\mu_0} \tag{4.48}
$$

wobei $\alpha_i = \frac{w_{i\bullet}}{w_{i\bullet} + \frac{\sigma^2}{\tau^2}}$ und $\widehat{\mu_0}$ wie in Gleichung (4.39).

Wir sind im letzten Kapitel den Beweis für Satz 3.8 schuldig geblieben. **Beweis Satz 3.8:** Das Modell von Bühlmann-Straub unterscheidet sich vom Bühlmann Modell nur durch die Hinzunahme des Volumenmaßes w_{ij}. Setzen wir diese Gewichte gleich eins für alle $i = 1, \ldots, I, j = 1, \ldots, J$ erhalten wir

$$\check{X}_i = \sum_{j=1}^{J} \frac{w_{ij}}{w_{i\bullet}} X_{ij} = \sum_{j=1}^{J} \frac{1}{J} X_{ij} = \overline{X_i}$$

$$\alpha_i = \frac{w_{i\bullet}}{w_{i\bullet} + \frac{\sigma^2}{\tau^2}} = \frac{J}{J + \frac{\sigma^2}{\tau^2}} = \alpha$$

$$\widehat{\mu_0} = \sum_{i=1}^{I} \frac{\alpha_i}{\alpha_\bullet} \check{X}_i = \sum_{i=1}^{I} \frac{1}{I} \overline{X_i} = \overline{\overline{X}}$$

□

Als nächstes berechnen wir quadratische Verlust $\mathrm{E}[(\widehat{\widehat{\mu(\Theta ß)}}^{hom} - \mu(\Theta_i))^2]$ des homogenen Credibility-Schätzers im Bühlmann-Straub-Modell. Satz 4.7(ii) gibt uns über die Abstandsverähltnisse bei Orthogonalprojektionen (hier die aus Gleichung (4.37)) folgendes an:

$$\mathrm{E}\left[\left(\widehat{\widehat{\mu(\Theta_i)}}^{hom} - \mu(\Theta_i)\right)^2\right]$$
$$= \underbrace{\mathrm{E}\left[\left(\widehat{\widehat{\mu(\Theta_i)}}^{hom} - \widehat{\widehat{\mu(\Theta_i)}}\right)^2\right]}_{(1)} + \underbrace{\mathrm{E}\left[\left(\widehat{\widehat{\mu(\Theta_i)}} - \mu(\Theta_i)\right)^2\right]}_{(2)}$$

Setzen wir in (1) die Schätzer ein vereinfacht sich der Term zu $(1-\alpha)^2 Var(\widehat{\mu_0})$. Dabei ist uns $Var(\widehat{\mu_0})$ (= quadr. Verlust von $\widehat{\mu_0}$)aus Gleichung (4.46) als $\tau^2/\alpha_\bullet$ bekannt.

(2) ist der bekannte quadratische Verlust des inhomogenen Credibility-Schätzers $(1-\alpha_i)\tau^2)$ (Satz 4.15)). Das führt uns direkt zu

Satz 4.19 *Der quadratische Verlust des homogenen Credibility-Schätzers im Bühlmann-Straub-Modell beträgt*

$$\mathrm{E}\left[\left(\widehat{\widehat{\mu(\Theta_i)}}^{hom} - \mu(\Theta_i)\right)^2\right] = \frac{(1-\alpha)^2\tau^2}{\alpha_\bullet} + (1-\alpha_i)\tau^2 = \tau(1-\alpha)\left(1 + \frac{1-\alpha_i}{\alpha_\bullet}\right) \tag{4.49}$$

Bereits ganz zu Anfang(1.1) hatten wir darauf aufmerksam gemacht, dass aus Sicht der Versicherer die Schadendeckung eine wichtige Eigenschaft von Prämien ist. Für die zu prognostizierende Periode bedeutet das, dass der Schätzer für $\mu(\Theta_i)$ erwartungstreu sein soll. Doch ich der Rückschau können wir objektiv feststellen, ob unsere gewählten Prämien diese Anforderungen erfüllen.

Sei w_{ij} hier neben dem Gewicht des einzelnen Schadenfalls auch ein Faktor, der entsprechend dem Gewicht (Volumenmaß) für die Berechnung der Versicherungsprämie herangezogen wird. Es gilt dann die

Satz 4.20 *Balance Eigenschaft: Im Modell von Bühlmann-Straub erfüllt der homogene Credibility-Schätzer die Eigenschaft:*

$$\sum_{i=1}^{I}\sum_{j=1}^{J} w_{ij}\,\widehat{\widehat{\mu(\Theta_i)}}^{hom} = \sum_{i=1}^{I}\sum_{j=1}^{J} w_{ij}X_{ij} \tag{4.50}$$

Beweis Satz 4.20:

$$\begin{aligned}
& & \sum_{i=1}^{I}\sum_{j=1}^{J} w_{ij}\left(\alpha_i\check{X}_i+(1-\alpha_i)\widehat{\mu_0}\right) &= \sum_{i=1}^{I}\sum_{j=1}^{J} w_{ij}X_{ij} \\
&\Leftrightarrow & \sum_{i=1}^{I}\left(-(1-\alpha_i)\check{X}_i+(1-\alpha_i)\widehat{\mu_0}\right) &= 0 \\
&\Leftrightarrow & \sum_{i=1}^{I} w_{i\bullet}(1-\alpha_i)\left(\widehat{\mu_0}-\check{X}_i\right) &= 0 \\
&\Leftrightarrow & \frac{\sigma^2}{\tau^2}\sum_{i=1}^{I}\alpha_i\left(\widehat{\mu_0}-\check{X}_i\right) &= 0 \\
&\Leftrightarrow & \alpha_\bullet\widehat{\mu_0}-\sum_{i=1}^{I}\alpha_i\check{X}_i &= 0
\end{aligned}$$

□

Wir machen darauf aufmerksam, dass für die Gültigkeit des Beweises von Satz 4.20 die genaue Wahl von σ^2 und τ^2 keinen Einfluss hat.

4.4 Schätzen der Modellparameter

Als Vorteil des homogenen Credibility-Schätzers haben wir bisher immer angegeben, dass er einen Schätzer für das Kollektivmittel μ_0 beinhaltet. Verwenden wir den inhomogenen Credibility-Schätzer wird μ_0 als bekannt vorausgesezt. Stellen wir uns eine Gruppe von Autofahrern vor, die Mitglied einer Versicherung sind, wird aber sofort klar, dass μ_0 nicht „bekannt " sein kann. In der Praxis müssen wir uns daher fast immer auf Schätzwerte (im empirischen Sinn) oder Einschätzungen (im Bayes-Sinn) verlassen.

Neben μ_0, für das wir bereits gute Schätzer gefunden haben, bleiben immer mindestens zwei Parameter näher zu bestimmen:

- σ^2: Für das Risiko i ist $\sigma^2(\Theta_i)$ die Varianz zwischen den Schadenforderungen der unterschiedlichen Perioden (normalisiert auf das Gewicht 1). $\sigma^2 := \mathrm{E}[\sigma^2(\Theta_i)]$ ist der Durchschnittswert dieser Varianz, gemittelt über das ganze Portefeuille.
- τ^2: Für die erwarteten Mittelwerte der Risiken i ist $\tau^2 := \mathrm{Var}(\mu(\Theta_i))$, ein Maß dafür, wie breit die unterschiedlichen individuellen Risikodispositionen gestreut sind.

Wir haben die Credibility-Modelle wiederholt als Zwei-Urnen Modelle charakterisiert. Im Rahmen dieser Beschreibung ist τ^2 die Varianz bei Ziehung aus der ersten, σ^2 bei Ziehung aus der zweiten Urne.

4.4.1 Ein erwartungstreuer und konsistenter Schätzer für σ^2

Ein bekannter Schätzer[5] für die Varaianz ist (erweitert um die Gewichtung w_{ij})

$$S_i := \frac{1}{J-1} \sum_{j=1}^{J} w_{ij} \left(X_{ij} - \breve{X}_i \right)$$

[5] etwa in [Als02], S.48

Für diesen modifizierten Schätzer wollen wir die bedingte und unbedingte Erwartungstreue zeigen:

$$
\begin{aligned}
S_i &= \frac{1}{J-1} \sum_{j=1}^{J} \left(X_{ij} - \mu(\Theta_i) + \mu(\Theta_i) - \check{X}_i\right)^2 \\
&= \frac{1}{J-1} \left(\sum_{j=1}^{J} w_{ij} \left(X_{ij} - \mu(\Theta_i)\right)^2 + \sum_{j=1}^{J} 2 w_{ij} Z_{ij} + w_{i\bullet} \left(\mu(\Theta_i) - \check{X}_i\right)^2 \right) \\
&= \frac{1}{J-1} \left(\sum_{j=1}^{J} w_{ij} \left(X_{ij} - \mu(\Theta_i)\right)^2 - w_{i\bullet} \left(\mu(\Theta_i) - \check{X}_i\right)^2 \right)
\end{aligned}
$$

Denn für den Zwischenterm $Z_i j$ gilt:

$$
\begin{aligned}
&\sum_{j=1}^{J} 2 w_{ij} \left(X_{ij} - \mu(\Theta_i)\right) \left(\mu(\Theta_i) - \check{X}_i\right) \\
&= \sum_{j=1}^{J} 2 w_{ij} \left(X_{ij}\mu(\Theta_i) + \mu(\Theta_i)\check{X}_i - X_{ij}\check{X}_i - \mu(\Theta_i)^2\right) \\
&= -2 w_{i\bullet} \left(\mu(\Theta_i)^2 - 2\check{X}_i \mu(\Theta_i) + \check{X}_i^{\,2}\right)
\end{aligned}
$$

Betrachten wir nun auf beiden Seiten den auf Θ_i bedingten Erwartungswert:

$$
\begin{aligned}
\mathrm{E}\left[S_i \middle| \Theta_i\right] &= \frac{1}{J-1} \left(\sum_{j=1}^{J} w_{ij} \underbrace{\mathrm{Var}\left(X_{ij} \middle| \Theta_i\right)}_{=\sigma^2(\Theta_i)} - w_{i\bullet} \sum_{j=1}^{J} \frac{w_{ij}}{w_{i\bullet}^2} w_{ij} \underbrace{\mathrm{Var}\left(X_{ij} \middle| \Theta_i\right)}_{=\sigma^2(\Theta_i)} \right) \\
&= \frac{1}{J-1} \left(J\sigma^2(\Theta_i) - \sigma^2(\Theta_i) \right) = \sigma^2(\Theta_i)
\end{aligned}
$$

Durch erneutes Anwenden des Erwartungswertes erhalten wir daraus direkt, dass für den unbedingten Erwartungswert gilt

$$
\mathrm{E}\left[S_i\right] = \mathrm{E}\left[\sigma^2(\Theta_i)\right] = \sigma^2
$$

Das bedeutet jedes S_i ist ein erwartungstreuer Schätzer für σ^2. Wir können also einen gewichteten und normierten Mittelwert der S_i benutzen, um σ^2 zu schätzen. Bisher haben wir die Gewichtsverteilung oft am Kehrwert des quadratischen Fehlers des Schätzers festgemacht. Diese Prozedur würde jedoch bei einem Schätzer für ein zweites Moment sehr kompliziert, da der quadratische Fehler dieses Schätzers auch vom vierten Moment der X_{ij} abhängen. Bühlmann und Gisler schlagen deshalb, motiviert durch eine Plausibilitätsüberlegung anhand der Normalverteilung, vor, die S_i alle gleich zu gewichten und gelangen so zu:

$$\widehat{\sigma^2} = \frac{1}{I} \sum_{i=1}^{I} \frac{1}{J-1} \sum_{j=1}^{J} w_{ij} \left(X_{ij} - \breve{X}_i\right)^2 \tag{4.51}$$

Wir gehen davon aus, dass die Varianzen der S_i durch einen Wert b beschränkt sind[6], also $\mathrm{Var}(S_i) \leq b \quad \forall 1 = 1, \ldots, I$. Aus Chebychefs Ungleichung folgt dann:

$$P\left[\left|\widehat{\sigma^2} - \sigma^2\right| > k\right] \leq \frac{\mathrm{Var}\left(\widehat{\sigma^2}\right)}{k} = \frac{\frac{1}{I^2}\sum_{i=1}^{I} \mathrm{Var}\left(S_i\right)}{k} \leq \frac{b}{Ik} \xrightarrow[I \to \infty]{} 0$$

Also konvergiert $\widehat{\sigma^2}$ gegen σ^2 in Wahrscheinlichkeit. Einen solchen Schätzer nennen wir konsistent.

Der Schätzer beruht auf der empirischen Summe der Fehlerquadrate innerhalb (=within, w) der einzelnen Risikogruppen. Für diese Größe schreiben wir kurz:

$$SS_w := \sum_{i=1}^{I} \sum_{j=1}^{J} w_{ij} \left(X_{ij} - \breve{X}_i\right)^2 \tag{4.52}$$

4.4.2 Ein erwartungstreuer (und konsistenter) Schätzer für τ^2

Um einen Schätzer für $\tau = \mathrm{Var}(\mu(\Theta - i))$ zu finden wollen wir[7] die empirische Summe der Fehlerquadrate zwischen (=between b) den Gruppen näher betrachten:

$$\begin{gathered} SS_b := \sum_{i=1}^{I} w_{i\bullet} \left(\breve{X}_i - \breve{\breve{X}}\right)^2 \\ \text{wobei } \breve{\breve{X}} := \sum_{i=1}^{I} \frac{w_{i\bullet}}{w_{\bullet\bullet}} \breve{X}_i \end{gathered} \tag{4.53}$$

[6] Die gegenteilige Annahme würde bedeuten, wir schätzen σ^2 eventuell mit unbeschränkt großem quadratischen Verlust

[7] dem Vorgehen von [KDDG09] S.218 folgend

Es ist

$$\begin{aligned}
\operatorname{Var}\left(\breve{\breve{X}}\right) &= \sum_{i=1}^{I} \frac{w_{i\bullet}^2}{w_{\bullet\bullet}^2} \operatorname{Var}\left(\breve{X}_i\right) \\
&= \sum_{i=1}^{I} \left(\frac{w_{i\bullet}^2}{w_{\bullet\bullet}^2} \tau^2 \frac{w_{i\bullet} + \frac{\sigma^2}{\tau^2}}{w_{i\bullet}} \right) \\
&= \tau^2 \sum_{i=1}^{I} \frac{w_{i\bullet}^2}{w_{\bullet\bullet}^2} + \frac{\sigma^2}{w_{\bullet\bullet}}
\end{aligned} \tag{4.54}$$

Wegen der Unabhänigkeit der $\breve{X}_i, \breve{X}_j \forall\, i \neq j$ ist

$$\begin{aligned}
\operatorname{Cov}\left(\breve{X}_i, \breve{\breve{X}}\right) &= \frac{w_{i\bullet}}{w_{\bullet\bullet}} \operatorname{Var}\left(\breve{X}_i\right) \\
&= \tau^2 \frac{w_{i\bullet}}{w_{\bullet\bullet}} + \frac{\sigma^2}{w_{\bullet\bullet}}
\end{aligned} \tag{4.55}$$

Ergänzen wir in der Klammer in (4.53) $\mu_0 - \mu_0$ und bilden den Erwartungswert erhalten wir

$$\begin{aligned}
\mathrm{E}\left[SS_b\right] &= \sum_{i=1}^{I} w_{i\bullet} \left[\operatorname{Var}\left(\breve{X}_i\right) + \operatorname{Var}\left(\breve{\breve{X}}\right) - 2 \operatorname{Cov}\left(\breve{X}_i, \breve{\breve{X}}\right) \right] \\
&= \sum_{i=1}^{I} w_{i\bullet} \left[\tau^2 + \frac{\sigma^2}{w_{i\bullet}} + \tau^2 \sum_{k=1}^{I} \frac{w_{k\bullet}^2}{w_{\bullet\bullet}^2} + \frac{\sigma^2}{w_{\bullet\bullet}} - 2 \left(\tau^2 \frac{w_{i\bullet}}{w_{\bullet\bullet}} + \frac{\sigma^2}{w_{\bullet\bullet}} \right) \right] \\
&= \tau^2 \sum_{i=1}^{I} w_{i\bullet} \left[1 + \sum_{k=1}^{I} \frac{w_{k\bullet}^2}{w_{\bullet\bullet}^2} - 2 \frac{w_{i\bullet}}{w_{\bullet\bullet}} \right] + \sigma^2 \sum_{i=1}^{I} w_{i\bullet} \left[\frac{1}{w_{i\bullet}} - \frac{1}{w_{\bullet\bullet}} \right] \\
&= \tau^2 \left[w_{\bullet\bullet} + w_{\bullet\bullet} \sum_{k=1}^{I} \frac{w_{k\bullet}^2}{w_{\bullet\bullet}^2} - 2 \sum_{i=1}^{I} w_{i\bullet} \frac{w_{i\bullet}}{w_{\bullet\bullet}} \right] + \sigma^2 (I-1) \\
&= \tau^2 \left[w_{\bullet\bullet} - \sum_{i=1}^{I} \frac{w_{i\bullet}^2}{w_{\bullet\bullet}} \right] + \sigma^2 (I-1)
\end{aligned}$$

Damit, und mir der Erwartungstreue von $\widehat{\sigma^2}$ ist klar, dass

$$\widehat{\widehat{\tau^2}} := \frac{\sum_{i=1}^{I} w_{i\bullet} \left(\breve{X}_i - \breve{\breve{X}} \right)^2 - \sigma^2 (I-1)}{w_{\bullet\bullet} - \sum_{i=1}^{I} \frac{w_{i\bullet}^2}{w_{\bullet\bullet}}}$$

ein erwartungstreuer Schätzer für τ^2 ist. Es ließe sich ferner zeigen, dass $\widehat{\widehat{\tau^2}}$ auch konsistent ist. Dafür müssen wir zusätzlich fordern $\sum_{i=1}^{I}\left(\frac{w_{i\bullet}}{w_{\bullet\bullet}}\right)^2 \xrightarrow[I\to\infty]{} 0$. Das heiës t, keines der Risiken i dominiert die anderen. Da wir die Eigenschaft nicht verwenden werden, unterbleibt der Beweis.

Durch die Konstruktion dieses Schätzers kann es vorkommen, dass $\widehat{\widehat{\tau^2}} < 0$. Um das zu vermeiden definieren wir

$$\widehat{\tau^2} := \max\left\{\widehat{\widehat{\tau^2}}, 0\right\} \tag{4.56}$$

So ist $\widehat{\tau^2}$ allerdings nichtmehr erwartungstreu. Zur Interpretation eines negativen $\widehat{\widehat{\tau^2}}$ (oder $\widehat{\tau^2}$=0) kommen wir noch kurz im nächsten Abschnitt.

4.4.3 Der empirische Credibility-Schätzer im Bühlmann-Straub-Modell

Setzen wir die gefundenen erwartungstreuen (und konsistenten) Schätzer nun in den homogenen Credibility-Schätzer (Gl. (4.48)) ein, erhalten wir

Definition 4.21 *Der empirische Credibility-Schätzer im Modell von Bäuhlmann-Straub ist gegeben durch*

$$\widehat{\widehat{\mu(\Theta_i)}}^{emp} = \widehat{\alpha}_i \check{X}_i + (1-\widehat{\alpha}_i)\,\widehat{\mu_0} \tag{4.57}$$

Dabei ist

$$\widehat{\alpha}_i := \frac{w_{i\bullet}}{w_{i\bullet} + \frac{\widehat{\sigma^2}}{\widehat{\tau^2}}}$$

und

$$\widehat{\mu_0} := \sum_{i=1}^{I} \frac{\widehat{\alpha}_i}{\widehat{\alpha_\bullet}} \check{X}_i \tag{4.58}$$

Ein negatives $\widehat{\widehat{\tau^2}}$ interpretieren wir in der Weise, dass $\mathrm{Var}(\mu(\Theta_i)) = 0$ (dass also bei $\widehat{\widehat{\tau^2}} < 0$ ein Schätzfehler vorliegt). Das bedeutet, die Daten geben es nicht her, einen Unterschied im Erwartungswert der X_{ij} für die unterschiedlichen $i = 1,\dots,I$ zu vermuten. In diesem Fall setzen wir $\alpha_i = 0 \,\forall\, i = i = 1,\dots,I$. Somit ist der Schätzer für das Kollektivmittel der Credibility-Schätzer für alle $\mu(\Theta_i), i = 1,\dots,I$. Wir brauchen dann allerdings auch eine Näherung für $\widehat{\widehat{\mu_0}}$.

Schon in der Formel für τ^2 verwenden wir mit $\check{\check{X}}$ einen Schätzer für μ_0, von dem wir wissen, dass er nicht optimal im Sinne der Credibility-Theorie ist. Denn mit $\widehat{\mu_0} := \sum_{i=1}^{I} \alpha_i/\alpha_{\bullet}\check{X}_i$ kennen wir bereits den Credibility-Schätzer für μ_0. Das Problem ist allerdings, dass wir für seine Definition über die α_i bereits auf den Strukturparameter τ^2 zugreifen müssen.

Alternativ zum Vorgehen in [BG05] (einen anderen erwartungstreuen Schätzer zu verwenden) schlagen [KDDG09] (S.221f) vor, zuerst mit einem groben Schätzwert z.B. für τ^2 zu beginnen. Anschließend wird mit dem Credibility-Schätzer für μ_0 die Schätzung für τ^2 wiederholt. Dieser Prozess kann vielfach wiederholt werden. Da der „Schätzer“ in diesem Fall nicht nur direkt auf den Ergebnissen eines statistischen Experiments beruht (oder einer Statistik davon), sondern weitere Berechnungen einbezieht, wird er auch Pseudo-Schätzer genannt. An dieser Stelle wird der starke Praxis-Bezug der Credibility-Theorie deutlich: Die Konvergenz dieses Prozesses (d.h. der Folgen von Schätzwerte $(\tau^{2(n)})_{n\in\mathbb{N}}, (\widehat{\mu_0^{(n)}})_{n\in\mathbb{N}}$) ist schwer zu beweisen. Doch lassen sich durch Computereinsatz damit schnell und robust Schätzwerte ermitteln.

Wie wir gesehen haben, gingen die genauen Werte von τ^2 und σ^2 nicht in den Beweis von Satz 4.20 ein. Die Balance-Eigenschaft gilt daher auch für den empirischen Credibility-Schätzer. Über dessen Erwartungstreue können wir wegen der Verwendung des Schätzers $\widehat{\tau^2}$ ohne weiteres keine genauen Angaben mehr machen. Aus der Balance-Eigenschaft erhalten wir aber immerhin:

Satz 4.22 *Gewichtet nach den Risikogruppen ist der empirische Credibility-Schätzer erwartungstreu:*

$$\mathrm{E}\left[\sum_{i=1}^{I} \frac{w_{i\bullet}}{w_{\bullet\bullet}} \widehat{\widehat{\mu(\Theta_i)}}^{emp}\right] = \mu_0 \tag{4.59}$$

Beweis Satz 4.22: Nach Satz 4.20 gilt:

$$\sum_{i=1}^{I}\sum_{j=1}^{J} w_{ij}\,\widehat{\mu(\Theta_i)}^{emp} = \sum_{i=1}^{I}\sum_{j=1}^{J} w_{ij}X_{ij}$$

$$\Leftrightarrow \qquad \sum_{i=1}^{I} \frac{w_{i\bullet}}{w_{\bullet\bullet}}\widehat{\mu(\Theta_i)}^{emp} = \sum_{i=1}^{I}\sum_{j=1}^{J} \frac{w_{ij}}{w_{\bullet\bullet}}X_{ij}$$

Erwartungswertbildung auf beiden Seiten führt zur gewünschten Aussage.

□

4.5 Ein Datensatz mit poisson-verteilten Zufallsgrößen

An einem Beispiel aus der KFZ-Haftpflicht-Versicherung[8] sollen die bis hierher gefundenen Ergebnisse abschließend einmal angewendet werden. Untersucht wird das Auftreten von (großen) Schadenereignissen in einem Versicherungsportefeuille, das mehrer Regionen umfasst.

Bekanntermaßen kann die Häufigkeit von Zufallsereignissen in einer bestimmten Periode oft gut durch die Poisson-Verteilung modelliert werden. Es sei daher N_{ij} die Anzahl der Schadenfälle in Region i im Jahr j. Als Volumenmaß haben wir die Zahl der „Jahreseinheiten" w_{ij}. Eine Jahreseinheit ist dabei ein Fahrzeug, das über ein Jahr versichert ist[9]. Wir kommen daher zu folgenden Modellannahmen:

Annahme 4.23 *Bedingt auf der Zufallsvariablen Θ_i sind die Schadenhäufigkeiten $N_{ij}, j = 1, \dots, J$ unabhängig und identisch verteilt gemäß einer Poisson-Verteilung zum Parameter $\lambda_{ij} := w_{ij}\Theta_i\lambda_0$.*

Annahme 4.24 *Für $N_i = (N_{i1}, \dots, N_{iJ})'$ sind die Tupel $(\Theta_1, N_1), (\Theta_2, N_2), \dots, (\Theta_I, N_I)$ unabhängig. Die $\Theta_i, i = 1, \dots, I$, sind unabhängig und identisch verteilt mit Erwartungswert 1.*

[8] Freundlicherweise wurde vom Landwirtschaftlichen Versicherungsverein Münster a.G. ein Datensatz zur Verfügung gestellt.

[9] es könnten auch zwei Fahrzeuge sein, die jeweils ein halbe jahr versichert sind etc. In der englischsprachigen Literatur wird der Ausdruck „years at risk" verwendet.

λ_0 nennen wir Tarif-Niveau.

Um Annahme 4.12 des Bühlmann-Straub-Modells zu erfüllen, betrachten wir die relativen Schadenhäufigkeiten in Region i und Jahr j, also $F_{ij} := N_{ij}/w_{ij}$. Für diese gilt, wie erforderlich

$$\mathrm{E}\left[F_{ij}\middle|\Theta_i\right] = \Theta_i\lambda_0$$
$$\mathrm{Var}\left(F_{ij}\middle|\Theta_i\right) = \frac{\Theta_i\lambda_0}{w_{ij}}$$

Wir können also nach einem Credibility-Schätzer für das Θ-Funktional μ suchen, das den Erwartungswert einer Poisson-Verteilung zum Parameter $\lambda_0\Theta$ ausgibt, also $\mu(\Theta_i) = \Theta_i\lambda_0 = \sigma^2(\Theta_i)$, da bei der Poisson-Verteilung die Varianz gleich dem Erwartungswert ist. Durch das Benennen der Verteilungsfamilie, zu der die N_{ij} gehören, erhalten wir also weitergehende Aussagen zu den Strukturparametern des Modells. Es gilt

$$\begin{aligned}
\lambda_0 &:= \mathrm{E}\left[\mu(\Theta_i)\right] \\
\sigma^2 &:= \mathrm{E}\left[\sigma^2(\Theta_i)\right] = \lambda_0 \\
\tau^2 &:= \mathrm{Var}\left(\mu(\Theta_i)\right) = \lambda_0^2\,\mathrm{Var}(\Theta_i) = \lambda_0^2
\end{aligned}$$

Satz 4.14 (inhomogener) bzw. Korollar 4.18 (homoger Credibility-Schätzer) liefern dann sofort (mit ˘ wie gehabt als Symbol für die gewichtete Summation über den weggelassenen Index)

Korollar 4.25 *Im durch die Annahmen 4.23 und 4.24 beschriebenen Modell ist der inhomogene Credibility-Schätzer gegeben durch*

$$\widehat{\widehat{\mu(\Theta_i)}} = \alpha_i\breve{F}_i + (1-\alpha_i)\,\lambda_0 \quad \textit{mit } \alpha_i = \frac{w_{i\bullet}}{w_{i\bullet} + \frac{\lambda_0}{\tau^2}} \tag{4.60}$$

Der homogene Credibility-Schätzer ist

$$\widehat{\widehat{\mu(\Theta_i)}}^{hom} = \alpha_i\breve{F}_i + (1-\alpha_i)\,\widehat{\widehat{\lambda}}_0 \quad \textit{mit } \widehat{\widehat{\lambda}}_0 = \sum_{i=1}^{I}\frac{\alpha_i}{\alpha_\bullet}\breve{F}_i \tag{4.61}$$

Um unseren Datensatz verarbeiten zu können, müssen wir nun noch die Strukturparameter schätzen. A priori wissen wir bereits $\lambda_0 = \sigma^2$.

Wir wenden nun das am Ende von Abschnitt 4.4.3 beschriebene stufenweise Verfahren an:

Es sei

$$c := \left(w_{\bullet\bullet} - \sum_{i=1}^{I} \frac{w_{i\bullet}^2}{w_{\bullet\bullet}} \right)^{-1}$$

Für den Start wählen wir

$$\widehat{\widehat{\lambda_0^{(0)}}} := \breve{\breve{F}}$$

Dann ist $\widehat{\widehat{\tau^2}}$ zu Beginn entsprechend Gleichung (4.4.2)

$$\widehat{\widehat{\tau^{2(0)}}} := c \left(\sum_{i=1}^{I} w_{i\bullet} \left(\breve{F}_i - \widehat{\widehat{\lambda_0^{(0)}}} \right)^2 - \widehat{\widehat{\lambda_0^{(0)}}} (I-1) \right)$$

und als Hilfsgröße für jedes $i = 1, \ldots, I$

$$\alpha_i^{(0)} := \frac{w_{i\bullet}}{w_{i\bullet} + \frac{\widehat{\widehat{\lambda_0^{(0)}}}}{\widehat{\widehat{\tau^{2(0)}}}}}$$

Setzen wir nun

$$\widehat{\widehat{\lambda_0^{(1)}}} := \sum_{i=1}^{I} \frac{\alpha_i^{(0)}}{\alpha_{i\bullet}^{(0)}} \breve{F}_i$$

für den Übergang zur nächsten Stufe können wir iterativ auch die anderen Werte neu berechnen.

Der Datensatz beinhaltet (u.a.) die Zahl der Jahreseinheiten in der KFZ-Haftpflicht-Versicherung der LVM in 27 Regionen. Für jede Region ist für den Untersuchungszeitraum (2008 bis drittes Quartal 2011) ausgewiesen, wieviel Schadenereignisse es gab und wieviele davon Großschäden (>50.000EUR) waren. Credibility-Theory ist hier ein gutes Analyseinstrument, da groß̈s e Schadenereignisse zwar selten auftreten, aber einen großen Einfluss auf die Finanzen eines Versicherungsunternehmens haben. Wie bereits an den Formeln für die Pseudo-Schätzer abzulesen war, genügen uns hier Daten, die über die gesamte Periode kumuliert sind.

Für die Auswertung verwenden wir die Pseudo-Schätzer aus dem fünften Durchlauf. Dann ist $\widehat{\widehat{\lambda_0}} = 0.0003358903$, $\widehat{\widehat{\tau^2}} = 0.0000006639488$[10].

[10]beide scheinen sehr schnell zu konvergieren. Bereits nach dem dritten Schritt verändern sich die Werte nichtmehr in den ersten sechs gültigen Stellen.

Der Variationskoeffizient von $F_i^{ind} = \mu(\Theta_i)$ kann hieraus abgelesen werden.

$$\widetilde{\kappa} := \frac{\lambda_0^2}{\tau^2} = \frac{1}{\mathrm{Var}(\Theta_i)} = \frac{1}{\mathrm{Var}(\Theta_i)}\left(CoVa\left(F^{ind}\right)\right)^{-2} \tag{4.62}$$

In unserem Beispiel beträgt $CoVa\left(F^{ind}\right)$ etwa $7{,}671\%$. Das bedeutet, dass der auf die Daten zurückzuführende Unterschied zwischen den Regionen eher klein ist.

Region	$w_{i\bullet}$	$N_{i\bullet}$	$F_{i\bullet}*$	$\widehat{\widehat{\mu(\Theta_i)}}^{hom} *$	α_i
1	77171	26	0,337	0,336	0,132
2	50271	26	0,517	0,352	0,090
3	206887	52	0,251	0,311	0,290
4	59643	22	0,369	0,339	0,105
5	154948	58	0,374	0,345	0,234
6	108102	46	0,426	0,352	0,176
7	13325	6	0,450	0,339	0,026
8	125565	34	0,271	0,323	0,199
9	118302	45	0,380	0,344	0,190
10	148283	48	0,324	0,333	0,227
11	130465	48	0,368	0,342	0,205
12	117010	35	0,299	0,329	0,188
13	149601	44	0,294	0,326	0,228
14	173147	63	0,364	0,343	0,255
15	18618	9	0,483	0,341	0,035
16	135070	51	0,378	0,345	0,211
17	91458	33	0,361	0,340	0,153
18	7915	1	0,126	0,333	0,015
19	22464	6	0,267	0,333	0,043
20	92368	29	0,314	0,333	0,154
21	47277	7	0,148	0,320	0,085
22	133270	48	0,360	0,341	0,209
23	20458	4	0,196	0,330	0,039
24	49010	12	0,245	0,328	0,088
25	74980	30	0,400	0,344	0,129
26	61586	21	0,341	0,336	0,109
27	39219	10	0,255	0,330	0,072

Tabelle 4.3: Auswertung des Datensatzes für große Schadenfälle

Tabelle 4.3 zeigt zum Abschluss die Auswertung unseres Datensatzes. Dabei sind die relativen Häufigkeiten und der Credibility-Schätzer (mit * markiert) der besseren Lesbarkeit halber mit 1000 multipliziert worden. Die Tabelle zeigt deutlich, wie der Credibility-Schätzer große (unwahrscheinliche) Ausschläge bei den Daten mit einem niedrigen Credibility-Gewicht belegt und einen Schätzwert nahe $\widehat{\widehat{\lambda}}_0$ annimmt.

Literaturverzeichnis

[Als02] ALSMEYER, Gerold: *Mathematische Statistik*. Münster : Skripten zur Mathematischen Statistik, Inst. f. Mathemat. Statistik, 2002

[BG05] BÜHLMANN, Hans ; GISLER, Alois: *A Course in Credibility Theory and its Applications*. Berlin, Heidelberg, New York : Springer, 2005

[Bic64] BICHSEL, Fritz: Erfahrungstarifierung in der Motorfahrzeug-Haftpflichtversicherung. In: *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker* 1 (1964), S. 119–130

[BS70] BÜHLMANN, Hans ; STRAUB, Erwin: Glaubwürdigkeit für Schadensätze. In: *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker* 1 (1970), S. 111–133

[Gür62] GÜRTLER, Max: Selektion und Antiselektion in der Versicherungswirtschaft. In: *Zeitschrift für Betriebswirtschaftslehre* 32 (1962), S. 605–614

[HS87] HEILMANN, Wolf-Rüdiger ; SCHRÖTER, Klaus: Eine Bemerkung über bedingte Wahrscheinlichkeiten, bedingte Erwartungswerte und bedingte Unabhängigkeit. In: *Blätter der Deutschen Gesellschaft für Versicherungs- und Finanzmathematik e.V.* 18 (1987), S. 119–126

[KDDG09] KAAS, Rob ; DENUIT, Michael ; DHAENE, Jan ; GOOVAERTS, Marc: *Modern actuarial risk theory : using R*. Berlin, Heidelberg, New York : Springer, 2009

[Kre99] KREMER, Erhard: *Applied Risk Theory*. Aachen : Shaker, 1999

[Meh62] MEHRING, Johannes: Die Schadenstruktur in der Kraftfahrt–Haftpflichtversicherung. In: *Blätter der Deutschen Gesellschaft* *für Versicherungs- und Finanzmathematik e.V.* 1 (1962), S. 23–41

[Sun99] SUNDT, Bjorn: *An introduction to non-life insurance mathematics*. Karlsruhe : VVW, 1999

Erklärung des Studierenden

Hiermit versichere ich, dass ich die vorliegende Arbeit mit dem Titel

Credibility Theory

selbständig verfasst habe, und dass ich keine anderen Quellen und Hilfsmittel als die angegebenen benutzt habe und dass die Stellen der Arbeit, die anderen Werken – auch elektronischen Medien – dem Wortlaut oder Sinn nach entnommen wurden, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht worden sind.

Ort/Datum Unterschrift