DAV
Deutsche
Aktuarvereinigung e.V.

Leitfaden für das Grundwissen

Angewandte Stochastik

Köln, Januar 2026

Prof. Dr. Torsten Becker, Prof. Dr. Christian Heumann, Dr. Richard Herrmann,
Prof. Dr. Viktor, Sandor, Dr. Dominik Schäfer, Dr. Fabian Winter

Präambel

Die Arbeitsgruppe [Name der Arbeitsgruppe] des Ausschusses [Name des Ausschusses] der Deutschen Aktuarvereinigung e. V. (DAV) hat den vorliegenden Ergebnisbericht erstellt.[1]

Anwendungsbereich

Der Ergebnisbericht behandelt Fragestellungen zu [möglichst konkrete, kurze Dar-stellung der Fragestellung] und betrifft Aktuare [in der Rolle als … (z. B. Verantwortlicher Aktuar, Sachverständiger, Aktuar bei einer Wirtschaftsprüfungsgesellschaft, Versicherungsmathematischen Funktion etc.) bei der Ausführung aktuarieller Aufgaben … (z. B. im Rahmen des Jahresabschlusses eines Lebensversicherers, zur Bewertung von fondsgebundenen Versicherungen mit Mindestgarantie etc.)] [Sofern zutreffend:] Der Anwendungsbereich umfasst die Produkte/Produktkategorien der …, die durch … gekennzeichnet sind.

Der Ergebnisbericht ist an die Mitglieder und Gremien der DAV zur Information über den Stand der Diskussion und die erzielten Erkenntnisse gerichtet und stellt keine berufsständisch legitimierte Position der DAV dar. [2]

Verabschiedung

Dieser Ergebnisbericht ist durch den Ausschuss [Name des Ausschusses] am [Datum] verabschiedet worden.

1 Der Ausschuss dankt der Arbeitsgruppe [Name der Arbeitsgruppe] ausdrücklich für die geleistete Arbeit, namentlich [Name 1] (Leitung), [Name 2], …

2 Die sachgemäße Anwendung des Ergebnisberichts erfordert aktuarielle Fachkenntnisse. Dieser Ergebnisbericht stellt deshalb keinen Ersatz für entsprechende professionelle aktuarielle Dienstleistungen dar. Aktuarielle Entscheidungen mit Auswirkungen auf persönliche Vorsorge und Absicherung, Kapitalanlage oder geschäftliche Aktivitäten sollten ausschließlich auf Basis der Beurteilung durch eine(n) qualifizierte(n) Aktuar DAV/Aktuarin DAV getroffen werden.

Vorbemerkung

Dieser Leitfaden skizziert Inhalte und verzichtet zugunsten der besseren Lesbarkeit teilweise auf die mathematische Genauigkeit bzw. Vollständigkeit.

Die exakten Formulierungen können in der Literatur nachgeschlagen werden.

Bei der Erstellung dises Leitfadens haben mitgewirkt:

Prof. Dr. Torsten Becker

Prof. Dr. Christian Heumann

Dr. Richard Herrmann

Prof. Dr. Viktor Sandor

Dr. Dominik Schäfer

Dr. Fabian Winter

Inhaltsverzeichnis

1 Deskriptive Statistik[1]

Grobübersicht:

Kerninhalte

- Ordne die Grundlagen statistischer Arbeit ein.
- Begründe die Bedeutung der Qualitätssicherung und Bereinigung von Daten als Grundlage statistischer Arbeit.
- Beschreibe die Besonderheiten im Umgang mit großen Datenmengen.
- Wende einfache Verfahren der deskriptiven Datenanalyse an

1.1 Grundlagen statistischer Arbeit

Kerninhalte

- Untersuchungseinheit
- Merkmal
- Merkmalstypen
- Datengewinnung
- Studiendesigns
- Datenquellen
- Datenvorverarbeitung

1.1.1 Untersuchungseinheit

Daten werden stets an bestimmten Objekten erhoben. Diese nennt man die **Untersuchungseinheiten** ω. Als **Grundgesamtheit** Ω oder **Population** betrachtet man die Menge aller Untersuchungseinheiten, über die Erkenntnisse gewonnen werden sollen. Die Grundgesamtheit kann endlich oder unendlich oder nur hypothetisch sein. Dies ist natürlich kontextspezifisch. Betrachtet man nur eine Teilmenge der Grundgesamtheit, so spricht man von **Teilgesamtheit** oder **Teilpopulation**. Diese Teilgesamtheiten sind oft **Stichproben**, die nach gewissen Kriterien aus der Grundgesamtheit durch einen Zufallsprozess ausgewählt werden und die ein möglichst repräsentatives Bild der Grundgesamtheit darstellen sollen.

1.1.2 Merkmal

Eigenschaften der Untersuchungseinheiten werden als **Merkmale** oder **statistische Variable** X bezeichnet[2]. Die statistischen Einheiten sind die **Merkmalsträger**. Es sind nur Merkmale sinnvoll, die verschiedene **Merkmalswerte** annehmen können. Diese Werte werden auch als **Merkmalsausprägungen** bezeichnet.

[1] Teile dieser Darstellung basieren auf Fahrmeir, L., Heumann, C., Künstler, R., Pigeot, I., Tutz, G., Statistik — Der Weg zur Datenanalyse, Springer, Berlin (2016).

[2] In der Disziplin des maschinellen Lernens werden Merkmale meist als **Features** bezeichnet

1.1.3 Einteilung der Merkmale

Die Einteilung der Merkmale kann durch verschiedene Typisierungen erfolgen. Die für die Anwendung statistischer Methoden wichtigste Einteilung verwendet das **Skalenniveau** eines Merkmals. Eine Unterteilung in **nominale** Merkmale, **ordinale** Merkmale und **metrische** oder **kardinalskalierte** Merkmale ist für statistische Zwecke ausreichend. Bei metrischen Merkmalen gibt es noch eine Feinaufteilung in **intervallskalierte** Merkmale und **verhältnisskalierte** Merkmale. Nominale und ordinale Merkmale (mit wenigen Ausprägungen) werden oft als **kategoriale** Merkmale bezeichnet.

Eine weitere Typisierung teilt die Merkmale in **qualitative** und **quantitative** Merkmale ein. Ein qualitatives Merkmal ist dabei ein Merkmal mit endlich vielen Ausprägungen, das höchstens ordinalskaliert ist. Ordinale Merkmale können aber auch schwach quantitativ sein, insbesondere wenn sie sehr viele Ausprägungen besitzen. Metrische Merkmale sind stets quantitativ.

Eine weitere Unterscheidung ist die Unterteilung in **diskrete** und **stetige** Merkmale. Diskrete Merkmale können eine endliche oder abzählbar unendliche Anzahl von Ausprägungen haben, stetige Merkmale eine überabzählbar unendliche Anzahl von Ausprägungen. In der Praxis sind alle stetigen Merkmale **quasi-stetig**, da die Messgenauigkeit nicht beliebig groß ist oder weil die Merkmale nur in feinen Abstufungen aufgezeichnet werden.

1.1.4 Datengewinnung

Typische Methoden zur Datengewinnung sind das geplante **Experiment**, die **Befragung** und die **Beobachtung**. Letztere Methode ist generisch und umfasst alle Arten von Datenaufzeichnungen, die manuell oder automatisch erfolgen können. Werden Daten in einem kontinuierlichen Prozess gewonnen, so spricht man von einem **Datenstrom**[3].

1.1.5 Studiendesigns

Bei einer **Querschnittsstudie** werden die Merkmale an den Untersuchungseinheiten zu einem bestimmten Zeitpunkt erhoben. Wird ein Merkmal einer einzelnen Untersuchungseinheit über einen längeren Zeitraum aufgezeichnet, spricht man von einer **Zeitreihe**. Werden Merkmale mehrerer Untersuchungseinheiten an mehreren Zeitpunkten erfasst, spricht man von einer **Panelstudie** oder **Längsschnittstudie**. Die resultierenden Daten heißen dann entsprechend **Querschnittsdaten**, **Paneldaten** oder **Längsschnittdaten**.

1.1.6 Datenquellen

Üblicherweise werden Daten in **Datenbanken**, **Datenseen**[4], aber auch häufig noch in einzelnen **Tabellen** aufbewahrt. Datenseen streben die Verwaltung aller Daten eines Unternehmens in einer zentralen Plattform an, um wichtige wiederkehrende Aufgaben wie Analysen, Visualisierungen, Berichte und Vorhersagen lösen zu können. In Datenseen werden sowohl strukturierte Daten, wie sie in gängigen relationalen Datenbanken gespeichert werden, als auch auch semi- oder unstrukturierte Daten, die in einzelnen Dateien unterschiedlichster Formate aufbewahrt werden, gespeichert.

[3] Datenströme werden durch die immer größere Verfügbarkeit von beispielsweise Online-Daten, Videodaten, Stimmdaten oder Sensordaten zunehmend bedeutender und sind ein Teil des sogenannten Big Data-Phänomen

[4] Ein Datensee ist ein zentraler Speicherungsort für Daten verschiedenster Form (Rohdaten, Textdaten, Bilddaten, strukturierte und transformierte Daten)

1.1.7 Datenvorverarbeitung

Die **Datenvorverarbeitung** ist integraler Bestandteil jeglicher statistischer Arbeit. Aus den Datenquellen müssen die Daten zunächst so aufbereitet werden, dass sie sinnvoll mit statistischen Methoden weiter ausgewertet werden können. In der Regel müssen zunächst adäquate Datenzusammenstellungen durch mehr oder weniger komplexe Datenbankabfragen erfolgen. Zunehmend sind Daten auch in sogenannten NoSQL Datenbanken oder Big Tables gespeichert oder treten gar als Datenströme auf. Mit zunehmender Datenmenge und Analysekomplexität gewinnen dabei (halb-)automatisierte Vorverarbeitungen an Bedeutung. Um Rohdaten aufzubereiten werden u.a. folgende Operationen[5] durchgeführt: **Sortierung**, **Aggregation**, **Selektion, Filterung und Löschung von Duplikaten**, **Transformationen** (Normalisierung, Standardisierung, Skalierung), **Binning**, **Gewichtung** und Erzeugung **abgeleiteter Merkmale**. Danach muss eine **Bereinigung** der Daten erfolgen. Insbesondere muss eine Strategie für **fehlende Daten** entwickelt werden, etwa durch **multiple Imputation**, wenn komplexere statistische Verfahren zum Einsatz kommen sollen. Für viele moderne statistische Verfahren, insbesondere des maschinellen Lernens, ist es notwendig, die Daten geeignet zu **partitionieren** in **Trainings-, Validierungs- und Testdaten**. Dazu werden zufällige, möglicherweise balancierte oder stratifizierte, Stichproben gezogen. Am Ende der Datenvorverarbeitung müssen die Daten für viele statistische Verfahren in Form einer **Datenmatrix** mit n Zeilen und p Spalten vorliegen. Die Zeilen entsprechen den Untersuchungseinheiten, die Spalten den betrachteten Merkmalen.

Lernergebnisse (B2)
Die Studierenden können die oben genannten Begriffe erläutern, eigene Beispiele nennen und gegebene Beispiele zuordnen.

1.2 Qualitätssicherung und Bereinigung von Daten

Kerninhalte

- Datenverfügbarkeit hinsichtlich des Ziels der Analyse
- Datenbereinigung
- Besonderheiten im Umgang mit großen Datenmengen

1.2.1 Datenverfügbarkeit

Es muss im Vorfeld der Analyse geklärt werden, ob die zur Verfügung stehenden Daten hinsichtlich des **Ziels der Analyse** geeignet sind und auch, ob sie überhaupt verwendet werden dürfen. Die Daten werden primär oft für ganz andere Zwecke erhoben. Zu prüfen ist auch, ob die Daten die **gewünschte Population** abbilden, auch in **zeitlicher** Hinsicht. Es muss ein **Verständnis** für die in den betrachteten Daten vorhandenen Merkmale hergestellt werden. Nur dann ist die Generierung neuer Merkmale aus den vorhandenen Merkmalen sinnvoll. Die **Kodierung** der Merkmale, also die Zuordnung von Zahlen zu den Ausprägungen, muss **nachvollziehbar** und **eindeutig** sein. Üblicherweise wird die Kodierung in einem sogenannten **Codebook** festgehalten.

1.2.2 Datenbereinigung

Für alle Merkmale müssen zunächst einfache Statistiken erstellt werden. Für nominale und ordinale Merkmale mit wenigen Ausprägungen genügt meist die **Häufigkeitsverteilung**, um **unplausible** oder **unlogische** Werte aufzuspüren.

[5] Im Englischen werden dafür die Begriffe *data wrangling* und *data munging* verwendet

Für metrische Merkmale und ordinale Merkmale mit vielen Ausprägungen sollten mindestens **Minimum** und **Maximum** berechnet werden, um zu überprüfen, das die Daten in einem sinnvollen **Bereich** liegen.

Eine **Ausreißer**- und **Extremwert**-Analyse unter Hinzuziehung grafischer Verfahren wie dem **Box-Plot** ist meist zusätzlich sinnvoll und erforderlich.

Schwieriger ist die Erkennung von **logischen Fehlern** über mehrere Merkmale hinweg. Hier können einfache **regelbasierte Überprüfungen** eingesetzt werden.

Es muss das Bewusstsein für die Problematik **fehlender Werte** vorhanden sein. Statistische Standardverfahren ignorieren Datenzeilen mit fehlenden Werten und führen daher meist nur eine sogenannte **Complete Case** Analyse durch. Fehlende Werte können aber viele Ursachen haben. Steht das Fehlen in unmittelbarem Zusammenhang mit dem Analyseziel, so ist zu überprüfen, ob die Complete Case Analyse **unverzerrte Ergebnisse** liefern kann. Strategien zur Ersetzung fehlender Werte durch einfache oder mehrfache **Imputation** müssen in Betracht gezogen werden. Vor- und Nachteile von Imputationen sind dabei im konkreten Fall gegeneinander abzuwägen.

1.2.3 Besonderheiten im Umgang mit großen Datenmengen

Mit Besonderheiten sind hier spezielle Probleme gemeint, die so in der Regel nur bei großen Datenmengen auftauchen.

Von **großen Datenmengen** spricht man dann, wenn die Daten nicht mehr komplett im Arbeitsspeicher eines üblichen Arbeitsplatzcomputers Platz finden. Das ist meist dann der Fall, wenn die Zahl der Datenzeilen oder/und die Zahl der betrachteten Merkmale sehr groß wird. Beispiele für den ersten Fall (”lange Tabelle”) sind beispielsweise Sensordaten. Viele Variablen oder Merkmale (”breite Tabelle”) liegen vor, wenn viel Information pro Beobachtung vorliegt, zum Beispiel Daten aus sozialen Medien. Speziell im Fall von Textdaten kann die Zahl der betrachteten Merkmale (eventuell unnötig) extrem groß werden: die einzelnen Merkmale sind dann zum Beispiel Wörter oder Wortkombinationen und zusätzlich sind viele Informationen in den Daten nicht relevant, wie zum Beispiel Punktuationszeichen oder sogenannte Stopwörter. Eine große Anzahl von Beobachtungen und Merkmalen kann zu zeitaufwendigen Operationen (z.B. Joins) führen, wenn diese Daten mit anderen Daten zusammengeführt werden sollen. **Unstrukturierte Daten**, insbesondere Bilder und Videos benötigen meist viel Speicherplatz, beispielsweise Bilder von KFZ-Schäden. Die Schwierigkeit ist dann, diese Daten für weitergehende Analysen in sinnvolle Merkmale zu transformieren. Generell führt eine zu wenig starke Normalisierung von Daten einerseits, aber auch die ”Nicht-Reduktion” auf relevante Informationen andererseits dazu, dass Daten nicht optimal gespeichert werden. Aus all den genannten Gründen benötigt es eine Strategie, wie welche Daten gespeichert werden müssen, welchen Speicherplatz, welche Hardware und welche Software man braucht.

Programme, die versuchen, die Daten komplett in den Arbeitsspeicher zu laden, können dann nicht mehr zuverlässig verwendet werden. Die verwendeten Programme müssen also in der Lage sein, die Daten portionsweise aus dem Speicher zu lesen und zu verarbeiten. Die Berechnungszeiten können selbst für einfache Analysen stark ansteigen. Viele statistische Methoden sind ineffizient für Datenmatrizen mit sehr vielen Zeilen und Spalten (Merkmale) und ungültig für den Fall, das mehr Merkmale als Fälle vorhanden sind. Viele fehlende Werte sind möglich, insbesondere bei Daten, die online und freiwillig erhoben werden. Handelt es sich um Bild- oder Textdaten, so sind diese nur un- oder semi-strukturiert. Eine Qualitätssicherung ist bei großen Datenmengen oft noch schwieriger zu erreichen als bei kleinen Datenmengen.

Lernergebnisse (B2-D2)
Die Bedeutung der Qualitätssicherung und Bereinigung der Daten kann an konkreten Fallbeispielen aufgezeigt werden. Die Probleme, die im Zusammenhang mit großen Datenmengen auftreten, können beschrieben werden. Methoden zur Erkennung von Datenfehlern sollen bekannt sein. Verschiedene Imputationsmethoden bei fehlenden Daten, sowie Vor- und Nachteile von Imputationen sollen bekannt sein. Die Studierenden können mögliche Vorgehensweisen zur Qualitätssicherung und Bereinigung von Daten beschreiben.

1.3 Einfache Verfahren der deskriptiven und induktiven Datenanalyse

Kerninhalte

- Häufigkeitsverteilungen und ihre grafische Darstellung
- Lagemaße, Quantile und Streuungsmaße
- Boxplot und Quantil-Quantil Diagramm
- Lorenzkurve und Gini-Koeffizien
- Zweidimensionale Häufigkeitsverteilungen und ihre grafische Darstellung
- Zusammenhangsmaße

1.3.1 Häufigkeitsverteilungen und ihre grafische Darstellung

Hier unterscheidet man **ungruppierte** und **gruppierte** Merkmale. Gruppierte Merkmale werden in benachbarte **Intervalle** oder **Klassen** eingeteilt. **Absolute** und **relative** Häufigkeiten werden in einer **Häufigkeitstabelle** numerisch dargestellt. Eine grafische Darstellung erfolgt für ungruppierte Merkmale mit Kreis-, Säulen-, Balken oder Stabdiagrammen. Für gruppierte Merkmale wird das **Histogramm** oder ein **Kerndichteschätzer** verwendet. Metrische Merkmale lassen sich auch in einem **Stamm-Blatt-Diagramm** aufbereiten. Verteilungen können dabei annähernd **symmetrisch** oder **schief** sein. Man unterscheidet **links**- und **rechtsschiefe** Verteilungen.

1.3.2 Beschreibung von Verteilungen durch Maßzahlen

Maßzahlen sollen die Verteilung möglichst kompakt mit wenigen Kennziffern beschreiben. Die beobachteten Daten eines Merkmals X werden durch den Vektor $\mathbf{x} = (x_1, \ldots, x_n)$ bezeichnet. Für nominale Merkmale, deren Merkmalsausprägungen nicht geordnet werden können, kann als **Lagemaß** nur der **Modus** (häufigster Wert) verwendet werden.

Können die Merkmalsausprägungen der Größe nach geordnet werden (ordinal und metrisch skalierte Merkmale), so nennt man die aufsteigend sortierten Werte $(x_{(1)}, \ldots, x_{(n)})$ der beobachteten Daten **Ordnungsstatistik**. Ab ordinalem Skalenniveau ist der **Median**

$$x_{med} = \begin{cases} x_{\left(\frac{n+1}{2}\right)} & \text{für } n \text{ ungerade} \\ \frac{1}{2}\left(x_{(n/2)} + x_{(n/2+1)}\right) & \text{für } n \text{ gerade} \end{cases}$$

sinnvoll definiert. Der Median ist ein gegen Ausreißer **robustes** Lagemaß Lagemaß. Für metrische Merkmale kann das **arithmetische Mittel**

$$\bar{x} = \frac{1}{n}(x_1 + \ldots + x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

oder das **geometrische Mittel**

$$\bar{x}_{geom} = (x_1 \cdot \ldots \cdot x_n)^{1/n}.$$

als Lagemaß verwendet werden. Letzteres kommt bei Wachstumsuntersuchungen und Zinsberechnungen zum Einsatz. Alle Lagemaße können auf **gruppierte Daten** erweitert werden.

Quantile teilen die aufsteigend sortierten Daten in Teilbereiche auf. Jeder Wert x_p mit $0 < p < 1$, für den mindestens ein Anteil p der Daten kleiner oder gleich x_p und mindestens ein Anteil $1-p$ größer oder gleich x_p ist, heißt *p-Quantil*. Es muss also gelten

$$\frac{\text{Anzahl } (x\text{-Werte} \leq x_p)}{n} \geq p \quad \text{und} \quad \frac{\text{Anzahl } (x\text{-Werte} \geq x_p)}{n} \geq 1-p\,.$$

Damit gilt für das *p-Quantil*:

$$\begin{aligned} x_p &= x_{([np]+1)}\,, \text{ wenn } np \text{ nicht ganzzahlig}\,, \\ x_p &\in [x_{(np)}, x_{(np+1)}]\,, \text{ wenn } np \text{ ganzzahlig}\,. \end{aligned}$$

Dabei ist $\lfloor np \rfloor$ die zu np nächste kleinere ganze Zahl. [6]

Streuungsmaße

Alle im Folgenden aufgeführten Streuungsmaße können ab ordinalem Skalenniveau verwendet werden.

Die **Spannweite** ist die Differenz $x_{(n)} - x_{(1)}$.

Der **Quartilsabstand** ist definiert als Differenz des 75%-Quantils und des 25%-Quantils

$$d_Q = x_{0.75} - x_{0.25}\,.$$

Die empirische **Varianz** s^2 und die empirische **Standardabweichung** s sind definiert als

$$s^2 = \frac{1}{n-1}[(x_1-\bar{x})^2 + \ldots + (x_n-\bar{x})^2] = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

und

$$s = \sqrt{s^2}\,.$$

Die **mittlere absolute Abweichung vom arithmetischen Mittel** ist

$$\text{MAD(Mittelwert)} = \frac{1}{n}[|x_1-\bar{x}| + \ldots + |x_n-\bar{x}|] = \frac{1}{n}\sum_{i=1}^{n}|x_i-\bar{x}|\,.$$

Analog ergibt sich die **mittlere absolute Abweichung vom Median** als

$$\text{MAD(Median)} = \frac{1}{n}\sum_{i=1}^{n}|x_i-x_{med}|\,.$$

1.3.3 Weitere grafische Darstellungen

Box-Plot

Der **Box-Plot** stellt die sogenannte **Fünf-Punkte-Zusammenfassung** grafisch dar. Der einfache Box-Plot verwendet Minimum, 25%-Quantil, Median, 75%-Quantil und Maximum zur Darstellung. Beim modifizierten Box-Plot werden die Linien außerhalb der Box nur dann

[6] Software verwendet oft andere Definitionen und die Berechnungen können zu anderen Werten führen.

bis zu Minimum bzw. Maximum gezogen, falls diese innerhalb eines Bereichs liegen, der durch verschiedene Berechnungen zustande kommen kann.[7]. Ansonsten gehen die Linien nur bis zum kleinsten bzw. größten Wert innerhalb des berechneten Bereichs, und die außerhalb liegenden Werte werden individuell eingezeichnet. Die Begriffe ”einfacher” und ”modifizierter” Box-Plot werden in der Literatur nicht einheitlich verwendet. Box-Plots können keine multimodalen Verteilungen erkennen. Diesem Mangel kann durch zusätzliche Verwendung von Histogrammen begegnet werden.

Quantil-Quantil-Plot

Ein **Quantil-Quantil-Plot**(**Q-Q-Plot**) stellt zwei Verteilungen grafisch gegenüber. Sind die Verteilungen ähnlich, liegen die Punktepaare des Q-Q-Plot auf der Winkelhalbierenden. Ist ein Merkmal eine lineare Transformation des anderen Merkmals, so liegen die Punktepaare auf einer Geraden. Man kann zwei Fälle unterscheiden. Im ersten Fall sollen die Verteilungen zweier Merkmale verglichen werden, in zweiten Fall soll die Verteilung eines Merkmals gegen eine theoretische Verteilung verglichen werden. Sind $\mathbf{x} = (x_{(1)}, \ldots, x_{(n)})$ und $\mathbf{y} = (y_{(1)}, \ldots, y_{(n)})$ zwei aufsteigend sortierte Datenreihen zweier Merkmale gleicher Länge, so ergibt sich der Q-Q-Plot als Plot der Paare $(x_{(i)}, y_{(i)})$. Im Fall, dass eine Datenreihe länger ist als die andere werden einfache Interpolationsmethoden angewandt. Beim Vergleich gegen eine theoretische Verteilung wird die aufsteigend geordnete Datenreihe $x_{(1)}, \ldots, x_{(n)}$ des betrachteten Merkmals für $i = 1, \ldots, n$ gegen die (ebenfalls aufsteigenden) $(i - 0.5)/n$-Quantile[8] der theoretischen Verteilung abgetragen.

Verwendet man als theoretische Verteilung die Standardnormalverteilung, so spricht von einem **Normal-Quantil-Plot** (NQ-Plot).

1.3.4 Zweidimensionale Häufigkeitsverteilungen und ihre grafische Darstellung

Die gemeinsame Verteilung von zwei **diskreten** Merkmalen X und Y, die nur relativ wenige Ausprägungen aufweisen, kann in einer **Kontingenztafel** dargestellt werden. Sind $a_1, \ldots, a_k$ die Ausprägungen des einen Merkmals und $b_1, \ldots, b_m$ die Ausprägungen des anderen Merkmals, so ist die $(k \times m)$-Kontingenztafel gegeben durch

	b_1	$\ldots$	b_m	
a_1	h_{11}	$\ldots$	h_{1m}	$h_{1\cdot}$
a_2	h_{21}	$\ldots$	h_{2m}	$h_{2\cdot}$
$\vdots$	$\vdots$		$\vdots$	$\vdots$
a_k	h_{k1}	$\ldots$	h_{km}	$h_{k\cdot}$
	$h_{\cdot 1}$	$\ldots$	$h_{\cdot m}$	n

Dabei sind

$h_{ij} = h(a_i, b_j)$	die absolute Häufigkeit der Kombination (a_i, b_j),
$h_{1\cdot}, \ldots, h_{k\cdot}$	die Randhäufigkeiten von X und
$h_{\cdot 1}, \ldots, h_{\cdot m}$	die Randhäufigkeiten von Y.

Sind beide Merkmale X und Y metrisch, so kann ein **Streudiagramm** oder, bei mehr als zwei metrischen Merkmalen, eine **Streudiagramm-Matrix** verwendet werden. Hierbei werden die Punktepaare von je zwei Merkmalen in ein Diagramm übertragen.

Ist ein Merkmal metrisch und ein anderes Merkmal diskret mit relativ wenigen Ausprägungen, kann die bedingte Verteilung des metrischen Merkmals gegeben das diskrete Merkmal durch einen Box-Plot pro Kategorie des diskreten Merkmals veranschaulicht werden.

[7] Häufig wird die Länge der Box, also der Quartilsabstand, mit 1,5 multipliziert und nach oben ab dem 75%-Quantil und nach unten ab dem 25%-Quantil abgetragen

[8] Andere verwendete Varianten sind u.a. i/n und $i/(n+1)$.

Für die Darstellung zweier bedingter Merkmale geschichtet nach diskreten Merkmalen mit relativ wenig Ausprägungen können sogenannte **Ko-Plots** verwendet werden. Es handelt sich dabei einfach um Streudiagramme, die für jede der durch die diskreten Merkmale definierten Schichten separat gezeichnet werden.

1.3.5 Empirische Zusammenhangsmaße

Für zwei metrische Merkmale X und Y ist der empirische **Korrelationskoeffizient** r nach Bravais-Pearson definiert durch

$$r = r_{XY} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2 \sum_{i=1}^{n}(y_i-\bar{y})^2}} = \frac{s_{XY}}{s_X s_Y}, \tag{1.1}$$

Dabei sind

$$s_X = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}, \qquad s_Y = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2}$$

die empirischen Standardabweichungen der Merkmale X bzw. Y und

$$s_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$$

die empirische **Kovarianz** bezeichnet.

Geht man von den Originalwerten zu den Rängen über und setzt diese in Gleichung (1.1) ein, so erhält man den **Rangkorrelationskoeffizienten nach Spearman**.

Lernergebnisse (C3)
Die Studierenden kennen einfache Verfahren der deskriptiven Datenanalyse. Sie sind in der Lage, die je nach Datensituation geeigneten Verfahren auszuwählen und anzuwenden.

2 Lebensdauermodelle

Untersucht werden Ereigniszeiten für das Auftreten von Ereignissen wie z.B. Tod, Schaden, usw. Lebensdauer bedeutet hier die Zeit bis zum Eintritt eines Ereignisses.

2.1 Grundlagen

Kerninhalte Definition der Begriffe

- Survivalfunktion
- Hazardrate
- Hazardfunktion
- Zensierung

Definition 2.1. *Sei $T \geq 0$ eine Zufallsvariable mit Verteilungsfunktion F.*

*(a) Die Funktion $S : \mathbb{R} \longrightarrow [0,1]$, $S(t) := 1 - F(t) = P(T > t)$ heißt **Survivalfunktion**.*

(b) Ist T stetig verteilt, dann heißt für $t \geq 0$

$$\lambda(t) := \lim_{h \to 0} \frac{P(T \leq t + h | T > t)}{h}$$

*die **Hazardrate**.*

*(c) Die Funktion $\Lambda : [0, \infty) \longrightarrow [0, \infty)$, $\Lambda(t) := \int_0^t \lambda(s)\, ds$ heißt **kumulierte Hazardfunktion**.*

Die Hazardrate ist die Änderungsrate von F zum Zeitpunkt t. Sie kann auch als bedingte Wahrscheinlichkeit interpretiert werden, dass das Ereignis im nächsten Augenblick eintritt, gegeben es ist bis t nicht eingetreten. Häufig verwendete Verteilungsmodelle sind die Exponential- und die Weibullverteilung sowie Transformationsmodelle (meist logarithmische Transformation). Die Hazardrate der Weibullverteilung $\mathcal{W}(\alpha, \lambda)$ ist $t \mapsto \lambda\alpha(\lambda t)^{\alpha-1}$. Mit $\alpha = 1$ handelt es sich um die Exponentialverteilung $\mathcal{E}(\lambda)$ mit Hazardrate konstant λ.

Ist f die Dichte von T, dann gelten die folgenden Zusammenhänge:

$$E(T) = \int_0^\infty S(t)\, dt \tag{2.1}$$

$$\lambda(t) = \frac{f(t)}{S(t)} \tag{2.2}$$

$$S(t) = \exp\left(- \int_0^t \lambda(s)\, ds \right). \tag{2.3}$$

Für die Entwicklung von Schätzern ist auch die folgende Dartstellung nützlich. Sei hierzu $t > 0$ und $0 =: t_0 < t_1 < \cdots < t_{n-1} < t_n := t$. Dann gilt

$$S(t) = \prod_{i=1}^{n} P(T > t_i | T > t_{i-1}). \tag{2.4}$$

Bei der Analyse von Lebensdauern wird man oft mit **zensierten Beobachtungen** konfrontiert. Im aktuariellen Umfeld ist am häufigsten die **Zensierung** von rechts anzutreffen, d.h. die Beobachtung wird beendet bevor das Ereignis (z.B. Tod) eintritt. Das Modell der zufälligen Zensierung wird dabei am häufigsten verwendet:

Angenommen es werden die Lebensdauern $T_1, \ldots, T_n$ von n Individuen beobachtet, $C_1, \ldots, C_n$ seien die Zensierungszeiten der Individuen. Die Zufallsvariablen T_i, C_i, $i = 1, \ldots, n$ seien unabhängig. Beobachtet werden $T_i^* := \min(T_i, C_i)$ und die Indikatorvariable $\Delta_i := 1_{\{T_i \leq C_i\}}$, die den Wert 0 annimmt, wenn die Beobachtung zensiert ist.

Einen anderen Zugang zu diesem Modell erhält man mit Hilfe der Zählprozesse

$$N(t) = \sum_{i=1}^{n} 1_{[0,t]}(T_i), \text{ Anzahl der Ausfälle in } [0, t]$$

$$Y(t) = \sum_{i=1}^{n} 1_{[t,\infty]}(T_i), \text{ Anzahl der Risiken unter Beobachtung in } t.$$

Lernergebnisse (B2)
Die Studierenden kennen die Definitionen der Begriffe Survivalfunktion und Hazardrate und die Zusammenhänge zwischen ihnen. Sie können diese für grundlegende Verteilungen und Survivalmodelle bestimmen und ineinander überführen. Sie können die Entstehung von zensierten Daten beschreiben.

2.2 Schätzverfahren

Kerninhalte

- Kaplan-Meier Schätzer
- Nelson-Aalen Schätzer
- Maximum Likelihood

Gegeben sei eine Stichprobe $t_1, \dots, t_n$ von Lebensdauern von denen bekannt ist, ob eine Zensierung vorliegt, also

$$\delta_i = \begin{cases} 1 & \text{falls } t_i \text{ unzensiert} \\ 0 & \text{sonst.} \end{cases}$$

Sei $0 =: t_{(0)} < t_{(1)} < \dots < t_{(m)}$ die Ordnungsstatistik der Ereigniszeitpunkte. Es gilt $m \leq n$, da hier nur die nicht zensierten Beobachtungen betrachtet werden. Sei d_j bzw. n_j, $j = 1, \dots, m$ die Anzahl der Ereignisse bzw. der Risiken unter Beobachtung in $t_{(j)}$.

2.2.1 Kaplan-Meier

Es handelt sich um einen Schätzer der Survivalfunktion. Aus den Daten wird

$$p_j := P(T > t_{(j)} | T > t_{(j-1)}),\ j = 1, \dots, m$$

geschätzt mittels

$$\hat{p}_j := 1 - \frac{d_j}{n_j}.$$

Mit (2.4) erhält man den **Kaplan-Meier-Schätzer**

$$\hat{S}(t) = \begin{cases} 1 & \text{falls } t < t_{(1)} \\ \prod_{j \mid t_{(j)} \leq t} \hat{p}_j & \text{sonst} \end{cases}$$

für $S(t)$. Eine Schätzung der Varianz dieses Schätzers erhält man mit der approximativen Formel von Greenwood:

$$\widehat{\mathrm{Var}}\left(\hat{S}(t)\right) = \hat{S}(t)^2 \sum_{j \mid t_{(j)} \leq t} \frac{d_j}{n_j(n_j - d_j)}. \tag{2.5}$$

Damit lassen sich auch approximativ normale, punktweise Konfidenzintervalle berechnen.

2.2.2 Nelson-Aalen

Der **Nelson-Aalen Schätzer** für die kumulierte Hazard-Funktion lautet

$$\hat{\Lambda}(t) = \begin{cases} 0 & t < t_{(1)} \\ \sum_{j \mid t_{(j)} \leq t} \frac{d_j}{n_j} & \text{sonst.} \end{cases}$$

Die Varianz des Nelson-Aalen Schätzers $\mathrm{Var}(\hat{\Lambda}(t))$ wird mittels dem Analogon von (2.5)

$$\widehat{\mathrm{Var}}(\hat{\Lambda}(t)) = \sum_{j \mid t_{(j)} \leq t} \frac{d_j(n_j - d_j)}{n_j^3}$$

geschätzt. Man findet auch $\frac{d_j}{n_j^2}$, statt $\frac{d_j(n_j - d_j)}{n_j^3}$. Über den Zusammenhang von Survivalfunktion und kumulierter Hazardfunktion (2.3) erhält man einen Schätzer für die Survivalfunktion:

$$\hat{S}(t) = \exp\left(-\hat{\Lambda}(t)\right).$$

2.2.3 Maximum-Likelihood

Sei F die Verteilungsfunktion von T mit Dichte f. Bei nicht informativer Zensierung, d.h. der zur Zensierung führende Mechanismus ist unabhängig von Einflussgrößen auf die Überlebenszeit, ist die Gesamtlikelihoodfunktion gegeben durch

$$L = \prod_{j=1}^{n} f(t_j)^{\delta_j} S(t_j)^{1-\delta_j} = \prod_{t_j \text{ unzensiert}} f(t_j) \prod_{t_j \text{ zensiert}} S(t_j). \tag{2.6}$$

Daraus erhält man gegenenfalls die Maximum-Likelihood-Gleichungen bzw. -Schätzer. Ist T exponentialverteilt $\mathcal{E}(\lambda)$ ergibt sich

$$\hat{\lambda} = \frac{\sum_{j=1}^{n} \delta_j}{\sum_{j=1}^{n} t_j}.$$

Lernergebnisse (C3)
Die Studierenden können aus gegebenen zensierten Daten Schätzer für die Survivalfunktion und die kumulierte Hazardfunktion mit dem Kaplan-Meier- und dem Nelson-Aalen-Schätzer bestimmen. Für grundlegende zensierte Verteilungen können die Studierenden die Maximum-Likelihood Gleichungen bestimmen und gegebenenfalls Schätzwerte für Parameter, Survivalfunktion und Hazardrate bestimmen.

2.3 Regressionsmodelle

Kerninhalte

- proportionale hazard Modelle
- accelarated failure Modelle
- Schätzmethoden

Im Folgenden seien $T_1, \ldots, T_n > 0$ unabhängige Lebensdauern von n Individuen mit Kovariablen $\mathbf{z}_i \in \mathbb{R}^p$, $i = 1, \ldots, n$.

2.3.1 Proportional-Hazard Regressions-Modell (Cox-Modell)

Der Einfluss der Kovariablen wird in der Hazardrate modelliert. Die Hazardrate $\lambda_i(t)$ von T_i zur Zeit t erfüllt

$$\lambda_i(t) = \lambda_0(t) \exp\left(\langle \mathbf{z}_i, \beta \rangle\right),\ t > 0.$$

Dabei kann die **Baseline-Hazardrate** λ_0 parametrisch oder nicht parametrisch sein.

Für die Darstellung der Schätzung seien $t_{(j)}$, d_j, $j = 1, \ldots, m$ wie im Abschnitt 2.2. Desweiteren sei $\mathbf{z}_{(j)}$ der Merkmalsvektor zur Beobachtung $t_{(j)}$ und $R(t_{(j)})$ die Menge der Individuen, die in $t_{(j)}$, $j = 1, \ldots, k$ unter Risiko stehen.

Die Parameter β werden über die Maximierung der **partiellen Likelihood** PL in (2.7) geschätzt. Sind die $t_{(j)}$, $j = 1, \ldots, m$ paarweise verschieden, dann wird definiert

$$PL(\beta, \mathbf{z}_1, \ldots, \mathbf{z}_n) := \prod_{j=1}^{m} \frac{\exp\left(\langle \mathbf{z}_{(j)}, \beta \rangle\right)}{\sum_{l \in R(t_l)} \exp\left(\langle \mathbf{z}_{(l)}, \beta \rangle\right)}. \tag{2.7}$$

Im Falle von **Bindungen**, d.h. es gibt identische Stichprobenwerte $t_k = t_{k'}$, $k \neq k'$, muss PL noch modifiziert werden.

Die Baseline-Hazardrate $\lambda_0(t)$ wird in einigen Programmpaketen als konstant zwischen den beobachteten Lebensdauern $t_{(1)} < \cdots < t_{(m)}$ geschätzt.

2.3.2 Transformationsmodelle

Man kann das Regressionsmodell

$$\ln T_i = \langle \beta, \mathbf{z}_i \rangle + \varepsilon_i,\ i = 1, \ldots, n$$

formulieren, mit den Regressionskoeffizienten $\beta \in \mathbb{R}^p$. Für die Zufallsvariablen gelte $\varepsilon_i \overset{\text{iid}}{\sim} \varepsilon_0$. Sei S_i, $i = 1, \ldots, n$ bzw. S_0 die Überlebensfunktion von T_i bzw. $\eta_0 := e^{\varepsilon_0}$. Es gilt

$$S_i(t) = S_0\left(t \exp\left(-\langle \mathbf{z}_i, \beta \rangle\right)\right).$$

Diese Modelle heißen **accelarated failure time models**. Je nach Wahl der Verteilung von ε_0 erhält man beispielsweise Log-logistische, Log-Normal- und Weibull-Modelle.

Die Schätzung von β wird mit Maximum Likelihood durchgeführt, vgl. (2.6). Bis auf die Zensierungsproblematik entsprechen diese Modelle den verallgemeinerten linearen Modelle.

Lernergebnisse (C3)
Die Studierenden können zu gegebenen Zielvariablen in Abhängigkeit von Kovariablen Regressionsmodelle aufstellen. Für grundlegende Verteilungen können sie daraus Hazardraten und Survivalfunktionen in Abhängigkeit von Merkmalen bestimmen. Die Studierenden kennen die Schätzverfahren und interpretieren die Ergebnisse in konkreten Anwendungsfällen. Sie können die Schätzer für die Hazardraten und Survivalfunktion aus den Ergebnissen bestimmen.

2.4 Ausscheideordnungen

Kerninhalte

- Schätzer für die Übergangswahrscheinlichkeiten des Zustandsmodells
- Ermittlung relativer Sterbehäufigkeiten
- Ausgleichsverfahren mit Hilfe von Hypothesentests beurteilen

In diesem Abschnitt werden Aspekte der Erstellung von Sterbetafeln behandelt. Die Begriffe Ausscheideordnung, Sterbetafel, Rechnungsgrundlagen usw. werden als bekannt vorausgesetzt, vgl. Leitfaden Versicherungsmathematik.

Sterbetafeln in der Versicherungswirtschaft sind auf den Versicherungszweck (Tod, Erleben, Invalidisierung, Reaktivierung, usw.) ausgerichtet und legen Versicherungskolletive zugrunde, die sich von der Gesamtbevölkerung unterscheiden.

Ermittelt werden hierbei Schätzwerte für die bedingten Wahrscheinlichkeiten

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

wobei T die Lebensdauer einer Person ist.

2.4.1 Ermittlung relativer Sterbehäufigkeiten

Es wird eine homogene Gruppe von n Personen des Alters x beobachtet. Person $i \in \{1, \ldots, n\}$ wird zwischen den Altern $x + a_i$ und $x + b_i$ mit $0 \leq a_i < b_i \leq 1$ beobachtet. Die Zeit $t_i := b_i - a_i$, die Person i als x-Jähriger unter Beobachtung verbringt, heißt Verweildauer und wird als bedingte Lebensdauer im Alter x interpretiert. Sei $I \subset \{1, \ldots, n\}$ die Menge der wegen Tod beendeten Beobachtungen, $D_x := |I|$ ist dann die Anzahl der beobachteten Todesfälle. Je nach Annahmen des Zustandsmodells ergeben sich als Momenten- bzw. ML-Schätzer für q_x

$$\hat{q}_x^{(1)} := \frac{D_x}{\sum_{i \notin I} t_i + \sum_{i \in I}(1 - a_i)}, \qquad \hat{q}_x^{(2)} := \frac{D_x}{\sum_{i \notin I} t_i + \frac{1}{2} D_x}, \qquad \hat{q}_x^{(3)} := 1 - \exp\left(-\frac{D_x}{\sum_{i \notin I} t_i}\right).$$

In der Praxis (Versicherungswirtschaft, statistisches Bundesamt) werden die Geburtsjahr-, Sterbejahr-, Sterbeziffern- und Verweildauermethode verwendet. Hierbei wird jeweils der Schätzer $\frac{D_x}{L_x}$ verwendet, wobei L_x die Anzahl der unter Risiko stehenden Personen im Alter x ist. Die Methoden unterscheiden sich jeweils hinsichtlich der zur Bestimmung von L_x verwendeten Ansätze. Bei der Verweildauermethode wird der Schätzer $\hat{q}_x^{(V)} := \frac{D_x}{\sum_{i \notin I} t_i + D_x}$ eingesetzt.

2.4.2 Ausgleichsverfahren

Die relativen Sterbehäufigkeiten enthalten zufallsbedingte Schwankungen. Um einen möglichst glatten und gleichzeitig wirklichkeitsnahen Verlauf der Sterbewahrscheinlickeiten in Abhängigkeit vom Alter x zu erhalten, werden die relativen Sterbehäufigkeiten $\hat{q}_x$ ausgeglichen.

Es handelt sich um numerische Verfahren, deren Durchführung nicht geprüft wird.

2.4.3 Hypothesentests zur Überprüfung der ausgeglichenen Sterbewahrscheinlichkeiten

Um zu überprüfen ob eine gegebene Sterbetafel für die Alter $x_0, \ldots, x_n$ für eine Grundgesamtheit angemessen ist, kann man statistische Testverfahren verwenden. Mit diesen Verfahren kann überprüft werden, ob ein Ausgleichsverfahren realistische Sterbehäufigkeiten liefert. Die Nullhypothese H_0 bzw. die Alternativhypothese H_1 lauten:

H_0: Die tatsächlichen und die ausgeglichenen Ausscheidewahrscheinlichkeiten stimmen überein.

H_1: Die tatsächlichen und die ausgeglichenen Ausscheidewahrscheinlichkeiten sind verschieden.

Für die Darstellung der verschiedenen Teststatistiken werden folgende Bezeichnungen verwendet:

l_x	die Anzahl der Lebenden des Alters x
$Z_{x,j} =$	$\begin{cases} 1 & \text{wenn } j\text{-te Person im Alter } x \text{ verstirbt} \\ 0 & \text{sonst} \end{cases}$, $j = 1, \ldots, l_x$
$Z_x := \sum_{j=1}^{l_x} Z_{x,j}$	Summe der beobachteten Toten für das Alter x
$E_x := E(Z_x)$	rechnungsmäßig erwartete Anzahl der Toten im x.

Die Teststatistik T des **Vorzeichentests** lautet

$$T = \sum_{x=x_0}^{x_n} 1_{\{Z_x > E_x\}}.$$

Unter H_0 gilt $T \sim B\left(n+1, \frac{1}{2}\right)$, so dass mit dem Binomialtest H_0 überprüft werden kann.

Die Teststatistik T des **Iterationstests** lautet

$$T = \sum_{x=x_1}^{x_n} 1_{\{Sign(Z_x - E_x) \neq Sign(Z_{x-1} - E_{x-1})\}}.$$

Unter H_0 gilt $T \sim B\left(n, \frac{1}{2}\right)$, so dass auch hier der Binomialtest verwendet werden kann.

Die Teststatistik des χ^2-**Tests** orientiert sich am χ^2-Anpassungstest und lautet:

$$T = \sum_{x=x_0}^{x_n} \frac{(Z_x - E_x)^2}{Var(Z_x)} = \sum_{x=x_0}^{x_n} \chi_x^2 \text{ mit } \chi_x := \frac{Z_x - E_x}{\sqrt{Var(Z_x)}}.$$

Unter Gültigkeit der Nullhypothese gilt näherungsweise $\chi_x \sim \mathcal{N}(0,1)$, damit ist die Teststatistik näherungsweise χ^2-verteilt mit $n+1$ Freiheitsgraden und man kann H_0 untersuchen.

2.4.4 Berücksichtigung von Risiken

Prinzipiell kann die Berücksichtigung auf zwei Arten erfolgen:

- durch Modifikation der Ausscheidewahrscheinlichkeiten
- durch Zuschläge auf die Deckungsrückstellung (d.h. auf Bewertungsebene).

Im Folgenden wird die Berücksichtigung des Zufallsrisikos bei Rentenversicherungen dargestellt, um ein Sicherheitsniveau $1-\alpha$ zu erreichen. Dazu wird ein Modellbestand betrachtet. Folgende Bezeichnungen werden im Weiteren verwendet:

L_x^M	die Lebenden des Alters x des Modellbestandes
T_x^M	die Zufallsvariable der im Alter x Gestorbenen im Modellbestand
V_x	die Deckungsrückstellung für einen Versicherten des Alters x
$u_{1-\alpha}$	das $1-\alpha$-Quantil der Standard-Normalverteilung
$s^\alpha \in \mathbb{R}$	relativer **Schwankungszu-** bzw. **abschlag** auf die Sterbewahrscheinlichkeit im Alter (oder der Altersklasse) x

Modifikation der Ausscheidewahrscheinlichkeiten

Der zu bestimmende Schwankungsabschlag s_α ergibt sich aus dem Ansatz

$$P\left(\sum_x T_x^M \geqslant \sum_x (q_x - s^\alpha q_x) L_x^M\right) \stackrel{!}{=} 1-\alpha$$

und unter der Annahme, dass $Z := \sum_x T_x^M$ näherungsweise normalverteilt ist

$$s^\alpha = u_{1-\alpha} \frac{\sqrt{Var(Z)}}{E(Z)} = u_{1-\alpha} \frac{\sqrt{\sum_x q_x(1-q_x)L_x^M}}{\sum_x q_x L_x^M}.$$

Betrachtung der Bewertungsebene

Der Schwankungsabschlag s_α wird so gewählt, dass die durch Tod im Modellbestand freiwerdende Deckungsrückstellung $\sum_x T_x^M V_x$, die mit den modfizierten Sterbewahrscheinlichkeiten berechnete frei werdende Deckungsrückstellung mit der Wahrscheinlichkeit $1-\alpha$ überschreitet:

$$P\left(\sum_x T_x^M V_x \geqslant \sum_x \left(q_x - s^\alpha q_x\right) L_x^M V_x\right) \stackrel{!}{=} 1-\alpha.$$

Ebenfalls unter der Normalverteilungsapproximation von $\sum_x T_x^M V_x$ folgt

$$s^\alpha = u_{1-\alpha} \frac{\sqrt{\sum_x q_x(1-q_x)L_x^M V_x^2}}{\sum_x q_x L_x^M V_x}.$$

Lernergebnisse (C3)
Die Studierenden sind in der Lage aus unterschiedlichen Annahmen zu Sterberaten Schätzer herzuleiten und zu interpretieren. Sie können rohe Sterbewahrscheinlichkeiten mit unterschiedlichen Verfahren bestimmen. Die Studierenden erkennen die Notwendigkeit von Ausgleichsverfahren. Sie können unterschiedliche Hypothesentests anwenden, um die Angemessenheit der in Ansatz gebrachten Sterbewahrscheinlichkeiten zu beurteilen. Die Studierenden können Zu- bzw. Abschläge für unterschiedliche Ansätzen bestimmen.

3 Abhängigkeiten und Copulas

3.1 Abhängigkeitsmaße

Kerninhalte

- Korrelationskoeffizient
- Rangkorrelationen
- Schätzer

3.1.1 Lineare Korrelation

Für Zufallsvariablen $X, Y : \Omega \longrightarrow \mathbb{R}$ heißt

$$\rho(X,Y) := \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = \frac{E(XY)-E(X)E(Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

der **Korrelationskoeffizient** und ist eine Maßzahl, die den (linearen) Zusammenhang zwischen X und Y quantifiziert. Es gilt $\rho(X,Y) \in [-1,1]$. Je näher $|\rho(X,Y)|$ an 1 ist, umso stärker ist der Zusammenhang, für unabhängige X, Y ist $\rho(X,Y) = 0$. Ist $(X,Y)^\top$ zwei dimensional normalverteilt, dann ist die Abhängigkeitsstruktur mit $\rho(X,Y)$ vollständig beschrieben.

Die Abhängigkeiten zwischen Zufallsvariablen ist aber meist zu komplex um mit nur einer Maßzahl erfasst zu werden.

3.1.2 Rangkorrelationen

Die Schätzer der folgenden Abhängigkeitsmaße werden aus den Rängen bivariater Stichproben bestimmt, daher spricht man von Rangkorrelationen.

Vergleicht man zwei unabhängige Realisierungen $(x_1,y_1)^\top$ und $(x_2,y_2)^\top$ von $(X,Y)^T$, dann wird bei einer positiven Abhängigkeit von X und Y die Steigung der Geraden durch die beiden Realisierungen positiv sein, also wird $(x_1-x_2)(y_1-y_2) > 0$ häufiger auftreten. Dies führt zur Definition von Kendalls tau, vgl. Definition 3.1 (a).

Bei einer positiven Abhängigkeit von X und Y sollte sich dies auch auf die Zufallsvariablen $F_X(X)$ und $F_Y(Y)$ auswirken. Spearmans rho misst die Korrelation zwischen ihnen, vgl. Definition 3.1 (b).

Definition 3.1. *(a) Sei $(X,Y)^\top$ ein Zufallsvektor und $(X',Y')^\top$ eine davon unabhängige Kopie. Dann ist **Kendalls tau** definiert als*

$$\begin{aligned}\rho_\tau(X,Y) &:= P((X-X')\cdot(Y-Y') > 0) - P((X-X')\cdot(Y-Y') < 0)\\ &= E[sign((X-X')\cdot(Y-Y'))].\end{aligned}$$

(b) *Seien X und Y Zufallsvariablen mit stetigen Verteilungsfunktionen F_X und F_Y. Dann ist* ***Spearmans rho*** *definiert als*

$$\rho_S(X,Y) := \rho(F_X(X), F_Y(Y)).$$

Beide so definierten Abhängigkeitsmaße sind symmetrisch, nehmen Werte in $[-1,1]$, verschwinden, wenn X, Y unabhängig sind und sind invariant gegenüber monoton wachsenden Transformationen.

3.1.3 Schätzer für die Korrelationen

Für eine bivariate Stichprobe $(x_i, y_i)^\top$, $i = 1, \dots, n$ ist der Schätzer für den Korrelationskoeffizienten in (1.1) gegeben.

Für die Schätzer der Rangkorrelationen werden die Ränge $rg(x_i)$, $rg(y_i)$, $i = 1, \dots, n$ benötigt. Spearmans rho kann wie in (1.1) geschätzt werden, nur dass man anstelle der Realisationen die Ränge verwendet:

$$r_S := \frac{\sum_{i=1}^{n} \left(rg(x_i) - \overline{rg_x}\right)\left(rg(y_i) - \overline{rg_y}\right)}{\sqrt{\sum_{i=1}^{n} \left(rg(x_i) - \overline{rg_x}\right)^2} \sqrt{\sum_{i=1}^{n} \left(rg(y_i) - \overline{rg_y}\right)^2}}. \tag{3.8}$$

Um Kendalls tau zu schätzen, wird angenommen, dass die bivariate Stichprobe bereits nach den Werten von x_i geordnet ist, also

$$rg(x_i) = i,\ i = 1, \dots n.$$

Mit $q_i := \sharp\left\{j > i \,|\, rg(y_i) > rg(y_j)\right\}$ ist der Schätzer gegeben durch

$$r_\tau := 1 - \frac{4\sum_{i=1}^{n} q_i}{n(n-1)}. \tag{3.9}$$

Lernergebnisse (C3)
Die Studierenden kennen die Definitionen der wichtigsten Abhängigkeitsmaße und können sie interpretieren. Sie sind in der Lage diese in konkreten Fällen aus gegebenen Daten zu schätzen.

3.2 Copulas

Kerninhalte

- Definition einer Copula
- Der Satz von Sklar
- Wichtige Beispiele
- Tailabhängigkeiten
- Parameterschätzung

Der Copulaansatz trennt die Randverteilungen und die Abhängigkeitsstruktur. Dies ist möglich mit dem Satz von Sklar, Satz 3.3.

3.2.1 Grundlagen

Definition 3.2 (Copula)**.** *Eine d-**dimensionale Copula** $C : [0,1]^d \longrightarrow [0,1]$ ist die Verteilungsfunktion eines Zufallsvektor $(U_1, \ldots, U_d)^\top$ mit $U_k \sim U[0,1]$, $k = 1, \ldots, d$.*

Man kann Copulas auch als Funktionen mit gewissen Monotonie-Eigenschaften definieren, so dass kein Rückgriff auf Zufallsvariablen notwendig ist.

Sind die U_k unabhängig, dann ergibt sich die **Unabhängigkeitscopula** $C(u_1, \ldots, u_d) := u_1 \cdot \ldots \cdot u_d$.

Satz 3.3 (Sklar)**.** *(a) Es sei F die Verteilungsfunktion eines Zufallsvektors $\mathbf{X} = (X_1, \ldots, X_d)^\top$ mit Randverteilungen $F_1, \ldots, F_d$. Dann existiert eine Copula $C : [0,1]^d \longrightarrow [0,1]$ mit*

$$F(\mathbf{x}) = C(F_1(x_1), \ldots, F_d(x_d)) \text{ für alle } \mathbf{x} \in \mathbb{R}^d. \tag{3.10}$$

Die Copula ist eindeutig, wenn $F_1, \ldots, F_d$ stetig sind.

(b) Sind $F_1, \ldots, F_d$ Verteilungsfunktionen von $X_1, \ldots, X_d$ und C eine d-dimensionale Copula, dann wird durch

$$F(x_1, \ldots, x_d) := C(F_1(x_1), \ldots, F_d(x_d))$$

eine Verteilungsfunktion mit Randverteilungen $F_1, \ldots, F_d$ definiert.

In Satz 3.3 (b) gilt bei stetigen F_k, $k = 1, \ldots, d$

$$C(u_1, \ldots, u_d) = F\left(F_1^{\leftarrow}(u_1), \ldots, F_d^{\leftarrow}(u_d)\right), \tag{3.11}$$

wobei $F_k^{\leftarrow}$ die Pseudoinverse von F_k ist, vgl. Definition 8.1.

Die Copula eines Zufallsvektors ändert sich unter streng monoton wachsenden Transformationen nicht. Genauer gilt

Satz 3.4. *Sei $\mathbf{X}$ ein Zufallsvektor mit stetigen Randverteilungen F_k, $k = 1, \ldots, d$ und $T_k : \mathbb{R} \longrightarrow \mathbb{R}$ streng monoton wachsende Abbildungen. Dann gilt $C_{(T_1(X_1), \ldots, T_d(X_d))} = C_{(X_1, \ldots, X_d)}$.*

3.2.2 Beispiele

Um die Darstellung einfach zu halten, wird für den Rest des Kapitels nur der zweidimensionale Fall, $d = 2$ betrachtet.

Gauß-Copula Die Copula der mehrdimensionalen Normalverteilung heißt **Gauß-Copula**. Ist $\mathbf{X} \sim \mathcal{N}_2(\mu, \mathbf{\Sigma})$, dann ist mit (3.11) und Satz 3.4 die Gauß-Copula gegeben durch

$$C^{\text{Ga}}(u_1, u_2) := \Phi_{2,\rho}(\phi^{-1}(u_1), \phi^{-1}(u_2)),$$

wobei $\Phi_{2,\rho}$ die Verteilungsfunktion von $\mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\right)$, $\rho := \text{Cov}(X_1, X_2)$ und ϕ die Verteilungsfunktion der Standardnormalverteilung ist.

t-Copula In analoger Weise wird die Copula der mehrdimensionalen t-Verteilung bestimmt. Ist $\mathbf{X} \sim t_2(\nu, \mu, \mathbf{\Sigma})$ bivariat t-verteilt mit ν Freiheitsgraden, dann ist mit (3.11) und Satz 3.4 die t-Copula gegeben durch

$$C^{\text{St}}(u_1, u_2) := t_{2,\rho}(t_\nu^{-1}(u_1), t_\nu^{-1}(u_1)),$$

mit $t_{2,\rho}$ die Verteilungsfunktion von $t_2\left(\nu, \begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\right)$, t_ν die Verteilungsfunktion der univariaten t-Verteilung mit ν Freiheitsgraden.

Gumbel-, Clayton- und Frank-Copula Die nun folgenden drei Copulas werden in expliziter Form angegeben, und sie gehören zur Familie der archimedischen Copulas. Sie sind durch

$$C(u_1, u_2) = \phi^{-1}(\phi(u_1) + \phi(u_2))$$

definiert, wobei $\phi : (0, 1] \longrightarrow [0, \infty)$ mit $\lim_{x\to 0} \phi(x) = \infty$, und $\phi(1) = 0$ eine streng monoton fallende, konvexe Funktion ist. Sie sind in Tabelle 3.2.2 zu finden.

Name	$C(u_1, u_2) =$	$\phi(x) =$	Bedingung
Gumbel C_θ^{Gu}	$\exp\left[-\left((-\ln u_1)^\theta + (-\ln u_2)^\theta\right)^{1/\theta}\right]$	$(-\ln x)^\theta$	$\theta \geq 1$
Clayton C_θ^{Cl}	$(u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta}$	$x^{-\theta} - 1$	$\theta > 0$
Frank C_θ^{Fr}	$-\frac{1}{\theta}\ln\left(1 + \frac{(e^{-\theta u_1}-1)(e^{-\theta u_2}-1)}{e^{-\theta}-1}\right)$	$-\ln\left(\frac{e^{-\theta x}-1}{e^{-\theta}-1}\right)$	$\theta \neq 0$

Tabelle 3.1: Beispiele von Copulas

3.2.3 Tailabhängigkeiten

Es werden nun Maßzahlen definiert, die Abhängigkeiten in den Tails quantifizieren.

Definition und Satz 3.5 (Tailabhängigkeiten)**.** *Seien X und Y stetige Zufallsvariablen mit Verteilungsfunktionen F_X und F_Y und Copula C. Dann heißt*

$$\lambda_U(X, Y) := \lim_{t\to 1-} P(Y > F_Y^{\leftarrow}(t) \,|\, X > F_X^{\leftarrow}(t)) = \lim_{q\to 1} \frac{1 - 2q + C(q, q)}{1 - q},$$

*der **Koeffizient der oberen Tailabhängigkeit**, falls dieser Grenzwert existiert. Analog heißt*

$$\lambda_L(X, Y) := \lim_{t\to 0+} P(Y \leq F_Y^{\leftarrow}(t) \,|\, X \leq F_X^{\leftarrow}(t)) = \lim_{q\to 0} \frac{C(q, q)}{q},$$

*der **Koeffizient der unteren Tailabhängigkeit**.*

Ein Wert von λ_U nahe eins deutet grob gesprochen an, dass Y mit hoher Wahrscheinlichkeit Werte am oberen Ende des Wertebereiches annimmt, wenn X das tut. Ist $\lambda_U = 0$, so sagt man, dass X und Y keine obere Tailabhängigkeit haben.

Für die behandelten Beispiele sind in Tabelle 3.2 Kendalls tau, Spearmens rho und die Tailabhängigkeiten in Abhängigkeit der Parameter dargestellt. **(∗)** bedeutet, dass keine geschlossene Formel bekannt ist.

3.2.4 Parameterschätzung

Die Rankorrelationen kann man aus den Daten schätzen, vgl. (3.8) und (3.9). Diese kann man dann verwenden um aus der Tabelle 3.2 Schätzer zu konstruieren, die in Tabelle 3.3 zu finden sind.

Es ist auch möglich mit Maximum Likelihood die Parameter schätzen, dazu muss man die bivariate Stichprobe (x_i, y_i), $i = 1, \ldots, n$ in **Pseudorealisationen** $\hat{\mathbf{u}}_i \in [0, 1]^2$ transformieren:

1. Schritt Bestimmen der empirischen Verteilungsfunktionen der Randverteilungen $\hat{F}_X, \hat{F}_Y$.

2. Schritt Transformation der Stichprobe

$$\hat{\mathbf{u}}_i = \left(\hat{F}_X(x_i), \hat{F}_Y(y_i)\right).$$

3. Schritt ML-Schätzung der Copula-Parameter mit den Pseudorealisationen $\hat{\mathbf{u}}_i$.

	ρ_τ	ρ_S	λ_L	λ_U
C_ρ^{Ga}	$\frac{2}{\pi}\arcsin(\rho)$	$\frac{6}{\pi}\arcsin\left(\frac{\rho}{2}\right)$	0	0
$C_{\rho,\nu}^{\text{St}}$	$\frac{2}{\pi}\arcsin(\rho)$	$(*)$	$2t_{\nu+1}\left(-\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}}\right)$	$=\lambda_L$
C_θ^{Cl}	$\frac{\theta}{\theta+2}$	$(*)$	$2^{-1/\theta}$	0
C_θ^{Gu}	$\frac{\theta-1}{\theta}$	$(*)$	0	$2-2^{1/\theta}$
C_θ^{Fr}	$\approx \frac{32}{9\theta}\ln\left(\cosh\left(\frac{\theta}{4}\right)\right)$	$\approx \frac{3}{2}\rho_\tau$	0	0

Tabelle 3.2: Kenngrößen für die Beispiele von Copulas

Copula	Parameterschätzer	Bemerkung
C_ρ^{Ga}	$\hat{\rho} = r_s$	$\rho \approx \rho_S$
$C_{\rho,\nu}^{\text{St}}$	$\hat{\rho} = \sin\left(\frac{\pi r_\tau}{2}\right)$	Auflösen von $r_\tau = \frac{2}{\pi}\arcsin\hat{\rho}$ nach $\hat{\rho}$
	$\hat{\nu}$	mit Maximum Likelihood.
C_θ^{Cl}	$\hat{\theta} = \frac{2r_\tau}{1-r_\tau}$	Auflösen von $r_\tau = \frac{\hat{\theta}}{\hat{\theta}+2}$ nach $\hat{\theta}$
C_θ^{Gu}	$\hat{\theta} = \frac{1}{1-r_\tau}$	Auflösen von $r_\tau = \frac{\hat{\theta}-1}{\hat{\theta}}$ nach $\hat{\theta}$
C_θ^{Fr}	$r_\tau = \frac{32}{9\hat{\theta}}\ln\left(\cosh\left(\frac{\hat{\theta}}{4}\right)\right)$	Auflösen nach $\hat{\theta}$

Tabelle 3.3: Schätzer für die Parameter von ausgewählten Copulas

3.2.5 Copulas in R

Im Statistik-Programm R sind Copula-Methoden in den Paketen Copula und QRM (Quantitative Risk Management) für die hier dargestellten Copulas implementiert.

Lernergebnisse (C3)
Die Studierenden kennen die grundlegenden Definitionen von Copulas und können sie mit dem Satz von Sklar interpretieren. Sie kennen die wesentlichen Beispiele von Copulas sowie deren Eigenschaften und können sie zur Beschreibung von Abhängigkeiten verwenden. Sie können mithilfe von R die Parameterschätzung durchführen und die Ergebnisse beurteilen.

4 Induktive Statistik[9]

Kerninhalte

- Unterscheide die wichtigsten eindimensionalen diskreten und stetigen Verteilungen und ihre Bedeutung für die Modellierung von Risiken. (B2)
- Beschreibe die wichtigsten mehrdimensionalen stetigen Verteilungen. (B2)
- Wende die asymptotischen Eigenschaften von Maximum-Likelihood-Schätzern an (inkl. Kenntnis von Eigenschaften ein- und mehrdimensionaler Exponentialfamilien, asymptotischer Normalität von ML-Schätzern, Fisher-Information, Konfidenzintervalle). (C3)
- Führe hierauf aufbauend Hypothesentests durch (Likelihood-Quotienten-Test, einfache und zusammengesetzte Hypothesentests). (C3)
- Erkläre die Struktur linearer und verallgemeinerter linearer Modelle (Designmatrix, linearer Prädiktor mit und ohne Wechselwirkungen, Linkfunktion, exponentielle Verteilungsfamilie, Erwartungswert- und Varianzfunktion, Devianz). (B2)
- Erkläre die wesentlichen Unterschiede zwischen Verfahren des maschinellen Lernens (ML) und statistischen Standardmodellen. Nenne Verfahren des überwachten und unüberwachten ML und mögliche Einsatzgebiete.

4.1 Verteilungen

Kerninhalte

- Verteilungen auf $\mathbb{R}_0^+$ oder $\mathbb{R}^+$
  - Beta-Verteilung
  - Burr-Verteilung
  - Inverse Gaußverteilung
  - Lognormal-Verteilung
  - Pareto-Verteilung
  - Weibull-Verteilung
- Mehrdimensionale Verteilungen
  - Multivariate Normalverteilung
  - Multinomial-Verteilung
- Extremwertverteilungen
  - Gumbel (Typ I)
  - Fréchet (Typ II)
  - Weibull (Typ III)

4.1.1 Verteilungen auf $\mathbb{R}_0^+$ oder $\mathbb{R}^+$

Linkssteile/rechtsschiefe Verteilungen für nicht-negative Zufallsvariablen oder Merkmale werden häufig für Risiken oder Schadenshöhen eingesetzt. Zwei prominente Vertreter sind die Lognormal- und Gamma-Verteilung.

[9] Teile dieser Darstellung basieren auf Fahrmeir, L., Heumann, C., Künstler, R., Pigeot, I., Tutz, G., Statistik — Der Weg zur Datenanalyse, Springer, Berlin (2016).

Die **Beta-Verteilung** $\mathcal{B}(\alpha, \beta)$, $\alpha > 0$, $\beta > 0$, ist auf dem Intervall $[0, 1]$ definiert und hat die Dichte

$$f(x) = \begin{cases} \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} & 0 \le x \le 1 \\ 0 & \text{sonst,} \end{cases}$$

wobei $B(\alpha, \beta)$ die Beta-Funktion bezeichnet. Es gilt: $\mathrm{E}(X) = \frac{\alpha}{\alpha+\beta}$ und $\mathrm{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

Die **Burr-Verteilung** oder **Burr Typ XII-Verteilung** besitzt die Dichte

$$f(x) = \begin{cases} \frac{ck}{\lambda}\left(\frac{x}{\lambda}\right)^{c-1}\left[1+\left(\frac{x}{\lambda}\right)^{c}\right]^{-k-1} & x \ge 0 \\ 0 & \text{sonst,} \end{cases}$$

mit $c > 0$, $k > 0$, $\lambda > 0$.

Die **inverse Gaußverteilung** oder **inverse Normalverteilung** $InvN(\lambda, \mu)$, $\lambda > 0$, $\mu > 0$, hat die Dichte

$$f(x) = \begin{cases} \left(\frac{\lambda}{2\pi x^3}\right)^{\frac{1}{2}} e^{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}} & x > 0 \\ 0 & \text{sonst.} \end{cases}$$

Es gilt: $\mathrm{E}(X) = \mu$ und $\mathrm{Var}(X) = \frac{\mu^3}{\lambda}$. Speziell gilt:

$$\mathrm{E}\left(\frac{1}{X}\right) = \frac{1}{\mu} + \frac{1}{\lambda} \quad \text{und} \quad \mathrm{Var}\left(\frac{1}{X}\right) = \frac{1}{\mu\lambda} + \frac{2}{\lambda^2}\,.$$

Die **Lognormal-Verteilung** $\mathcal{LN}(\mu, \sigma^2)$ lässt sich als Transformation einer normalverteilten Zufallsvariable darstellen. Ist X normalverteilt $N(\mu, \sigma^2)$, so ist $Y = \exp(X)$ lognormalverteilt mit $\mathrm{E}(Y) = \exp(\mu + \sigma^2/2)$ und $\mathrm{Var}(X) = \exp(2\mu + \sigma^2)(\exp(\sigma^2) - 1)$. Die Dichtefunktion kann über den Dichtetransformationssatz abgeleitet werden.

Die **Pareto-Verteilung** $\mathrm{Pa}(\alpha, d)$ wird für die Modellierung von Großschäden verwendet, welche einen Mindestschaden von d überschreiten. Die Dichte ist

$$f(x) = \begin{cases} \frac{\alpha d^\alpha}{x^{\alpha+1}} & \text{für } x \ge d \\ 0 & \text{sonst} \end{cases}$$

für $d > 0$ und $\alpha > 0$. Es gilt dann: der Erwartungswert existiert für $\alpha > 1$ und lautet dann $\mathrm{E}(X) = \frac{d\alpha}{\alpha-1}$. Die Varianz existiert für $\alpha > 2$ und ist dann $\mathrm{Var}(X) = \left(\frac{d}{\alpha-1}\right)^2 \frac{\alpha}{\alpha-2}$. Ist X $\mathcal{P}a(\alpha, d)$, so ist die transformierte Zufallsvariable $Y = X/d$ $\mathrm{Pa}(\alpha)$. Durch einen **Shift** erhält man die Pareto-Verteilung auf der positive reellen Achse einschließlich der 0:

$$f(x) = \begin{cases} \frac{\alpha\beta^\alpha}{(x+\beta)^{\alpha+1}} & \text{für } x \ge 0 \\ 0 & \text{sonst} \end{cases}$$

für $\beta > 0$ und $\alpha > 0$. D.h., eine komplette Schadensverteilung und nicht nur Schäden größer als einem Mindestschaden können modelliert werden.

Eine weitere Verteilung, die in diesem Kontext benutzt wird, ist die **Weibull-Verteilung** $\mathcal{W}(\beta, \delta)$ mit Dichtefunktion

$$f(x) = \begin{cases} \frac{\delta}{\beta^\delta} x^{\delta-1} e^{(-x/\beta)^\delta} & \text{für } x \ge 0 \\ 0 & \text{sonst} \end{cases}$$

für $\beta > 0$ und $\delta > 0$.

In vielen Fällen werden auch **Transformationen** von Zufallsvariablen durchgeführt, beispielsweise zur Berücksichtigung von **Inflationseffekten**, welche sowohl Schadenshöhen als auch Schadenshäufigkeiten beeinflussen können.

4.1.2 Mehrdimensionale Verteilungen

Zu den mehrdimensionalen Verteilungen zählen die **multivariate Normalverteilung** und die **Multinomalverteilung**.

Die multivariate Normalverteilung $N_p(\mu, \Sigma)$ beschreibt die Verteilung eines p-dimensionalen Zufallsvektors $X = (X_1, .., X_p)$ von Zufallsvariablen die i.a. nicht stochastisch unabhängig sein müssen. Die multivariate Normalverteilung ist durch den Erwartungswertvektor $\mu = (\mu_1, \ldots, \mu_p)$ und die symmetrische, positiv definite[10] $p \times p$ Kovarianzmatrix $\Sigma = (\sigma_{jk})_{1 \leq j,k \leq p}$ bestimmt. Die Dichtefunktion lautet ($x = (x_1, \ldots, x_p)$)

$$f(x) = \frac{1}{\sqrt{(2\pi)^p \det(\Sigma)}} \exp\left(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right) .$$

Dabei gilt:

$$\begin{aligned} \mu_j &= \mathrm{E}(X_j) \, , j = 1, \ldots, p \\ \sigma_{jj} &= \mathrm{Var}(X_j) \, , j = 1, \ldots, p \\ \sigma_{jk} &= \mathrm{Cov}(X_j, X_k) \, , j, k = 1, \ldots, p \, , j \neq k \end{aligned}$$

Die multivariate Normalverteilung besitzt die günstige Eigenschaft, dass alle Randverteilungen und alle bedingten Verteilungen wiederum Normalverteilungen sind, d.h.

$$X_j \sim N(\mu_j, \sigma_{jj}) \, , j = 1, \ldots, p$$

und für jede Aufteilung von X in zwei Teilvektoren $X = (X^1, X^2)$, wobei o.w.E. $X^1 = (X_1, X_2, \ldots, X_l)$ und $X^2 = (X_{l+1}, \ldots, X_p)$, $1 \leq l \leq p-1$, gilt

$$\begin{pmatrix} X^1 \\ X^2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu^1 \\ \mu^2 \end{pmatrix}, \begin{pmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{pmatrix}\right)$$

und

$$X^1 | X^2 \sim N\left(\mu^1 + \Sigma^{12}(\Sigma^{22})^{-1}(X^2 - \mu^2), \Sigma^{11} - \Sigma^{12}(\Sigma^{22})^{-1}\Sigma^{21}\right) .$$

Für die Verteilung metrischer Zufallsvariablen wird die multivariate Normalverteilung oftmals als näherungsweise Approximation verwendet. Ihre andere Bedeutung liegt in der asymptotischen Verteilung der Maximum-Likelihood Schätzung mehrdimensionaler Parameter, die die Bestimmung von Konfidenzintervallen und statistische Hypothesentests ermöglicht. Nützlich ist zudem folgende Eigenschaft. Ist X multivariat normalvertelt, $X \sim N(\mu, \Sigma)$, dann ist jede Lineartransformation ebenfalls multivariat normalverteilt, d.h. ist A eine $q \times p$-Matrix mit vollem Spaltenrang[11] und b ein $q \times 1$-Vektor, dann ist

$$Y = AX + b \sim N_q(A\mu + b, A\Sigma A') .$$

Die Multinomialverteilung kann als Verteilung über eine endliche Menge von Ausprägungen angewendet werden. Sie kann für eine diskrete Zufallsvariable, aber auch für die mehrdimensionale Verteilung von diskreten, abhängigen Zufallsvariablen verwendet werden. Dabei wird jede mögliche Kombination von Werten als eine eigene Ausprägung aufgefasst. Sind $a_1, \ldots, a_k$ die möglichen Ausprägungen mit $p_j = P(X = a_j)$ und werden n Zufallsvariablen realisiert, so ist

$$X = (X_1, \ldots, X_k) \sim \mathrm{Multinom}(n; p_1, \ldots, p_k) .$$

Dabei ist X_j die Anzahl der Werte mit realisierter Ausprägung a_j, $n = \sum_{j=1}^k X_j$, $\sum_{j=1}^k p_j = 1$ und

$$P(X_1 = x_1, \ldots, X_k = x_k) = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k} .$$

Wegen der Restriktion $\sum_{j=1}^k p_j = 1$ besitzt die Verteilung nur $(k-1)$ freie Parameter.

[10] Positive Definitheit garantiert die Invertierbarkeit der Matrix

[11] Der Fall $q > p$ führt auf jeden Fall zu einer singulären oder degenerierten Normalverteilung, deren normale Inverse nicht existiert

4.1.3 Extremwertverteilungen

Extremwertverteilungen beschreiben die Verteilung des Maximums einer Folge von iid Zufallsvariablen mit Verteilungsfunktion F.

Die **verallgemeinerte Extremwertverteilung** $GEV(\mu, \sigma, \xi)$ enthält als Spezialfälle die **Gumbel-Verteilung** (Typ I Extremwertverteilung), die **Fréchet-Verteilung** (Typ II Extremwertverteilung) und **inverse Weibull-Verteilung** (Typ III Extremwertverteilung).

Es seien $X_1, \ldots, X_n$ eine Folge von unabhängig und identisch verteilten Zufallsvariablen mit Verteilungsfunktion F und $M_n = \max(X_1, \ldots, X_n)$. Die exakte Verteilung des Maximums ist dann $P(M_n \leq z) = F(z)^n$. Abhängig von der Verteilung von M_n ist die asymptotische Verteilung für $n \to \infty$ eine der drei genannten Verteilungen: Gumbel für exponentielle, Weibull für leichte und Fréchet für schwere Tails der Verteilung von M_n.

Lernergebnisse (B2)
Die Studierenden sind mit den genannten Verteilungen vertraut. Sie können relevante Anwendungen aus der Versicherungswirtschaft aufzeigen.

4.2 Maximum-Likelihood-Schätzer

Kerninhalte

- Maximum-Likelihood Schätzung
- Maximum-Likelihood Schätzung in ein- und mehrdimensionalen Exponentialfamilien
- Eigenschaften (Invarianz, Delta-Methode, Konsistenz, asymptotische Normalität, Konfidenzintervalle, Tests)

4.2.1 Maximum-Likelihood Schätzung

Die Maximum-Likelihood Methode ist eine allgemeine Methodik zur Konstruktion von Schätzern und Schätzfunktionen für die Parameter einer Verteilung oder eines parametrischen statistischen Modells. Beispiele sind der Parameter p einer Binomialverteilung, die Varianz σ^2 einer Normalverteilung oder der Koeffizientenvektor $\boldsymbol{\beta}$ eines logistischen Regressionsmodells (siehe Abschnitt 4.3.2). Im einfachsten Fall betrachtet man eine Stichprobe von unabhängig und identisch verteilten Zufallsvariablen $(X_1, \ldots, X_n)$ mit einer durch einen Parameter θ parametrisierten Dichtefunktion $f(x|\theta)$. Nach Beobachtung der konkreten Stichprobe $\mathbf{x} = (x_1, \ldots, x_n)$ ist die Likelihoodfunktion eine Funktion proportional zum Produkt der Dichten

$$L(\theta) \propto f(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta) \,,$$

wobei $f(x_1, \ldots, x_n|\theta)$ die gemeinsame Dichtefunktion ist, das Produkt sich aus der Unabhängigkeitsannahme ergibt und die Funktion für feste $(x_1, \ldots, x_n)$ als Funktion von θ aufgefasst wird.

Die Maximum-Likelihood Schätzung für θ, $\hat{\theta}_{ML}$, ergibt sich dann als

$$\hat{\theta}_{ML} = \arg\max_{\theta} L(\theta) \,.$$

Die Interpretation ist anschaulich: wähle dasjenige $\hat{\theta}_{ML}$ als Schätzung, welches die Likelihood der realisierten Stichprobe maximiert. Im Fall diskreter Zufallsvariablen, bei denen die Dichte einer Wahrscheinlichkeitsfunktion entspricht, ist $\hat{\theta}_{ML}$ die Schätzung, welche die Wahrscheinlichkeit der realisierten Stichprobe maximiert und damit am plausibelsten ist. Im

allgemeinen Fall (Dichten sind keine Wahrscheinlichkeiten) ist dieser plausibelste Wert derjenige, der die Dichte maximiert. Dies setzt implizit voraus, dass die Likelihood ein eindeutiges globales Maximum besitzt. Numerische Verfahren, insbesondere im Fall eines mehrdimensionalen Parameters θ, können meist nur lokale Maxima auffinden. Alternativ lässt sich die logarithmierte Likelihood (Log-Likelihood Funktion)

$$l(\theta) = \log L(\theta) = c + \sum_{i=1}^{n} \log f(x_i|\theta)$$

(mit Konstante c) zur Maximierung verwenden, d.h.

$$\hat{\theta}_{ML} = \arg\max_{\theta} l(\theta) \ .$$

Da der Logarithmus eine streng monoton wachsende Funktion ist, nehmen $L(\theta)$ und $l(\theta)$ das Maximum an der gleichen Stelle an. Die Konstante c ist von θ unabhängig und kann beim Lösen des Maximierungsproblems ignoriert werden. Betrachtet man nur die Terme in $L(\theta)$ bzw. $l(\theta)$, die von θ abhängen, spricht man vom Kern der Likelihood. Das Maximierungsproblem wird praktisch meist so gelöst, dass die Log-Likelihood nach θ abgeleitet wird, und dann die Nullstellen dieser Ableitung gesucht werden. Die Ableitung

$$s(\theta) = \frac{\partial}{\partial \theta} l(\theta) \tag{4.1}$$

wird als **Score-Funktion** bezeichnet. Der ML Schätzer erfüllt dann die Bedingung[12].

$$s(\hat{\theta}) = 0 \ .$$

Streng genommen ist für ein lokales Maximum noch die Bedingung zweiter Ordnung,

$$\frac{\partial^2}{\partial \theta^2} l(\theta) < 0$$

zu überprüfen.[13] In der Praxis wird dies jedoch meist als erfüllt vorausgesetzt. Die zweite Ableitung bzw. die zweiten partiellen Ableitungen sind aber noch aus einem anderen Grund interessant. Sie geben die Krümmung der Likelihood im Maximum an. Je stärker diese Krümmung ist, desto präziser kann θ geschätzt werden, d.h. umso kleiner ist die Varianz der Schätzung von θ. Die Krümmung steht also in umgekehrt proportionaler Beziehung zur Varianz $\mathrm{Var}(\hat{\theta}_{ML})$.

Die **Fisher-Information** ist ein Maß für die Information, die in einer Stichprobe über einen zu schätzenden Parameter enthalten ist und steht in enger Beziehung zur Varianz. Man unterscheidet die **erwartete** Fisher-Information und die **beobachtete** Fisher-Information. Die erwartete Information wird in der Regel definiert als Varianz der Score-Funktion (4.1):

$$I(\theta) = \mathrm{Var}(s(\theta)) \ .$$

Unter bestimmten Regularitätsbedingungen ist die erwartete Information der negative Erwartungswert der zweiten Ableitung:

$$I(\theta) = \mathrm{E}\left(-\frac{\partial^2}{\partial \theta^2} l(\theta) \right) \ .$$

Alle Auswertungen erfolgen am „wahren“, unbekannten Parameter θ. Die Information muss also geschätzt werden durch Auswerten am ML-Schätzer.

Die beobachtete Information ist direkt die negative zweite Ableitung, ausgewertet am ML-Schätzer.

[12] Es gibt einfache Fälle wie die Gleichverteilung bei der der Träger von θ abhängt und das Maximum nicht durch Ableiten gefunden werden kann

[13] Für mehrdimensionales θ muss die Hesse-Matrix negativ definit sein.

Im Fall eines mehrdimensionalen Parameters θ ist die Fisher-Information $I(\theta)$ eine symmetrische Matrix mit Elementen

$$I(\theta)_{ij} = -\mathrm{E}\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j} l(\theta)\right) = \mathrm{Cov}(s(\theta)) \ .$$

Die beobachtete Informationsmatrix im mehrdimensionalen Fall ist die negative Hesse-Matrix, ausgewertet am ML-Schätzer.

Da

$$\mathrm{E}(s(\theta)) = 0$$

gilt, ist auch

$$I(\theta) = Var(s(\theta)) = \mathrm{E}\left[\left(\frac{\partial}{\partial\theta} l(\theta)\right)^2\right]$$

und analog im mehrdimensionalen Fall.

Die Information ist **additiv** für unabhängige Experimente. Es gilt daher

$$I(\theta)_{(X_1,\ldots,X_n)} = nI(\theta)_{X_1}$$

und, falls $(Y_1, \ldots, Y_m)$ eine weitere iid Stichprobe unabhängig von $(X_1, \ldots, X_n)$ ist, auch

$$I(\theta)_{(X_1,\ldots,X_n),(Y_1,\ldots,Y_m)} = I(\theta)_{(X_1,\ldots,X_n)} + I(\theta)_{(Y_1,\ldots,Y_m)} \ .$$

Die Fisher-Information ist zentral für die Schranke nach Rao-Cramér, die die Varianz eines **unverzerrten** Schätzers $\hat{\theta}$ für θ angibt. Es gilt:

$$\mathrm{Var}(\hat{\theta}) \geq \frac{1}{I(\theta)} \ ,$$

bwz., im mehrdimensionalen Fall,

$$\mathrm{Cov}(\hat{\theta}) \geq I(\theta)^{-1} \ .$$

Hier versteht sich $\geq$ im Sinne der Löwner-Halbordnung, d.h. $\mathrm{Cov}(\hat{\theta}) - I(\theta)^{-1}$ ist positiv semidefinit.

4.2.2 Maximum-Likelihood Schätzung in Exponentialfamilien

Eine Verteilungsfamilie gehört zur d-parametrigen Exponentialfamilie, wenn sich die Dichte (oder Wahrscheinlichkeitsfunktion im diskreten Fall) schreiben lässt als

$$f(x|\theta) = h(x)e^{\theta' t(x) - c(\theta)} \ , \theta \in \Theta \subseteq \mathbb{R}^d \ .$$

Hierbei wird angenommen, dass alle d Dimensionen notwendig sind, d.h. die **suffizienten Statistiken** $(t_1(x), \ldots, t_d(x))$ sind linear unabhängig. Diese Darstellungsform wählt θ als sog. kanonischen Parameter. Es gilt:

$$\frac{\partial c(\theta)}{\partial\theta_i} = \mathrm{E}_\theta\left(t_i(X)\right) \ .$$

Damit ergeben sich die Maximum-Likelihood Schätzgleichungen mit

$$l(\theta) = \theta' t(x) - c(\theta)$$

als

$$\frac{\partial l(\theta)}{\theta_i} = t_i(x) - \mathrm{E}_\theta\left(t_i(X)\right) = 0 \ .$$

Die Fisher-Information enthält die Elemente

$$I(\theta_{ij}) = -\frac{\partial^2 l(\theta)}{\partial\theta_i\partial\theta_j} = \mathrm{Cov}(t_i(x), t_j(x)) \ .$$

Die Likelihood ist in diesem Fall strikt konkav und es existiert ein eindeutiges Maximum.

4.2.3 Eigenschaften der ML-Schätzung

Bei den Eigenschaften der Maximum-Likelihood werden i.a. die **asymptotischen Eigenschaften** betrachtet, da die Eigenschaften für endlichen Stichprobenumfang nicht explizit herleitbar sind (z.B. die Verteilung). Die ML-Schätzung ist **konsistent**, **asymptotisch normalverteilt** mit der kleinst möglichen (asymptotischen) Varianz. Sie ist daher **asymptotisch effizient**.[14] Dazu betrachtet man die log-Likelihood Funktion in Abhängigkeit des Stichprobenumfangs n,

$$l_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \log f(x_i|\theta)$$

und den daraus abgeleiteten ML-Schätzer $\hat{\theta}_n$. Dann gilt: $\hat{\theta}_n$ konvergiert in Wahrscheinlichkeit gegen θ und ist damit konsistent.[15] Weiterhin lässt sich zeigen, dass für $n \to \infty$

$$\sqrt{n}(\hat{\theta}_n - \theta) \to N(0, I_1(\theta)^{-1}) \ ,$$

wobei $I_1(\theta)$ die Informationsmatrix berechnet für eine einzelne Beobachtung darstellt. Eine weitere Darstellung ist

$$\sqrt{n}(\hat{\theta}_n - \theta) \to N\left(0, \left[\frac{1}{n} I_n(\theta)\right]^{-1}\right) \ ,$$

die auch auf den Fall bedingt unabhängiger, aber nicht identisch verteilter Beobachtungen angewendet werden kann, zum Beispiel für generalisierte lineare Modelle. $I_n(\theta)$ ist die Information der gesamten Stichprobe. Eine Schätzung der Varianz erfolgt durch Auswerten der Information an der Stelle $\hat{\theta}$.

Basierend auf der asymptotischen Verteilung lassen sich Konfidenzintervalle und Hypothesentests für θ konstruieren.

Einfache Konfidenzintervalle für einen skalaren Parameter θ zum Niveau $1-\alpha$ ergeben sich als

$$\hat{\theta}_{ML} \pm z_{1-\frac{\alpha}{2}} \hat{\sigma}_{\hat{\theta}_{ML}} \ .$$

Dabei ist $z_{1-\frac{\alpha}{2}}$ das $(1-\frac{\alpha}{2})$-Quantil der Standardnormalverteilung und $\hat{\sigma}_{\hat{\theta}_{ML}}$ eine Schätzung für die Standardabweichung von $\hat{\theta}_{ML}$ basierend auf der inversen Information.

Hypothesentests können gemäß dem Dualitätsprinzip ebenfalls abgeleitet werden. Für die einfache Nullhypothese $H_0 : \theta = 0$ gegen die Alternative $H_1 : \theta \neq 0$ kann die Testgröße

$$Z = \frac{\hat{\theta}_{ML}}{\hat{\sigma}_{\hat{\theta}_{ML}}}$$

verwendet werden. H_0 wird dann abgelehnt, wenn $Z > |z_{1-\frac{\alpha}{2}}|$.

Für zusammengesetzte Hypothesen und Alternativen, $H_0 : \theta \in \Theta_0$, $H_1 : \theta \in \Theta_1$, wobei $\Theta_0 \subset \Theta$, $\Theta_1 \subset \Theta$ mit $\Theta_0 \cup \Theta_1 = \Theta$, $\Theta_0 \cap \Theta_1 = \emptyset$, lässt sich der **Likelihood-Quotienten-Test** mit Teststatistik

$$\Lambda(\mathbf{x}) = \frac{sup_{\theta \in \Theta_0} f(x_1, \ldots, x_n|\theta)}{\sup_{\theta \in \Theta} f(x_1, \ldots, x_n|\theta)}$$

formulieren.

Eine nützliche Eigenschaft der Maximum-Likelihood-Schätzung ist das **Invarianzprinzip**. Ist $\hat{\theta}_{ML}$ eine Maximum-Likelihood-Schätzung für θ und h eine ein-eindeutige Abbildung, dann gilt für den Maximum-Likelihood-Schätzer $\hat{\xi}_{ML}$ von $\xi = h(\theta)$: $\hat{\xi}_{ML} = h(\hat{\theta}_{ML})$.

[14] Alle Eigenschaften setzen natürlich die Gültigkeit des angenommenen parametrischen Modells voraus.
[15] Der ML-Schätzer ist in vielen Fällen sogar stark konsistent.

Die **Delta-Regel** erlaubt die Berechnung von asymptotischen Varianzen von Funktionen von asymptotisch normalverteilten Schätzern. Ist $\hat{\theta}_n$ asymptotisch normalverteilt mit Varianz-Kovarianzmatrix $V(\theta)$ und ist

$$h : \mathbb{R}^p \to \mathbb{R}^k \ , \quad k \leq p$$

eine Abbildung, so gilt für $n \to \infty$:

$$\sqrt{n}(h(\hat{\theta}) - h(\theta)) \to N\left(0, H(\theta)V(\theta)H(\theta)'\right) \ ,$$

wobei

$$H(\theta)_{ij} = \frac{\partial h_i(\theta)}{\partial \theta_j} \ , \quad 1 \leq i \leq k, 1 \leq j \leq p \ .$$

Dies gilt für alle θ, für die $h(\theta)$ komponentenweise stetig partiell differenzierbar ist.

Lernergebnisse (C3)
Die Studierenden sollen sowohl die Grundidee beschreiben als auch konkrete Berechnungen von Maximum-Likelihood Schätzern durchführen können. Dazu gehören die Bestimmung der asymptotischen Verteilung in konkreten Fällen, sowie die Anwendung für Hypothesentests und Konfidenzintervalle.

4.3 Lineare und verallgemeinerte lineare Modelle

Kerninhalte

- Lineares Modell in Matrixschreibweise
- Flexibilität des linearen Prädiktors (Transformationen, Wechselwirkungen)
- Verallgemeinertes lineares Modell für Exponentialfamilien (Linkfunktion, Erwartungswert- und Varianzfunktion, Devianz)
- Anwendungen auf Klassifikation, Regression (mit Spezialfall Varianzanalyse)
- Anwendung im Software-Tool (hier: R)

4.3.1 Lineares Regressionsmodell

Im Fall des linearen Regressionsmodells unterscheidet man drei Fälle.

Das **einfache lineare Regressionsmodell** behandelt den Fall eines abhängigen **Zielmerkmals** [16] Y und einer einzelnen **Einflussgröße**[17] X.

Das **multiple lineare Regressionsmodell** erweitert das einfache lineare Regressionsmodell um weitere Einflussgrößen.

Das **multivariate lineare Regressionsmodell** erweitert das multiple Regressionsmodell um weitere, möglicherweise korrelierte, Zielmerkmale.

Im folgenden soll der häufigste in der Praxis auftretende Fall des multiplen linearen Regressionsmodells beschrieben werden.

Die Daten sind gegeben in Form eines (Spalten-)Vektors eines **Zielmerkmals** Y und k **Einflussgrößen** $X_1, \ldots, X_k$. Die Einflussgrößen können in Form einer Matrix dargestellt werden.

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & & & \vdots \\ 1 & x_{n1} & \ldots & x_{nk} \end{pmatrix}.$$

[16] Andere Begriffe sind Zielvariable, abhängige Variable, Response-Variable, Regressand, Target-Variable, Label
[17] Andere Bezeichnungen sind unabhängige Variable, Kovariable, Kovariate, Regressor

Die Einflussgrößen können sowohl die Originalmerkmale, als auch davon abgeleitete Merkmale erhalten. Dazu gehören spezielle Kodierungen für nominale und ordinale Merkmale (z.B. Dummy-Kodierung) und transformierte metrische Merkmale (zum Beispiel Potenzen oder Logarithmus), siege auch Abschnitt 4.3.3. Deshalb nennt man **X** auch **Designmatrix**. Die Matrix **X** enthält zusätzlich in der ersten Spalte die Werte der künstlichen Variable $X_0 \equiv 1$, die restlichen Spalten enthalten die Werte zu den Variablen $X_1, \ldots, X_k$. Die Darstellung entspricht der Verarbeitung in entsprechender Software. Eine einzelne Zeile enthält die Beobachtungen eines Falles mit dem Wert für die Zielvariable, y_i, und dem Vektor der Kovariablen $x_i = (x_{i1}, \ldots, x_{ik})'$, $i = 1, \ldots, n$. Das lineare Modell lässt sich dann kompakt darstellen als

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}\,, \quad \mathrm{E}(\boldsymbol{\epsilon}) = \mathbf{0}\,, \mathrm{Var}(\boldsymbol{\epsilon}) = \sigma^2\mathbf{E}$$

wobei $\boldsymbol{\beta}$ der $(k+1) \times 1$)-Vektor der Regressionskoeffizienten und $\boldsymbol{\epsilon}$ der $(n \times 1)$-Vektor der Fehlervariablen ist.[18].

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}, \quad \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}.$$

Die **Schätzung** der Koeffizienten $\boldsymbol{\beta}$ und der Varianz σ^2 erfolgt in der Regel durch die **Methode der kleinsten Quadrate (KQ)**:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\mathbf{X}^\mathsf{T}\mathbf{y}\,, \tag{4.2}$$

$$\hat{\sigma}^2 = \frac{1}{n-(k+1)} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{4.3}$$

mit

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \ldots \hat{\beta}_k x_{ik}\,. \tag{4.4}$$

Die Werte $\hat{y}_i$ werden als **Prognosen** bezeichnet.

Die KQ-Schätzung setzt voraus, dass die Matrix **X** vollen Spaltenrang $(k+1)$ besitzt. Ansonsten ist die **Produktsummenmatrix** $\mathbf{X}^\mathsf{T}\mathbf{X}$ nicht invertierbar.

Als Gütemaße (goodness-of-fit) des Modells dienen das **Bestimmtheitsmaß** R^2 und das **adjustierte Bestimmtheitsmaß** R^2_{adj}, welche sich beide aus der **Streuungszerlegung**(ANOVA) ableiten lassen.

Ein wichtiger Aspekt der Modellinterpretation in der Praxis ist die Unterscheidung zwischen statistisch signifikanten Merkmalen (Merkmale, deren Koeffizienten signifikant von Null verschieden sind) und relevanten Merkmalen, also Merkmalen, die große Erklärungskraft für das Zielmerkmal haben, d.h. die einen großen Anteil der Streuung des Zielmerkmals erklären. Gerade im Fall von großen Datenmengen ergeben sich schnell signifikante Schätzungen, obwohl die Relevanz nicht gegeben ist.

4.3.2 Logistisches Regressionsmodell

Die Zielvariable Y ist hier angenommen als eine **Bernoulli-Variable**, wobei

$$\pi_i = P(Y_i = 1), \qquad 1 - \pi_i = P(Y_i = 0)\,. \tag{4.5}$$

für $i = 1, \ldots, n$. Die Auftretenswahrscheinlichkeit π_i hängt wiederum von den beobachteten Werten der Einflussgrößen, $x_{i1}, \ldots, x_{ik}$, ab. Das logistische Regressionsmodell lautet

$$\pi_i = \frac{\exp(\beta_0 + x_{i1}\beta_1 + \cdots + x_{ik}\beta_k)}{1 + \exp(\beta_0 + x_{i1}\beta_1 + \cdots + x_{ik}\beta_k)}\,. \tag{4.6}$$

[18]Die Fehlervariablen sind angenommen als unabhängige Zufallsvariablen mit Mittelwert 0 und konstanter Varianz

Mit der logistischen Funktion $h(z) = \exp(z)/\big(1+\exp(z)\big)$ erhält man die einfache Form

$$\pi_i = h(\beta_0 + x_{i1}\beta_1 + \cdots + x_{ik}\beta_k)\,. \tag{4.7}$$

Der Erwartungswert von Y_i, d.h. die Wahrscheinlichkeit π_i, hängt nun nicht direkt linear von den Einflussgrößen ab, sondern erst nach Transformation durch die Funktion h. Das Modell ist auch darstellbar als

$$\log \frac{\pi_i}{1-\pi_i} = \beta_0 + x_{i1}\beta_1 + \cdots + x_{ik}\beta_k\,, \tag{4.8}$$

wobei $\pi_i/(1-\pi_i)$ die Chancen („odds“) und $\log\big(\pi_i/(1-\pi_i)\big)$ die logarithmischen Chancen („Logits“) darstellen. Mit der letzten Darstellung lassen sich die Parameter einfach interpretieren: β_j ist diejenige Veränderung in Logits, die bei Zunahme von x_j um eine Einheit (von x_{ij} nach $x_{ij}+1$) bei festgehaltenen restlichen Kovariablen auftritt. Die Schätzung erfolgt gewöhnlich durch die **Maximum-Likelihood Methode**.

4.3.3 Flexibilität des linearen Prädiktors

Der lineare Prädiktor

$$\eta_i = \beta_0 + x_{i1}\beta_1 + \cdots + x_{ik}\beta_k$$

kann sehr flexibel gestaltet werden und auch nichtlineare Funktionen, zum Beispiel Polynome höheren Grades oder stückweise Polynome enthalten. Es muss lediglich die **Linearität in den Parametern** gewährleistet sein. Kategoriale Variablen könne mit Hilfe der **Dummy-Kodierung** oder der **Effekt-Kodierung** in den Prädiktor integriert werden. Transformationen von ursprünglichen Variablen oder neue aus mehreren Variablen abgeleitete Variablen können ebenfalls verwendet werden. Wechselwirkungen zwischen Variablen sind ebenfalls möglich, zum Beispiel als Produkt von zwei Variablen. D.h. im allgemeinen wird aus den ursprünglich k Kovariablen für jede Beobachtung ein **Design-Vektor** $z_i = z(x_i)$ mit $x_i = (x_{i1}, \ldots, x_{ik})$ bestimmt, der alle Variablen enthält, also inklusive Transformationen, kategorialen Dummyvariablen, Wechselwirkungen und sonstigen abgeleiteten Variablen. D.h. die Dimension von z_i kann erheblich größer als die von x_i sein und der Prädiktor lautet dann

$$\eta_i = \beta_0 + z_{i1}\beta_1 + \cdots + z_{il}\beta_l\,, l >= k\,.$$

4.3.4 Verallgemeinertes lineares Modell für Exponentialfamilien

Die logistische Regression ist eigentlich nur ein Spezialfall eines verallgemeinerten linearen Modells (GLM). Ein GLM wird durch folgende Annahmen charakterisiert:

Verteilungsannahme. Gegeben x_i, sind die y_i bedingt unabhängig und die bedingte Verteilung von y_i gegeben x_i gehört zu einer einfachen Exponentialfamilie mit bedingtem Erwartungswert $\mathrm{E}(y_i|x_i) = \mu_i$ und eventuell einem Skalenparameter ϕ, der nicht von i abhängt.

Strukturelle Annahme. Der Erwartungswert μ_i wird durch eine streng monoton wachsende und hinreichend glatte Funktion h in Beziehung zum linearen Prädiktor $\eta_i = z_i'\beta$ gesetzt:

$$\mu_i = h(\eta_i) = h(z_i'\beta).$$

h wird **Response-Funktion** genannt. Die Umkehrfunktion zu h wird **Link-Funktion** g genannt:

$$g(\mu_i) = \eta_i = z_i'\beta\,.$$

Damit ist ein GLM durch drei Komponenten charakterisiert: (i) die Art der Exponentialfamilie, (ii) die Response- bzw. Link-Funktion und (iii) den Design-Vektor z_i.

Die Maximum-Likelihood Schätzung für β erfolgt durch ein iteratives Verfahren (Newton-Raphson oder Fisher-Scoring). Als goodness-of-fit Maß dient die Devianz. Konfidenzintervalle und Tests beruhen auf der asymptotischen Normalität der Koeffizientenschätzungen.

4.3.5 Fallstricke: Variablenselektion, Extrapolation, Prognosefehler

Im Fall vieler Merkmale und der Annahme, dass ein relevanter Teil dieser Merkmale nicht relevant ist, werden oft automatische Verfahren der schrittweisen Variablenselektion eingesetzt (basierend auf R^2, R^2_{adj} und/oder dem Akaike Informationskriterium AIC). Man folgt hier dem **Prinzip der Sparsamkeit**, d.h. kleinere Modelle mit ähnlicher Güte wie größere Modelle werden größeren Modellen gegenüber bevorzugt.

Standardfehler, p-Werte und Konfidenzintervalle sind nach Abschluss der Selektionsverfahren i.a. inkorrekt und zu optimistisch, da die Selektion selbst ein stochastischer Prozess ist, der mit Unsicherheit behaftet ist. Alternativen sind penalisierte Schätzungen, zum Beispiel LASSO (least absolute shrinkage and selection operator), welche eine implizite Variablenselektion durchführen. Generell sollten bei Modellen, bei denen die Erklärung der Variabilität der Zielgröße im Vordergrund steht, Variablen möglichst nach inhaltlichen Gesichtspunkten und nicht automatisch selektiert/deselektiert werden.

Extrapolationen des Modells in mögliche Datenbereiche, die nicht durch die Daten abgedeckt sind, sind meist fragwürdig und liefern schlechte Prognosen.

Eine adäquate Abschätzung des wahren **Prognosefehlers** und damit der Prognosegüte eines Modells setzt in der Regel das Vorhandensein einer unabhängigen Teststichprobe oder die Anwendung von Kreuzvalidierung oder Resampling-Verfahren voraus. Der **in-sample** Prognosefehler ist immer eine Unterschätzung des wahren Prognosefehlers, der sich bei Anwendung auf neue Daten (**out-of-sample**) ergibt.

4.3.6 Anwendungen

- Cross-selling und Churn (Kundenbindung, Abwanderung des Kunden, Kündigung des Kunden)
- Up-selling und underwriting loadings
- Up-lift (Messung des zusätzlichen Effekts einer Intervention)
- Claim frequency and claims scoring
- Claimed amount
- Pricing
- Customer (Lifetime) Value
- Disease Management Program Evaluation
- Rating

4.3.7 Anwendung im Software-Tool

Zum Beispiel Funktionen *lm* und *glm* in R oder *proc genmod* in SAS.

Lernergebnisse (C3)
Die Studierenden können die Theorie zu linearen und verallgemeinerten linearen Modellen ausführlich erklären und die Modelle interpretieren. Sie kennen die relevanten Schätzverfahren, die dafür eingesetzt werden. Die Problematik der Extrapolation und Variablenselektion sowie das Prinzip der Sparsamkeit muss bekannt sein und erklärt werden können. Signifikanz soll unterschieden werden können von Relevanz. Der Unterschied von in-sample und out-of-sample Prognosefehlern muss erklärt werden können und die Konsequenzen für die Praxis – insbesondere zur Vermeidung von zu optimistischen Prognosen – beschrieben werden können. Berechnungen mit linearen und verallgemeinerten linearen Modellen sollen praktisch durchgeführt und die Ergebnisse interpretiert werden können.

4.4 Maschinelles Lernen

Kerninhalte

- Maschinelles Lernen (ML) versus statistischer Standardmodelle
- Überwachtes und unüberwachtes ML
- Mögliche Einsatzgebiete

4.4.1 Maschinelles Lernen (ML) versus statistischer Standardmodelle

Maschinelle Lernverfahren sind Verfahren, die durch datenangepasste Vorgehensweisen ohne den Hintergrund eines konkreten (parametrischen) statistischen Modells (z.B. eines linearen Modells) Zusammenhänge in den Daten erkennen können (sofern diese vorhanden sind). ML Verfahren können oft in Form eines (iterativen) Algorithmus dargestellt werden. Im Gegensatz zu statistischen Standardmodellen liefern ML Verfahren oft keine geschätzten Parameter und deren geschätzte Varianzen, so dass Inferenz nur durch Resampling-Verfahren (Bootstrap, Subsampling) eingeschränkt möglich ist.

4.4.2 Überwachtes und unüberwachtes ML

Überwachtes Lernen setzt das Vorhandensein eines Zielmerkmals voraus. Bei metrischen Zielmerkmalen entspricht ein ML Verfahren einer Regression, bei diskretem Merkmal einer Klassifikation. Ziel ist die Minimierung einer **Verlustfunktion**. Diese kann situationsbedingt sein. Daher sind ML Verfahren sehr flexibel anwendbar. Überwachte ML Verfahren zielen in der Regel auf Prognosen ab. Zur realistischen Einschätzung des Prognosefehlers werden die Daten in Trainings- und Testdaten aufgeteilt, wobei das Verfahren mittels der Trainingsdaten trainiert und der Fehler auf den Testdaten evaluiert wird. Alternativ wird eine Kreuzvalidierung durchgeführt. Komplexere Verfahren, bei denen zur Optimierung der Prognosequalität eine umfangreiche Optimierung von Tuning-Parametern erforderlich ist, benötigen oft zusätzliche Schritte. In diesem Fall können zusätzliche Validierungsdaten notwendig sein oder beispielsweise eine genestete Kreuzvalidierung.

Unüberwachtes Lernen dagegen arbeitet ohne Zielmerkmal und versucht Strukturen und Muster in den Daten zu erkennnen und gegebenenfalls ähnliche Objekte zu strukturieren. Ein typisches unüberwachtes Lernverfahren ist die **Clusteranalyse**.

4.4.3 Mögliche Einsatzgebiete

ML Verfahren werden einerseits dann verwendet, wenn die Prognose von zukünftigen Zuständen im Fokus steht. Zukunft kann dabei kurzfristig (Zustand innerhalb der nächsten paar Millisekunden) bis langfristig (in ein paar Jahren) definiert sein. Andererseits können ML Verfahren in der Versicherung für die Verbesserung von Prozessen durch die Strukturierung unstrukturierter Daten (Beispiel: Google Bildersuche) sorgen. Dabei sind die Verfahren in der Lage wichtiges Zusatzwissen mit zu berücksichtigen.

Lernergebnisse (B2)
Die Studierenden können die Unterschiede zwischen statistischen Standardmodellen und maschinellen Lernverfahren erläutern. Sie kennen den Unterschied zwischen überwachtem und unüberwachtem Lernen. Das Vorgehen zur Bestimmung eines Prognosefehlers im Fall von überwachtem Lernen kann präzise beschrieben werden. Sie können mögliche Einsatzgebiete aufzeigen.

5 Zeitreihenanalyse[19]

Kerninhalte

- Beschreibe die Grundidee des klassischen Modellansatzes einer univariaten Zeitreihe. (B2)
- Bestimme die Stationariät einer Zeitreihe (strikt und schwach) anhand von praxisrelevanten Beispielen. (B2)
- Beschreibe die grundlegenden Zeitreihenmodelle (AR, MA, ARMA, ARIMA). (B2)
- Wende Zeitreihen im Softwaretool an. (C3)

Ziele der Zeitreihenanalyse:

- Beschreiben der Charakteristiken und der Dynamik von Zeitreihen
- Zerlegung der Zeitreihen in Komponenten (Trend, Saison, etc.)
- Glättung von Zeitreihen
- Prognosen

5.1 Allgemeine Zeitreihenmodelle

Kerninhalte

- Grundmodell und Einbettung in die Theorie der stochastischen Prozesse
- Trend
- Saisonale Komponente
- Weißes Rauschen

Eine **univariate Zeitreihe** ist eine Folge von Beobachtungen $(y_1, y_2, \ldots, y_T)$ oder in kurzer Notation $\{y_t\}_{t=1}^{T}$. Im Folgenden nehmen wir an, dass die Beobachtungen zu äquidistanten Zeitpunkten erfolgen und dass es sich um wiederholte Beobachtungen einer metrisch skalierten Variable handelt. Man geht weiter davon aus, dass die einzelnen Beobachtungen korreliert sind. Das Grundmodell geht meist von einer **additiven Zusammensetzung** der Zeitreihe aus verschiedenen Komponenten aus. Ein einfaches **Trendmodell** ist

$$y_t = d_t + u_t \,, \quad t = 1, 2, \ldots, n \,.$$

Dabei ist d_t eine deterministische Trendkomponente, die die langfristige Entwicklung der Zeitreihe (Beispiel Sterblichkeitstrend) beschreibt, beispielsweise eine in t lineare Funktion und u_t eine (irreguläre) stochastische, stationäre Komponente.

Weitere Komponenten können saisonale und zyklische Komponenten sein. Saisonale Komponenten sollen dabei eher kurzfristige periodische Schwankungen bekannter Periodenlänge beschreiben, zyklische Komponenten eher längerfristige periodische Schwankungen unbekannter Periodenlänge. Ein Modell mit einer einfachen saisonalen Komponente s_t wäre dann

$$y_t = d_t + s_t + u_t \,, \quad t = 1, 2, \ldots, n \,.$$

Für die stochastische Komponente u_t wird meist **weißes Rauschen** angenommen, d.h. die u_t sind unabhängig und identisch $N(0, \sigma^2)$-verteilt mit konstanter Varianz σ^2.

Typische Saisonkomponenten beschreiben periodische Schwankungen basierend auf Monatswerten oder Quartalswerten.

[19] Teile dieser Darstellung basieren auf Pruscha, H., Statistisches Methodenbuch — Verfahren, Fallstudien, Programmcodes, Springer, Berlin (2006).

Das Ziel besteht dann in der Extraktion der einzelnen Komponenten aus der Zeitreihe, also d_t und s_t im obigen Beispiel. In der Regel geht man zweistufig vor. Zunächst werden Schätzungen $\hat{d}_t$ und $\hat{s}_t$ bestimmt, anschließend die Residuenzeitreihe

$$\hat{u}_t = y_t - \hat{d}_t - \hat{s}_t$$

gebildet. Falls $\hat{u}_t$ keinen Trend und keine Periode mehr enthält, wird $\hat{u}_t$ als gefundene stochastische Komponente betrachtet. Ist dies nicht der Fall, werden Trend und zyklische Komponenten neu bestimmt. Für die trend- und zyklus-bereinigte Zeitreihe wird dann ein stochastisches Zeitreihenmodell angepasst. Dazu wird vorausgesetzt, dass die Zeitreihe **stationär** ist, siehe (5.1)-(5.3).

Alternativ zur Bestimmung der Residuen (und damit zur Entfernung von Trend und Periode) lässt sich die Zeitreihe in vielen Fällen durch **Differenzenbildung** stationär machen. Liegt z.B. eine Zeitreihe nur mit Trendkomponente vor, so sind einfache Methoden der Trendbereinigung die Bildung erster Differenzen

$$\tilde{y}_t = y_t - y_{t-1}$$

im Fall eines linearen Trends, oder die Verwendung von zweiten Differenzen

$$\tilde{y}_t = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2})$$

im Fall eines quadratischen Trends.

Für die **Trendbestimmung** können Polynome

$$y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \ldots \beta_q t^q + u_t \ , \quad t = 1, \ldots, n$$

oder gleitende Durchschnitte (siehe Abschnitt 5.2) verwendet werden. Im ersten Fall ist die Trendkomponente dann gegeben durch

$$\hat{d}_t = \hat{\beta}_0 + \hat{\beta}_1 t + \hat{\beta}_2 t^2 + \ldots \hat{\beta}_q t^q \ .$$

Für die Bestimmung der **Saisonkomponente** wird zunächst eine Trendbereinigung vorgenommen, die die Saisonkomponente möglichst wenig beeinflussen soll. Für Monatsdaten bietet sich z.B. ein 12-er Durchschnitt an. Für Periodenlänge p und einer **starren** Saisonfigur nimmt man an, dass

$$s_t = s_{t+p}$$

mit der Zentrierung $\sum_{t=1}^{p} s_t = 0$.

Die stochastische Komponente einer Zeitreihe kann als Realisation eines zeitdiskreten stochastischen Prozesses mit kontinuierlichem Zustandsraum aufgefasst werden (siehe Abschnitt 6.1). Der stochastische Prozess ist also der sogenannte datengenerierende Prozess, die beobachtete Zeitreihe eine (einzige) Realisation dieses Prozesses zusammen mit den deterministischen Komponenten. In diesem Sinne unterscheidet sich die Zeitreihenanalyse von üblichen statistischen Analysen, da lediglich eine einzige Realisation des Prozesses beobachtet wird.

Um den stochastischen Prozess charakterisieren zu können (zum Beispiel Erwartungswert, Varianz, Auto-Kovarianzfunktion), beschränkt man sich auf stationäre stochastische Prozesse. Eine wichtige allgemeine Klasse von stationären Prozessen sind die ARIMA (autoregressive integrated moving average) Prozesse. Diese enthalten als Spezialfall AR (autoregressive) und MA (moving average) Prozesse, siehe Abschnitt 5.2. Nützlich für die Darstellung dieser verschiedenen Modelle sind der **Lag-Operator**

$$\begin{aligned} Ly_t &= y_{t-1} \\ L^2 y_t &= Ly_{t-1} = y_{t-2} \\ L^p y_t &= y_{t-p} \ . \end{aligned}$$

und der **Differenzbildungsoperator**

$$\nabla y_t = y_t - y_{t-1} \ , \quad t = 1, \ldots, n \ ,$$

der ebenfalls wiederholt angewendet werden kann, z.B. (für $t \geq 3$)

$$\begin{aligned} \nabla^2 y_t &= \nabla(\nabla y_t) = \nabla y_t - \nabla y_{t-1} \\ &= (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2} \ . \end{aligned}$$

Eine Zeitreihe ist charakterisiert durch ihre Erwartungswertfunktion

$$\mathrm{E}(Y_t) = \mu(t) \ ,$$

durch ihre Kovarianzfunktion

$$\mathrm{Cov}(Y_t, Y_{t+l}) = \gamma(t, t+l) = \mathrm{E}[(Y_t - \mu_t)(Y_{t+l} - \mu_{t+l})] \ , l = 0, 1, \ldots \ ,$$

und durch die Autokorrelationsfunktion

$$\rho(t, t+l) = \frac{\gamma(t, t+l)}{\sqrt{\gamma(t,t)\gamma(t+l, t+l)}} \ , l = 0, 1, \ldots \ .$$

Die Zeitreihe heißt dann **schwach stationär**, falls für alle $t, l = 0, 1, \ldots$

$$\mu(t) = \mu \tag{5.1}$$

$$\gamma(t, t+l) = \gamma(l) \tag{5.2}$$

$$\rho(t, t+l) = \rho(l) = \frac{\gamma(l)}{\sqrt{\gamma(0)\gamma(0)}} = \frac{\gamma(l)}{\gamma(0)} = \frac{\gamma(l)}{\sigma^2} \ , \tag{5.3}$$

mit $\sigma^2 = \gamma(0)$. $\gamma(l)$ und $\rho(l)$ sind dann die Autokovarianz- bzw. Autokorrelationsfunktion, die lediglich vom Lag l abhängen. Es gilt $\rho(0) = 1$.

Ein typisches Beispiel einer nichstationären Zeitreihe ist der **random walk**

$$y_t = y_{t-1} + u_t \ ,$$

mit weißem Rauschen u_t. Mit zunehmender Zeit wird dabei die Varianz immer größer.

Eine Schätzung der Autokorrelationsfunktion im stationären Fall erfolgt durch die empirische Autokorrelationsfunktion

$$r(l) = \frac{\sum_{t=1}^{T-l}(y_{t+l} - \bar{y})(y_t - \bar{y})}{\sum_{t=1}^{T}(y_t - \bar{y})^2} \ .$$

Trägt man die empirischen Autokorrelationsfunktionen $r(l)$ über $l = 0, 1, \ldots$ ($r(0) = 1$) auf, enthält man das **Korrelogramm**. Unter gewissen Voraussetzungen ist $r(l)$ ein konsistent und asymptotisch normalverteilter Schätzer für $\rho(l)$.

Weitere interessante Kenngrößen sind die **partiellen Autokorrelationen** (PACF) $\pi(l)$, die die Korrelation zwischen y_t und y_{t+l} gegeben die Werte $y_{t+1}, \ldots, y_{t+l-1}$ an den Zwischenzeitpunkten messen. Betrachtet man die Vorhersage der Regression

$$\hat{y}_t = \hat{\alpha}_1 y_{t-1} + \ldots \hat{\alpha}_l y_{t-l} \ ,$$

so ist $\hat{\alpha}_l$ eine Schätzung für $\pi(l)$ ($l \geq 2$).

Lernergebnisse (B2)
Die Studierenden kennen das klassische Zeitreihenmodell und können die Komponenten und Eigenschaften erklären. Sie können den Unterschied zwischen Autokorrelationsfunktion und partieller Autokorrelationsfunktion erklären und die empirischen Schätzungen berechnen. Anwendungsbereiche in der Versicherungswirtschaft können benannt und erklärt werden.

5.2 Erweiterungen des klassischen Modells

Kerninhalte

- Gleitende Durchschnitte
- Autoregressive Modelle
- Moving average Modelle
- ARMA und ARIMA Modelle

Gleitende Durchschnitte können für die **Glättung** einer Zeitreihe und zur **Saisonbereinigung** verwendet werden. Ein gleitender Durchschnitt ungerader Ordnung wird an einem bestimmten Zeitpunkt t so gebildet:

$$\tilde{y}_t = \frac{1}{2q+1} \sum_{l=-q}^{q} y_{t+l} \, , q \in \mathbb{N} \, .$$

Ein gleitender Durchschnitt gerader Ordnung an einem bestimmten Zeitpunkt t ist ein gewichtetes arithmetisches Mittel

$$\tilde{y}_t = \frac{1}{2q} \left(\frac{1}{2} y_{t-q} + \sum_{l=-(q-1)}^{q-1} y_{t+l} + \frac{1}{2} y_{t+q} \right) , q \in \mathbb{N} \, .$$

Für beliebige Zeitreihendaten ist die Glättung umso stärker, je größer q ist.

Für Quartalsdaten mit Saison bietet sich ein gleitender Durchschnitt gerader Ordnung mit $q = 2$ an, für Monatsdaten mit Saison ein gleitender Durchschnitt gerader Ordnung mit $q = 6$.

Modelle für stationäre Zeitreihen erlauben eine weitergehende Analyse von Zeitreihen und die Verwendung von Zeitreihen für Prognosen in die Zukunft. Im Folgenden betrachten wir o.w.E. Zeitreihen mit Erwartungswert $\mu_t = 0$.

Ein Modellansatz sind **autoregressive Modelle** der Form

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots \alpha_p y_{t-p} + u_t \, ,$$

u_t weißes Rauschen. Man spricht in diesem Fall von einem AR(p)-Modell. Mittels des Lag-Operators lässt sich dieses Modell schreiben als

$$(1 - \alpha_1 L - \alpha_2 L^2 - \ldots \alpha_p L^p) y_t = u_t \, .$$

Die Stationarität eines AR(p)-Prozesses ist in der Praxis schwierig überprüfbar, formal muss die Bedingung erfüllt sein, dass alle Nullstellen des Polynoms

$$a(\lambda) = 1 - \alpha_1 \lambda - \ldots - \alpha_p \lambda^p$$

betragsmäßig größer als 1 sind (für $\alpha_0 = 1$). Der Koeffizient α_p eines AR(p)-Prozesses entspricht der partiellen Autokorrelation von y_t und y_{t+p} nach Bereinigung um die Zwischenbeobachtungen $(y_{t+1}, \ldots, y_{t+p-1})$.

Ein **moving average** Modell ist gegeben durch

$$y_t = \beta_1 u_{t-1} + \beta_2 u_{t-2} + \ldots + \beta_q u_{t-q} + u_t \, .$$

Jede Komponente u_t ist eine Realisation von weißem Rauschen. Man spricht von einem MA(q)-Prozess. Ein MA-Prozess ist immer stationär.

Ein ARMA(p, q)-Prozess ist gegeben durch

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots \alpha_p y_{t-p} + u_t + \beta_1 u_{t-1} + \beta_2 u_{t-2} + \ldots + \beta_q u_{t-q} \, .$$

Die Stationarität des ARMA-Prozesses hängt nur vom AR-Teil ab.

Ein ARIMA(p, d, q)-Prozess ergibt sich, wenn die d-ten Differenzen $\nabla^d y_t$ der Zeitreihe einem ARMA(p, q)-Prozess folgen. Die Differenzordnung wird so gewählt, dass die d-ten Differenzen keinen Trend aufweisen.

Die Parameter(Ordnungen) p und q werden durch die Autokorrelationsfunktion und die partielle Autokorrelationsfunktion bestimmt.

Die Schätzung der Parameter α_i, $i = 1, \ldots, p$ und β_j, $j = 1, \ldots, q$ erfolgt über die Residuen und die Minimierung der Residuenquadratsumme (MA, ARMA) oder unter Benutzung des Korrelogramms (AR). Die Koeffizienten eines AR-Prozesses lassen sich beispielsweise mittels der **Yule-Walker-Gleichungen** schätzen.

Lernergebnisse (B2)
Die Studierenden können die wichtigsten Modelle der Zeitreihenanalyse definieren und die Parameter interpretieren. Für einfache autoregressive Prozesse (AR(1), AR(2)) kann die Stationarität bewiesen werden. Zustandsraummodelle können als Alternative für nichtstationäre Zeitreihen benannt werden.

5.3 Spezial- und Mischformen

Kerninhalte

- Zeitreihen mit Kovariablen
- Bivariate Zeitreihen
- Zustandsraummodelle

Zeitreihen liegen oft zusammen mit weiteren Kovariablen x_t vor. Diese wiederum können konstant über die Zeit oder zeitvariierend sein. Liegt eine Zeitreihe mit einer exogenen Kovariablenzeitreihe vor, so lässt sich der ARMAX(p, q, s)-Ansatz verwenden, der aus einem AR, einem MA und der exogenen Reihe x_t besteht:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots \alpha_p y_{t-p} + u_t + \beta_1 u_{t-1} + \beta_2 u_{t-2} + \ldots + \beta_q u_{t-q} + \sum_{j=1}^{s} \theta_j x_{t-j} \, .$$

Ein Spezialfall tritt auf, wenn sehr viele Zeitreihen (Längsschnittdaten) vorliegen, beispielsweise jährliche Kosten aller Verträge über mehrere Jahre von einem Versicherungsbestand. Klassische Zeitreihenmodelle nicht hier nicht unbedingt geeignet. Für Prognosezwecke können dann beispielsweise bedingte generalisierte lineare Modell besser geeignet sein. Dabei werden sowohl vergangene Werte der Zeitreihe, als auch Kovariablen im Prädiktor aufgenommen. Der Prädiktor η_t hat für eine skalare Kovariable also etwa die Gestalt

$$\eta_t = \alpha_0 + \alpha_1 y_{t-1} + \ldots + \alpha_p y_{t-p} + \beta_t x_t + \ldots + \beta_s x_{t-s} \, .$$

Dieser Ansatz erlaubt sogar die Schätzung binärer und kategorialer Zeitreihen.

Liegt eine einzelne bivariate Zeitreihe (x_t, y_t) vor, so interessiert, neben den üblichen Kenngrößen welche für jede einzelne Zeitreihe berechnet werden können, die **Kreuz-Kovarianzfunktion** und die **Kreuz-Autokorrelationsfunktion**

$$\begin{aligned} \mathrm{Cov}(x_t, y_{t+l}) &= \gamma_{xy}(l) \\ \mathrm{Cor}(x_t, y_{t+l}) &= \rho_{xy}(l) \end{aligned}$$

mit den Eigenschaften $\gamma_{xy}(-l) = \gamma_{yx}(l)$ und $\rho_{xy}(-l) = \rho_{yx}(l)$. Für l lässt man hier die ganzen Zahlen $\mathbb{Z}$ zu. Beide Funktionen können über den Wertebereich von $\mathbb{Z}$ grafisch dargestellt werden (Kreuz-Korrelogramm). Die Kreuz-Korrelationsfunktion ist dabei weder symmetrisch,

noch nimmt sie für $l = 0$ den Wert 1 an. Eine Schätzung erfolgt durch die empirische Kreuz-Kovarianzfunktion.

Allgemein existieren mit VARIMA, VAR und VARIMAX Verallgemeinerungen von ARIMA, AR und ARMAX auf multivariate Zeitreihen.

Zustandsraummodelle sind eine erfolgreich angewendete Modellklasse auch für nichtstationäre Zeitreihen. Während in ARIMA-Modellen die Trend- und periodischen Komponenten als nuisance-Parameter aufgefasst werden, werden sie in Zustandsraummodellen explizit modelliert. Eine kompakte Darstellung ist

$$\begin{aligned} y_t &= z_t'\alpha_t + \varepsilon_t\,, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2) \\ \alpha_t &= \mathbf{F}_t\alpha_{t-1} + \mathbf{R}_t\eta_t\,, \quad \eta_t \sim \mathcal{N}(0, \mathbf{Q_t})\,. \end{aligned}$$

Dabei sind z_t, α_t, η_t Vektoren und $\mathbf{F}_t, \mathbf{R}_t$ Matrizen. Die Zustände $\alpha_\mathbf{t}$, die Varianz σ_ε^2, sowie die Varianz-Kovarianzmatrix $\mathbf{Q_t}$ sind das Ziel der Schätzung. ARIMA-Modelle können in die Theorie der Zustandsraummodelle eingebettet werden. Zustandsraummodelle erlauben ebenfalls eine einfache Einbeziehung von Kovariablen im Designvektor z_t.

Lernergebnisse (B2)
Die Studierenden können die verschiedenen Situationen, in denen abhängige Beobachtungen in Form vom Zeitreihen zusammen mit Kovariablen auftauchen, erläutern. Bivariate Zeitreihen als Verallgemeinerung sind bekannt und die zusätzlichen interessierenden Parameter können formal dargestellt und interpretiert werden.

5.4 Modelldiagnostik und Prognose

Kerninhalte

- Modelldiagnostik
- Residualanalyse
- Prognoseverfahren
- Grafische Darstellung und grafische Methoden der Analyse

Die Bestimmung der Parameter eines ARIMA-Modells wird durch die sogenannte **Box-Jenkins-Methode** in drei Schritten vorgenommen:

(a) Identifikation des Modells. Nach Bereinigung einer Trendkomponente und saisonaler Effekte (durch geeignete Differenzenbildung) werden die Ordnungen p und q durch die Autokorrelationsfunktion und die partielle Autokorrelationsfunktion bestimmt.

(b) Schätzung der Parameter durch Maximum-Likelihood oder die nichtlineare KQ-Methode.

(c) Überprüfung des Modells. Hierfür werden die Residuen analysiert. Diese sollten eine konstante Mittelwerts- und Varianzfunktion über die Zeit aufweisen und möglichst unkorreliert sein. Hilfreich sind hierzu grafische Darstellungen: Residual-Plots, ACF-Plot und PACF-Plot. Autokorrelation erster Ordnung kann mit dem **Durbin-Watson-Test** getestet werden. Erfüllen die Residuen nicht die Anforderungen, so geht man zurück zum ersten Schritt.

Auf der Basis des schließlich ausgewählten Modells kann dann eine Prognose über den zukünftigen Verlauf der Zeitreihe durchgeführt werden. Hierbei betrachtet man die Vorhersagefunktion $y_n(l)$, die die l-Schritt Prognose ($l = 1, 2, \ldots,$) auf Basis der bekannten Werte $y_1, \ldots, y_n$ angibt. Die beste lineare Vorhersage ist dabei der bedingte Erwartungswert $\hat{y}_{n+l} = \mathrm{E}(y_{n+l}|y_1, \ldots, y_n)$, welcher den mittleren quadratischen Vorhersagefehler

$$\mathrm{E}\,(\hat{y}_{n+l} - y_{n+l})^2$$

minimiert. Für ARIMA-Modelle stehen entsprechende schrittweise Berechnungsformeln zur Verfügung. Zusätzlich lassen sich **Prognosefehler** und **Prognoseintervalle** berechnen.

Zustandsraummodelle werden ebenfalls durch mehrere Schritte geschätzt (Glättung, Filterung, Vorhersage) und können dann zur Vorhersage verwendet werden.

Lernergebnisse (C3)
Die Studierenden können die Ordnung von MA– und AR–Prozessen anhand der Autokorrelationsfunktion und der partiellen Autokorrelationsfunktion bestimmen. Punktprognosen von MA –und AR–Prozessen können durchgeführt werden.

6 Stochastische Prozesse[20]

6.1 Beschreibung stochastischer Prozesse

Kerninhalte

- Definition von stochastischen Prozessen
- Definition der Stationarität
- Beispiele stochastischer Prozesse

Risiken, die sich dynamisch im Zeitablauf entwickeln, beschreibt man üblicherweise durch stochastische Prozesse . Ein **stochastischer Prozess** ist eine Familie $\{X_t\}_{t\in T}$ von Zufallsvariablen $X_t : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$ über einen Zeitbereich $T \subseteq [0, \infty)$. Die Menge $S := \bigcup_{t\in T} X_t(\Omega)$ heißt **Zustandsraum**.

Falls $T = \{0, \dots, n\}$ oder $T = \mathbb{N}_0$ bzw. $T = [0, \infty)$ oder $T = [a, b]$, dann heißt der stochastische Prozess **zeitdiskret** bzw. **zeitstetig**. bzw. Die Realisierungen des stochastischen Prozesses sind seine **Pfade** $t \mapsto X_t(\omega)$ für $\omega \in \Omega$. Im Wesentlichen werden stochastische Prozesse mit diskretem (also endlichem oder abzählbarem) oder kontinuierlichem Zustandsraum betrachtet.

Die Existenz von stochastischen Prozessen bei vorgegebenen endlich dimensionalen Randverteilungen ist Inhalt des **Existenzsatzes von Kolmogorov**. Ein stochastischer Prozess $\{X_t\}_{t\in T}$ heißt **stationär**, wenn für alle $h > 0$, alle $n \in \mathbb{N}$ und alle $t_1, \dots, t_n \in T$ mit $t_1 + h, \dots, t_n + h \in T$

$$(X_{t_1+h}, \dots, X_{t_n+h}) \stackrel{d}{=} (X_{t_1}, \dots, X_{t_n})$$

gilt, d.h. seine Verteilung ändert sich im Verlauf der Zeit nicht. Hierbei bedeutet $\stackrel{d}{=}$, dass dieselbe Verteilung vorliegt. Stationarität wird bei der Beschreibung der wichtigsten stochastischen Prozesse (Poisson-Prozess, Brownsche Bewegung) benötigt. Trivialerweise ist ein stochastischer Prozess $\{X_n\}_{n\in\mathbb{N}}$ mit identisch verteilten, unabhängigen X_n stationär.

Übergänge in der Personenversicherung können mit zeitstetigen und zeitdiskreten stochastischen Prozessen mit endlich vielen Zuständen modelliert werden und können mit Hilfe von gerichteten Graphen visualisiert werden. In der Schadenversicherung und der Finanzmathematik werden darüber hinaus stochastische Prozesse mit kontinuierlichem Zustandsraum verwendet.

Lernergebnisse (B2)
Die Studierenden kennen die Definition von stochastischen Prozessen (SP), können zeitdiskrete, zeitstetige SP unterscheiden und die Stationarität eines gegebenen Prozesses beurteilen. Sie sind in der Lage Beispiele von SP zu benennen.

[20] Teile dieser Darstellung basieren auf Becker, T., Herrmann, R., Sandor, V., Schäfer, D., Wellisch, U. : Stochastische Risikomodellierung und statistische Methoden, 2. Auflage, Springer, Berlin (2024).

6.2 Endliche Markov-Ketten

Kerninhalte

- Definition Markov-Ketten und homogenen Markov-Ketten
- Asymptotik

Endliche **Markov-Ketten** sind stochastische Prozesse $\{X_n\}_{n\in\mathbb{N}_0}$ in diskreter Zeit und endlichem Zustandsraum $S = \{1, \ldots, m\}$ mit der **Markoveigenschaft**: ihre zukünftige Entwicklung wird nur von der Gegenwart, nicht aber von der Vergangenheit bestimmt (man spricht auch von Gedächtnislosigkeit), d.h. für alle $n \in \mathbb{N}$ und alle $x_0, \ldots, x_n \in S$ gilt

$$P(X_n = x_n | X_0 = x_0, \ldots, X_{n-1} = x_{n-1}) = P(X_n = x_n | X_{n-1} = x_{n-1}).$$

Die Übergangswahrscheinlichkeiten $p_{ij}(n) := P(X_n = j | X_{n-1} = i), n \in \mathbb{N}$ lassen sich dann in **Übergangsmatrizen**

$$\mathbf{P}(n) := \big(p_{ij}(n)\big)_{i,j=1,\ldots,m} \in \mathbb{R}^{m\times m}$$

zusammenfassen. Man setzt $\mathbf{P}(0) := \mathbf{E}$, die Einheitsmatrix. Der Zeilenvektor

$$\mathbf{p}(n) := (P(X_n = 1), \ldots, P(X_n = m))$$

beschreibt die Verteilung der Markov-Kette zum Zeitpunkt $n \in \mathbb{N}_0$.

Aus obigen Definitionen folgt unmittelbar für die Übergangsmatrizen und die Verteilung der Markov-Kette:

$$\begin{aligned} \mathbf{P}(X_n = j | X_0 = i) &= [\mathbf{P}(1) \cdot \mathbf{P}(2) \cdot \ldots \cdot \mathbf{P}(n)]_{ij} \\ \mathbf{p}(n) = \mathbf{p}(n-1) \cdot \mathbf{P}(n) &= \mathbf{p}(0) \cdot \mathbf{P}(1) \cdot \mathbf{P}(2) \cdot \ldots \cdot \mathbf{P}(n). \end{aligned} \tag{6.4}$$

6.2.1 Homogene endliche Markov-Ketten

Andern sich die Übergangswahrscheinlichkeiten im Zeitverlauf nicht, spricht man von einer homogenen Markov-Kette. Eine endliche Markov-Kette heißt **homogen**, wenn für alle $n \in \mathbb{N}$

$$p_{ij}(n) =: p_{ij}$$

gilt. Nach (6.4) gilt wegen $\mathbf{P} := \mathbf{P}(1) = \mathbf{P}(2) = \ldots$

$$\mathbf{p}(n) = \mathbf{p}(0) \cdot \mathbf{P}^n,\ n \in \mathbb{N}.$$

6.2.2 Langzeitverhalten endlicher homogener Markov-Ketten

Es wird nun der Grenzübergang $n \to \infty$ untersucht, genauer die Aussage:

$$\text{Es gibt } \rho^* \in [0,1]^m \text{ mit } \lim_{n\to\infty} p(n) = \rho^* \text{ und } \sum_{i=1}^{m} \rho_i^* = 1. \tag{6.5}$$

Satz 6.1 (Konvergenzsatz für endliche homogene Markov-Ketten)**.** *Sei $\lambda = 1$ der einzige Eigenwert von $\mathbf{P}$ mit Betrag 1 und habe die (algebraische) Vielfachheit 1. Dann gilt (6.5) für jede Ausgangsverteilung $\mathbf{p}(0)$. Die asymptotische Verteilung $\mathbf{p}^*$ erfüllt für alle n die Gleichung*

$$\mathbf{p}^* = \mathbf{p}^* \cdot \mathbf{P}^n. \tag{6.6}$$

Aufgrund von (6.6) bezeichnet man $\mathbf{p}^*$ auch als **stationäre Verteilung** der Markov-Kette. Wird nämlich $\mathbf{p}(0) := \mathbf{p}^*$ als Ausgangsverteilung gewählt, so ist die Markov-Kette stationär. Die stationäre Verteilung erhält man als den auf Zeilensumme 1 normierte Linkseigenvektor von $\mathbf{P}$ zum Eigenwert 1.

Anmerkung 6.2. Ein weiteres Kriterium für die Konvergenz von $\mathbf{P}^n$ ist das folgende: Gibt es $k \in \mathbb{N}$ so dass $\mathbf{P}^k \in (0,1)^{m\times m}$, dann existiert die asymptotische Verteilung $\mathbf{p}^*$. Die Bedingung $\mathbf{P}^k \in (0,1)^{m\times m}$ bedeutet, dass man von jedem Zustand i in jeden Zustand j in k Schritten mit einer positiven Wahrscheinlichkeit kommt.

Lernergebnisse (C3)
Die Studierenden kennen die Definition von endlichen Markov Ketten (MK) und homogenen endlichen MK. Sie kennen den Konvergenzsatz für homogene endliche MK und können in konkreten Beispielen die Übergangsmatrix einer MK bestimmen. Die Studierenden können in konkreten Beispielen die Voraussetzungen des Konvergenzsatzes überprüfen, die stationäre Verteilung bestimmen und die Ergebnisse interpretieren.

6.3 Endliche, homogene Markov-Prozesse

Kerninhalte

- Definition von endlichen, homogenen Markov-Prozessen
- Definition der Intensitätsmatrix
- Wichtige Eigenschaften und Interpretation der Intensitätsmatrix
- Asymptotik

Für viele aktuarielle Anwendungen wird ein stetiger Zeitbereich betrachtet. Endliche Markov-Prozesse sind eine Verallgemeinerung der endlichen Markov-Ketten.

6.3.1 Endliche homogene Markov-Prozesse

Ein stochastischer Prozess $\{X_t\}_{t\geq 0}$ heißt **endlicher Markov-Prozess**, wenn für jede Folge von Zeitpunkten $0 \leq t_0 < t_1 < t_2 < \dots$ durch $\{X_{t_i}\}_{i\in\mathbb{N}_0}$ eine endliche Markov-Kette mit Zustandsraum $S = \{1,\dots,m\}$ gegeben ist. Der Markov-Prozess heißt zudem **homogen**, falls die Übergangswahrscheinlichkeiten $P(X_{t+h} = j | X_t = i)$ für alle $i, j \in S$ nur von h abhängen.

Für einen endlichen homogenen Markov-Prozess existiert somit eine Familie $\{\mathbf{P}(t)\}_{t\geq 0}$ stochastischer $m \times m$-Matrizen (**Übergangsmatrizen**) mit Komponenten

$$p_{ij}(t) := P(X_t = j | X_0 = i),\ t \geq 0.$$

Der Zeilenvektor $\mathbf{p}(t) := (P(X_t = 1), \dots, P(X_t = m))$ ist die Verteilung des Markov-Prozesses zum Zeitpunkt $t \geq 0$. Es gilt

$$\mathbf{p}(t) = \mathbf{p}(0) \cdot \mathbf{P}(t),\ t \geq 0.$$

Die Übergangsmatrizen erfüllen die **Chapman-Kolmogorov-Gleichung**

$$\mathbf{P}(t+s) = \mathbf{P}(t) \cdot \mathbf{P}(s) \text{ für alle } s, t \geq 0. \tag{6.7}$$

Setzt man zusätzlich voraus, dass $[0,\infty) \ni t \mapsto \mathbf{P}(t) \in R^{m\times m}$ stetig ist, dann existiert eine Matrix $\mathbf{Q} \in \mathbb{R}^{m\times m}$ mit

$$\mathbf{P}(t) = e^{\mathbf{Q}t} = \sum_{k=0}^{\infty} \frac{t^k}{k!}\mathbf{Q}^k,\ (t \geq 0) \tag{6.8}$$

wobei man $\mathbf{Q}^0 := \mathbf{E}$ setzt. $\mathbf{Q}$ heißt **Fundamentalmatrix** bzw. **Intensitätsmatrix** von $\mathbf{P}(t)$.

Anmerkung 6.3. Der Zusammenhang der Gleichungen (6.8) und (6.7) ist die Matrixversion des gleichen Zusammenhangs in einer Dimension. Die stetigen Lösungen $f \neq 0$ der Funktionalgleichung $f(x+y) = f(x) \cdot f(y)$ sind genau die Funktionen für die $f(x) = \exp(x\alpha), x \in \mathbb{R}$ gilt.

Die wichtigsten mit der Fundamentalmatrix $\mathbf{Q} = (q_{ij})_{ij}$ verbundenen Eigenschaften sind:

(a) Die Fundamentalmatrix $\mathbf{Q}$ erhält am durch Ableiten von $P(t)$ an der Stelle $t = 0$:

$$\frac{d}{dt}\mathbf{P}(t)\Big|_{t=0} = \frac{d}{dt}\sum_{k=0}^{\infty}\frac{t^k}{k!}\mathbf{Q}^k\Big|_{t=0} = \sum_{k=0}^{\infty}\frac{d}{dt}\frac{t^k}{k!}\Big|_{t=0}\mathbf{Q}^k = \mathbf{Q}. \tag{6.9}$$

Allgemein gilt

$$\frac{d}{dt}\mathbf{P}(t) \quad = \quad \mathbf{Q}\cdot\mathbf{P}(t),\ t > 0.$$

Aufgrund dieser Beziehung bezeichnet man die Elemente q_{ij} der Fundamentalmatrix $\mathbf{Q} = \big(q_{ij}\big)_{ij}$ auch als **Übergangsraten**.

(b) Ist $\mathbf{Q}$ diagonalisierbar, d.h. gibt es eine Darstellung $\mathbf{Q} = \mathbf{A}\mathbf{D}\mathbf{A}^{-1}$ mit einer Diagonalmatrix $\mathbf{D} = diag(\nu_1, \ldots, \nu_m)$ und einer invertierbaren Matrix $\mathbf{A}$, dann gilt

$$\mathbf{P}(t) \quad = \quad \mathbf{A}\cdot diag(\exp(t\nu_1), \ldots, \exp(t\nu_m))\cdot\mathbf{A}^{-1}. \tag{6.10}$$

(c) Aus $\sum_{j=1}^{m} p_{ij}(t) = 1$ für alle $i = 1, \ldots, m$ und (6.9) folgt für alle $i, j = 1, \ldots, m$

$$q_{ii} \le 0,\ q_{ij} \ge 0 \text{ für } i \ne j \text{ und } \sum_{j=1}^{m} q_{ij} = 0.$$

Gilt für eine Matrix $\mathbf{Q} \in \mathbb{R}^{m\times m}$

$$q_{ii} < 0,\ q_{ij} \ge 0 \text{ für } i \ne j \text{ und } \sum_{j=1}^{m} q_{ij} = 0,$$

dann gibt es einen endlichen homogenen Markov-Prozess mit Übergangsmatrizen $\mathbf{P}(t) = \exp(t\mathbf{Q})$. Er kann folgenderweise interpretiert werden:

- Die Verweildauer T_i im Zustand i ist $\mathcal{E}(-q_{ii})$ verteilt.
- Tritt eine Zustandsänderung ein, dann ist $\frac{-q_{ij}}{q_{ii}}$ die Wahrscheinlichkeit des Übergangs von i nach j.
- Man kann einen Zustand i zu einem **absorbierenden Zustand** machen, indem man in der Fundamentalmatrix in der i-ten Zeile eine Nullzeile setzt. Ist nämlich $q_{ij} = 0$ für alle j, so ergibt (6.8) $p_{ii}^{(t)} = 1$ und $p_{ij}^{(t)} = 0$ für $i \ne j$. Der Zustand i wird also, einmal angenommen, mit Wahrscheinlichkeit 1 nicht mehr verlassen.

6.3.2 Langzeitverhalten endlicher homogener Markov-Prozesse

Analog zum Langzeitverhalten endlicher, homogener Markov-Ketten gilt folgendes Ergebnis:

Satz 6.4 (Konvergenzsatz für endliche homogene Markov-Prozesse)**.** *Für die Fundamentalmatrix* $\mathbf{Q}$ *eines endlichen homogenen Markov-Prozesses sei* $\nu = 0$ *der einzige Eigenwert mit Realteil 0 und habe die (algebraische) Vielfachheit 1. Dann existiert die asymptotische Verteilung*

$$\mathbf{p}^* := \lim_{t\to\infty}\mathbf{p}(t)$$

für jede Ausgangsverteilung $\mathbf{p}(0)$ *und genügt für alle* t *der Gleichung*

$$\mathbf{p}^* = \mathbf{p}^*\cdot\mathbf{P}(t). \tag{6.11}$$

In der Situation von Satz 6.4 wird der „konvergente“ Markov-Prozess mitunter auch als **ergodisch** bezeichnet. Wegen (6.11) ist $\{X_t\}_{t\geq 0}$ mit Anfansverteilung $\mathbf{p}^*$ ein stationärer Prozess. Deshalb heßt $\mathbf{p}^*$ auch hier **stationäre Verteilung** des Prozesses. Von besonderer praktischer Bedeutung ist die Folgerung, dass

$$\mathbf{0} = \mathbf{p}^* \cdot \mathbf{Q}, \tag{6.12}$$

gilt, so dass die asymptotische Verteilung $\mathbf{p}^*$ unter den Voraussetzungen des Konvergenzsatzes der auf Zeilensumme 1 normierten Linkseigenvektor der Fundamentalmatrix $\mathbf{Q}$ zum Eigenwert 0 ist.

In Anwendungsbeispielen aus der Versicherung können mit Hilfe der stationären Verteilungen differenzierte Prämien bestimmt werden.

Lernergebnisse (C3)
Die Studierenden kennen die Definition von endlichen, homogenen Markov-Prozessen (MP) und unterscheiden sie von Markov-Ketten. Sie kennen die Bedeutung der Intensitätsmatrix, können sie bestimmen und interpretieren. Die Studierenden können in konkreten Beispielen die Voraussetzungen des Konvergenzsatzes überprüfen, die stationäre Verteilung bestimmen und die Ergebnisse interpretieren.

6.4 Ausgewählte Markov-Prozesse

Kerninhalte

- Definition des homogenen Markov-Prozesses
- Definition und Eigenschaften
  - des Poisson-Prozesses
  - der Brownschen Bewegung
  - der geometrischen Brownschen Bewegung

In aktuariellen Fragestellungen treten regelmäßig Prozesse auf, deren mögliche Zustände aus einer abzählbaren Menge stammen oder sogar ein Kontinuum innerhalb der reellen Zahlen annehmen können. Aus diesem Grund wird das Konzept der Markov-Prozesse auf reellwertige Prozesse ausgedehnt.

6.4.1 Homogene Markov-Prozesse

Markov-Prozesse in den reellen Zahlen können wieder anhand der Eigenschaft der Gedächtnislosigkeit definiert werden.

Ein reellwertiger stochastischer Prozess $\{X_t\}_{t\geq 0}$ mit Startwert $X_0 = x_0$ heißt **Markov-Prozess**, wenn für jede Folge von Zeitpunkten $0 \leq t_1 < t_2 < \ldots$, jede Folge $x_{t_i} \in \mathbb{R}$ und jedes $B \subseteq \mathbb{R}$

$$P(X_{t_n} \in B | X_{t_1} = x_{t_1}, \ldots, X_{t_{n-1}} = x_{t_{n-1}}) = P(X_{t_n} \in B | X_{t_{n-1}} = x_{t_{n-1}}) \tag{6.13}$$

gilt. Er heißt **homogener Markov-Prozess**, wenn die die Übergangswahrscheinlichkeiten auf der rechten Seite von (6.13) außer von $x_{t_{n-1}}$ und B nur von $t_n - t_{n-1}$ abhängen.

Die Existenz von homogenen Markov-Prozessen ist gesichert, auch die **Chapman-Kolmogorov-Gleichung** kann formuliert werden. Beides geht aber über den Umfang des Faches hinaus.

Im Folgenden werden die charakterisierenden Eigenschaften ausgewählter stochastischer Prozesse jeweils explizit angegeben. Hierbei spielen die Begriffe stationäre bzw. unabhängige Zuwächse eine entscheidende Rolle:

Ein stochastischer Prozess $\{X_t\}_{t\geq 0}$ besitzt

- **stationäre Zuwächse**, wenn gilt:

$$X_t - X_s \overset{\mathrm{d}}{=} X_{t+h} - X_{s+h} \text{ für alle } 0 \leq s \leq t \text{ und alle } h > 0$$

- **unabhängige Zuwächse**, wenn für alle t_i, $i = 0, \ldots, n$, $n \in \mathbb{N}$ mit $0 = t_0 < t_1 < \cdots < t_n$ die Zuwächse $X_{t_i} - X_{t_{i-1}}$, $i = 1, \ldots n$ unabhängig sind.

6.4.2 Der Poisson-Prozess

Ein **Poisson-Prozesses** mit Intensität $\lambda > 0$ ist ein stochastischer Prozess $\{N_t\}_{t \geq 0}$ mit Zustandsraum $\mathbb{N}_0$, der durch folgende Eigenschaften charakterisiert ist:

- $N_0 = 0$ fast sicher,
- $\{N_t\}_{t \geq 0}$ besitzt stationäre und unahbhängige Zuwächse,
- N_t ist Poisson $\mathcal{P}(\lambda t)$-verteilt.

Ist N_t die Anzahl der im Zeitintervall $[0, t]$ aufgetretenen Ereignisse (z.B. Schadenereignisse) und sind die Zeitdauern zwischen Ereignissen unabhängig und $\mathcal{E}(\lambda)$-verteilt, dann ist $\{N_t\}$ ein Poisson-Prozess mit Intensität $\lambda > 0$. Weitere Eigenschaften von Poisson-Prozessen sind:

Überlagerungseigenschaft Sind $\{N_t\}_{t \geq 0}, \{M_t\}_{t \geq 0}$ voneinander unabhängige Poisson-Prozesse mit Intensitäten λ, μ, dann ist $\{N_t + M_t\}_{t \geq 0}$ ein Poisson-Prozess mit Intensität $\lambda + \mu$.

Gleichverteilungseigenschaft Ist $N = N_b - N_a$ die Anzahl der Ereignisse im Intervall $(a, b]$, dann sind die Zeitpunkte, in denen die Ereignisse eintreten, gleichverteilt in $(a, b]$.

Für eine Folge $\{X_n\}_{n \in \mathbb{N}}$ unabhängiger, identisch verteilter Zufallsvariablen, definiert man den zusammengesetzten **Poisson-Prozess**

$$S_t := \sum_{i=1}^{N_t} X_i.$$

Dies ist Grundlage des kollektiven Modells der Versicherung, wobei N_t die Anzahl der Schäden im Zeitintervall $[0, t]$, $X_i \geq 0$ die Schadenhöhen und S_t der kumulierte Schaden im Zeitintervall $[0, t]$.

6.4.3 Die Brownsche Bewegung

Ein **Wiener-Prozess** (auch **Standard-Brownsche Bewegung**) ist ein stochastischer Prozess $\{W_t\}_{t \geq 0}$ mit Zustandsraum $\mathbb{R}$, der durch folgende Eigenschaften charakterisiert ist:

- $W_0 = 0$ fast sicher,
- $\{W_t\}_{t \geq 0}$ besitzt stationäre und unahbhängige Zuwächse,
- W_t ist normalverteilt $\mathcal{N}(0, t)$.

Es ergibt sich:

- $W_t - W_s$ ist für alle $0 \leq s \leq t$ normalverteilt $\mathcal{N}(0, t - s)$,
- $\mathrm{Cov}(W_t, W_s) = \min(s, t)$ für alle $s, t \geq 0$,
- Für alle $0 \leq t_1 < \cdots < t_n$ ist $(W_{t_1}, \ldots, W_{t_n})$ multivariat normalverteilt.

Setzt man für $B_0, \mu \in \mathbb{R}$, $\sigma > 0$

$$B_t := B_0 + \mu t + \sigma W_t,\ t \geq 0,$$

so erhält man die **Brownsche Bewegung B_t mit Drift μ und Volatilität σ**. Für sie gilt $B_t \sim \mathcal{N}(\mu t, \sigma^2 t)$. Die unabhängigen Zuwächse sind für $t > s \geq 0$ normalverteilt, also $B_t - B_s \sim \mathcal{N}(\mu(t-s), \sigma^2(t-s))$.

Aus der Brownschen Bewegung mit Drift kann die **geometrische Brownsche Bewegung** $S_t := S_0 \cdot \exp(B_t)$ abgeleitet werden, welche das klassische Modell für Kursentwicklungen an Aktienmärkten darstellt. Hier ist also S_t lognormal-verteilt.

Lernergebnisse (B2-B4)
Die Studierenden kennen die Definition des homogenen Markov-Prozesses, kennen die wichtigsten Eigenschaften des Poisson-Prozesses und können sie zur Lösung konkreter Fragestellungen einsetzen. Sie kennen die charakterisierenden Eigenschaften des Wiener Prozesses sowie der Brownschen Bewegungen.

6.5 Grundlagen der stochastischen Differenzialrechnung und Itō-Kalkül

Kerninhalte

- Definition der stochastische Integration
- Definition und Beispiele für Itō-Prozesse
- Itō-Formel
- Grundlegende Rechenregeln der stochastischen Differenzialrechnung

Die **stochastische Differenzialrechnung** führt die Inkremente dX_t eines stochastischen Prozesses X_t während der infinitesimalen Zeitdauer dt auf eine zeitliche Drift D_t und eine auf den Inkrementen dW_t eines Wienerprozesses basierenden Volatilitätskomponente V_t zurück. Intuitiv drückt man dies in der Schreibweise

$$dX_t = D_t dt + V_t dW_t \tag{6.14}$$

aus. Dabei wird vorausgesetzt, dass die Zufallsvariablen D_t und V_t gewissen Regularitätsbedingungen genügen.

Naheliegend wäre es zunächst, (6.14) pfadweise in dem Sinn zu lesen, dass für festes $\omega \in \Omega$ die Ableitung des Pfades $X_t(\omega)$ nach der Zeit durch

$$dX_t(\omega)/dt = D_t(\omega) + V_t(\omega) dW_t(\omega)/dt$$

gegeben ist. Diese Lesart ist jedoch insofern nicht zulässig, als die Pfade eines Wiener-Prozesses (mit Wahrscheinlichkeit 1) nirgends differenzierbar sind, also $dW_t(\omega)/dt$ nicht sinnvoll existiert. Stattdessen begreift man (6.14) als differenzielle Notation einer auf stochastischen Integralen beruhenden Darstellung

$$X_t = X_0 + \int_0^t D_s ds + \int_0^t V_s dW_s. \tag{6.15}$$

Die Integration erfolgt dabei bezüglich ds bzw. dW_s im folgenden Sinn (unter gewissen Regularitätsbedingungen an die Integranden):

- Integrale $\int_0^t Y_s ds$ werden durch pfadweise Integration im Lebesgue-Stieltjes-Sinn definiert:
$$\int_0^t Y_s ds(\omega) := \lim_{n \to 0} \sum_{i=1}^{n} Y_{s_i}(\omega)(s_i - s_{i-1}),$$
wobei immer feiner werdende Partitionen $0 = s_0 < s_1 < \ldots < s_n = t$ von $[0, t]$ gewählt werden.

- Integrale $\int_0^t Y_s dW_s$ werden konstruiert, indem man den zu integrierenden Prozess Y_s zunächst im L^2-Sinn durch stückweise konstante Prozesse $Y_s^* := \sum_{i=1}^n Y_{s_i}^* \mathbf{1}_{(s_{i-1},s_i]}(s)$ approximiert und den L^2-Grenzwert von

$$\int Y_s^* dW_s := \sum_{i=1}^n Y_{s_i}^* (W_{s_i} - W_{s_{i-1}})$$

für $n \to \infty$ bildet.

Stochastische Prozesse X_t, die „differenzierbar“ in dem Sinn sind, dass sie eine Darstellung gemäß (6.14) bzw. (6.15) besitzen, werden als **Itō-Prozesse** (nach Kiyoshi Itō, 1915-2008) bezeichnet. Hierunter fällt unter anderem die Brownsche Bewegung mit Drift, welche die differenzielle Darstellung

$$dX_t = \mu dt + \sigma dW_t$$

besitzt, bzw. gemäß (6.15) die Integraldarstellung

$$X_t = X_0 + \int_0^t \mu ds + \int_0^t \sigma dW_s = X_0 + \mu t + \sigma(W_t - W_0) = X_0 + \mu t + \sigma W_t. \tag{6.16}$$

Der Itō-Kalkül etabliert die Differenziationsregeln für Itō-Prozesse. Wichtigster Grundstein ist dabei die **Itō-Formel**. Diese besagt, dass die Transformation eines Itō-Prozesses (6.14) mittels $Y_t := f(t, X_t)$ wieder auf einen Itō-Prozess bezüglich desselben Wienerprozesses führt. Dieser hat die Darstellung

$$dY_t = \left(\frac{\partial f}{\partial t} + \frac{\partial f}{\partial X_t} D_t + \frac{1}{2} \frac{\partial^2 f}{\partial X_t^2} V_t^2 \right) dt + \left(\frac{\partial f}{\partial X_t} V_t \right) dW_t. \tag{6.17}$$

Man beachte dabei, dass (6.17) im Vergleich zur Kettenregel der deterministischen Differenzialrechnung einen zusätzlichen Term $(1/2)(\partial^2 f/\partial X_t^2)V_t^2$ enthält.

Die Itō-Formel stellt ein wichtiges Instrument zur Lösung stochastischer Differenzialgleichungen dar (vgl. Abschnitt 6.6). Für die Itō-Formel (6.17) gibt es zudem eine mehrdimensionale Fassung, mit deren Hilfe sich weitere Rechenregeln der stochastischen Differenzialrechnung ergeben, unter anderem die **Produktregel**. Diese besagt, dass das Produkt zweier Itō-Prozesse $dX_t^{(i)} = D_t^{(i)} dt + V_t^{(i)} dW_t \, (i = 1, 2)$ wieder Itō-Prozess ist mit

$$d(X_t^{(1)} X_t^{(2)}) = X_t^{(1)} dX_t^{(2)} + X_t^{(2)} dX_t^{(1)} + V_t^{(1)} V_t^{(2)} dt \tag{6.18}$$

Auch hier ergibt sich im Vergleich zur deterministischen Differenzialrechnung ein zusätzlicher, von der Volatilität abhängiger Term.

Lernergebnisse (B2-B3)
Die Studierenden kennen die Definition von Itō-Prozessen. Sie können deren differenzielle wie auch die Integralschreibweise interpretieren und verschiedene Beispiele für einfache Itō-Prozesse erläutern. Die Studierenden kennen zudem die Itō-Formel und können sie hinsichtlich ihrer Aussage interpretieren und anwenden. Als Beispiel einer weiterführenden, aus der Itō-Formel folgenden Differenziationsregel ist den Studierenden die Produktregel bekannt.

6.6 Stochastische Differenzialgleichungen

Kerninhalte

- Stochastische Differenzialgleichungen
- Existenz- und Eindeutigkeitssatz, starke Lösung
- Lösungsansätze für stochastische Differenzialgleichungen (Transformation, Variation der Konstanten, Simulation)

Stochastische Differenzialgleichungen ergeben sich aus (6.14), wenn Drift und Volatilität Funktionen von t und X_t sind:

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, \tag{6.19}$$

bzw. in Integralform notiert:

$$X_t = X_0 + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s. \tag{6.20}$$

Dabei kommt eine Anfangsbedingung $X_0 = x$ hinzu.

Der **Existenz- und Eindeutigkeitssatz** von Picard-Lindenlöf für die Lösungen deterministischer Differenzialgleichungen hat eine Entsprechung für stochastische Differenzialgleichungen dahingehend, dass (6.20) unter Regularitätsbedingungen an $b(t, x)$ und $\sigma(t, x)$ eine (fast sicher) eindeutige Lösung in Form eines stetigen Prozesses X_t mit endlicher Varianz $Var(X_t)$ besitzt. X_t bezeichnet man dann als **starke Lösung** der stochastischen Differenzialgleichung (6.19).

Für die Lösung von stochastischen Differenzialgleichungen gibt es keine allgemeingültige Lösungsstrategie und in der Regel auch keine „geschlossene Form“ (d.h. eine Darstellung wie in (6.15)). Anwendung finden unter anderem folgende Ansätze:

(a) **Transformationen**: Hierbei überführt man die ursprüngliche stochastische Differenzialgleichung mit einer Funktion f und der Itō-Formel (6.17) in eine stochastische Differenzialgleichung für die Zufallsvariable $f(X_t)$, welche eine bekannte Lösung besitzt.

So erhält man z.B. aus der Differenzialgleichung $dX_t = \mu X_t dt + \sigma X_t dW_t$ mit der log-Transformation $Y_t := \ln X_t$ die geometrische Brownsche Bewegung mit Drift $X_t = X_0 \cdot \exp((\mu - 1/2 \cdot \sigma^2)t + \sigma W_t)$.

(b) **Variation der Konstanten**: Anstelle der linearen inhomogenen stochastischen Differenzialgleichung

$$dX_t = (b_1(t)X_t + b_2(t))dt + (\sigma_1(t)X_t + \sigma_2(t))dW_t$$

betrachtet man die homogene stochastische Differenzialgleichung

$$dZ_t = b_1(t)Z_t dt + \sigma_1(t)Z_t dW_t,$$

deren Lösung Z_t durch log-Transformation ermittelt werden kann und bis auf eine multiplikative Konstante eindeutig ist. Für X_t macht man dann den Ansatz $X_t := Y_t \cdot Z_t$, bei dem die multiplikative Konstante durch einen stochastischen Prozess Y_t „variiert“ wird. Mit Hilfe der Produktregel (6.18) ermittelt man für Y_t die differenziellen Darstellung

$$dY_t = \frac{b_2(t) - \sigma_1(t)\sigma_2(t)}{Z_t}dt + \frac{\sigma_2(t)}{Z_t}dW_t.$$

Mit dieser Methode ergibt sich z.B. aus der stochastischen Differenzialgleichung $dX_t = -\alpha X_t dt + \sigma dW_t$ der Ornstein-Uhlenbeck-Prozess

$$X_t = X_0 \cdot \exp(-\alpha t) + \sigma \int_0^t \exp(\alpha(s - t))dW_s.$$

(c) **Simulation**: Hierbei werden die Inkremente des Prozesses dX_t über diskreten Zeitschritten simuliert (vgl. Abschnitt 8.4).

Lernergebnisse (C3)
Die Studierenden kennen die allgemeine Form stochastischer Differenzialgleichungen und können diese interpretieren. Ihnen ist der Existenzsatz für starke Lösungen stochastischer Differenzialgleichungen bekannt. Die Studierenden sind zudem in der Lage, die stochastischen Differenzialgleichungen aus obigen und weiteren Beispielen mittels grundlegender Lösungsansätze zu lösen.

7 Credibility

Credibility-Modelle werden in der aktuariellen Praxis bei der Beurteilung sehr individueller Risiken mit zum Teil nicht beobachtbaren Risikomerkmalen verwendet z.B. in der Gruppen-Lebensversicherung, bei der Versicherung von Autoflotten, usw. Ziel ist es bei der Beurteilung sowohl die individuelle als auch die kollektive Erfahrung zu berücksichtigen.

Gegeben seien m Risiken mit den Zielvariablen $X_1, \dots, X_m$. Die Risikomerkmale werden als zufällig angenommen und durch einen **Strukturparameter** Θ modelliert. Beobachtet werden n Realisierungen $x_{i1}, \dots, x_{in}$ des i-ten Risikos, $i = 1, \dots, m$. Mit Hilfe der Credibility-Theorie wird die individuelle Information $x_{i1}, \dots, x_{in}$ mit der kollektiven Information x_{ij}, $i = 1, \dots, m$, $j = 1, \dots, n$ mit sinnvoller Gewichtung kombiniert um zu einer stabilen und dennoch individuellen Beurteilung zu gelangen. Meist geht es darum Erwartungswerte zu schätzen.

7.1 Das Bayes'sche Modell

Kerninhalte

- Strukturverteilung, a-priori und a-posteriori Verteilung
- konjugierte Verteilungen
- Exakte und linearisierte Credibility-Prämie und Beispiele

7.1.1 A-priori und a-posteriori Verteilung

Für ein Einzelrisiko X realisieren sich die Risiken in zwei Stufen:

(a) Realisierung θ einer Zufallsvariablen Θ mit Dichte f_Θ, dem Strukturparameter,

(b) Daraus ergeben sich Schäden $X_1, \dots, X_n$, die iid verteilt sind, gegeben $\Theta = \theta$ mit Dichte $f_{X|\Theta=\theta}$, bedingtem Erwartungswert $\mu(\Theta) := E(X|\Theta)$ und bedingter Varianz $\mathrm{Var}(X|\Theta) = \sigma^2(\Theta)$. Im Folgenden wird mit $\mathbf{X}$ der Zufallsvektor $(X_1, \dots, X_n)^\top$ bezeichnet.

Die Verteilung des Strukturparameters Θ heißt **a-priori Verteilung**. Sie wird aufgrund der Beobachtung $\mathbf{x} := (x_1, \dots, x_n)$ verbessert zur **a-posteriori Verteilung**. Mit Hilfe von **konjugierten Verteilungsfamilien** kann man aus der a-priori Dichte f_Θ die aposteriori Dichte $f_{\Theta|\mathbf{X}=\mathbf{x}}$ explizit bestimmen, einige Fälle sind in der Tabelle 7.1 aufgeführt.

Verteilung $P_{X\|\Theta=\theta}$	a-priori-Verteilung von Θ	a-posteriori-Verteilung $P_{\Theta\|\mathbf{X}=\mathbf{x}}$
$B(m, \theta)$	$\mathcal{B}(a, b)$	$\mathcal{B}(n\overline{x} + a, nm - n\overline{x} + b)$
$NB(\beta, \theta)$	$\mathcal{B}(a, b)$	$\mathcal{B}(n\beta + a, n\overline{x} + b)$
$\mathcal{P}(\theta)$	$\Gamma(\alpha, \lambda)$	$\Gamma(\alpha + n\overline{x}, \lambda + n)$
$\mathcal{E}(\theta)$	$\Gamma(\alpha, \lambda)$	$\Gamma(\alpha + n, \lambda + n\overline{x})$

Tabelle 7.1: Ausgewählte a-priori und a-posteriori Verteilungen mit $\overline{x} := \frac{1}{n}\sum_{i=1}^n x_i$.

7.1.2 Die exakte Credibility Prämie

Da der Strukturparameter θ unbekannt ist, kann die individuelle Prämie $E(X|\Theta=\theta)=\mu(\theta)$ nicht bestimmt werden, man kann sie jedoch mit Hilfe der Beobachtung $\mathbf{X}=\mathbf{x}$ bezüglich der quadratischen Abweichung approximieren. Die **exakte Credibility-Prämie** H^* ist gegeben durch $H^* := E(\mu(\Theta)|\mathbf{X})$ und erfüllt

$$E\left[\left(H^*-\mu(\Theta)\right)^2\right]=\min\{E\left[\left(h(\mathbf{X})-\mu(\Theta)\right)^2\right]\,|\,h:\mathbb{R}^n\longrightarrow\mathbb{R}\text{ messbar }\}. \tag{7.1}$$

Im Falle der konjugierten Verteilungsfamilien in Tabelle 7.1 ergeben sich die exakten Credibility-Prämien in Tabelle 7.2. Man verwendet die iterierte Erwartung um die kollektive Prämie $E(X)=E(E(X|\Theta))$ zu betimmen.

$P_{X\|\Theta=\theta}$	$\mu(\theta)$	f_Θ	$E(X)$	Credibility-Prämie H^*
$B(m,\theta)$	$m\theta$	$\mathcal{B}(a,b)$	$\frac{am}{a+b}$	$\frac{nm}{a+b+nm}\cdot\overline{X}+\frac{a+b}{a+b+nm}\cdot E(X)$
$NB(\beta,\theta)$	$\beta(1-\theta)/\theta$	$\mathcal{B}(a,b)$	$\frac{\beta b}{a-1}$	$\frac{n\beta}{a+n\beta-1}\cdot\overline{X}+\frac{a-1}{a+n\beta-1}\cdot E(X)$
$\mathcal{P}(\theta)$	θ	$\Gamma(\alpha,\lambda)$	$\frac{\alpha}{\lambda}$	$\frac{n}{\lambda+n}\cdot\overline{X}+\frac{\lambda}{\lambda+n}\cdot E(X)$
$\mathcal{E}(\theta)$	$1/\theta$	$\Gamma(\alpha,\lambda)$	$\frac{\lambda}{\alpha-1}$	$\frac{n}{\alpha+n-1}\cdot\overline{X}+\frac{\alpha-1}{\alpha+n-1}\cdot E(X)$

Tabelle 7.2: Exakte Credibility-Prämien zu Tabelle 7.1

7.1.3 Linearisierte Credibility

In der Tabelle 7.2 ergibt sich die Credibility-Prämie als gewichtetes Mittel von $E(X)$ und $\overline{X}=\frac{1}{n}\sum_{i=1}^n X_i$, d.h. als gewichtetes Mittel der „kollektiven Prämie“ $E(X)$ und individueller Prämie in der Form

$$H^*=z_n\overline{X}+(1-z_n)E(X). \tag{7.2}$$

Die ist im Allgemeinen jedoch nicht der Fall. Die **linearisierte Credibility-Prämie** H^{**} ergibt sich in Analogie zu (7.1) indem die Approximation von $\mu(\theta)$ über affin lineare Funktionen gebildet wird:

$$E\left[\left(H^{**}-\mu(\Theta)\right)^2\right]=\min\{E\left[\left(h(\mathbf{X})-\mu(\Theta)\right)^2\right]\,|\,h:\mathbb{R}^n\longrightarrow\mathbb{R},h(\mathbf{x})=z\overline{x}+(1-z)E(X)\}.$$

Es ergibt sich

$$H^{**}=z_n\overline{X}+(1-z_n)E(X)\text{ mit} \tag{7.3}$$

$$z_n=\frac{Var(\mu(\Theta))}{\frac{1}{n}E(\sigma^2(\Theta))+Var(\mu(\Theta))}=\frac{n}{n+\frac{E(\sigma^2(\Theta))}{Var(\mu(\Theta))}} \tag{7.4}$$

Das Gewicht z_n heißt **Credibility-Faktor**. Wegen $\lim_{n\to\infty}z_n=1$ nimmt für große Stichprobenumfänge die Glaubwürdigkeit der individuellen Erfahrung $\overline{X}$ zu.

Lernergebnisse (C3)
Die Studierenden interpretieren das Bayes´sche Modell als zweistufiges Zufallsexperiment und können in ausgewählten Beispielen, unter anderem im Fall von konjugierten Verteilungen, die exakte Credibility-Prämie bestimmen. Sie verstehen wie sich der Ansatz der linearisierten Credibility-Prämien ergibt und können diese ermitteln. Die Studierenden können den Credibility-Faktor interpretieren.

7.2 Das Bühlmann-Straub-Modell

Kerninhalte

- Voraussetzungen des Bühlmann-Straub-Modells (BS)
- Credibility-Faktoren im BS-Modell
- Schätzer im BS-Modell

Das Bühlmann-Straub-Modell überträgt die in Abschnitt 7.1.3 dargestellten Bayes'schen Modelle für ein Einzelrisiko auf ein Modell für Kollektive.

Wie vorher realisieren sich die Schäden für jedes Einzelrisiko $i = 1, 2, \ldots$ in zwei Stufen:

(a) Für jedes Einzelrisiko $i = 1, 2, \ldots$ ist der Strukturparameter θ_i die Realisation einer Zufallsvariablen Θ_i.

(b) Für jedes Einzelrisiko i ergeben sich die Schäden $X_{ij}, j = 1, \ldots, n$ (z.B. die Schadenquote über n Jahre) als unabhängige Realisationen einer Zufallsvariablen X_i mit

$$E(X_{ij}|\Theta_i = \theta_i) = \mu(\theta_i) \text{ und } \mathrm{Var}(X_{ij}|\Theta_i = \theta_i) = \frac{\sigma^2(\theta_i)}{w_{ij}}$$

wobei $w_{ij} > 0$ ein Gewicht (z.B. Prämienvolumen) ist.

(c) Es gilt $\Theta_i \overset{\text{iid}}{\sim} \Theta$.

(d) Bezeichnet $\mathbf{X}_i := (X_{i1}, \ldots, X_{in})$ den Zufallsvektor der beobachteten Schäden, sind die Paare $(\mathbf{X}_1, \Theta_1)$, $(\mathbf{X}_2, \Theta_2)$, ... unabhängig.

Aufgrund dieser Annahmen hängen die Erwartungswerte von X_{ij} und $\sigma(\Theta_i)$ sowie die Varianzen von $\mu(\Theta_i)$ nicht von den Indizes ab. Es werden die Bezeichnungen

$$\mu := E(\mu(\Theta_i)) \qquad \sigma^2 := E(\sigma^2(\Theta_i)) \qquad \tau^2 = \mathrm{Var}(\mu(\Theta_i))$$

vewendet. Man kann μ als globales Mittel, σ^2 als Mittel der individuellen Varianzen und τ^2 als Varianz der individuellen Erwartungswerte interpretieren.

Für die Beobachtungen von Risiko i betrachten wir nun das gewichtete Mittel:

$$\overline{X}_i =: \frac{\sum_{j=1}^n w_{ij} X_{ij}}{\sum_{j=1}^n w_{ij}},\ i = 1, 2, \ldots$$

Analog zu Abschnitt 7.1.3 ist die **Credibility-Prämie nach Bühlmann-Straub** H_i^{**} für das Risiko i definiert als die Lösung von

$$E\left[\left(H_i^{**} - \mu(\Theta_i)\right)^2\right] = \min\left\{E\left[(h(\mathbf{X}_i) - \mu(\Theta_i))^2\right] \middle| h : \mathbb{R}^n \longrightarrow \mathbb{R}, h(\mathbf{X}_i) = z_i\overline{X}_i + (1 - z_i)\mu\right\}.$$

Es ergibt sich

$$H^{**} = z_i\overline{X}_i + (1 - z_i)\mu \text{ mit} \tag{7.5}$$

$$z_i = \frac{w_{i\bullet}}{w_{i\bullet} + \frac{\sigma^2}{\tau^2}} \text{ mit } w_{i\bullet} := \sum_{j=1}^n w_{ij}. \tag{7.6}$$

Das **Bühlmann-Modell**, das ein Vorläufer des Bühlmann-Straub Modells ist, ergibt sich aus $w_{ij} = 1$ für alle i, j. Aus (7.6) erhält man im Bühlmann-Modell von i unabhängige Credibility-Faktoren

$$z_i = \frac{n}{n + \frac{\sigma^2}{\tau^2}}.$$

Aus (7.5) und (7.6) erkennt man:

(a) Je größer der Stichprobenumfang n bzw. die Summe $w_{i\bullet}$ der Gewichte ist, desto größer ist das Gewicht, mit dem $\overline{X}_i$ in die linearisierte Credibility-Prämie eingehen kann.

(b) Je größer die Varianz $\mathrm{Var}(\mu(\Theta))$ zwischen den Risiken ist, desto stärker unterscheiden sich die Risiken im Kollektiv. Diese Unterschiede begründen die Notwendigkeit, in der Prämienermittlung die individuelle Schadenerfahrung $\overline{X}_i$ höher zu gewichten.

(c) Je größer die Varianz $\sigma^2(\theta)$ des Einzelrisikos bei gegebenem Strukturparameter $\Theta = \theta$ ist, desto höher ist die Unsicherheit, mit der $\overline{X}_i$ behaftet ist, und desto geringer ist das zugehörige Gewicht bei der Ermittlung der linearisierten Credibility-Prämie.

In der praktischen Anwendung müssen μ, σ^2 und τ^2 aus den Daten von m Risiken geschätzt werden. Dafür gibt es in der Literatur mehrere Ansätze. Eine Möglichkeit ist der sogenannte Bichsel-Straub-Schätzer. Mit

$$\overline{X}_i := \frac{\sum_{j=1}^{n} w_{ij} X_{ij}}{\sum_{j=1}^{n} w_{ij}} \text{ und } \hat{\sigma}_i^2 := \frac{1}{n-1} \sum_{j=1}^{n} w_{ij} (X_{ij} - \overline{X}_i)^2$$

erhält man

$$\hat{\sigma}^2 := \frac{1}{m} \sum_{i=1}^{m} \hat{\sigma}_i^2.$$

Die beiden anderen Parameter schätzt man mit Hilfe der iterativen Lösung des Gleichungssystems für $z_1, \ldots, z_n$ bestehend aus (7.6) und

$$\hat{\mu} := \frac{\sum_{i=1}^{m} z_i \cdot \overline{X}_i}{\sum_{i=1}^{m} z_i}$$

$$\hat{\tau}^2 := \frac{1}{m-1} \sum_{i=1}^{m} z_i \cdot (\overline{X}_i - \hat{\mu})^2.$$

Im Fall des Bühlmann-Modells vereinfachen sich die Schätzer, und man kann den Credibility-Faktor durch

$$\hat{Z} = 1 - \frac{\hat{\sigma}^2}{n \hat{\nu}^2}$$

schätzen, wobei

$$\hat{\mu} = \frac{1}{m} \sum_{i=1}^{m} \overline{X}_i \text{ und } \hat{\nu}^2 = \frac{1}{m-1} \sum_{i=1}^{m} (\overline{X}_i - \hat{\mu})^2$$

Schätzer von $E(\overline{X})$ und $\mathrm{Var}(\overline{X}) = n\tau^2 + \sigma^2$ sind.

Lernergebnisse (B2)
Die Studierenden kennen den Aufbau des BS-Modells. Sie können dessen Modellannahmen und die sich ergebenden Credibility-Faktoren und Credibility-Prämien interpretieren. Außerdem verstehen sie den Aufbau der Schätzer der Parameter im Bühlmann- und BS-Modell und können diese zur Bestimmung der Credibility-Prämie anwenden.

8 Monte-Carlo-Simulation

8.1 Prinzip und Grundlagen der Methode

Monte-Carlo Simulationen werden vielfältig eingesetzt, z.B. für die quantitative Analyse komplexer Modelle für Aktiv- und Passivseite von Versicherungsbilanzen. Aufgrund der hohen Komplexität ist keine analytische Behandlung möglich. Deshalb werden eine Vielzahl von Szenarien, die auf den entwickelten Modellen beruhen, erzeugt. Dies geschieht indem die gegebenenfalls abhängigen Einzelrisiken simuliert und dann zusammengeführt werden. Daraus werden Erwartungswerte, Varianzen, Gesamtschadenverteilungen, Value at Risk, Kapitalallokationen usw. geschätzt.

Auch Resamplingmethoden wie Bootstrap und Kreuzvalidierung basieren auf der Simulation des Zufallsexperiments „Ziehen mit bzw. ohne Zurücklegen“. Bei allen Simulationsverfahren sind auf $[0,1]$ gleichverteilte Zufallszahlen der Ausgangspunkt. Diese werden geeignet transformiert um die Realisationen der Modelle zu gewinnen.

Die MC Methode basiert auf dem starken Gesetz der großen Zahlen, wonach das arithmetische Mittel $S_n := \frac{1}{n}(X_1 + \cdots + X_n)$ von Zufallsvariablen $X_1, X_2, \ldots \overset{iid}{\sim} X$ fast sicher gegen $E(X)$ konvergiert. Für eine Folge (x_k) von Realisierungen der X_k ist also $\frac{1}{n}(x_1 + \cdots + x_n)$ eine Näherung von $E(X)$. Da

$$\mathrm{Var}(S_n) = \frac{1}{\sqrt{n}} \cdot \mathrm{Var}(X)$$

ist die Güte dieser Näherung proportioinal zu $\frac{1}{\sqrt{n}}$.

Lernergebnisse (B2)
Die Studierenden kennen die Gründe, die den Einsatz von Monte-Carlo Methoden (MC) notwendig machen und die Einsatzgebiete in der Versicherungs- und Finanzwirtschaft. Sie wissen, dass die MC Methode auf dem Gesetz der großen Zahlen basiert und kennen die Größenordnung des Standardfehlers einer MC Simulation.

8.2 Inversionsmethode

Kerninhalte

- Definition der Pseudoinversen
- Funktionsweise der Inversionsmethode
- Verteilungen für die die Inversionsmethode direkt anwendbar ist

Für die Methode ist die Pseudoinverse (auch verallgemeinerte Inverse) entscheidend.

Definition 8.1 (Pseudoinverse)**.** *Ist $F : \mathbb{R} \longrightarrow [0,1]$ die Verteilungsfunktion einer Zufallsvariablen $X : \Omega \to \mathbb{R}$, dann heißt $F^{\leftarrow} : (0,1) \longrightarrow \mathbb{R}$,*

$$F^{\leftarrow}(u) := \inf\{x \in \mathbb{R} : F(x) \geq u\}, \quad 0 < u < 1$$

*die **Pseudoinverse** von F.*

Ist F stetig und streng monoton auf dem Träger von X, dann stimmen $F^{\leftarrow}$ und F^{-1} überein. Die Inversionsmethode beruht auf folgendem Satz:

Satz 8.2 (Inversionsmethode)**.** *Sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable mit Verteilungsfunktion $F : \mathbb{R} \longrightarrow [0,1]$, Pseudoinverser $F^{\leftarrow} : (0,1) \longrightarrow \mathbb{R}$ und $U \sim \mathcal{U}[0,1]$. Sei $Y := F^{\leftarrow}(U)$. Dann gilt $Y \sim F$, d.h. die Zufallsvariable Y ist wie X verteilt.*

Eine Realisierung x der Zufallsvariablen X kann mit der Inversionsmethode wie folgt simuliert werden:

1. Schritt Erzeuge eine Zufallszahl u aus dem Intervall $(0, 1)$.

2. Schritt Setze $x := F^{\leftarrow}(u)$.

Dieser Algorithmus kann sowohl für stetige als auch für diskrete Zufallsvariablen verwendet werden. Kann man die Umkehrfunktion der Verteilungsfunktion analytisch berechnen, dann ist die Simulation einer Realisation unproblematisch. In der folgenden Tabelle ist $U \sim \mathcal{U}[0, 1]$.

Verteilung	Generator
Exponentialverteilung $\mathcal{E}(\lambda)$	$X = -\dfrac{\ln U}{\lambda}$
Weibull $\mathcal{W}(\alpha)$	$X = (-\ln U)^{1/\alpha}$
Fréchet $\mathcal{F}(\alpha)$	$X = (-\ln U)^{-1/\alpha}$
Pareto $\mathcal{P}a(\alpha)$	$X = U^{-1/\alpha} - 1$
Logistische Verteilung $\mathcal{L}$	$X = \ln\left(\dfrac{U}{1-U}\right)$

Über Faltung und Transformationen erhält man weitere Simulationsmöglichkeiten, Beispiele finden sich in der folgenden Tabelle. Hierzu ist $U \sim \mathcal{U}[0, 1]$ bzw. $U_k \overset{\text{iid}}{\sim} \mathcal{U}[0, 1]$.

Verteilung	Generator	Prinzip
Gleichverteilung $\mathcal{U}[a, b]$	$X = a + (b-a)U$	
Gumbel $\mathcal{G}$	$X = -\ln(-\ln U)$	logarithmierte $\mathcal{E}(1)$- Verteilung
Log-logistische Verteilung $\mathcal{LL}$	$X = \dfrac{U}{1-U}$	Exponential einer $\mathcal{L}$-Verteilung
Erlang $\mathcal{E}(n, \lambda)$, $\lambda > 0, n \in \mathbb{N}$	$X = -\sum_{k=1}^{n} \dfrac{\ln(U_k)}{\lambda}$	Faltung von $\mathcal{E}(\lambda)$-Verteilungen

Lernergebnisse (C3)
Die Studierenden kennen die Definition der Pseudoinversen und können sie anwenden. Sie können Realisationen von diskreten und stetigen Zufallsvariablen mit der Inversionsmethode bestimmen und verstehen die Funktionsweise von einfachen Transformationen.

8.3 Spezielle Verfahren

Kerninhalte

- Simulation der wichtigsten diskreten und stetigen Zufallsvariablen
- Simulation der mehrdimensionalen Normalverteilung
- Simulation von Abhängigkeiten mit Copulas

Es existieren zahlreiche Simulationsmethoden, die auf spezifische Verteilungen zugeschnitten sind. Oft liegt die Transformationsformel für Dichten diesen Verfahren zugrunde.

8.3.1 Diskrete Verteilungen

In der folgenden Tabelle sind die Algorithmen für die wichtigsten diskreten Verteilungen enthalten. Dabei sind $U \sim \mathcal{U}[0,1]$ bzw. $U_1, U_2, \ldots \overset{\text{iid}}{\sim} \mathcal{U}[0,1]$.

Verteilung	Generator
Bernoulli $B(1,p)$	$X = \lfloor U + p \rfloor$
Binomial $B(n,p)$	$X = \sum_{k=1}^{n} \lfloor U_k + p \rfloor$
Geometrisch $NB(1,p)$	$X = \left\lfloor \frac{\ln U}{\ln(1-p)} \right\rfloor$
Negativ Binomial $NB(n,p)$	$X = \sum_{k=1}^{n} \left\lfloor \frac{\ln U_k}{\ln(1-p)} \right\rfloor$
Poisson $\mathcal{P}(\lambda)$	$X = \inf\left\{ n \in \mathbb{N}_0 : -\frac{1}{\lambda} \sum_{k=1}^{n+1} \ln U_k > 1 \right\}$

Die Binomialverteilung $B(n,p)$ und die Negativ Binomialverteilung $NB(n,p)$ entstehen als Summen von n unabhängigen Bernoulli- bzw. geometrisch verteilten Zufallsvariablen.

Für die Simulation der Poisson-Verteilung nutzt man aus, dass die Zwischenankunftszeiten eines Poisson-Prozesses exponentialverteilt sind. Man generiert so lange exponentialverteilte Zufallszahlen, bis die Summe zum ersten Mal den Wert 1 überschreitet. Dann wird die Realisation x =Anzahl der Summanden minus eins gesetzt.

8.3.2 Normalverteilung

Standardnormalverteilte Zufallszahlen werden nach dem Box-Muller Verfahren generiert:

- Erzeuge unabhängige $U_1, U_2 \sim \mathcal{U}[0,1]$
- Setze
$$X = \sqrt{-2 \ln U_1} \cos(2\pi U_2), \quad Y = \sqrt{-2 \ln U_1} \sin(2\pi U_2).$$

X, Y sind unabhängig und $\mathcal{N}(0,1)$ verteilt. Dies folgt aus der Transformationsformel für Dichten.

8.3.3 Auf der Normalverteilung basierende Verteilungen

Stehen Realisationen $X \sim \mathcal{N}(0,1)$ bzw. $X_i \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$, $i = 1, \ldots n+1$ zur Verfügung, dann können weitere Verteilungen wie folgt simuliert werden:

Verteilung	Transformation
Normalverteilung $\mathcal{N}(\mu, \sigma^2)$	$\mu + \sigma X$
Chi-Quadrat χ^2_n	$\sum_{k=1}^{n} X_k^2$
Student t_n	$\frac{\sqrt{n} X_{n+1}}{\sqrt{\sum_{k=1}^{n} X_k^2}}$
Log-Normal $\mathcal{LN}(\mu, \sigma^2)$	$e^{\mu + \sigma X}$

Die Verteilungseigenschaften folgen hier aus der Konstruktion bzw. den Eigenschaften der Verteilungen.

8.3.4 Mehrdimensionale Normalverteilung

Ein Zufallsvektor $\mathbf{X} \sim \mathcal{N}_n(\mu, \mathbf{\Sigma})$ mit Erwartungswert $\mu \in \mathbb{R}^n$ und einer positiv definiten symmetrischen Matrix $\mathbf{\Sigma} \in \mathbb{R}^{n \times n}$ wird mit folgendem Verfahren erzeugt:

- Bestimme eine nicht singuläre Matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ mit $\mathbf{A}\mathbf{A}^\top = \mathbf{\Sigma}$, z.B. mit der Cholesky-Zerlegung.
- Simuliere $\mathbf{Y} = (Y_1, \ldots, Y_n)^\top$ mit $Y_k \overset{\text{iid}}{\sim} \mathcal{N}(0, 1)$.
- Setze $\mathbf{X} = \mu + \mathbf{A}\mathbf{Y}$.

Dann gilt $\mathbf{X} \sim \mathcal{N}_n(\mu, \mathbf{\Sigma})$.

Anmerkung 8.3. (a) Im Fall $n = 2$ und Korrelationskoeffizient ρ erhält man das folgende Verfahren:

- Simuliere $Y_1, Y_2 \overset{\text{iid}}{\sim} N(0, 1)$
- $X_1 := Y_1$, $X_2 := \rho Y_1 + \sqrt{1 - \rho^2} Y_2$.

(b) Bei gegebenem $\mathbf{\Sigma}$ kann man mit dem obigen Verfahren aus jedem Zufallsvektor $\mathbf{Y}$ mit unabhängigen Komponenten, Erwartungswert $E(\mathbf{Y}) = 0$, $V(\mathbf{Y}) = \mathbf{E}$ mittels

$$\mathbf{X} := \mu + \mathbf{A}\mathbf{Y}$$

einen Zufallsvektor mit Ervartungswert $E(\mathbf{X}) = \mu$ und Kovarianzmatrix $V(\mathbf{X}) = \mathbf{\Sigma}$ erzeugen. Verteilungsaussagen über die Randverteilungen sind im Allgemeinen nicht möglich, man kann also nicht Randverteilung und Korrelationskoeffizienten vorgeben.

(c) Derselbe Algorithmus kann für die Simulation der mehrdimensionalen t-Verteilung verwendet werden.

8.3.5 Simulation von Zufallsvektoren

Zusammenhänge zwischen Zufallsvariablen kann man mit Hilfe von Copulas und der Inversionsmethode simulieren:

1. Schritt Parametrisierung der Randverteilungen $F_1, \ldots, F_n$ und der Copula C

2. Schritt Erzeuge $(U_1, \ldots, U_n) \sim C$

3. Schritt Setze $\mathbf{X} := (F_1^{\leftarrow}(U_1), \ldots, F_n^{\leftarrow}(U_n))$

Der 2. Schritt wird für ausgewählte Copulas mit $n = 2$ skizziert.

Die Gauß-Copula mit Korrelationskoeffizient $\rho \in (-1, 1)$ kann wie folgt simuliert werden:

- Erzeuge einen Vektor $(X_1, X_2)^\top$ aus einer $\mathcal{N}_2((0, 0)^\top, \mathbf{\Sigma})$-Verteilung mit $\mathbf{\Sigma} = \rho^2 \mathbf{E}$.
- Bilde den Vektor $(U_1, U_2)^\top := (\Phi(X_1), \Phi(X_2))^\top$, wobei Φ die Verteilungsfunktion der Standardnormalverteilung ist.

Für archimedische Copulas wie die Gumbel-, Frank oder Clayton-Copula ergibt sich der folgende Algorithmus:

- Erzeuge $X_1, X_2 \overset{\text{iid}}{\sim} U[0, 1]$.

- Löse die Gleichung

$$X_2 = \frac{(\phi^{-1})'((\phi(X_1) + \phi(X_3))}{(\phi^{-1})'(\phi(X_3))} \tag{8.1}$$

  nach X_3 auf. Dies ist eventuell nicht analytisch möglich.

- Bilde den Vektor $(U_1, U_2)^\top := (X_1, X_3)^\top$.

Im Fall der Clayton-Copula bzw. der Frank-Copula kann man (8.1) explizit lösen:

$$X_3 = \left(X_1^{-\theta}\left(X_2^{-\frac{\theta}{\theta+1}} - 1\right) + 1\right)^{-1/\theta} \text{ bzw. } X_3 = -\frac{1}{\theta}\ln\left(1 + \frac{X_2(1-e^{-\theta})}{X_2(e^{-\theta X_1} - 1) - e^{-\theta X_1}}\right).$$

Zum Beispiel bietet das Statistik-Programm R im Paket QRM Funktionen zur Simulation der hier vorgestellten Copulas. So erzeugt man mit `rcopula.gauss, rcopula.clayton, rcopula.frank` die entsprechenden Zufallszahlen.

Lernergebnisse (C2)
Die Studierenden verstehen die Funktionsweise der Verfahren und können deren Verteilungseigenschaften nachweisen. Sie wissen, wie man Normalverteilungen und nichtlineare Abhängigkeitsstrukturen mit Copulas simuliert.

8.4 Numerik stochastischer Differenzialgleichungen

Kerninhalte

- Euler-Methode

Wir betrachten den Itō-Prozess der Form

$$dX_t = D(X_t, t)dt + V(X_t, t)dW_t,\ X_0 = x_0 \tag{8.2}$$

mit der Drift $D(x, t)$, der Volatilität $V(x, t)$ und dem Inkrement dW_t eines Wienerprozesses. Es sollen Pfade für $\{X_t\}_{t\in[0,T]}$ simuliert werden. Sei $\Delta t := \frac{T}{n}$, $n \in \mathbb{N}$ und $t_i = i\Delta t$, $i = 0, \ldots n$. Nun wird (8.2) mit der Euler-Methode diskretisiert:

$$\begin{aligned}\hat{X}_0 &:= x_0\\ \hat{X}_{t_{i+1}} &:= \hat{X}_{t_i} + D(\hat{X}_{t_i}, t_i)\Delta t + V(\hat{X}_{t_i}, t_i)\varepsilon_{i+1}\sqrt{\Delta t},\ i = 0, \ldots, n-1\end{aligned}$$

mit $\varepsilon_i \overset{\text{iid}}{\sim} N(0, 1)$. Die Simulation eines Pfads erhält man durch die Simulation der ε_i.

Lernergebnisse (C3)
Die Studierenden verstehen wie man Approximationen der Pfade von Itō-Prozessen erhält. Sie können die Euler-Approximation bei konkreten stochastischen Differenzialgleichungen umsetzen.

9 Literaturverzeichnis

9.1 Deskriptive Statistik

[1] Becker, T., Herrmann, R., Sandor, V., Schäfer, D., Wellisch, U. : Stochastische Risikomodellierung und statistische Methoden, 2. Auflage, Springer, Berlin (2024), insbesondere Kapitel 2.

[2] Fahrmeir, L., Heumann, C., Künstler, R., Pigeot, I., Tutz, G.: Statistik - Der Weg zur Datenanalyse, Springer, Berlin (2016).

[3] Ohlsson, E., Johansson, B.: Non-Life Insurance Princing with Generalized Linear Models. Springer, Berlin (2010).

9.2 Lebensdauermodelle

[1] Becker, T., Herrmann, R., Sandor, V., Schäfer, D., Wellisch, U. : Stochastische Risikomodellierung und statistische Methoden, 2. Auflage, Springer, Berlin (2024), insbesondere Kapitel 9.

[2] DAV Richtlinie, Biometrische Rechnungsgrundlagen bei Pensionskassen und Pensionsfonds, 28.1.2019

[3] Herleitung der DAV-Sterbetafel 2004 R für Rentenversicherungen, DAV-Unterarbeitsgruppe Rentnersterblichkeit, Blätter der DGVFM, Band XXVII, Heft 2, Oktober 2005 (21-22)

[4] Herrmann, R. (2006): Value-at-Risk, Tail Value-at-Risk und Schadenverteilung in der Personenversicherung, Blätter der DGVFM, Bd. XXVII, Heft 4

[5] Kainhofer, R., Predota, M., Schmock, U. (2006): The New Austrian Annuity Valuation Table AVÖ 2005R, Mitteilungen der AVÖ, Heft 13 (75-76)

[6] Kakies, P., Behrens, H.-G., Loebus, H., Oehlers-Vogel, B., Zschoyan, B.: Methodik von Sterblichkeitsuntersuchungen, Schriftenreihe Angewandte Versicherungsmathematik Heft 15, Verlag Versicherungswirtschaft e.V., Karlsruhe (1985), (Auszüge aus den Kapiteln 1-3: 15-28; 54-59; 86-87; 92; 103; 108; 114)

[7] Schmithals, B., Schütz, E. U. (1995): Herleitung der DAV-Sterbetafel 1994 R für Rentenversicherungen. Blätter der DGVM, Band XXII, Heft 1.

[8] DAV-Unterarbeitsgruppe Haftpflicht-Unfallrenten des HUK-Ausschusses, Herleitung der DAV-Sterbetafel 2006 HUR, 2006

[9] Lee, R. D., Carter, L. (1992): Modeling and forecasting the time series of US mortality. Journal of the American Statistical Association 87, 659-671.

[10] Toutenburg, H., Heumann, C. (2007): Induktive Statistik, Kapitel 12, Springer

[11] Fahrmeier, L., Hamerle, A. (1996): Multivariate statistische Verfahren, Kapitel 7, De-Gruyter

9.3 Abhängigkeiten und Copulas

[1] Becker, T., Herrmann, R., Sandor, V., Schäfer, D., Wellisch, U. : Stochastische Risikomodellierung und statistische Methoden, 2. Auflage, Springer, Berlin (2024)

[2] McNeil, A., Frey, R., Embrechts, P. (2015): Quantitative Risk Management. Princton University Press, NY.

[3] Mai, J.F., Scherer, M. (2014): Financial Engineering with Copulas Explained. Palgrave Macmillan, NY.

[4] Nelsen, R.B. (2006): An Introduction to Copulas. Springer, NY.

9.4 Induktive Statistik

[1] Becker, T., Herrmann, R., Sandor, V., Schäfer, D., Wellisch, U. : Stochastische Risikomodellierung und statistische Methoden, 2. Auflage, Springer, Berlin (2024)

[2] Fahrmeir, L., Heumann, C., Künstler, R., Pigeot, I., Tutz, G.: Statistik - Der Weg zur Datenanalyse, Springer, Berlin (2016).

[3] Fahrmeir, L., Kneib, T., Lang, S. (2009): Regression. Springer Berlin.

[4] Hastie, T., Tibshirani, R., Friedman, J. (2009): The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second Edition. Springer New York.

9.5 Zeitreihenanalyse

[1] Pruscha, H. (2006): Statistisches Methodenbuch: Verfahren, Fallstudien, Programmcodes. Springer Berlin.

[2] Schlittgen, R., Streitberg, B. H. J. (2001): Zeitreihenanalyse. 9. Auflage. Oldenbourg München.

9.6 Stochastische Prozesse

[1] Becker, T., Herrmann, R., Sandor, V., Schäfer, D., Wellisch, U. : Stochastische Risikomodellierung und statistische Methoden, 2. Auflage, Springer, Berlin (2024)

[2] Rolski,T. , Schmidli, H., Schmidt, V., Teugels J. (1999): Stochastic Processes for Insurance and Finance. Wiley, N.Y.

[3] Korn, R., Korn, E. (1999): Optionsbewertung und Portfolio-Optimierung. Vieweg, Braunschweig.

9.7 Credibility

[1] Bühlmann, H., Gisler, A. (2005): A Course in Credibility Theory and its Applications. Springer, Berlin.

[2] Becker, T., Herrmann, R., Sandor, V., Schäfer, D., Wellisch, U. : Stochastische Risikomodellierung und statistische Methoden, 2. Auflage, Springer, Berlin (2024)

9.8 Monte-Carlo-Simulation

[1] Becker, T., Herrmann, R., Sandor, V., Schäfer, D., Wellisch, U. : Stochastische Risikomodellierung und statistische Methoden, 2. Auflage, Springer, Berlin (2024)

[2] Glassermann, P. (2004): Monte Carlo Methods in Financial Engineering. Springer, N.Y.

[3] Korn, R., Korn, E., Kroisandt, G. (2010): Monte Carlo Methods and Models in Finance and Insurance. Chapmann Hall, London.

[4] Robert, C., Casella, G. (2010): Introducing Monte Carlo Methods with R. Springer, N.Y.

[5] Mai, J.F. , Scherer, M. (2012): Simulating Copulas. Imperial College Press, London.

Index