



FH Nordakademie - Hochschule der Wirtschaft

Bachelorarbeit
im Studiengang B.Sc. Wirtschaftsinformatik

Arbeitstitel

Maschinelle Grenzbetragsanalyse anhand
mathematischer Methoden im
Compliance-Umfeld eines
Finanzdienstleisters

Zeitraum : 08. Dezember 2014 - 09. Februar 2015

Gutachter : StB Prof. Dr. Gehrke, Nick

CoGutachter : Prof. Dr. Zimmermann, Frank

Student : Johannßen, Dirk

Matr.-Nr. : 4857

Firma : PPI AG

Betreuer : Dr. Rieß, Stefan

Eidesstattliche Erklärung

Hiermit versichere ich, die vorliegende Arbeit selbstständig und unter ausschließlicher Verwendung der angegebenen Literatur und Hilfsmittel erstellt zu haben.

Die Arbeit wurde bisher in gleicher oder ähnlicher Form keiner anderen Prüfbehörde vorgelegt und auch nicht veröffentlicht.

06.02.2015, Elmshorn

Dirk Johannßen

Inhaltsverzeichnis

Abbildungsverzeichnis	VI
Tabellenverzeichnis	VII
Abkürzungsverzeichnis	VIII
1. Thematische Einführung, Problemstellung und Vorgehen	1
1.1. Einführung und Problemstellung	1
1.2. Vorgehen und Aufbau der Bachelorthesis	1
2. Fachliche Grundlagen	3
2.1. Begriffe im Compliance-Umfeld	3
2.2. Grenzbetragsanalyse im Compliance-Umfeld	4
2.3. Analysemethodenauswahl	5
2.3.1. Vorstellung der exemplarischen Geschäftsprozesse Kon-	
toeröffnung und Kreditgewährung	5
2.3.2. Identifizierte Red Flags in den Geschäftsprozessen	7
2.3.3. Unterscheidung in automatisierbare und manuelle Ana-	
lysemöglichkeiten	7
2.3.4. Ableitung von Kategorien	8
2.4. Forschungsstand	11
2.5. Hypothesenbildung und Forschungsfrage	12
3. Betrugsszenarien	13
3.1. Betrug durch Mitarbeiter	13
3.1.1. Betrug durch Mitarbeiter: Der „interne Feind“	13
3.1.2. Entdeckung gefälschter Zahlen durch Carslaw, 1988	14
3.1.3. Umgehung der Mitarbeiter-Identifikation, Miami 2002	15
3.2. Bilanzfälschung	15
3.2.1. Der Begriff Bilanzdelikt	16
3.2.2. Das Schneeballsystem Jürgen Schneiders, 1994	16
4. Mathematische Methoden	17
4.1. Prozentuale Grenzen	17
4.2. Unscharfe Logik	17
4.2.1. Der Begriff und Motivation der <i>unscharfen Logik</i>	17
4.2.2. Fuzzy-Mengen	18
4.2.3. Zugehörigkeitsfunktionen	19
4.2.4. Logische Operationen auf Fuzzy-Mengen	19
4.2.5. Fuzzifizierung, Defuzzifizierung	20
4.3. Stochastik	21
4.3.1. Abgrenzung relevanter Stochastik	21
4.3.2. Mittelwerte und Streuungsmaß	21

4.3.3.	Erwartungswert, Varianz und Standardabweichung	22
4.3.4.	Chi-Quadrat-Verteilung und Chi-Quadrat-Test	24
4.4.	Ermittlung des Maximums einer Menge	24
4.4.1.	Maschinelle Umsetzbarkeit	25
4.5.	Benfordsches Gesetz	25
4.5.1.	Mathematische Darstellung	25
4.5.2.	Bedeutung für die Grenzbetragsanalyse	26
4.5.3.	Maschinelle Umsetzbarkeit	26
5.	Analyse der Szenarien	27
5.1.	Datenbestand	27
5.2.	Analyse gefälschter Datenmengen nach Carslaw, 1988	28
5.3.	Analyse des Betrugs in Miami, 2002	30
6.	Bewertungsmodell	37
6.1.	Eignungskriterien	37
6.2.	Erstellung eines Bewertungsmodells	37
6.2.1.	Ermittlung eines Maximums	38
6.2.2.	Standardabweichung	38
6.2.3.	Absoluter Warnhinweis nach Schwellenwert und Bewertungsmodell	38
6.3.	Motivation und Eigenschaften	38
6.4.	Maschinelle Abbildung	39
6.4.1.	Maximum, arithmetisches Mittel, Varianz und Standardabweichung	39
6.4.2.	Bewertungsmodell als Pseudocode	40
6.5.	Anwendung des Bewertungsmodells	41
7.	Zusammenfassung, Fazit und Ausblick	42
7.1.	Resümee der Kernaspekte	42
7.2.	Hypothesen und Forschungsfrage	43
7.3.	Fazit der gewonnenen Erkenntnisse	44
7.4.	Ausblick und zukünftige Entwicklung	45
	Literaturverzeichnis	X
	Onlinequellen-Verzeichnis	XII
	Interne Quellen	XIV
	Inhalt der CD	XV
Anhang		XVI
A.	Prototyp	XVI
A.1.	Motivation	XVI
A.2.	GUI	XVI
B.	Benfordsches Gesetz	XX
B.1.	Historische Entwicklung	XX
B.2.	Erkenntnisgewinnung aus Newcombs Studien	XX
B.3.	Testdatenerstellung zu Carslaw, 1988	XXI
B.4.	Originale vs. manipulierte Daten Carslaws	XXIII

C.	Octave	XXV
	C.1. Anbindung einer Java-Bibliothek	XXV
	C.2. Beispiel-Code der Nutzung von jOctave	XXV
D.	Fuzzylogik	XXVII
	D.1. Anbindung an Java	XXVII
	D.2. Grafische Darstellung der Defuzzifizierung	XXVII
E.	Beispiel-Fuzzy-Code mit grafischen Abbildungen	XXIX
	E.1. Mathematische Eigenschaften der t-Norm	XXIX
	E.2. FCL-Anweisungen zu einer Fuzzy-Berechnung	XXIX
F.	FCL der Fuzzy-Analyse des Miami Betrugs	XXXIII
G.	Jürgen Schneider Untersuchung	XXXVI

Abbildungsverzeichnis

2.1.	Vereinfachte Darstellung aller Prozessschritte des Kontoeröffnungsprozesses	6
2.2.	Vereinfachte Darstellung aller Prozessschritte des Kreditbearbeitungsprozesses	6
2.3.	Kategorisierte Red Flags	10
4.1.	T-Norm „AND“ und „MAX“	20
4.2.	Wahrscheinlichkeiten eines Wertes innerhalb des Sigma-Intervalls, wobei x die Anzahl der σ ist	23
5.1.	Originale Testdaten auf Basis von empirischen Daten zu tausend in der Carslaw-Untersuchung	29
5.2.	Manipulierte Testdaten auf Basis von empirischen Daten zu tausend in der Carslaw-Untersuchung	30
5.3.	Stichproben von Feierabend-Zeiten	32
5.4.	Fuzzifizierung und defuzzifizierung am Beispiel der Regeln 1 und 11, sowie der Eingabeparameter 8,5h und 15% MA anwesend	36
A1.	Das Hauptmenü der Oberfläche des Prototyps	XVI
A2.	Einführungsdialog zur Auswahl einer CSV-Datei	XVII
A3.	Nutzerdialog für Analysekonfigurationen	XVII
A4.	Einfache Eingabe eines zu prüfenden Z-Index des Prototyps	XVIII
A5.	Darstellung der Resultate aus Prüfroutinen	XIX
D1.	Fuzzifizierung und Defuzzifizierung	XXVIII
E1.	Die Service-Qualität aus dem Beispiel grafisch dargestellt	XXXI
E2.	Die Qualität der Speisen aus dem Beispiel grafisch dargestellt	XXXI
E3.	Das resultierende Trinkgeld je nach Service- und Speisequalität aus dem Beispiel	XXXII
G1.	Kredite Schneiders und der Grundgesamtheit	XXXVIII
G2.	Aufbereitete Daten in logarithmischer Darstellung der Häufigkeiten als Histogramm	XXXIX

Tabellenverzeichnis

2.1. Red Flags im Kontoeröffnungsprozess	7
2.2. Red Flags im Kreditbearbeitungsprozess	8
5.1. Auswertung der Minuten bis zum Feierabend erfasster Mitarbeiter ab der Uhrzeit 17 Uhr	31
5.2. Auswertung der Minuten bis zum Feierabend ab 17 Uhr	33
5.3. Wahrscheinlichkeiten doloser Handlungsweisen der verdächtigten Mitar- beiterin	35
B1. Auszug aus Datenerhebungen Newcombs	XX
B2. Wahrscheinlichkeitsverteilung der ersten Ziffer aller untersuchten Zahlen .	XXI
B3. Manipulierte Einwohnerzahlen Deutscher Städte des Jahres 1950 in Ein- heiten zu tausend	XXII
B4. Minuten bis zum Feierabend ab 17 Uhr, genutzt in Kapitel 5.3 zur Un- tersuchung auf Verteilung nach dem Benfordschen Gesetz	XXIII
B5. Auswertung der Originaldaten	XXIII
B6. Auswertung der manipulierten Daten	XXIV
G1. Hypothetische Daten des Jürgen-Schneider Betrugs	XXXVII
G2. Hypothetische Daten des Jürgen-Schneider Betrugs	XXXVII
G3. Aufbereitete Daten der Kreditverteilungen Schneiders	XL
G4. Fuzzy-Untersuchung zum Fall Jürgen Schneider	XLI
G5. σ -Abweichungen Jürgen Schneider	XLII

Abkürzungsverzeichnis

API	A pplication P rogramming I nterface (dt. Programmierschnittstelle)
BDSG	Bundesdatenschutzgesetz
FCL	F uzzy C ontrol L anguage (dt. Sprache zur Kontrolle von Unschärfe)
GOB	G rundsätze O rdnungsmäßiger B uchführung
GUI	G raphical U ser I nterface (dt. Grafische Benutzeroberfläche)
IEC	I nternational E lectrical C ommission
JDK	J ava D evelopment K it
ZQA	Z entraler Q uartilsabstand

„Humans are allergic to change. They love to say, 'We've always done it this way.' I try to fight that. That's why I have a clock on my wall that runs counter-clockwise.“

von Grace Hopper, Flottillenadmiral und Computerpionierin.

1. Thematische Einführung, Problemstellung und Vorgehen

Zu Beginn wird in diesem Abschnitt die Motivation für das Thema der Grenzbetragsanalyse im Compliance-Umfeld eines Finanzdienstleisters skizziert. Es stellt die wirtschaftliche Relevanz und das Schadenspotential von betrügerischen Handlungen dar und verdeutlicht den Zugzwang für Wirtschaftsunternehmen, sich mit dem Thema zu beschäftigen.

Ferner wird dieses Kapitel eine kurze Übersicht über den Aufbau der Thesis bieten.

1.1. Einführung und Problemstellung

In einer Studie der Association of Certified Fraud Examiners (ACFE) aus dem Jahr 2012, gehen CFOs aller befragten Firmen davon aus, dass sie etwa 5% ihrer Einnahmen durch betrügerisches oder doloses Verhalten, auch Fraud genannt, einbüßen müssen. Gemessen am Weltbruttosozialprodukt 2011, würde dies laut der ACFE eine Summe von 3,5 Billionen US-Dollar bedeuten.¹

Der National Criminal Intelligence Service (NCIS) Großbritanniens ordnet Fraud unter die sieben größten Bedrohungen des Landes, nach Feuerwaffen und Drogentransport. Ferner wird geschätzt, dass die Hälfte aller von der britischen Polizei untersuchten Taten dem Bereich Fraud zuzuordnen sind, doch der zugeteilte Bereich der Polizei mit nur etwa drei Prozent demgegenüber gering ausfällt.²

Die gewaltige Höhe der Schadenssumme, sowie der Anteil der Taten im Fraud-Bereich, bei einem ungleich geringeren Maß an beschäftigten Beamten, veranschaulicht exemplarisch die Compliance-Landschaft vieler Unternehmen und die zentrale Problemstellung. Zwar sind sich die meisten Akteure der akuten Bedrohung durch Fraud bewusst, doch werden kaum Mittel für die Bekämpfung zur Verfügung gestellt. Ein Grund dafür ist, dass die Betrüger kreativ und innovativ, die Methoden der Ermittler und ermittelnden Unternehmen jedoch begrenzt sind. Einen Fraud-Fall zu detektieren, also diesen selbst ohne Anfangsverdacht zu bemerken, stellt viele Akteure vor großen Herausforderungen.

1.2. Vorgehen und Aufbau der Bachelorthesis

Zunächst werden einleitend in Kapitel 2 die Begriffe Compliance und Grenzbetrag mitsamt weiteren Fachtermini geklärt, um den wirtschaftlichen Kontext einzuführen. Im Anschluss daran erfolgt der Grund für die Auswahl des Grenzbetrags als Analyseobjekt, der Stand der Forschung, die Bildung von Hypothesen und darauf aufbauend die eigentliche Forschungsfrage.

¹Vgl. [AC12]

²Vgl. [Do13]

Nach Erläuterung der Begriffe befasst sich der Kern der Thesis anfangs in Kapitel 3 mit der Frage, wie betrügerische Handlungen in der Wirtschaft aussehen können, und führt einige Szenarien auf.

Kapitel 4 behandelt mathematische Grundlagen, die bei der Grenzbetragsanalyse verwendet werden. Die maschinelle Abbildung der mathematischen Grundlagen ist ebenfalls Gegenstand des Kapitels.

Kapitel 5 zeigt die Anwendung der mathematischen Methoden auf die vorgestellten Szenarien. Dabei wird aufgezeigt, wie Analysen mit der Fragestellung nach einem Grenzbetrag oder Limit aussehen könnten und zu welchen Aussagen die Ergebnisse gekommen wären. Dabei wird zunächst die Problematik der Datenbeschaffung erklärt. Für diese Thesis werden Testdaten erstellt und analysiert.

Untersuchungen von Betrugsszenarien in Kapitel 5 können eine Menge von Warntausagen zur Folge haben. Diese können sich sogar widersprechen. Für die Wirtschaft von Bedeutung ist die Frage, wie eine Menge von Warntausagen zu einer einzigen Aussage über das Schadenspotential aggregiert werden kann. Dies ist Aufgabe eines Bewertungsmodells, welches in Kapitel 6 erarbeitet wird.

Abschließend werden in Kapitel 7 die Kernaussagen und -erkenntnisse rekapituliert. Ferner wird ein kritisches Fazit der Analysen und des Bewertungsmodells gezogen, um anschließend einen Ausblick auf die weitere Vertiefung der Thematik und dieser Thesis zu geben.

Im Verlauf der Thesis werden vertiefende Informationen referenziert, die ab Anhang A aufzufinden sind. Dies können Konfigurationsdateien, Tabellen, Beschreibungen, Daten oder Abbildungen sein.

2. Fachliche Grundlagen

Um sich den eigentlichen Fragestellungen dieser Bachelorthesis zu nähern, werden zunächst relevante Begriffe und Gegenstände erläutert. Die Forschungsfrage befasst sich mit zwei grundlegenden Themen. Dies sind zum einen Compliance und zum anderen der Begriff Grenzbetrag.

2.1. Begriffe im Compliance-Umfeld

Zunächst wird die Definition oder das allgemeine Verständnis des Begriffs Compliance und dessen Umfeld eingeführt. Zudem behandelt dieser Abschnitt die Begriffe Betrug (engl. Fraud), Red Flags und umschreibt die juristische Relevanz von Compliance für ein wirtschaftliches Unternehmen.

Compliance

Eine einheitliche Definition des Begriffs Compliance ist in der Literatur nicht zu finden. Vielmehr gibt es verschiedene Auffassungen und Erklärungen, um was es sich bei dem Begriff handelt. „Compliance bezeichnet die Gesamtheit aller zumutbaren Maßnahmen, die das gesetz- und regelkonforme Verhalten eines Unternehmens [...] und seiner Mitarbeiter im Hinblick auf alle gesetzlichen Ge- und Verbote begründen.“³

An dieser vagen Umschreibung ist hervorzuheben dass die *zumutbaren Maßnahmen* letztlich ein durch den Zeitgeist, die Unternehmensstruktur etc. geprägter Begriff ist. Er ist von Fall zu Fall variabel und nicht klar definiert.

Betrug (engl. Fraud)

Auch für den Begriff Fraud gibt es keine einheitliche Definition. Allgemein spricht man jedoch von Fraud, wenn folgende Elemente gegeben sind:

„Eine Repräsentation über einen materiellen Gegenstand, welche falsch ist, und dies bewusst oder trotz des Wissens darum, die zudem vom Opfer geglaubt wird, was der Täter ausnutzt, um dem Opfer Schaden zuzufügen.“⁴

Red Flag

Als Red Flags werden Warnsignale oder -hinweise bezeichnet, die Grund zu der Annahme doloser Handlungen oder Betrug (engl. Fraud) geben.⁵

³[Qe11] S. 13

⁴Vgl. [Al11] S. 7

⁵Vgl. [HS10] S. 234

Fraud Detection

Bei Fraud Detection handelt es sich um alle Maßnahmen zur Erkennung von bereits getätigten deliktischen Handlungen. Somit ist Fraud Detection vergangenheitsbezogen. Erreicht werden kann Fraud Detection vor allem durch investigative Untersuchungen und interne Revisionen.⁶

Fraud Prevention

Fraud Prevention ist ein Teil der Bekämpfungsmaßnahmen von Fraud. Dabei handelt es sich um die Vorbeugung vor zukünftigen deliktischen Handlungen. Unterschieden wird dabei in primäre und sekundäre Maßnahmen. Die primären Maßnahmen sollen gesellschaftliche Werte stärken und Menschen für das Thema Fraud sensibilisieren. Die sekundären Maßnahmen zielen darauf ab, Täter abzuschrecken.⁷

2.2. Grenzbetragsanalyse im Compliance-Umfeld

Praktisch vor Augen führen kann man sich die Relevanz der Grenzbetragsanalyse im Compliance-Umfeld durch ein Beispiel, aufgeführt in einer Bachelorthesis zu Fraud⁸. Bei diesem Beispiel werden Kreditanträge ab dem Überschreiten eines Grenzbetrags an ein anderes Gremium des Finanzdienstleisters übergeben, welches daraufhin umfangreichere Prüfungen durchführt. Ein Betrüger, welcher von solch einer Schwelle oder solch einem Grenzbetrag wüsste, würde sodann versuchen, stets unter diesem Wert für Kreditanträge zu bleiben. Idealerweise soll die Geldsumme aus einem betrügerischen Kredit hoch sein. Daher wird der Betrüger seine Anträge nahe bei, aber stets unter dem Grenzbetrag stellen.⁹

Grenzbetrag

Andere Begriffe für den Grenzbetrag sind Limit oder Grenzwert. Abzugrenzen ist der Grenzbetrag im Compliance-Umfeld vom Grenzwert dem Verständnis der Mathematik nach. Dabei wird der mathematische Grenzwert folgendermaßen definiert:

„Eine Folge (x_n) reeller oder komplexer Zahlen heißt konvergent mit *Grenzwert* $a \in \mathbb{R}$ bzw. $a \in \mathbb{C}$, wenn es für alle $\epsilon \in \mathbb{R}$ mit $\epsilon > 0$ eine natürliche Zahl $N \in \mathbb{N}$ gibt, sodass für alle $n \in \mathbb{N}$ mit $n \geq N$ gilt: $|x_n - a| < \epsilon$ “¹⁰

Für den Grenzbetrag im Compliance-Umfeld ergibt sich ferner, dass dessen Verletzung einer Eskalation bedarf.¹¹ Veranschaulichen lässt sich dies exemplarisch an einer Spekulationsgrenze, bis zu welcher ein angestellter Börsenmakler an der Börse handeln darf.

⁶Vgl. [Ho08] S. 488

⁷Vgl. [Ho08] S. 81

⁸Vgl. [Hi14] S. 30

⁹Vgl. [Hi14] S. 30

¹⁰[PS09] S. 166

¹¹Vgl. [JLZ13] S. 170

Grenzbetragsanalyse

Die titelgebende Grenzbetragsanalyse ist eine grammatikalische Komposition der Worte *Grenzbetrag* und *Analyse*, der systematischen Untersuchung, bei dem das zu untersuchende Objekt in Elemente zerlegt, welche anschließend geordnet und ausgewertet werden.¹²

Die Grenzbetragsanalyse umfasst daher die Gesamtheit aller Methoden, welche dabei helfen, die Frage nach Verletzung eines Grenzbetrags zu untersuchen und zu beantworten.

2.3. Analysemethodenauswahl

Nachfolgend wird die Entscheidungsfindung zur Auswahl der Grenzbetragsanalyse als Analysemethode skizziert. Es wird eine Ausarbeitung zweier Geschäftsprozesse eines Finanzdienstleisters dargestellt und aufgezeigt, wie aus solchen Geschäftsprozessen Red Flags abgeleitet werden können.

Nach einer Kategorisierung der Red Flags wird herausgestellt, welche Analysemöglichkeiten sich je Kategorie ergeben können, um letztlich eine Empfehlung zur Vertiefung auszusprechen.

2.3.1. Vorstellung der exemplarischen Geschäftsprozesse Kontoeröffnung und Kreditgewährung

In einer Bachelorthesis aus 2014¹³, werden zwei Geschäftsprozesse eines Finanzinstituts exemplarisch und vereinfacht dargestellt. An ihnen sollen interne Red Flags, also Warnhinweise auf dolose Handlungen seitens der Mitarbeiter des Finanzinstituts, hergeleitet werden. Bei diesen Red Flags handelt es sich um eine Auswahl aus einer Menge aller denkbaren Warnhinweise.

Kontoeröffnungsprozess Besteht ein *Kundenwunsch zur Kontoeröffnung*, so erfolgt zunächst ein *beratendes Gespräch* mit einem Mitarbeiter des Instituts. Anschließend wird von jenem Mitarbeiter die *Verfügungsberechtigung festgelegt*, also vor allem Bonitätsprüfungen. Als nächstes werden auch die *Legitimationspapiere des Kunden geprüft*. Sind alle Prüfungen positiv, wird ein *Konto eröffnet*. Damit ist der Kontoeröffnungsprozess abgeschlossen. Dargestellt ist dies in Abbildung 2.1. Innerhalb dieser Prozessschritte aus der Abbildung kann es zu dolosen Handlungen kommen, die durch Red Flags potentiell erkennbar sind. Mögliche Red Flags werden nachfolgend in Abschnitt 2.3.2 aufgeführt.

¹²Vgl. [Du14]

¹³Vgl. [Hi14] S. 21

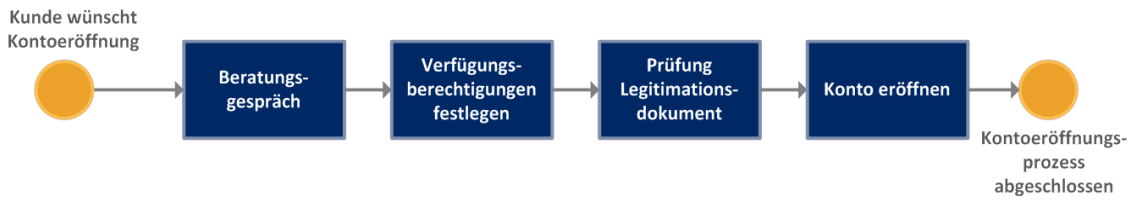


Abbildung 2.1.: Vereinfachte Darstellung aller Prozessschritte des Kontoeröffnungsprozesses ¹⁴

Kreditgewährungsprozess Auch beim Prozess zur Kreditgewährung wird der Prozess ausgelöst durch den entsprechenden *Kundenwunsch*. Zunächst erfolgt ebenfalls ein *beratendes Gespräch*, dem sich eine *Selbstauskunft* seitens des Kunden anschließt. Zur Sicherheit werden über den Kunden *weitere Informationen* eingeholt. Es folgen die *Prüfung auf Kreditfähigkeit und -würdigkeit*. Auf Basis der eingeholten Informationen wird eine *Kreditentscheidung* gefällt, nach deren positiver Ausgang der *Kredit bereitgestellt* wird. Nach einem vereinbarten Schema hat der Kunde in Folge *Zins- und Tilgungszahlungen* zu leisten. Letztlich werden nach Beendigung der letzten Zahlungen die SCHUFA informiert und die *Sicherheiten freigegeben*. Damit ist der Prozess abgeschlossen. Dargestellt ist dies in Abbildung 2.2. Innerhalb der Prozessschritte aus der Abbildung kann es zu dolosen Handlungen kommen, die durch Red Flags potentiell erkennbar sind. Mögliche Red Flags werden nachfolgend in Abschnitt 2.3.2 aufgeführt.

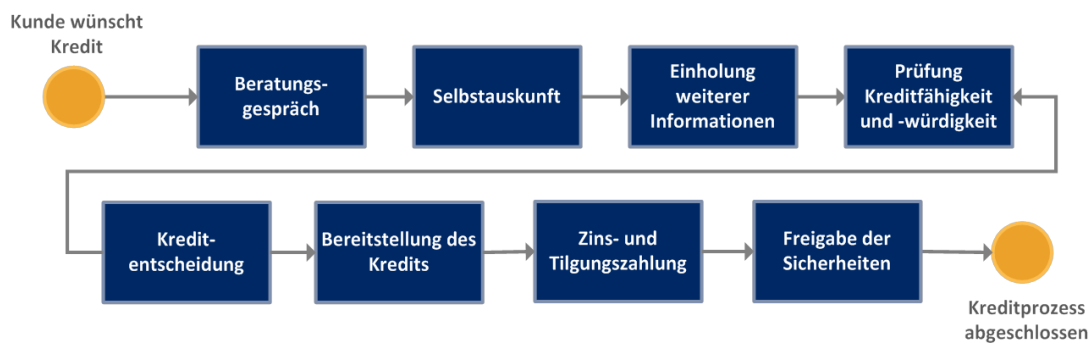


Abbildung 2.2.: Vereinfachte Darstellung aller Prozessschritte des Kreditbearbeitungsprozesses ¹⁵

¹⁴Vgl. [Hi14] S. 21

¹⁵ Vgl. [Hi14] S. 23

2.3.2. Identifizierte Red Flags in den Geschäftsprozessen

Weiter werden in der Ausarbeitung einige Red Flags exemplarisch abgeleitet. Diese stellen nur eine kleine Auswahl dar, da die Gesamtheit aller Red Flags zu umfangreich ausfällt. Die in der Bachelorthesis aufgelisteten Red Flags werden dargestellt in Tabelle 2.1 und Tabelle 2.2.

Prozessschritt	Red Flag
Beratungsgespräch	Überdurchschnittliche Neukundengewinnung; Unplausible Produktwahl (beispielsweise Festgelder langer Laufzeiten im hohen Alter)
Verfügungsberechtigung festlegen	Vollmacht Bankangestellter über Kundenkonto
Prüfung Legitimationsdokumente	Unplausible, gefälschte, fehlerhafte Identifikationsdokumente; Proxyausweis als Identitätsdokument
Konto eröffnen	Kontoeröffnung außerhalb der Geschäftszeiten

Tabelle 2.1.: Red Flags im Kontoeröffnungsprozess ¹⁶

An den herausgestellten Red Flags aus Tabelle 2.1 zeigt sich bereits, dass häufig die Frage nach der Verletzung eines bestimmten Wertes im Vordergrund steht. So stellt sich der Grenzbetrag des Red Flags *Kontoeröffnung außerhalb der Geschäftszeiten* als Schließzeiten einer Filiale dar. Auch die *überdurchschnittliche Neukundengewinnung* ist eine Frage nach einem Grenzbetrag, nämlich der als normal oder unnormal einzustufenden Anzahl an Neukunden pro Bemessungszeitraum. Hier setzt die Grenzbetragsanalyse an, die im Laufe dieser Thesis von Bedeutung ist.

2.3.3. Unterscheidung in automatisierbare und manuelle Analysemöglichkeiten

Red Flags wie *Bearbeitung außerhalb d. Geschäftszeiten* oder *zu frühe Freigabe von Sicherheiten* verlangen nicht nur nach absoluter Bestimmung von Regelverletzungen, sondern auch nach deren Ausmaßen. So ist die Bearbeitung außerhalb der Geschäftszeiten von wenigen Sekunden fachlich unbedenklich, während eine solche Bearbeitung Stunden nach Geschäftsschluss ein deutlicher Warnhinweis ist.

Sowohl die Frage nach Regelverletzung, als auch die Frage nach dem Grad einer Regelverletzung lassen sich automatisieren. Die Frage nach dem Grad einer Regelverletzung lässt sich durch eine Grenzbetragsanalyse beantworten. Dabei werden zwei Werte verglichen: Der Grenzwert und ein Wert darunter oder darüber, welcher beliebig nahe an diesen Grenzwert heran reicht. Mathematisch lassen sich Aussagen dazu treffen, wie nahe ein Wert an einen Grenzwert heranreichen darf, und ab wann eine Warnung oder Teilwarnung ausgesprochen wird.

¹⁶[Hi14] S. 23

¹⁷[Hi14] S. 28

Prozessschritt	Red Flag
Beratungsgespräch	Unplausible Produktwahl (etwa Kredite langer Laufzeiten für kurzfristige Investitionen)
Selbstauskunft	Unplausible, unvollständige Angaben
Einholung weiterer Informationen (Sicherheiten, Bürgschaft)	Gefälschte, fehlerhafte, unplausible Unterlagen
Prüfung der Kreditfähigkeit & Kreditwürdigkeit	Umgehung des Vier-Augen-Prinzips
Kreditentscheidung	Umgehung des Vier-Augen-Prinzips (Doppelvotierung); Berechtigung für Kompetenzbestimmung umgangen (Verschleierung von Kreditnehmereinheit, Kreditsumme gering gehalten)
Bereitstellung des Kredits	Wertstellung des Kredits zu Gunsten eines Kontos, das nicht dem Kreditnehmer gehört; Fehlende Transparenz in der Verwendung von Kreditmitteln
Zins- und Tilgungszahlungen	Kapitaldiensterbringung außerhalb der Geschäftszeiten; Plötzliche Kredittilgung und -Ablösung; Rückständige Zins- und Tilgungszahlungen; Bankseitig veranlasste Regulierungsbuchungen
Freigabe der Sicherheiten	Keine Freigabe erfolgt

Tabelle 2.2.: Red Flags im Kreditbearbeitungsprozess ¹⁷

Der Charakter der Grenzwertanalyse bedingt ferner, dass Überschreitungen des Grenzwertes in jedem Fall Warnungen zur Folge haben. Somit sind viele automatisierbare Prüfungen auf Red Flags durch Grenzwertanalysen abdeckbar. Es lässt sich die Empfehlung aussprechen, die Grenzbetragsanalyse tiefreichend zu erschließen, um Chancen zu nutzen.

2.3.4. Ableitung von Kategorien

Wie die Tabellen 2.1 und 2.2 zeigen, kann es für Geschäftsprozesse und deren Prozessschritte viele Red Flags geben. Da ein Finanzdienstleister vom Gesetzgeber aufgefordert wird alle bekannten zu prüfen, kann sich eine gewaltige Menge durchzuführender Analysen ergeben. Dies bindet viele Ressourcen. Um die Komplexität der zu bewältigenden Menge an Red Flags zu verringern, empfiehlt es sich, Kategorien von Red Flags gemeinsamer Charakteristika zu bilden, welche ähnlich untersucht und behandelt werden können. So würden Einzelbetrachtungen von Red Flags in nächster Instanz entfallen.

Für eine interne Präsentation innerhalb der PPI AG, wurden auf der Bachelorthesis aufbauend Kategorien gebildet. Zwei wesentliche Kategorien von Red Flags wurden identifiziert: Kennzahlen und Prozessabläufe. Die Kennzahlen zeichnen sich dadurch aus, dass diese mit numerischen Werten erfassbar, und häufig dem mathematischen Umfeld

zuzuordnen sind. Prozessabläufe, hingegen, sind im fachlichen Umfeld angesiedelt und lassen sich zumeist manuell bewerten. Die meisten in Abschnitt 2.3.2 genannten Red Flags lassen sich entweder der einen oder anderen Kategorie zuordnen. Wenige weisen Charakteristika von sowohl einer Kennzahl, als auch einem Prozessablauf auf.

Weiter unterteilt in die zwei Prozesse Kontoeröffnung und Prozessbearbeitung, ergibt sich die in Abbildung 2.3 dargestellte Einordnung von Red Flags. Die Spalten *Kennzahlen* und *Prozessabläufe* stellen die Kategorien dar. *Kreditbearbeitungsprozess* und *Kontoeröffnungsprozess* sind Bereiche, deren Red Flags den Kategorien zugeordnet werden.

Um den Gegenstand eines Red Flags besser fassen zu können, werden zwei Red Flags exemplarisch als Erläuterungen einer Ausarbeitung entnommen:¹⁸

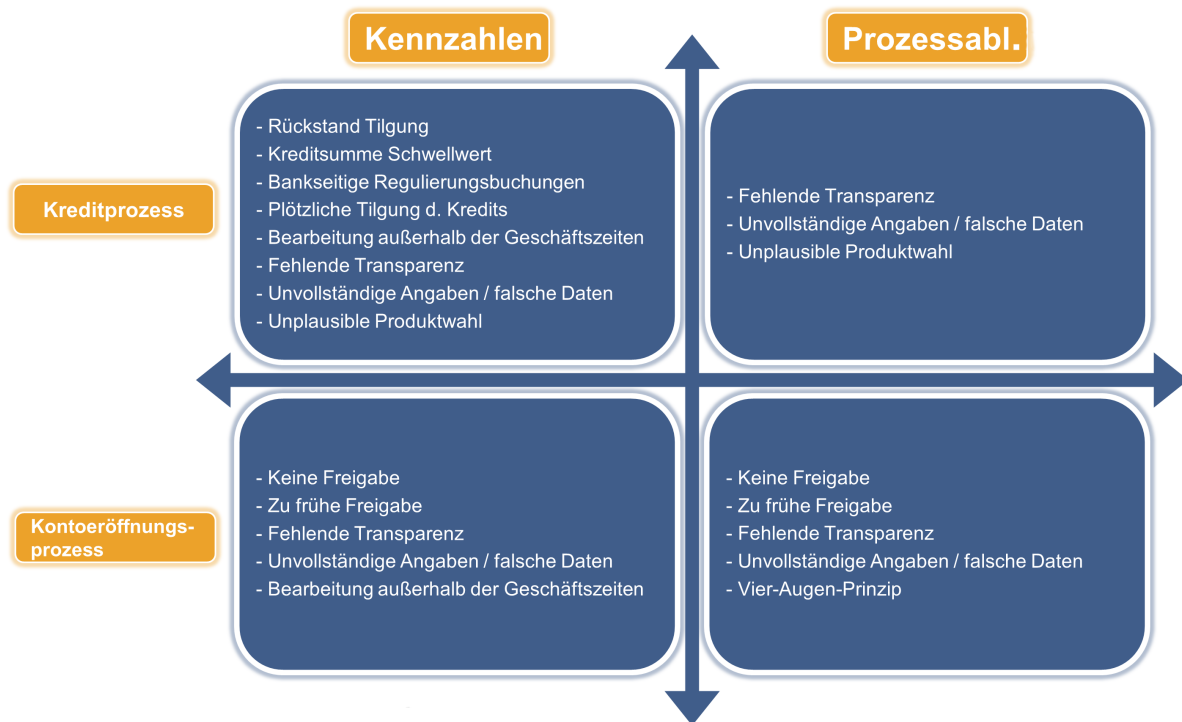
Bearbeitung außerhalb der Geschäftszeit

Vor allem die auffällige Bearbeitung eines Kontoeröffnungsprozesses außerhalb der Geschäftszeiten eines Finanzdienstleisters gibt Hinweise auf dolose Handlungen. So ist anzunehmen, dass Intransparenz und umgangene Sicherungsmechanismen, wie das Vier-Augen-Prinzip - das Prüfen einer Geschäftstätigkeit eines Mitarbeiters durch eine zweite Person - ausgenutzt werden sollen, um betrügerische Handlungen entweder des Kunden oder des Mitarbeiters zu ermöglichen.

Unplausible Produktwahl

Für Mitarbeiter ist die Neukundengewinnung oft in der zu erreichenden Anzahl vorgegeben. Wird diese Anzahl nicht erreicht, drohen entweder negative Konsequenzen, also beispielsweise das Ausbleiben von Boni oder flexiblen Gehaltszahlungen, oder der Karriereaufstieg wird vom Mitarbeiter als für gefährdet empfunden. Daher ist das Bestreben oft, eine hohe Neukundengewinnung aufweisen zu können. Nach Definition eines Red Flags in 2, stellt die Neukundengewinnung entgegen der Interessen der Kunden Fraud dar. Beispielhaft könnte dies so vollzogen werden, dass einer alten Frau, dem Alter ihrer Lebenserwartung nahe, Produkte mit Renditen langer Laufzeiten verkauft werden. In diesem Fall liegt der Verdacht nahe, dass der Mitarbeiter betrügerisch oder dolos gehandelt hat.

¹⁸[Jo14a] S. 4

Abbildung 2.3.: Kategorisierte Red Flags ¹⁹

Anforderungen an die Kategorien sind, möglichst alle Red Flags abdecken zu können. Gleichsam soll die Anzahl an verschiedenen Kategorien gering gehalten werden - wie am Beispiel der zwei Kategorien aufgezeigt. Ferner sollen vorhandene Analysen die Kategorien bearbeiten können, sodass nicht mehr die einzelnen Red Flags auf Analysen untersucht, sondern geeignete Analysen pro Kategorie festgelegt werden können. Es soll also von den einzelnen Red Flags abstrahiert werden, um Aufwand und damit Kosten einzusparen. Folgende Kategorien haben sich dabei herausgestellt:

Kennzahl, Kreditprozess Die Red Flags *Fehlende Transparenz*, *falsche Daten* und *unplausible Produktwahl* lassen sich nur schwerlich automatisiert oder mathematisch prüfen. Sie sind je Fall subjektiv durch Menschen zu prüfen aufgrund dessen, dass deren Maß nicht in feste Werte und feste Routinen überführt werden können. Anders sieht dies für die Red Flags *Bankseitige Regulierungsbuchungen* und *Plötzliche Tilgung des Kredits* aus, welche sich rasch in bedenklich oder unbedenklich unterteilen lassen. Mit wenig Aufwand algorithmisch prüfen lassen sich auch die Red Flags *Bearbeitung außerhalb der Geschäftszeiten*, *Rückstand Tilgung* und *Kreditsumme Schwellwert*, da vor allem die Frage nach dem Maß der Regelverletzung im Fokus steht.

¹⁹Quelle: Vgl. [MJ14] S. 14

Kennzahl, Kontoeröffnung *Fehlende Transparenz, falsche Daten* und *unplausible Produktwahl* sind subjektiver Natur und lassen sich nur schlecht mathematisch erfassen.

Prozessablauf, Kreditprozess Auch in dieser Kategorie lassen sich die Red Flags *Unvollständige Angaben, falsche Daten* und *fehlende Transparenz* nur schwer mathematisch prüfen. Die Red Flags *keine Freigabe, zu frühe Freigabe* und *Bearbeitung außerhalb d. Geschäftszeiten* müssen daraufhin geprüft werden, wie sehr - zu welchem Grade - eine Regelverletzung vorliegt. Mehrfach kategorisiert sind einige Red Flags deshalb, weil sie sowohl Prozessabläufe repräsentieren, als auch Kennzahlen darstellen. Ein Beispiel dafür ist die *zu frühe Freigabe*, die in Sachen Zeitpunkt eine Kennzahl, in Sachen Aufgaben zeitgleich ein Prozess ist.

Prozessablauf, Kontoeröffnung Die Red Flags unterteilen sich ebenso wie in der Kategorie der Kennzahlen im Kreditprozess. Einzig die *Verletzung des Vier-Augen-Prinzips* erweitert die Kategorie und lässt sich nur durch zusätzliche Maßnahmen algorithmisch prüfen.

2.4. Forschungsstand

Um den Überblick darüber zu erhalten, welche Ausführungen und Aspekte dieser Thesis in der Forschung bereits behandelt wurden und welche neu erschlossen werden, wird nachfolgend der aktuelle Stand der Forschung in den Fachwerken referenziert. Außerdem wird kurz ausgeführt, wie die Fachwerke die Kernaspekte dieser Thesis betreffen.

SINGLETON, Tommie W., and Aaron J. Singleton. *Fraud Auditing and Forensic Accounting*. Auflage: 4. Wiley, 2010.

Methodische Betrugsanalyse von Grenzbeträgen anhand des Benfordschen Gesetz.

BRANCIK, Kenneth C. *Inside Computer Fraud: An In-Depth Framework for Detecting and Defending Against Insider IT Attacks*. Boca Raton: Auerbach Pubn, 2007.

Maschinelle Betrugsanalyse im Bereich der Informatik auf Basis eines vorgestellten Frameworks.

NIGRINI, Mark, and Joseph T. Wells. *Benford's Law: Applications for Forensic Accounting, Auditing, and Fraud Detection*. Auflage: 1. Wiley, 2012.

Praktische Erarbeitung des Benfordschen Gesetz.

CINGOLANI, Pable; Alcalá-Fdez, Jesús. *jFuzzyLogic: A Java Library to Design Fuzzy Logic Controllers According to the Standard for Fuzzy Control Programming*. Zeitschriftenartikel, 2013

MARIÉ, Sylvain. *An Octave Extension for RapidMiner*. Zeitschriftenartikel. 2009. Abhandlung über technologische Möglichkeiten, sowie die Anbindung der Software zur numerischen Lösung mathematischer Probleme mit Namen Octave.

2.5. Hypothesenbildung und Forschungsfrage

Der in Abschnitt 2.4 aufgeführte Stand der Forschung stellt die Grundlage dar für das Ableiten von Hypothesen, die in dieser Bachelorthesis einer genaueren Betrachtung unterzogen werden sollen.

1. **Hypothese** *Prozentuale Grenzen sind für eine Grenzbetragsanalyse ungeeignet.*
2. **Hypothese** *Mathematische Gesetzmäßigkeiten können aussagekräftige Hinweise auf Manipulationen von Finanzdatenbeständen liefern.*
3. **Hypothese** *Softwaretechnologische Analysemethoden in Form der Fuzzylogik erlauben differenzierte Aussagen zum Grade eines Warnhinweises auf Grenzbetragsverletzungen in höherer Qualität, als es die klassische Aussagenlogik ermöglicht.*
4. **Hypothese** *Statistische Auswertungen zu Abweichungen eines Zahlenwertes in einer Grundgesamtheit stellen eine qualifizierte Aussage zu einem möglichen Betrugsfall dar.*
5. **Hypothese** *Einzelne Methoden sind ohne Verknüpfung zu weiteren Analysen wenig aussagekräftig, um qualifizierte Warnhinweise aussprechen zu können.²⁰*

Dies führt zu folgender Forschungsfrage:

Führt das Aggregieren multipler Aussagen zum Betrugspotenzial und Abbildung auf ein Bewertungsmodell zu einer qualifizierteren Aussage über Fraud-Warnhinweise, als die Betrachtung der einzelnen Ergebnisse?

²⁰Unverändert entnommen aus [Jo14b] S. 5

3. Betrugsszenarien

Nach der theoretischen Aufbereitung möglicher Analysemethoden, werden in diesem Kapitel ausgewählte exemplarische Betrugsfälle aus der Wirtschaft aufgeführt. Alle haben sich real ereignet und sind daher praxisnahe Motive für die theoretische Erschließung des Forschungsgegenstands.

Auf Basis der hier aufgeführten Beispiele, wird im Anschluss in Kapitel 5 aufgezeigt, wie solche Betrugsfälle durch die Grenzbetragsanalyse hätten detektiert werden-, und vor allem in Zukunft wirksam unterdrückt werden können.

Die hier vorgestellten Szenarien stellen nur einen kleinen Ausschnitt denkbarer Fraud-Arten dar. In der Realität finden sich mannigfaltige Vorgehensweisen von Betrügern, die den Rahmen einer Bachelorthesis übersteigen würden.

Es kann nicht bewiesen werden, dass Daten frei von Betrug sind. Lediglich lässt sich aussagen, dass die berücksichtigten Red Flags nicht nach vorhandenen Prüfmethode aufgetreten sind, aber nicht, dass es keinen weiteren Betrug im Datensatz aufzufinden gibt.

3.1. Betrug durch Mitarbeiter

Das erste Szenario beschäftigt sich mit Fraud oder Betrug durch die Mitarbeiter eines Finanzinstituts. Vor allem der Bereich der internen Revision hat derlei Fälle zum Fokus.

3.1.1. Betrug durch Mitarbeiter: Der „interne Feind“

Zwar ist als konservativer Finanzdienstleister immer davon auszugehen, dass Fraud durch Kunden betrieben wird. Versicherungen versuchen daher, gezielt wirtschaftlichen Betrug durch Kunden aufzudecken und vorzubeugen.

Allerdings geht von den Mitarbeitern eines Finanzdienstleisters ein nicht zu unterschätzendes Gefahrenpotential aus. Da die internen Prozesse von Mitarbeitern, auch mit deren Schlupflöchern und Schwächen, gekannt werden, haben die Mitarbeiter oft und über lange Zeit die Gelegenheit, kreative Wege des Frauds zu finden. Oft sind dolose Handlungen wohl überlegt und sorgfältig geplant, sodass Datenauffälligkeiten weniger wahrscheinlich zutage treten, als bei einem Betrug im Affekt.

Die Ursachen für Fraud durch Mitarbeiter sind vielfältig, da das Spektrum an Motiven breit ist. Vor allem große Betrugsfälle vergangener Jahre zeigten, dass oft die betriebsinterne Vorgabe zur Erreichung von anspruchsvollen Zielen und interner Vertriebsdruck Mitarbeiter zum Betrug verleiten können. Auch das Vorhandensein von Prozesslücken und eher sorglose Handhabung von Kontrollstrukturen, können Fraud begünstigen.

Nebst den Möglichkeiten, alle Mitarbeiter zu sensibilisieren, muss sich der Finanzdienstleister generell im Klaren darüber sein, dass dolose Handlungen auch intern ausgeübt

werden und daher ein großes Schadenspotential darstellen. Somit sind Maßnahmen zur Fraud-Prävention auch nach innen zu richten.²¹

3.1.2. Entdeckung gefälschter Zahlen durch Carslaw, 1988

In seinem Werk *Anomalies in Income Numbers: Evidence of Goal Oriented Behavior*²² begründete 1988 der Mathematiker Charles Carslaw die Betrugsanalyse durch ein Phänomen, genannt das Benfordsche Gesetz, welches in Kapitel 4.5 genauer dargestellt wird.

Bei dem Phänomen handelt es sich um ein Gesetz, welches besagt, dass das Vorkommen der Zahlen 0 bis 9, als erste Ziffer einer Zahl, festen Verteilungen folgen. So kommt die Zahl 1 als erste Ziffer einer Zahl zu etwa 30%, die Zahl 2 zu 17,6% und die Zahl 3 zu 12,5% vor. Es gilt das Gesetz der großen Zahlen, das besagt, dass die relative Häufigkeit mit steigender Anzahl an Ausführungen eines Zufallsexperiment akkurater wird. Erkannt wurde es durch den Astronom und Mathematiker Simon Newcomb und später von Frank Benford erneut publiziert.²³

Zwar mag diese Verteilung auf den ersten Blick zu pauschal zu sein, um ein echtes Gesetz, also anwendbar auf fast alle empirischen Daten, zu sein. Doch bestätigen verschiedene Untersuchungen, dass das Gesetz generell bestand hat. Es trifft vor allem auf natürlich vorkommende Zahlen zu, wie der Einwohnerzahl von Städten, Länge von Flüssen oder Finanzzahlen. Der Begriff *Gesetz* ist dabei nicht akkurat, da das Benfordsche Gesetz einem Fehlerintervall von etwa 11 % unterliegt.²⁴

Vor allem eine empirische Studie Carslaws erstaunte Wissenschaftler. Dabei betrachtete er durch Manager publizierte Nettogewinne. Es zeigte sich, dass vor allem die letzten Stellen der veröffentlichten Zahlen erstaunlich große Häufigkeiten der Ziffer 0 aufwiesen, doch die Ziffer 9 weit unter dem durch das Benfordsche Gesetz erwarteten Wert blieb.

Weitere Untersuchungen ergaben schließlich, dass sich Managern der psychologischen Schwelle eines gebrochenen zu einem glatten Geldbetrag beeinflussen ließen.²⁵

Darstellen lässt sich dies durch die Zahl 19,9 Mio. €, welche von Managern häufig zu 20,0 Mio. € aufgerundet wurde. Dies mag auf den ersten Blick unerheblich erscheinen, doch stellen solche Manipulationen ein erhebliches Schadenspotential dar, da derlei Manipulationen und damit Betrug von vielen Menschen gleichsam ausgeführt werden können.

²¹Vgl. [JLZ13] S. 16

²²Vgl. [Ca88] S. 321-327

²³Vgl. [Be38] S. 551 - 572

²⁴[Te14]

²⁵[Ho08] S. 555

3.1.3. Umgehung der Mitarbeiter-Identifikation, Miami 2002

Eine Compliance-Expertin der PPI AG wurde 2002 Zeugin eines realen Falls eines Betrugs skandals in Miami.²⁶ Eine langjährige Angestellte eines Finanzinstituts genoss eine gute Reputation unter ihren Kollegen und Kolleginnen. Das Vertrauensverhältnis war dementsprechend gestärkt. Die Mitarbeiterin war bekannt für ihre Arbeitszeiten bis spät in den Abend hinein, aufgrund dessen, dass sie durch die Betreuung ihres Kindes an Vormittagen zumeist spät bei der Arbeit erschien. Ihren Kollegen teilte die Frau des Öfteren mit, ihre Identifikationskarte verlegt zu haben, und erbat das Ausleihen der Karten ihrer Kollegen.

Die Identifikationskarten waren eine elektronische Umsetzung des Vier-Augen-Prinzips. Dabei soll eine gegenseitige Kontrolle etabliert werden, bei welcher jede sensible Transaktion eines Angestellten durch einen zweiten Angestellten kontrolliert wird. So soll Betrug vermieden werden (Fraud Prevention, Abschnitt 2.1).

Durch die bereitwillige Verleihung der Identifikationskarten an diese Mitarbeiterin, war es der Mitarbeiterin möglich, spät am Abend, als alle anderen Kollegen bereits im Feierabend waren, dolose Transaktionen auf das eigene Konto zu tätigen. Dabei nutzte sie ihre eigene Identifikationskarte und die Identifikationskarte eines Kollegen, und konnte so das Vier-Augen-Prinzip umgangen.

Auffallen ist der Betrug letztlich durch ungewöhnlich hohe Ausgaben der Mitarbeiterin im Privatbereich, der ihren Vermögensverhältnissen nicht zu entsprechen schienen. Letztlich erbeutete sie mehr als 80.000 US-Dollar.²⁷

Betrachtet man die Schwachstellen des Systems, fällt einem besonders ins Auge, dass die Mitarbeiterin ihre sozialen Kompetenzen ausgenutzt hat, um sich die Identifikationskarten anderer Mitarbeiter zu beschaffen. Dadurch ist das sonst sichere Vier-Augen-Prinzip ausgehebelt worden. Wenn Mitarbeiter nicht dafür sensibilisiert werden, Kontrollstrukturen einzuhalten, und etwa eine Identifikationskarte ausschließlich zur eigenen Authentifizierung zu nutzen, ergeben sich oft hohe Risiken eines - eventuell unbemerkten - Betrugs.

Auffällig war jedoch weiter, dass die Mitarbeiterin weit außerhalb der üblichen Geschäftszeiten agierte. Dies hätte ein guter Anhaltspunkt für eine Untersuchung sein können.

3.2. Bilanzfälschung

Nachfolgend werden wirtschaftskriminelle Szenarien dargestellt, welche den Gegenstand einer Bilanz zum Fokus haben. Zu Beginn wird der Begriff Bilanzdelikt erläutert, um anschließend reale Fälle der Bilanzfälschung aufzuzeigen.

²⁶[Ca14]

²⁷[Ca14]

3.2.1. Der Begriff Bilanzdelikt

Nachfolgend werden die wichtigsten Begriffe des Gegenstandes Bilanzdelikt genauer erläutert.

Wirtschaftskriminalität

Die das Bilanzrecht betreffenden Wirtschaftsdelikte sind der Wirtschaftskriminalität zuzuordnen. Von Wirtschaftskriminalität wird gesprochen, wenn ein Verstoß gegen die Rechtsnorm vorliegt, etwa bei einem Vertrauensmissbrauch, bei dem das Opfer den Täter nicht als solchen kennt und durch diesen monetär geschädigt wird und bei welchem der Täter das Unwissen des Opfers billigend ausnutzt.²⁸

Bilanzrecht

Das Deutsche Bilanzrecht gibt Richtlinien zur Erstellung einer Bilanz vor. Diese gegen Richtlinien verstoßende Falschaussagen in einer Bilanz können bewusst, also absichtlich, oder unbewusst, also aus Unwissen oder fachlichem Unvermögen, getätigt werden. Unbewusste Falschaussagen, dazu gehören auch unbeabsichtigte Fehler im Rechnungswesen, werden nicht den Wirtschaftsdelikten zugeordnet - sind vielmehr Menschlicher Makel -, während bewusste Falschaussagen das Bilanzrecht betreffende Wirtschaftsdelikte darstellen und damit zur Wirtschaftskriminalität zählen. Somit gilt es dies bei einem Aufkommen zu untersuchen. Dies wird dann Bilanzdelikt genannt.²⁹

3.2.2. Das Schneeballsystem Jürgen Schneiders, 1994

Ein Schneeballsystem nutzte Jürgen Schneider in den 90er Jahren des vergangenen Jahrhunderts, um sich betrügerisch zu bereichern. Schneider war ein Bauunternehmer, welcher sich darauf spezialisierte, Altbauten in Innenstadtlagen zu sanieren, und sie zu Luxushäusern auszubauen.

Von Banken erfragte Schneider weit höhere Darlehen, als die Projekte eigentlich benötigten. Die Differenz daraus verbuchte er auf Privatkonten. Um hohe Zins- und Darlehenskosten zahlen zu können, nahm Schneider bei unterschiedlichen Geldhäusern immer wieder neue Darlehen auf, und zahlte damit alte Schulden.

Erst Verbindlichkeiten von etwa 5,6 Mrd. DM ließen die Dresdner Bank Betrug vermuten, und die Geschäfte prüfen. Schneider setzte sich ins Ausland ab.

²⁸Vgl. [PH05] S. 19ff

²⁹Vgl. [PH05] S. 21

4. Mathematische Methoden

In diesem Abschnitt erfolgt die Vorstellung ausgewählter mathematischer Analysemethoden, welche geeignet sind, eine Grenzbetragsanalyse durchzuführen. Namentlich sind dies prozentuale Grenzen, die unscharfe Logik, Stochastik und das Benfordsche Gesetz.

4.1. Prozentuale Grenzen

Bei prozentualen Grenzen handelt es sich um die Angabe eines Relativwertes zu einem Limit oder Grenzbetrag, ab welchem zuvor festgelegte Aussagen getroffen werden.

So wird von den Unternehmen der Grenzbetrag in seiner Höhe identifiziert, und anschließend ein maximaler Prozentsatz festgelegt. Wenn diese Prozentgrenze zum Grenzbetrag erreicht wird, erfolgen Warnhinweise. Zur Veranschaulichung wird angenommen, dass ein Grenzbetrag bei 20.000€ festgelegt wird. Anschließend werden 99,5% als relative Grenze festgelegt. Somit wird ab einem Wert von 19.900€ eine genauere Prüfung vollzogen.

Dieses Vorgehen ist problematisch. Die Gefahr von Fraud geht häufig von den Mitarbeitern eines Finanzinstituts aus, wie in Abschnitt 3.1 beschrieben wird. Die Schwachstelle besteht also darin, dass Mitarbeiter interne Bestimmungen nach außen kommunizieren, oder diese für dolose Handlungen nutzen. Wenn eine *harte Grenze* oder eine *prozentuale Grenze*, also ein fester Wert, zu einem identifizierten Grenzbetrag festgelegt wird, dann wurde damit nichts anderes erreicht, als eine neue Grenze zu erzeugen, wodurch Betrüger unter dieser Grenze bleiben. Ferner kann diese selbstredend ebenso ausgenutzt oder nach außen kommuniziert werden.

Daher sind solche *harten-* oder *prozentuale Grenzen* als Prüfmittel ungeeignet, werden jedoch in der Wirtschaft aus Einfachheitsgründen oft genutzt.³⁰ Besser geeignet ist die Fuzzylogik.

4.2. Unscharfe Logik

In diesem Abschnitt wird die unscharfe Logik, auch Fuzzylogik genannt, beleuchtet. Dabei wird zuerst der Begriff *Fuzzylogik* erläutert, um anschließend die mathematischen Grundlagen aufzuzeigen. Letztlich schließt der Abschnitt mit der Erklärung einer möglichen maschinellen Nutzung der unscharfen Logik.

4.2.1. Der Begriff und Motivation der unscharfen Logik

Die Fuzzylogik ist eine Theorie in Form der klassischen Logik, bei welcher Grade oder Zwischenwerte der Wahrheitswerte *wahr* und *falsch* bestimmt werden, um realitätsnahe Aussagen behandeln zu können.³¹

³⁰[Ca14]

³¹Vgl. [Kr13] S. 1 ff.

Oft werden nicht nur die Wahrheitswerte der klassischen Logik *wahr* und *falsch* benötigt, sondern etwa auch Werte wie *überwiegend wahr*, *überwiegend falsch* oder *wenig wahr* und *wenig falsch*. Es wären also Zwischenwerte wünschenswert, die eher einen Wahrheitsgrad aufweisen als einen festen Wahrheitswert. Dies ist der grundlegende Gedanke der unscharfen Logik. Sie kennt Zwischenwerte und -grade.³²

Logik ist die Disziplin, die Wahrheitswerte von Aussagen zu bestimmen.³³ Die Logik gliedert sich in die Unterdisziplinen *Aussagenlogik*, *Prädikatenlogik erster Stufe* und *Prädikatenlogik zweiter Stufe*. In der Aussagenlogik werden Elementaraussagen mit so genannten Junktoren, aussagenlogischen Verknüpfungen, verbunden, und, auf Basis von Axiomen, durch Auflösung, ein Wahrheitswert - entweder *wahr* oder *falsch* - ermittelt. Die Aussagenlogik reicht noch wesentlich weiter als dies, doch reichen diese Grundlagen für das grobe Verständnis im Kontext der unscharfen Logik oder Fuzzylogik aus.

4.2.2. Fuzzy-Mengen

Bei der Fuzzylogik handelt es sich um ein System der mehrwertigen Aussagenlogik. Bei solchen Systemen existieren mehr Wahrheitswerte als die der klassischen Logik.

Bei normalen Mengen, ist eine Menge definiert durch die in ihr enthaltenen Objekte. Objekte sind also in einer Menge enthalten oder sie sind es nicht.

$Z = \{10, 11, 12, \dots, 97, 98, 99\} \in \mathbb{N}$ ist ein Beispiel einer üblichen Menge. Sie enthält in diesem Beispiel nur zweistellige Zahlen der Zahlenmenge \mathbb{N} . Es kann auch ein Prädikat zur Beschreibung genutzt werden: $Z = \{n \in \mathbb{N} | \text{zweistellig}(n)\}$.

Fuzzy-Mengen (engl. Fuzzy-Sets) sind davon klar abzugrenzen dahingehend, dass die Fuzzy-Mengen beschreiben, welche Art von Objekten sich in ihnen befinden sollen. Die Objekte selber erhalten daraufhin einen Zugehörigkeitsgrad. Sie gehören also *zu einem bestimmten Grade* einer Fuzzy-Menge an.³⁴

Um zu bestimmen, ob und zu welchem Grad ein Objekt zu einer Fuzzy-Menge gehört, werden Zugehörigkeitsfunktionen definiert. Dabei werden Objekte der Fuzzy-Menge auf das so genannte Einheitsintervall $[0, 1]$ abgebildet. Genau definiert wird eine Fuzzy-Menge folgendermaßen:

„Eine Fuzzy-Menge μ von X ist eine Funktion von der Referenzmenge X in das Einheitsintervall, d.h.

$$\mu : X \rightarrow [0, 1].$$

$F(X)$ bezeichne die Menge aller Fuzzy-Mengen von X .³⁵

³²Vgl. [Kr13] S. 2 f.

³³Vgl. [Zi07] S. 43 ff.

³⁴Vgl. [Kr13] S. 8 ff.

³⁵[Kr13] S. 10

Visualisiert werden kann dies an der Abbildung 4.1, bei welcher die beiden Dreiecks-Kurven Fuzzy-Mengen darstellen, die durch Zugehörigkeitsfunktionen beschrieben wurden. Die linke, rot markierte Dreiecks-Fläche stellt den Bereich dar, welcher von zwei Zugehörigkeitsfunktionen beschrieben wird. Elemente dieser Fläche gehören zu einem gewissen Grade der einen- und zu einem gewissen Grade der anderen Fuzzy-Menge an.

4.2.3. Zugehörigkeitsfunktionen

Um zu bestimmen, welche Objekte zu welchem Grade welcher Fuzzy-Menge zugehörig sind, bedarf es der genaueren Definition der Funktion. Es wird eine *Zugehörigkeitsfunktion* benötigt.

Bei nur wenigen möglichen Zustandsobjekten, lässt sich eine Aufzählung nutzen, um die Zugehörigkeitsfunktion zu beschreiben. Als bald die Menge jedoch sehr groß wird, empfiehlt sich eine allgemeinere Darstellung der Zugehörigkeitsfunktion durch Definition eines Funktionsterms mit Parametern.

Die Bezeichnung *warm* ist für die klassische Logik problematisch, da unterschiedlich aufgefasst wird, was *warm*. So können 18°C, je nach Auslegung, als *etwas warm* angesehen werden. Als monoton nicht fallende Funktion ließe sich dies exemplarisch folgendermaßen darstellen:

$$\mu_{a,b}(x) = \begin{cases} 0 & \text{falls } x \leq a \\ \frac{x-a}{b-a} & \text{falls } a \leq x \leq b \\ 1 & \text{falls } x \geq b \end{cases}$$

Bildlich veranschaulichen lässt sich eine Zugehörigkeitsfunktion an der Abbildung 4.1, durch den Verlauf der rot gefärbten Fläche - genauer den Inhalt unter der Kurve. Elemente werden gemäß dem Grad des Zutreffens einer Fläche zugeordnet, wobei sie auch unscharf zu mehreren Flächen gehören können, welches der rot markierten Dreiecksfläche entspräche.

4.2.4. Logische Operationen auf Fuzzy-Mengen

Wenn nun mehrere Fuzzy-Mengen definiert sind, lassen sich darauf die Operationen *Vereinigung* \cup und *Durchschnitt* \cap anwenden.

Seien die Fuzzy-Mengen A und B Teilmengen der Fuzzy-Menge X. \cap und \cup der Fuzzy-Mengen A und B sind definiert als:

$$(A \cup B)(x) = \max(A(x), B(x)); x \in X \text{ und}$$

$$(A \cap B)(x) = \min(A(x), B(x)); x \in X$$

Generell sind Fuzzy-Mengen mathematisch behandelbar wie übliche Mengen, weshalb die Operationen AND, OR und NOT ebenso anwendbar sind.

Einfacher ausgedrückt bedeutet dies, dass der Wert der Funktion $(\mu \cap \mu')(x)$ entweder $\mu(x)$ oder $\mu'(x)$ ist, je nachdem, welcher Funktionswert für das x größer ist. Weiterhin bleibt die Voraussetzung, dass μ und μ' Fuzzy-Mengen sind. Die mathematischen Eigenschaften der Junktoren befinden sich in Anhang E.1.

Anstelle der Operatoren *Vereinigung* \cup und *Durchschnitt* \cap , werden in der Fuzzylogik die Operatoren *t-norm* \top und *t-conorm* \perp genutzt.

An dieser Stelle reicht es aus, die Grundkonzepte des *min* und *max* der Fuzzy-Operationen zu verstehen, und zu wissen, dass die meisten Fuzzy-Berechnungen nur auf Basis dieser beiden Operationen vollzogen werden. Grafisch betrachtet lässt sich in Abbildung 4.1 erkennen, dass sich verschiedene t-norm-Operationen wie AND oder MAX auf Schnittflächen der Fuzzy-Mengen beziehen.

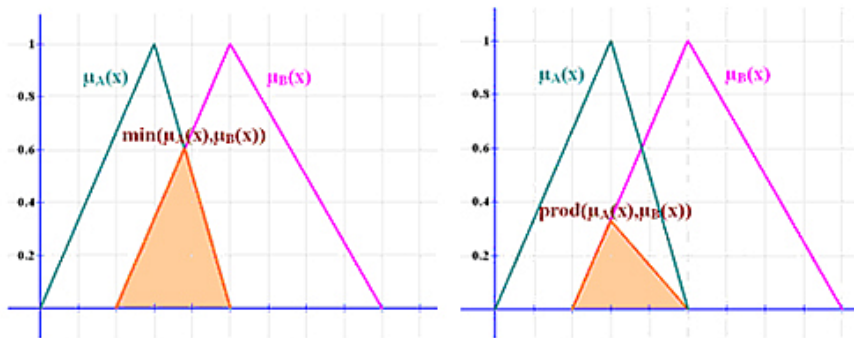


Abbildung 4.1.: T-Norm „AND“ und „MAX“³⁶

4.2.5. Fuzzifizierung, Defuzzifizierung

Wie bereits zuvor aufgeführt, sind Fuzzy-Mengen dadurch bestimmt, dass Elemente, nicht wie bei klassischen Mengen, entweder dazu gehören oder nicht, sondern zu einem gewissen Grade zu einer Fuzzy-Menge gehören. Dies wird bestimmt durch eine Zugehörigkeitsfunktion. Daher gibt es bei mehreren zutreffenden Zugehörigkeitsfunktionen, grafisch dargestellt, Überschneidungen der Fuzzy-Mengen. Dies sind in dem Fall Häufungspunkte der Mengen, und daher stärker gewichtet. Dies ist, grob ausgedrückt, der Ansatz der Fuzzylogik. Für einen in die Funktion gegebenen Wert, lässt sich danach eine Aussage im Intervall $[0, 1]$ darüber treffen, bis zu welchem Grade der Wert wahr oder falsch ist.

Für die Praxis sind dabei die Ausdrücke *Fuzzifizierung* und *Defuzzifizierung* von Bedeutung. *Fuzzifizierung* ist dabei die Überführung in Zugehörigkeitsgrade, zur Bestimmung

³⁶[Dm14]

eines Fuzzy-Wertes. Die *Defuzzifizierung* ist dann die Aussage im Intervall $[0, 1]$, welchem Wahrheitsgrad dies entspricht. Die Defuzzifizierung geschieht durch Summierung aller Ausgabeflächen. Mit diesem Wert - exemplarisch 0.3 -, ließe sich dann weiter rechnen. Die Ermittlung eines Ergebniswertes wird zur Steuerung benötigt. Grafisch dargestellt ist das Prinzip im Anhang D.2.

4.3. Stochastik

In diesem Kapitel wird zum einen der Begriff der Stochastik genauer betrachtet, ferner zudem das Gebiet der beschreibenden Statistik genauer erklärt. Dabei liegt der Fokus auf den mathematischen Grundzügen, mit einer kurzen Betrachtung der maschinellen Umsetzbarkeit der mathematischen Formeln.

4.3.1. Abgrenzung relevanter Stochastik

Die Stochastik umfasst die Unterkategorien der Wahrscheinlichkeitstheorie und der Statistik. Die Wahrscheinlichkeitstheorie ist am Rande relevant für die Untersuchung eines Grenzbetrags. Von weit größerer Bedeutung ist jedoch die Statistik. Sie kommt im Zuge der Prüfung auf Betrug weit häufiger zum Einsatz. Dies ist damit begründet, dass sie leicht anzuwenden ist, bei gleichzeitig verständlichen und etablierten Ergebnissen und Aussagen.

Statistik wird definiert als „Entwicklung und Anwendung von Methoden zur Erhebung, Aufbereitung, Analyse und Interpretation von Daten“.³⁷ Es wird angenommen, dass Datensätze bereits aufbereitet vorliegen.

Von einem breiten Spektrum der Stochastik, werden exemplarisch in dieser Thesis das Histogramm, der Mittelwert mit Median, arithmetischen- und geometrischem Mittel, die Varianz und Standardabweichung einer genaueren Betrachtung unterzogen.

Vor allem wenn Häufigkeitsverteilungen in Klassen angegeben werden können, eignet sich das Histogramm besonders zur Darstellung. In einem kartesischen Koordinatensystem stellt die Abszisse die Klassen dar, die Ordinate trägt die Klassenhäufigkeit.

4.3.2. Mittelwerte und Streuungsmaß

Ein Mittelwert hat die Aufgabe, das Zentrum einer Häufigkeitsverteilung als alleinstehenden Wert zu bestimmen. Dabei ergeben sich in der beschreibenden Statistik verschiedene Mittelwerte mit verschiedenen Aussagen.

Ein simpler Mittelwert ist der *Median*. Sei (x_1, x_2, \dots, x_n) eine geordnete Stichprobe. Dann ist der Median \tilde{x} ³⁸:

$$f(n) = \begin{cases} x_{\frac{n+1}{2}} & n \text{ ungerade} \\ \frac{1}{2} (x_{\frac{n}{2}} + x_{\frac{n}{2} + 1}) & n \text{ gerade.} \end{cases}$$

³⁷[Bo12] S. 1

³⁸Vgl. [Bo12] S. 72ff

Das *arithmetische Mittel*, bezeichnet mit \bar{x} , stellt den allgemeingebäuchlichen Durchschnitt dar. Der Unterschied zum Median ist, dass der Median den mittleren Wert einer Menge darstellt, das arithmetische Mittel jedoch alle Merkmalswerte mit den Merkmalsträgern multipliziert, die Produkte summiert, und diese durch die Anzahl der Werte dividiert.

Sei x ein Merkmalswert, i eine Zählvariable und n die Anzahl aller Werte. Dann berechnet sich \bar{x} folgendermaßen³⁹:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Das arithmetische Mittel der beschreibenden Statistik wird in der Wirtschaft oft genutzt, lässt sich jedoch durch Ausreißer schnell zu falschen Aussagen verändern. So kann sich bei einer homogenen Menge der Wert des arithmetischen Mittels durch einen einzigen Ausreißer stark verändern und somit kaum mehr eine Aussage zur Grundgesamtheit erlauben. Jede Bewertung des arithmetischen Mittels ohne Berücksichtigung dieser Eigenschaft könnte zu falschen Interpretationen der Grundgesamtheit führen. Somit ist das arithmetische Mittel kritisch zu betrachten.

Sowohl der Median, als auch das arithmetische Mittel weisen die Schwäche auf, Ausreißer entweder wenig darzustellen oder gar durch solche in der Gesamtaussage verfälscht zu werden.

Ein weiteres wesentliches Merkmal einer Grundgesamtheit von Daten, ist der Grad der *Streuung*. Er beschreibt, wie konzentriert die Merkmalsträger beisammen liegen. So steht eine in Relation zum Mittelwert hohe Streuung dafür, dass viele Werte vom Mittelwert abweichen, sich *streuen*.

Die Spannweite ist der Abstand zwischen dem größten und dem kleinsten Wert mit $R = x_{[n]} - x_{[a]}$, wobei n die Menge der Merkmalsträger ist, a der Index des Minimums und die Menge aller x geordnet. Quartile teilen die Gesamtmenge in mehrere Abschnitte. Üblich ist dabei die Unterteilung in drei Quartile und somit in vier Viertel. Der zentrale Quartilsabstand (ZQA) berechnet sich mit $ZQA = Q_3 - Q_1$. Da die oberen und unteren 25% aller Werte unbeachtet bleiben, bietet dieses Streuungsmaß den Vorteil, das Vorhandensein von Ausreißern zu beschreiben.⁴⁰

4.3.3. Erwartungswert, Varianz und Standardabweichung

Mit μ (μ) wird in der Statistik der Erwartungswert bezeichnet. Er beschreibt die Zahl, die eine Zufallsvariable im Mittel annimmt. Von Bedeutung ist der Erwartungswert insbesondere für den Chi-Quadrat-Test (χ^2 -Test) und der Bestimmung von Wahrscheinlichkeiten der deskriptiven Statistik.⁴¹

³⁹Vgl. [Bo12] S. 79

⁴⁰Vgl. [Bo12] S. 92

⁴¹Vgl. [Bo12] S. 98 ff.

Mit sigma^2 (σ^2) wird die Varianz, mit sigma (σ) die Standardabweichung bezeichnet. Bei der Varianz handelt es sich um die Summe aller Quadrate der Entfernungen zum Mittelwert. Mathematisch ausgedrückt, heißt dies⁴²:

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2,$$

wobei n die Anzahl aller Merkmalswerte x ist, \bar{x} das arithmetische Mittel darstellt und i die Laufvariable der Summe.

Eine andere Form der Darstellung ist wie folgt:

Sei A ein diskreter, also zählbarer, Wertebereich. Die Varianz ist:

$$\text{Var}(X) = \sum_{x \in A} (x - \mu)^2 P(X = x), \text{ mit } P(X = x) = \text{Wahrscheinlichkeit, dass } X = x \text{ und } \mu = \sum_{x \in A} x P(X = x) = \text{Erwartungswert von } X.$$

Zwar wird in der Literatur kritisiert, dass die Standardabweichung für die Statistik von geringem Wert ist, da sie Ausreißer unzureichend berücksichtigt, doch wird sie in der Wirtschaft sehr häufig bei Analysen eingesetzt. Dieser Sachverhalt wird auch für die *Normalverteilung* angemerkt. Die mittlere Abweichung sei demnach weit besser geeignet als die Standardabweichung, da diese auch Ausreißer stärker bemisst.⁴³

Die *Normalverteilung* hat als Eingabe das arithmetische Mittel μ und die Standardabweichung σ .

Die Wahrscheinlichkeit für eine $N(\mu; \sigma^2)$ -verteilte Zufallsgröße X mit einem Wert innerhalb eines Intervalls $[\mu - 1\sigma; \mu + 1\sigma]$ liegt bei etwa 68%, für einen Wert innerhalb des 2σ -Intervalls $[\mu - 2\sigma; \mu + 2\sigma]$ bei 95% und Werte innerhalb des 3σ -Intervalls $[\mu - 3\sigma; \mu + 3\sigma]$ liegt bei mehr als 99%, wie es in Abbildung 4.2 dargestellt ist. Mithilfe dieses Wissens ist es möglich, Aussagen zu Wahrscheinlichkeiten eines Wertes innerhalb einer normalverteilten Menge zu treffen.

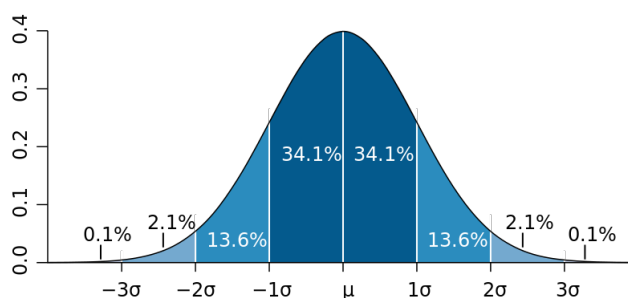


Abbildung 4.2.: Wahrscheinlichkeiten eines Wertes innerhalb des $[\mu - x * \sigma; \mu + x * \sigma]$ -Intervalls, wobei x die Anzahl der σ ist ⁴⁴

⁴²[Bo12] S. 97

⁴³[Bo12] S. 95

⁴⁴[Wi15b]

4.3.4. Chi-Quadrat-Verteilung und Chi-Quadrat-Test

Der χ^2 -Test ermöglicht es, die Vereinbarkeit eines vermuteten funktionalen Zusammenhangs einer Zahlenmenge zur tatsächlich anzutreffenden Verteilung zu prüfen. Dafür muss die zu untersuchende Menge jedoch χ^2 -verteilt sein.

Von besonderem Interesse für die Betrugsanalyse ist der χ^2 -Test auch deshalb, weil er als gerichtliches Mittel in einem Urteil des Finanzgerichts Münster anerkannt wurde. Bei einer mathematisch-statistischen Prüfung durch Chi-Quadrat, erwiesen sich die Ergebnisse ähnlich akkurat, wie aufwändig durchgeführte elektronische Analysen, weshalb das Gericht den generellen Gebrauch des Chi-Quadrat-Tests guthieß.⁴⁵

Für die χ^2 -Verteilung ist einzig der Parameter n , der *Freiheitsgrad* notwendig, wobei $n \in \mathbb{N}$. Der Freiheitsgrad ist die Anzahl aller unabhängigen Beobachtungswerte abzüglich der Anzahl geschätzter Parameter.⁴⁶

Um eine Stichprobe zu simulieren und deren Varianz zu prüfen, wird eine Anzahl m von Zufallszahlen Z_i einer Grundgesamtheit gewählt. Anschließend werden diese Zahlen quadriert und die Summe aller Z_i gebildet.

Ferner ist es möglich, ein Toleranzintervall zu bestimmen, bis zu welchen Abweichungen Werte als plausibel angesehen werden können. Mathematisch wird dies folgendermaßen ausgedrückt:

$\chi^2 = \sum_{j=1}^m \frac{(N_j - n_{0j})^2}{n_{0j}}$, wobei N_j die beobachtete Häufigkeit ist und n_{0j} die erwartete Häufigkeit mit $n_{0j} = p_{0j}n$. Dabei ist p_{0j} die Wahrscheinlichkeit dass eine Ausprägung von dem Merkmal X tatsächlich die vermutete Verteilung repräsentiert.

4.4. Ermittlung des Maximums einer Menge

Das Maximum einer Menge ist jenes Element, welches in dessen Wertigkeit allen anderen Elementen der Menge höher gestellt ist.

x ist größtes Element von M : $\iff \forall y \in M : y \leq x$.

Als rekursive Funktion:

$$\max(x_1, x_2, \dots, x_n) = \max(x_1, \max(x_2, \dots, x_n))$$

als mathematische Funktion mit zwei Variablen:

$$\max(x_1, x_2) = \frac{x_1 + x_2 + |x_1 - x_2|}{2}$$

⁴⁵Vgl. ES14

⁴⁶Vgl. [Ma15]

4.4.1. Maschinelle Umsetzbarkeit

Statistische Analysen manuell durchzuführen ist, mühsam. Die Bachelorthesis beschäftigt sich zentral mit der Frage, wie Analysen automatisiert werden können. In dem Szenario der Ausarbeitung eines Java-Programms, ließe sich zu diesen Zwecken die freie Software *GNU Octave* nutzen. Octave zeigt sich dabei ähnlich mächtig wie die kommerziellen Software *MATLAB*[®]. Sie ist als Stand-Alone-Anwendung frei verfü- und kommerziell nutzbar, besitzt jedoch zudem ein Application Programming Interface (API, dt. Programmierschnittstelle), sodass die Bibliothek für Java-Anbindungen genutzt werden kann. Genutzt wird die aufgezeigte Möglichkeit maschineller Anbindung für die Erarbeitung eines Bewertungsmodells in Kapitel 6. Eine Anleitung zu Octave findet sich im Anhang C.

4.5. Benfordsches Gesetz

Nachfolgend wird das Benfordsche Gesetz dargestellt, wonach anschließend die mathematischen Grundlagen geklärt werden. Letztlich schließt der Abschnitt mit der maschinellen Nutzbarkeit des Benfordschen Gesetz ab.

Das Benfordsche Gesetz basiert auf einer numerisch empirischen Gesetzmäßigkeit, die die Verteilung von Ziffern in empirischen Datenmengen mit berechenbaren Wahrscheinlichkeiten verknüpft.⁴⁷ Hintergründe zum Benfordschen Gesetz befinden sich im Anhang B.

4.5.1. Mathematische Darstellung

Im Anhang B wird beschrieben, wie Simon Newcomb beim Nachschlagen von Logarithmentafeln entdeckte, dass die Seiten aller mit der Ziffer 1 beginnenden Zahlen weit abgegriffener waren, als andere Seiten. Erhebungen verschiedener empirischer Daten führten Newcomb zu Erkenntnissen, aus denen sich zwei Formeln herleiten ließen, die die Wahrscheinlichkeit der Häufigkeit einer jeden Ziffer berechnen lassen. Die erste Formel bezieht sich auf die jeweils erste Ziffer einer Zahl, die zweite Formel ist eine allgemeinere Darstellung für alle Zifferpositionen. Die Formel für die Ermittlung der Wahrscheinlichkeit für die erste Ziffer stellt sich folgendermaßen dar:

$$W(D_1) = \log_{10}\left(1 + \frac{1}{D_1}\right); D_1 \in \{1, 2, \dots, 9\}$$

Für die erste Ziffer D_1 einer Zahl würde die obere Formel folgendermaßen aussehen⁴⁸:

$$W(D_1 = 1) = \log_{10}\left(1 + \frac{1}{1}\right) = \log_{10}2 \approx 0,301029995 \approx 30,10\%.$$

Die Allgemeine Formel für alle weiteren Ziffern gestaltet sich folgendermaßen⁴⁹:

⁴⁷Vgl. [Sc10] S. 3

⁴⁸Vgl. [RB07] S. 27

⁴⁹Vgl. [Sc10] S. 4

$W(D_1 = d - 1, \dots, D_k = d_k) = \log_{10} \left[1 + \left(\sum_{i=1}^k d_i \times 10^{k-i} \right)^{-1} \right]$, wobei d_j die Position einer Ziffer innerhalb einer Zahl darstellt.

Angewandt auf die ersten zwei Ziffern einer Zahl ließe sich die obige Formel folgendermaßen anwenden⁵⁰: $W(D_2) = \sum_{D_1=1}^9 \log_{10} \left(1 + \frac{1}{10 \times D_1 + D_2} \right)$; $D_1 \in \{1, 2, \dots, 9\}$ und $D_2 \in \{0, 1, \dots, 9\}$

4.5.2. Bedeutung für die Grenzbetragsanalyse

Die in Abschnitt 4.5.1 aufgezeigten Grundlagen bedeuten im Grunde nichts anderes, als dass bei einer Betrachtung einer empirische Datenmenge die Verteilung der Zahlen an ihren Ziffernpositionen der Tabelle B2 entsprechen sollten.

Wenn Betrüger einen Datensatz manipulieren, um eigene Ziele durchzusetzen und die realen Zahlen zu verschleiern, so werden die meisten Betrüger Zahlen erhofft willkürlich gemäß einer Gleichverteilung wählen, um den Anschein zu erwecken, reale, zufällige Zahlenwerte auszuweisen. Dies ist jedoch ein Trugschluss, da das Benfordsche Gesetz aussagt, dass Ziffern empirischer Zahlen nicht gleichverteilt sind. Abweichungen von der Verteilung aus Tabelle B1 deuten also auf Zahlenmanipulationen hin.

Nachdem Carslaw seine Ergebnisse publizierte, wurde das Benfordsche Gesetz zu einem der Hauptinstrumente in der Erkennung von Betrug durch Manipulationen von Zahlen.⁵¹

4.5.3. Maschinelle Umsetzbarkeit

Die Programme MATLAB und Octave bieten alle stochastischen Operationen, um die Relativverteilung der Ziffern einer gegebenen Datenmenge mit den Vorhersagen des Benfordschen Gesetzes zu vergleichen und durch χ^2 -Test zu ermitteln, ob die Datenmenge mit bestimmten Funktionen vereinbar ist.

In Anhang A ist exemplarisch aufgeführt, wie sich Octave in ein Java-Programm anbinden lässt.

⁵⁰Vgl. [RB07] S. 27

⁵¹Vgl. [Sc10] S. 3

5. Analyse der Szenarien

Nachdem im vorangegangenen Kapitel 4 die notwendigen mathematischen Grundlagen zur Durchführung einer Grenzbetragsanalyse dargestellt wurden, werden die Szenarien von Wirtschaftsbetrug aus Abschnitt 3 mit diesen mathematischen Methoden untersucht.

Dies geschieht in diesem Kapitel am Beispiel selbsterstellter Modelldaten, welche die Betrugsszenarien repräsentieren. Dieses Datenmaterial soll anhand mathematischer Methoden auf Grenzbetragsverletzungen untersucht werden.

Da sich die Vorgehensweise einer Untersuchung im Fall Jürgen Schneider ähnlich gestaltet wie im Fall Miami, wurde die Untersuchung und Bewertung des Betrugsfalls Jürgen Schneider in den Anhang G verschoben.

5.1. Datenbestand

Für die in Abschnitt 4 aufgeführten Analysen stehen keine realen Daten aus der Praxis zur Verfügung. Die Gründe dafür werden nachfolgend aufgeführt.

Das Bundesdatenschutzgesetz (BDSG) hat die Aufgabe die Persönlichkeitsrechte Einzelner vor falschem Umgang mit ihren Daten zu schützen. Auch die unbegrenzte Erhebung oder Speicherung von persönlichen Daten, sowie die Verwendung und Weitergabe eben dieser wird durch das BDSG geregelt.

Geschädigte Finanzinstitute möchten einen Betrug aus Angst vor Reputationsschäden verstecken, haben aber das Interesse daran, dass dies nicht erneut passiert. So kann es sein, dass Kunden das Vertrauen zum Finanzinstitut verlieren oder Anleger ihre Gelder abziehen. Daher werden Schäden lediglich intern oder mit Partnern durch Interesse an Forschung kommuniziert.⁵²

Prävention vor Nachahmern

Geschädigte Finanzinstitute, Fraud-Experten und Wissenschaftler versuchen betrügerische Methoden geheim zu halten, um Nachahmer zu vermeiden. Zwar kann in der Regel davon ausgegangen werden, dass publizierte Betrugsmethoden den Compliance-Experten bekannt sind und in Finanzinstituten auf diese geprüft wird. Dennoch kann die Vorstellung Nachahmer dazu motivieren, sich alternative Betrugsmethoden auszudenken. Daher herrscht im Compliance-Umfeld allgemein eine starke Geheimhaltung aller Informationen. Forschungsergebnisse werden in Teilen publik gemacht, um Exempel zu statuieren und so potentielle zukünftige Täter abzuschrecken.⁵³

⁵²[Ca14]

⁵³Vgl. [Me12] S. 42

Mangelnde Nachprüfbarkeit

Stünden authentische Daten aus operativen Geschäftsfeldern eines Finanzdienstleisters zur Verfügung, so bestünde eine Schwierigkeit darin, zu verifizieren oder falsifizieren, ob die angewandten Analysen zu richtigen Aussagen gelangt sind.

Wird während einer Betrugsanalyse ein potentieller Betrug identifiziert, gilt es im nächsten Schritt, Forensic Experts mit der Überprüfung des Betrugsverdachts zu beauftragen. Forensic Experts sind Spezialisten des Betrugswesens und der forensischen Kriminologie zugeordnet. Ohne das Urteil der Forensic Experts, kann nicht sicher festgestellt werden, ob ein Betrugsverdacht zutrifft oder nicht. Die Untersuchung durch Forensic Experts kostet Finanzdienstleister jedoch viele Finanz- und Zeitressourcen, weshalb diese häufig nur bei deutlichen Indizien auf Betrug eingesetzt werden, um Verdachtsfälle kriminalistisch untersuchen.

Einzig wenn das Zustandekommen eines Datensatzes transparent wäre und unter diese Datensätze manipulierte Datensätze gemischt würden, ließen sich sichere Aussagen über Analyseergebnisse tätigen, so eine Compliance-Expertin.⁵⁴ Das Erstellen von Testdaten gehört im Compliance-Umfeld zur gängigen Praxis.

5.2. Analyse gefälschter Datenmengen nach Carslaw, 1988

Obwohl der Carslaw-Betrug zu einem der am häufigsten referenzierten Fälle der Fachliteratur gehört, sind keine Daten zum Betrug frei zugänglich. Lediglich Umschreibungen dessen, was Carslaw bei seinen Untersuchungen an Auffälligkeiten über die Ergebniszahlen vieler Manager entdeckte, finden sich in der Fachlektüre. Daher kann der Versuch lediglich beschrieben und nachgestellt werden. Beschrieben wird der Versuchsaufbau in Abschnitt 3.1.2. Um den Versuch nachzustellen, wurde eine zufällig ausgewählte, empirische Datenmenge genutzt, in diesem Fall die Einwohnerzahl der Städte Deutschlands im Jahr 1950 in Einheiten zu tausend.⁵⁵

Diese Daten werden für diese Thesis erfasst und die Häufigkeit der ersten zwei Ziffern ausgewertet. Sie werden den nach Benford zu erwartenden Werten gegenübergestellt. Das Resultat ist in Abbildung 5.1 und 5.2 dargestellt. Dabei werden 81 Zahlen betrachtet. Werden die Fehlerbalken in die Betrachtung einbezogen, so wird deutlich, dass die beobachteten Werte von den erwarteten Werten wenig abweichen. Die einzige auffällige Abweichung ist die der Ziffer sieben.

⁵⁴[Ca14]

⁵⁵Vgl. [Wi14a]

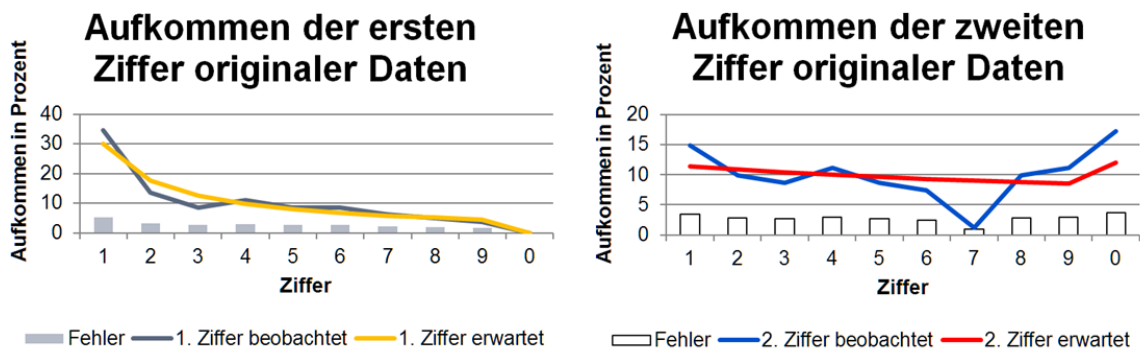


Abbildung 5.1.: Originale Testdaten auf Basis von empirischen Daten zu tausend in der Carslaw-Untersuchung

Carslaw hat mit seiner Auswertung gezeigt, dass Manager ihre Bilanzen von gebrochenen zu ungebrochenen Zahlen aufbesserten, also manipulierten. Dies zeigte sich daran, dass die Ziffer 9 in den untersuchten Datensätzen zu selten aufzufinden war, die Ziffer 0 jedoch ungewöhnlich oft.

Um das Verhalten der Manager nachzustellen, wird für die Daten der Einwohnerzahlen Deutscher Städte von 1950 ebenso verfahren. Wann immer ein gebrochener Wert vorliegt, wird diese Zahl aufgerundet. Diese manipulierten Daten werden ebenfalls ausgewertet. Das Ergebnis wird in Abbildung 5.2 gezeigt. Die Datentabellen dazu befinden sich im Anhang B.

Die meisten Ziffern zeigen in den manipulierten Daten ähnliche Abweichungen zum Benfordschen Gesetz wie bereits die originalen Daten. Auffällig ist jedoch die Abweichung der zweiten Ziffer im Aufkommen der Ziffern 9, 7 und vor allem der 0. Die erwarteten Häufigkeit der Ziffer 9 mit 2,47 zur beobachteten Häufigkeit von 8,50 Prozent weist statistische Signifikanz auf. Die Ziffer 0 wurde zu 11% erwartet, trat jedoch zu 22% auf. Dies ist in der Tat eine Auffälligkeit und spiegelt die Manipulationen wieder. Die Daten dazu sind im Anhang B zu finden.

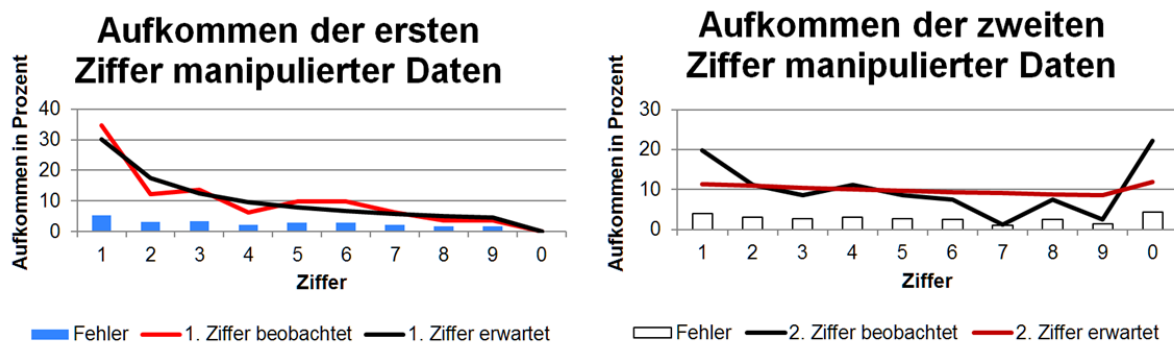


Abbildung 5.2.: Manipulierte Testdaten auf Basis von empirischen Daten zu tausend in der Carslaw-Untersuchung

Erhebungen kommen zu der Aussage, dass das Benfordsche Gesetz mit einer Wahrscheinlichkeit von etwa 89 Prozent zutrifft. Somit gilt auch dass das Benfordsche Gesetz zu etwa 11 Prozent empirische Daten falsch beschreibt und nicht zutrifft. Dies muss in der Aussage zu einem Red Flag berücksichtigt werden.⁵⁶

Die beobachtete Häufigkeit der Ziffer 0 beläuft sich auf ≈ 22 Prozent und steht der erwarteten Häufigkeit der Ziffer 0 von ≈ 11 Prozent gegenüber. Die Ziffer 0 ist also annähernd doppelt so oft aufgetreten wie es erwartet wurde. Das Benfordsche Gesetz kann mit einer Wahrscheinlichkeit von etwa 11 Prozent unzutreffend sein. Die Höhe der Warnung kann mit 100 Prozent beziffert werden, da die Ziffer 0 mehr als doppelt so oft beobachtet werden konnte, als es das Gesetz vorschreibt. Dabei wird das Mittel aus dem schlechtesten Fall der Aussagekraft Benfords von 89 Prozent mit dem bestehenden Verdacht von 100 Prozent herangezogen. Daher lässt sich aussagen, dass eine Empfehlung von $(\frac{89+100}{2}) = 94,5$ Prozent besteht, die Angaben der Manager kritisch zu hinterfragen.⁵⁷

Da davon auszugehen ist, dass bereits die Grundmenge der Daten gefälscht ist, würden weitere statistische Untersuchungen keine verlässlichen Ergebnisse mehr liefern und unsinnig sein.

5.3. Analyse des Betrugs in Miami, 2002

Ein Arbeitgeber hat den Verdacht, dass eine Mitarbeiterin dolos handelt. Die vom System erfassten Arbeitszeiten dieser Mitarbeiterin erstrecken sich stellenweise bis tief in die Nacht. Immer wieder arbeitet sie jedoch auch zu Kernzeiten. Den Tipp auf etwaigen Betrug erhielt der Arbeitgeber durch einen anderen Mitarbeiter. Trotz aller Bemühungen

⁵⁶[Te14]

⁵⁷[Te14]

von Analysten stammen die meisten entscheidenden Hinweise auf einen Betrug in der Regel von Kollegen dolos handelnder Mitarbeiter. Der Hinweis eines Mitarbeiters ist also typisch für das Aufkommen eines anfänglichen Betrugsverdachts. Begründet wird die Arbeitszeit der potentiellen Betrügerin mit dem Kümern um Kinder an Vormittagen. Andere Mitarbeiterinnen tun ähnliches. Daher möchte der Arbeitgeber wissen, welche Arbeitszeiten als normal und welche als nicht normal eingestuft werden können. Um zu prüfen ob die Arbeitsdaten der Mitarbeiterin bedenklich sind, wird zuerst der Datenbestand der übrigen Belegschaft analysiert.

Bei den exemplarischen Arbeitszeiten der Mitarbeiterin handelt es sich um die Feierabendzeiten in Stunden ab der morgendlichen Öffnung des Betriebs, in diesem Fall 8 Uhr werktags. Zehn Stichproben ergaben die Menge $S = \{8.5, 9, 8.7, 10, 9.4, 10.1, 9, 8.7, 9.5, 10.5\}$

Zuerst muss geprüft werden, ob die Daten eventuell manipuliert wurden. Dem Arbeitgeber liegen durch ein automatisches Stechuh-System die Arbeitszeitdaten aller Mitarbeiter vor. Zuerst möchte der Arbeitgeber prüfen, ob die automatisiert erfassten Zeiten echt sind, oder diese manipuliert wurden. Dafür nutzt der Arbeitgeber Minutenzahlen der Zeit ab 17 Uhr, beginnend nach 10 Minuten, da dies die Stoßzeit der Feierabende sind und sie laut eines Analytikers nach dem Benfordschen Gesetz verteilt sein müssten. Zudem fällt jede Minute ab 10 Minuten bis 99 Minuten zweistellig aus. Erfasst wurden dabei 70 Mitarbeiter. Die ersten beiden Ziffern aller gemessenen Zahlen werden untersucht. Tabelle 5.1 stellt die Erhebung dar. Die Basisdaten sind in Anhang B zu finden.

1. Ziffer				2. Ziffer			
#	Anz.	%	Erw.	#	Anz.	%	Erw.
1	21	30	30,1	1	8	11,42	11,39
2	12	17,14	17,61	2	8	11,42	10,88
3	9	12,85	12,49	3	7	10	10,43
4	7	10	9,69	4	7	10	10,03
5	6	8,57	7,92	5	7	10	9,67
6	5	7,14	6,69	6	7	10	9,34
7	4	5,71	5,80	7	6	8,57	9,04
8	4	5,71	5,12	8	6	8,57	8,76
9	3	4,29	4,58	9	6	8,57	8,50
0	-	-	-	0	8	8	11,42
Σ	69,99	101,41	100	Σ	69,65	99,97	100,01

Tabelle 5.1.: Auswertung der Minuten bis zum Feierabend erfasster Mitarbeiter ab der Uhrzeit 17 Uhr

Die Benford-Untersuchung zeigt, dass die Grundmenge nicht gefälscht zu sein scheint und die Messdaten der digitalen Stechuh der Realität entsprechen. Dies zeigen die herausgearbeiteten beobachteten Ziffer-Mengen, die nahe bei ihren Erwartungswerten

liegen. Nachfolgend können also weitere Analysen erfolgen.

Einige Stichproben der üblichen Feierabendzeiten ab morgendlicher Öffnung des Betriebs zeigt die Abbildung 5.3. Die Abbildung betrachtend, liegt die Vermutung nahe, dass die Arbeitszeiten einer Normalverteilung folgen. Dabei fällt besonders ins Auge, dass die Dichte der Feierabende 8 Stunden nach Öffnung des Unternehmens erfolgen. Somit ist davon auszugehen dass der Großteil der Mitarbeiter pünktlich zu Öffnungszeiten des Unternehmens ihre Arbeitstätigkeit aufnehmen, keine oder nur kurze Pausen machen, und nach 8 Stunden Arbeit das Unternehmen wieder verlassen.

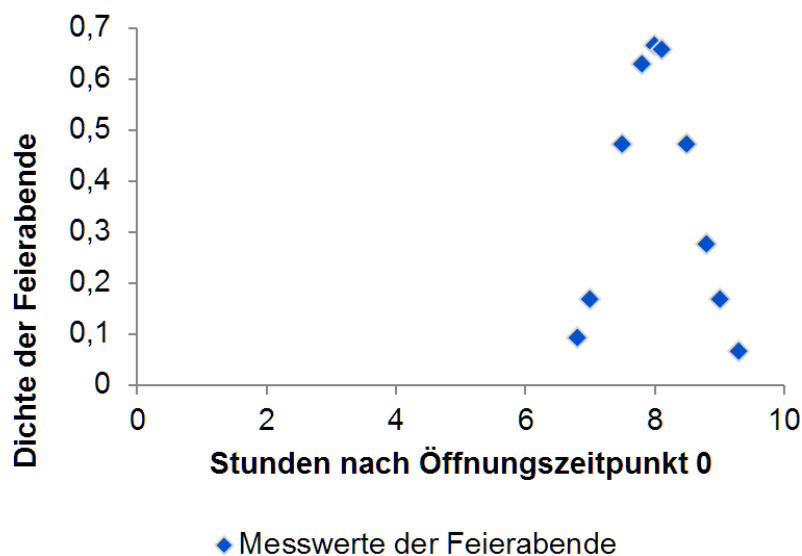


Abbildung 5.3.: Stichproben von Feierabend-Zeiten

Um diese Hypothesen zu prüfen, wird der χ^2 -Test, beschrieben in Abschnitt 4.3.4, eingesetzt. Der Arbeitgeber gibt an, dass die Mitarbeiter in der Regel 8 Stunden nach Öffnung des Betriebes Feierabend machen. Einige Mitarbeiter machen eine halbstündige Mittagspause, einige bleiben wenig länger oder kommen morgens etwas früher. Die Abweichung dürfte laut Arbeitgeber etwas mehr als 30 Minuten, also 0,5 Stunden betragen. Daher wird für den χ^2 -Test die Varianz von $\approx 0,6$, geschätzt. Durch Betrachtung der Abbildung 5.3 und der normalen Arbeitszeit, wird ein Erwartungswert μ von 8 angenommen. Erfasst wurden 399 Mitarbeiter. Die Daten sind in Tabelle 5.2 aufgeführt.

Auf Basis von Stichproben nahe der vermuteten Häufung wurden Klassen als Intervalle gebildet und festgestellt, dass sich die Anzahl der Feierabendzeiten stark symmetrisch verhalten.

Klasse	Intervall		Beob. Häuf.	Wahrschl.	Erwart. Häuf.
j	über	bis	n_j (aus N_j)	P_{0j}	n_{0j}
1	6	6,5	4	0,0058	2,3142
2	6,5	7	19	0,0413	16,478
3	7	7,5	63	0,1525	60,847
4	7,5	8	112	0,3	119,7
5	8	8,5	115	0,297	118,50
6	8,5	9	63	0,1525	60,847
7	9	9,5	19	0,0413	16,478
8	9,5	10	4	0,0058	2,3142
Σ			399	1	397,48

Tabelle 5.2.: Auswertung der Minuten bis zum Feierabend ab 17 Uhr

Dabei berechnet sich die Wahrscheinlichkeit P_{0j} exemplarisch folgendermaßen:
 $P_{0j} = P(6 < X < 6,5 | H_0) = P(-\frac{6-\mu}{\sigma} < \frac{X-\mu}{\sigma} \leq \frac{6,5-\mu}{\sigma} | H_0) = 0,0062 - 0,0004 = 0,0058$,
wobei H_0 die zu prüfende Nullhypothese darstellt. Die Nullhypothese ist die Hypothese, dass ein Zusammenhang besteht und die vermutete Funktion mit der untersuchten Datenmenge vereinbar ist. Dies bedeutet, dass davon ausgegangen werden darf, dass die Daten normalverteilt sind. Auf Basis dieser Erkenntnis können weitere statistische Prüfungen erfolgen.

Ferner berechnet sich darauffolgend n_{0j} derart:

$$n_{0j} = P_{0j} * n, \text{ angewandt auf die erste Berechnung: } 0,0058 * 399 = 2,3142.$$

Nachdem die Tabelle 5.2 aufgestellt wurde, lässt sich χ^2 folgendermaßen bestimmen:

$$\chi^2 = \sum_{j=1}^m \frac{(N_j - n_{0j})^2}{n_{0j}}, \text{ wobei } N_j \text{ die Menge der in einer Stichprobe beobachteter Häufigkeiten ist und } n_{0j} \text{ die unter } H_0 \text{ zu erwartende Häufigkeit.}$$

Der χ^2 -Test ergibt folgendes:

$$\chi^2 = \frac{(4-2,3142)^2}{2,3142} + \dots + \frac{(4-2,314)^2}{2,314} = 3,979$$

Der so genannte Freiheitsgrad ist ein Maß für die mögliche Abweichung einer χ^2 -Untersuchung. Er bestimmt sich nach der Anzahl der vorhandenen Zufallsvariablen abzüglich unsicherer Kategorien.

Die Tabelle 5.2 besitzt die Kategorie n_j mit 8 Zufallsvariablen. Es ergibt sich der Freiheitsgrad von $df = (8 - 1 - 2) = 5$, da zuvor σ und \bar{x} geschätzt wurden. Mit einem gewählten Signifikanzniveau von $\alpha = 0,05$ ergibt sich eine Testprüfgröße von $\chi(0,95; 5) = 9,49$. Es gilt ferner für die Untersuchung $\chi^2 = 3,979 < 9,49$, was bedeutet, dass die Nullhypothese angenommen wird und vermutet werden kann, dass in der Tat

die Normalverteilung vorliegt. Der P-Wert, ein Maß für den Test auf eine Hypothese, bestätigt mit $P - Wert < 0.00001$, dass das Ergebnis signifikant ist für $\alpha = 0,05$.⁵⁸

Die Arbeitsdaten der Mitarbeiterin liegen stellenweise weit außerhalb dieser Normalverteilung. Allerdings finden sich etwa sieben bis zehn Prozent der Arbeitszeiten der verdächtigten Mitarbeiterin im normalen Bereich. Aus diesem Grund wird eine Warnung von 93 Prozent angenommen.

Vielfache der Standardabweichungen

Auf Basis der Erkenntnis um die Normalverteilung der übrigen Mitarbeiterdaten lassen sich nun Aussagen über die Wahrscheinlichkeiten zu den Arbeitszeiten der Mitarbeiterin treffen. In Abbildung 4.2 sind Wahrscheinlichkeiten von σ -Abweichungen abgebildet.

Die Mitarbeiterin arbeitete auffallend regelmäßig bis zu zwei Stunden nach Kernarbeitszeit, und überschritt selbst diesen Wert häufig. Der Wert von σ beträgt 0,6. Bezogen auf das Beispiel der Öffnungszeiten in Minuten entspräche dies 36 Minuten. 3σ -Abweichungen von μ entspräche demnach 108 Minuten oder 1,8 Stunden. Belief sich sodann die Zeit bis zum Feierabend der Mitarbeiterin auf 10 Stunden nach regulärer Öffnung, so ist dies mit mehr als 3σ als sehr ungewöhnlich einzustufen.

In der Tabelle 5.3 führt die Spalte *Min. nach 8h* auf, wie vielen Minuten nach 8 Stunden Arbeitszeit nach Öffnung dies entspricht. Somit entspräche der Feierabend 8,5 Stunden nach Öffnung des Unternehmens, 30 Minuten nach 8 Stunden Arbeitszeit nach Öffnung des Unternehmens.

Die Spalte $\mu \pm \sigma$ beziffert die Anzahl an σ -Differenzen zum Erwartungswert μ . Die Spalte $P(\mu \pm \sigma)$ gibt an, welcher Wahrscheinlichkeit diese Differenz entspricht. Für die Fuzzy-Untersuchung wurde als weiterer Faktor einbezogen, wie viel Prozent der Mitarbeiter sich noch im Unternehmen befand zur Feierabendzeit. Die letzte Spalte, betitelt mit *Fuzzy*, gibt das Warnpotential der Fuzzy-Untersuchung des nächsten Abschnitts an.

Nr.	h bis Feierabend	Min. nach 8h	$\mu \pm \sigma$	$P(\mu \pm \sigma)$	% MA	Fuzzy
1	8,5	30	0,83 σ	0,795	15	75
2	9	60	1,67 σ	0,952	5	95
3	8,7	42	1,167 σ	0,878	15	80
4	10	120	3,33 σ	0,999	0	96
5	9,4	84	2,33 σ	0,990	5	96
6	10,1	126	3,5 σ	0,999	0	96
7	9	60	1,67 σ	0,952	5	96
8	8,7	42	1,167 σ	0,878	15	80
9	9,5	90	2,5 σ	0,994	5	96
10	10,4	144	4 σ	0,999	0	96

⁵⁸Vgl. [Ma15b]

Tabelle 5.3.: Wahrscheinlichkeiten doloser Handlungsweisen der verdächtigten Mitarbeiterin

Nach dieser Einschätzung lässt sich für die Arbeitszeiten der Mitarbeiterin von mehr als $\mu + 3\sigma$ eine Warnung von 100% aussprechen. Das Arithmetische Mittel aller $\mu \pm \sigma$ -Abweichungen beträgt 93,8%, was daher als Warnwert angenommen wird.

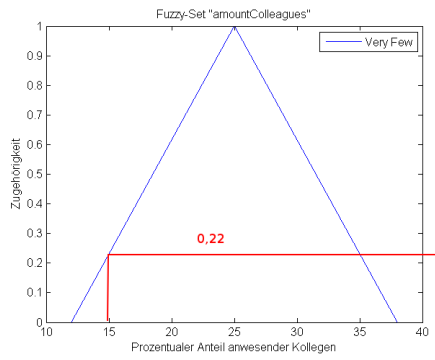
Fuzzy-Untersuchung

Weiter bietet sich eine Fuzzy-Untersuchung an, da für den *Mitarbeiterbetrug in Miami* Einschätzungen über ungewöhnliche Arbeitszeiten vorliegen. Die Eingabeparameter sind die Stunden bis Feierabend und die korrespondierende prozentuale Menge an verbliebenen Mitarbeitern aus der Tabelle 5.3, in welcher bereits die Fuzzy-Resultate in der letzten Spalte eingetragen sind. Der Ausgabeparameter ist ein Warnwert in Prozent.

Die Art und Weise, wie die aus der Fuzzy-Control-Language-Anweisung (FCL-Anweisung) definierten Regeln fuzzifiziert und defuzzifiziert werden, ist in Abbildung 5.4 zu sehen. In der Abbildung wurden die Regeln 1 und 11 exemplarisch mit den Eingabeparametern 8,5 Stunden Feierabend nach Öffnung des Unternehmens bei noch 15 % anwesenden Mitarbeitern durchgeführt. Dabei löst sich eine AND-Beziehung durch das Anwenden des Maximums auf. Im Beispiel ergibt sich aus der Regel 1, wie in Abbildung 5.4 zu sehen ist, dass die Warnung für die Anzahl der Mitarbeiter bei 0,22 liegt. Die Warnung durch die späte Arbeitsstunde liegt jedoch bei 0,75. In dem Fall wird das Maximum der beiden Teilaussagen für die Gesamtaussage genutzt, also 0,75.

Genau so wird mit allen Regeln verfahren. Im Beispiel wird ebenfalls eine weitere Regel der Nummer 11 aufgezeigt mit gleichen Eingabeparametern, um die Defuzzifizierung mehrerer Ergebnisse aufzuzeigen. Nachdem alle Regeln auf diese Weise ausgewertet wurden, werden die Flächen der Warn-Fuzzy-Menge, wie dargestellt, zusammengefügt. Die Gesamtfläche ergibt dann den defuzzifizierten Warnwert, welcher auch in der Tabelle 5.3 eingetragen ist.

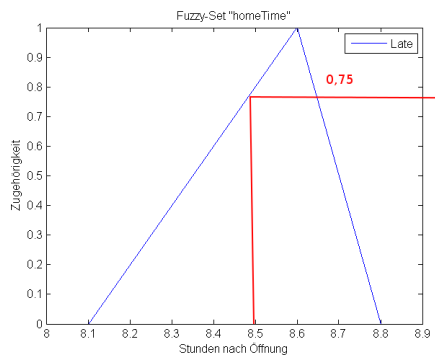
Die komplette FCL-Anweisung zur Untersuchung findet sich im Anhang F. Das arithmetische Mittel der Fuzzy-Warnungen beläuft sich auf 81%. Somit haben wir die Menge der Warnungen aller vorangegangenen Analysen von $w = \{93, 93.8, 81\}$ in [%].



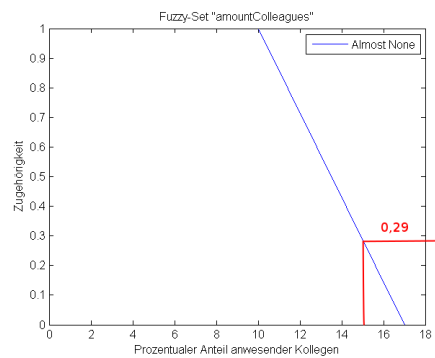
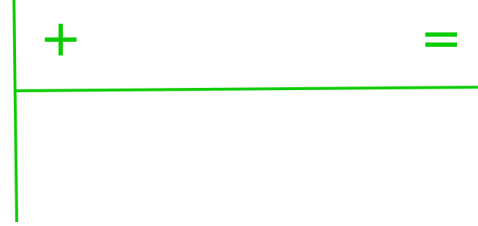
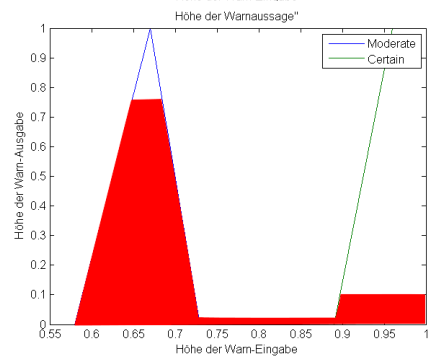
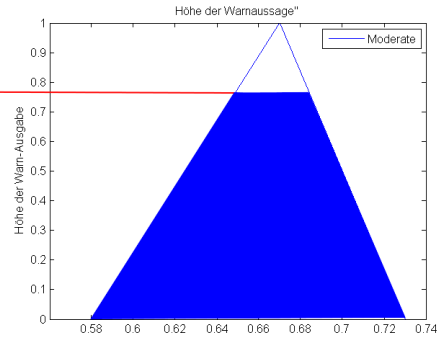
h nach Öffnung: 8,5

%MA anwesend: 15

Rule 1 = IF homeTime IS late
 AND amountColleagues IS veryFew
 THEN warningLevel IS moderate;



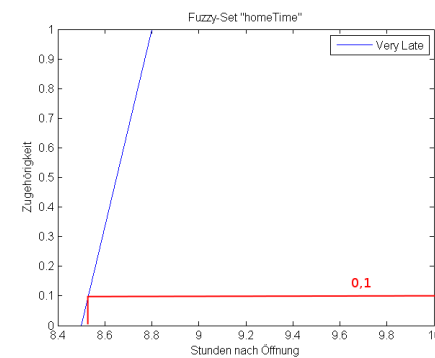
0,22 AND 0,75 = 0,75



h nach Öffnung: 8,5

% MA anwesend: 15

RULE 11 = IF homeTime IS veryLate
 AND amountColleahous IS almostNone
 THEN warningLevel IS certain;



0,29 AND 0,1 = 0,1

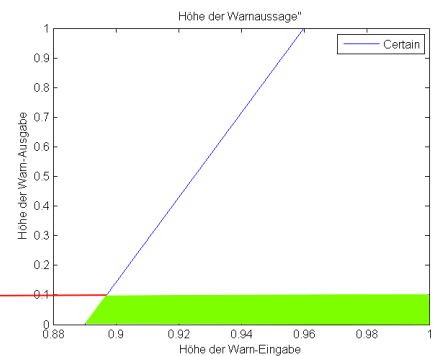


Abbildung 5.4.: Fuzzifizierung und defuzzifizierung am Beispiel der Regeln 1 und 11, sowie der Eingabeparameter 8,5h und 15% MA anwesend

6. Bewertungsmodell

Ziel ist es, mehrere Aussagen zu einer Gesamtaussage zu aggregieren. Dies soll mit einem Bewertungsmodell geschehen. Zuerst werden die Kriterien an ein solches Bewertungsmodell aufgeführt. Das Niveau der Aussage orientiert sich am wirtschaftlichen Bedarf. Dabei wird explizit darauf eingegangen, in wie fern sich die Kriterien von üblichen statistischen Erhebungen unterscheiden.

Darauffolgend werden mathematische Rechnungen aufgezeigt, die zur Erfüllung genannter Kriterien beitragen oder notwendig sind.

Ferner wird das aufgestellte Bewertungsmodell zu Teilen maschinell abgebildet, um aufzuzeigen, wie eine Lösung dafür aussehen kann, wobei dies lediglich ein Exempel vieler möglicher Lösungen darstellt. Letztlich wird das Bewertungsmodell an einer Untersuchung aus Kapitel 5 angewendet. Den Abschluss des Kapitels bildet die kritische Bewertung des aufgestellten Bewertungsmodells.

6.1. Eignungskriterien

Nachfolgend werden Eignungskriterien an das Bewertungsmodell aufgestellt. Diese richten sich nach den Vorstellungen durch die Wirtschaft eines Bewertungsmodells als Gesamtaussage aus Fraud-Untersuchungen anhand verschiedene Analysemethoden und soll eine hohe Praxisrelevanz aufweisen.

Für ein Finanzinstitut sind viele Analyseergebnisse, die sich eventuell widersprechen, wenig hilfreich. Stattdessen besteht Bedarf an einer einzigen Aussage zum Betrugspotential, ermittelt aus einer Menge von Einzelaussagen.

Selbst wenn die Mittelwerte aller Aussagen zum Warnniveau niedrig ausfallen, soll eine außergewöhnlich hohe Wahrscheinlichkeit selbst einer einzigen Analysemethode dazu führen, dass eine absolute Gesamtwarnung ausgesprochen werden, sodass empfohlen wird, den entsprechenden Datensatz genauer und manuell einer Untersuchung zu unterziehen.

Viele statistische Methoden beachten das Auftreten von Ausreißern nicht auf die Weise, wie sie für die Gesamtaussage zum Betrugspotenzial der Szenarien benötigt werden. Die statistischen Methoden verändern zwar bei Ausreißern die Gesamtaussage von vielen Mittelwerten, die Ausreißer sind im Zuge einer Grenzbetragsanalyse jedoch von besonderer Bedeutung: Sie stellen einen ernststen Warnhinweis dar. Sind alle Werte generell niedrig, ein Wert dabei jedoch hoch, kann es passieren, dass dieser Wert in der Gesamtmenge eine für diese Zwecke zu geringe Auswirkung hat. Das Bewertungsmodell soll daher besonders sensibel auf Extremwerte reagieren.

6.2. Erstellung eines Bewertungsmodells

Dieser Abschnitt befasst sich mit der mathematischen und maschinellen Konzeption und Umsetzung eines Bewertungsmodells, basierend auf den aufgestellten Eignungskriterien.

6.2.1. Ermittlung eines Maximums

Ein wichtiger Bestandteil des Bewertungsmodells im folgenden Verlauf des Kapitels, ist die Bestimmung des Maximums einer Menge. Es ist als höchster Wert aller Aussagen zum Betrugspotential von besonderem Interesse und wird daher im Bewertungsmodell gesondert behandelt. Die mathematischen Grundlagen dazu sind zu finden in Abschnitt 4.4.

6.2.2. Standardabweichung

Die Standardabweichung beschreibt die Streuung einer Zufallsvariable um einen Erwartungswert und berechnet sich als Quadratwurzel der Varianz. Für das Bewertungsmodell ist wichtig zu beachten, wie weit Werte von einem Mittel streuen. Da das Modell das Betrugspotential im Zweifel hoch bewerten soll, wird die Standardabweichung, als schlimmster anzunehmender Fall, stets zum arithmetischen Mittel hinzu addiert.

Für die Berechnung der Standardabweichung wird die Varianz benötigt. Die Varianz ist ein statistisches Streuungsmaß und beschreibt die Abweichung einer Zufallsvariable X zum Erwartungswert. Die Berechnung der Varianz und Standardabweichung befinden sich in Kapitel 4.

6.2.3. Absoluter Warnhinweis nach Schwellenwert und Bewertungsmodell

Nach Erarbeitung der besonderen Anforderungen an ein Bewertungsmodell und Erläuterung wichtiger Bestandteil eines solchen Bewertungsmodells, wird das erarbeitete Bewertungsmodell in diesem Abschnitt vorgestellt.

Das Bewertungsmodell stellt die Kernleistung dieser Thesis dar. Zwar mag es trivial wirken, doch unterscheidet es sich in seinen Aussagen grundlegend von bloßen Mittelwerten. Daher sind im nachfolgenden Abschnitt 6.3 die Gründe für die einzelnen Bestandteile des Modells und die Eigenschaften aufgeführt.

Sei M eine Menge von Aussagewerten

$$\forall x \in M : \{x \in \mathbb{R} | 0 \leq x \leq 1\} = [0, 1]$$

und $n = |M|$, d.h. n ist die Anzahl aller Elemente von M .

Dann gilt für $n \geq 2$ der Warnwert w_x :

$$w_x = \begin{cases} 1 & \text{falls } \exists x \in M : x \geq 0,95 \\ \frac{(\bar{x} + \sigma) + \max(M)}{2} & \text{sonst} \end{cases}$$

6.3. Motivation und Eigenschaften

Die erste Eigenschaft des Modells ist die Aussage zum Betrugspotential von 1, bzw. 100%, sofern in der Menge von Aussagewerten $\forall x \in M : \{x \in \mathbb{R} | 0 \leq x \leq 1\} = [0, 1]$ mindestens ein Wert befindlich ist, der größer oder gleich 0,95 ist. Dies geschieht in der ersten Zeile der Kondition durch: 1 falls $\exists x \in M : x \geq 0,95$.

Der Grund für dieses Verhalten des Modells, ist die hohe Signifikanz einer einzigen, hohen Betrugsaussage. Erreicht oder übersteigt ein Warnwert die Marke von 95%, so wird eine absolute Warnung von 100% ausgesprochen, damit eine Prüfung unabdingbar empfohlen wird.

Sind alle Elemente der Menge M unter 0.95, errechnet sich der Warnwert gemäß der zweiten Zeile der Funktion: $\frac{(\bar{x}+\sigma)+\max(M)}{2}$. Dabei wird zunächst das arithmetische Mittel \bar{x} gebildet, wie im Abschnitt 4.3 beschrieben. Der Grund für die Nutzung ist die Annahme, dass ein Betrug dazu führt, dass viele Warnwerte ähnlich hoch ausfallen⁵⁹. Somit ist das Mittel der Menge M von Interesse.

Ferner wird auf \bar{x} zudem die Standardabweichung σ addiert. Begründet wird dies mit einer Fehlertoleranz der Menge aller Warnwerte als Standardabweichung. Diese Fehlertoleranz wird entweder als Wurzel aller Zufallsvariablen oder als Standardabweichung angenommen.⁶⁰ Um vom schlechtesten Fall des größten Fehleraufkommens auszugehen, wird dieser Wert der Standardabweichung zu \bar{x} addiert.

Ein weiterer wichtiger Wert ist das maximale Element der Menge M . Das Ergebnis der Summe aus dem arithmetischen Mittel, der Standardabweichung und dem größten Elements wird halbiert. Auf diese Weise wird bewirkt, dass das Maximum der Menge M einen großen Einfluss auf die Gesamtaussage des Modells hat. Auch dies entspricht dem konservativen Charakter des Modells, eher einen Verdacht zu äußern.

Wird dieses Modell auf eine sehr kleine Menge von Aussagewerten angewandt, so wird die Gesamtaussage trivial. Ferner ist zu beachten, dass die Menge aller Warnwerte einem Fehler unterworfen ist und somit stets eine Unsicherheit durch Fehler der Menge M existiert, was sich bereits bei einer hohen Standardabweichung vermuten lässt.

6.4. Maschinelle Abbildung

Wie das hier erarbeitete Bewertungsmodell maschinell abgebildet werden könnte, ist Gegenstand dieses Abschnitts. Dabei lehnt sich die exemplarische Darstellung stark an die Programmiersprache Java an, kann jedoch in jedweder Programmiersprache erfolgen.

6.4.1. Maximum, arithmetisches Mittel, Varianz und Standardabweichung

Das Maximum einer gegebenen Menge zu ermitteln, ist ein Standard der meisten Programmiersprachen und in der Regel durch Bibliotheken abgebildet. Wird die Programmiersprache Java genutzt, so sind Lambda-Ausdrücke ab der Version 8 Teil des Java Development Kits (JDK), wodurch selbst die funktionale Abbildung des Findens eines Maximums ohne großen Aufwand möglich wäre.⁶¹

⁵⁹[Ca14]

⁶⁰Vgl. [Ph14]

⁶¹[Dr15]

Statistische Berechnungen können ohne Zusatzbibliotheken in einigen Programmiersprachen aufwändig sein. Hilfreiche Funktionen zu statistischen Berechnungen bietet das kostenpflichtige Programm MATLAB. Statistische Untersuchungen lassen sich ferner auch durch die an Java anbindbare Bibliothek der freien Software *GNU Octave* abbilden, die einen ähnlichen Funktionsumfang wie MATLAB bietet.

Dafür müssen in Java entsprechende Bibliotheken importiert werden. Anschließend wird eine *Octave Factory* erzeugt, welche Octave-Objekte handhaben kann und die Kommunikation der Octave-Engine und Java steuert. Mathematische Funktionen werden als für Octave verständliche Strings definiert und diese an die Octave-Factory gereicht. Die erhaltenen Octave-Objekte beinhalten die Ergebnisse, welche anschließend in Java-Objekte geparkt und weiterverarbeitet werden können. Eine exemplarische Anwendung der Java-Octave-Bibliothek findet sich im Anhang C unter C.

6.4.2. Bewertungsmodell als Pseudocode

Auf Basis der in diesem Abschnitt beschriebenen maschinellen Abbildung ließe sich das Bewertungsmodell mithilfe von Methodenaufrufen folgendermaßen gestalten, wobei Hilfsfunktionen wie `getAritmethicalMean(Set preds)` erst durch Bibliotheken wie Octave abgebildet oder selbst erstellt werden müssten, sollten sie nicht bereits in Utils vorhanden sein:

```
double warningNotice(Set premonitions){
    double arithmeticalMean = getAritmethicalMean(premonitions);
    double standardDeviation = getStandardDeviation(premonitions);
    double maximum = getMaximum(premonitions);

    function warning{
        for(double i : premonitions){
            if(i >= 0.95){
                return 1.0;
            }else{
                return (((arithmeticalMean + standardDeviation)+ maximum)/2)
            }
        }
    }
}
```

Es wurde gezeigt, dass sowohl die vorgestellten Analysen aus Kapitel 4, als auch das hier erarbeitete Bewertungsmodell maschinell abgebildet werden können. Dies bringt den entscheidenden Vorteil mit sich, dass ein Mal abgebildete Analysen fortan automatisiert durchgeführt werden können. So ist es denkbar und möglich, ein erweiterbares Programm zu entwickeln, dem eines Kataloges gleich immer weitere Analysen hinzugefügt werden. Auf diese Weise besteht die Chance ein effektives und effizientes Werkzeug zum Detektieren von Betrug zu erzeugen.

6.5. Anwendung des Bewertungsmodells

In diesem Abschnitt wird abschließend anhand der durchgeführten Analysen aus Kapitel 5 das Bewertungsmodell praktisch erprobt.

Mitarbeiterbetrug in Miami

Angewendet wird das Bewertungsmodell an der in Abschnitt 5.3 erarbeitete Menge von Warnwerten mit $M = \{93, 93.8, 81\}$ in [%], welche die Ergebnisse der verschiedenen Analysen des Miami-Betrugs aus dem Kapitel 5.3 sind. Nach Bewertungsmodell werden die Standardabweichung und das Maximum benötigt. Das Maximum dieser Menge ist 93,8. Das Arithmetische Mittel ist $\bar{x} = 89,2\bar{6}\%$, die Varianz $Var(M) = 34,275$ und die Standardabweichung $\sigma_X = \sqrt{34.275} = 5,8544$. Demnach ergibt sich für das Bewertungsmodell:

$$w_x = \begin{cases} 1 & \text{falls } \exists x \in M : x \geq 0,95 \\ \frac{(89,2+5,8544)+93,8}{2} & \text{sonst} \end{cases}$$

Somit ergibt sich eine Gesamtwahrscheinlichkeit von $w = 94,425\% \approx 94\%$, was einem deutlichen Red Flag entspricht. Werte ab 80% können als bedenklich, Werte ab 95% als sehr deutlich interpretiert werden.

Wäre das Mittel zur Evaluierung der Menge $M = \{93, 93.8, 81\}$ in [%] der einfache Durchschnitt gewesen, so ergäbe sich eine Warnhöhe von 89,26%. Ein risikobereiter Manager könnte für sich selbst einen Mindestwert von 90% zur Deeskelation definieren und so diesen Wert ignorieren. Das Bewertungsmodell, hingegen, wäre zu einer anderen, höheren Aussage von 94,425% gekommen.

7. Zusammenfassung, Fazit und Ausblick

Die wesentlichen Aussagen dieser Thesis, gewonnene Erkenntnisse und ein Ausblick auf die thematische Fortführung, sind Gegenstände dieses Kapitels.

7.1. Resümee der Kernaspekte

Ausgangspunkt der Thesis ist der Umstand, dass dolose Handlungen für die Finanzwirtschaft eine große Bedrohung darstellen. Dabei wird Fraud zu einem Großteil von Mitarbeitern, zum Beispiel eines Finanzdienstleisters, begangen. Die Frage nach einer Grenzbetragsverletzung ist für die Finanzwirtschaft von Relevanz. Die aktuelle Forschung bietet viele verschiedene Möglichkeiten, auf eine Grenzbetragsverletzung zu prüfen. Allerdings fehlt es an Modellen, um zu Gesamtaussagen aus mehreren Analyseergebnissen zu kommen.

Dolose Handlungen, bei denen die Grenzbetragsanalyse von Relevanz ist, treten auf verschiedene Weise auf. Betrüger sind den Analytikern oft viele Schritte voraus und betreiben Fraud auf kreative Weise. Durch die große Menge an täglichen Daten, ist die maschinelle Grenzbetragsanalyse gegenüber der manuellen unabdingbar und zukunftsweisend. Allerdings stellt es Analysten vor Herausforderungen eine Grenzbetragsanalyse zu automatisieren. Für Menschen intuitive Konzepte der Abschätzung eines Gefahrengrades, sind mathematisch nur schwer in Regeln abbildbar.

Das Benfordsche Gesetz, welches besagt dass die Ziffern von Zahlen empirischer Daten festen Häufigkeiten entsprechen, kann wertvolle Aussagen über etwaige Manipulationen an Datenbeständen empirischer Art liefern, mithilfe derer eine Vorprüfung, nicht nur auf Analysetauglichkeit, sondern auch auf Betrug, erfolgen kann. So ergab die Prüfung auf das Zutreffen des Benfordschen Gesetz der ausgewiesenen Nettojahresgewinne durch Manager, dass bereits die Grundmenge manipuliert gewesen zu sein schien, weshalb von weiteren Analysen abgesehen wurde.

Die Statistik ermöglicht es vor allem strukturierte Daten zu untersuchen und Unterschiede individueller Daten zur Gesamtmenge ausfindig zu machen. Dies zeigte sich vor allem durch χ^2 -Tests auf Normalverteilungen, nach deren Bestätigung der Verträglichkeit von vermuteten Funktionen zu beobachteten Daten, statistische Untersuchungen zu Standardabweichungen und deren Wahrscheinlichkeiten erfolgen konnten.

Die unscharfe Logik ist in der Lage, oft in der Finanzwirtschaft anzutreffende Abstufungen von *wahr* und *falsch* maschinell abzubilden und sich die mehrwertigen Konzepte der Logik zunutze zu machen. Dies zeigte sich am Beispiel des Betrugs in Miami. Durch vorherige Schätzungen des Arbeitgebers zu gewöhnlichen und ungewöhnlichen Arbeitszeiten konnten Regeln formuliert werden, die Arbeitszeiten in Relation zur Anwesenheit von Mitarbeitern auf Betrugspotenzial prüfen konnten. Diese Regeln wurden für die verdächtige Mitarbeiterin zur Untersuchung genutzt.

An einem realen Beispiel wurde exemplarisch aufgezeigt, dass eine maschinelle Grenzbetragsanalyse anhand mathematischer Methoden im Compliance-Umfeld eines Finanzdienstleisters sowohl möglich, als auch sinnvoll ist. Ferner wurde ein Bewertungsmodell aufgestellt, welches sensibel und maschinell abbildbar mehrere Teilaussagen zu einer einzigen Aussage aggregiert.

7.2. Hypothesen und Forschungsfrage

In diesem Abschnitt sollen die aufgestellten Hypothesen kritisch geprüft, und eine Antwort auf die Forschungsfrage aus Kapitel 1 formuliert werden.

1. Hypothese *Prozentuale Grenzen sind für eine Grenzbetragsanalyse ungeeignet.*

Die Betrugsszenarien aus Kapitel 3 haben deutlich gemacht, dass mit Wissen über eine Grenze, durch minimales Abweichen davon, die Grenze umgangen werden kann. Ferner zeigte das Kapitel 4, welche mathematischen Vorteile andere Analysen gegenüber der Prüfung auf harte Grenzen haben.

2. Hypothese *Mathematische Gesetzmäßigkeiten können aussagekräftige Hinweise auf Manipulationen von Finanzdatenbeständen liefern.*

Das Betrugsszenario, welches Carslaw 1988 aufdeckte (beschrieben in Abschnitt 3.1.2), konnte einzig durch das Benfordsche Gesetz aufgedeckt werden. Dieses besagt, dass die Verteilungen von Ziffern aller Zahlen einer empirischen Datenmenge festen Wahrscheinlichkeiten folgen. Praktisch angewandt wurde dies im Kapitel 5.

3. Hypothese *Softwaretechnologische Analysemethoden in Form der Fuzzylogik erlauben differenzierte Aussagen zum Grade eines Warnhinweises auf Grenzbetragsverletzungen in höherer Qualität, als es die klassische Aussagenlogik ermöglicht.*

Die Analysen des Miami-Betrugs in Abschnitt 5.3 und des Schneider-Betrugs im Anhang G haben aufgezeigt, dass für ein Vorgehen bei einem Betrug mehrere mögliche Analysen durchgeführt werden können. Einige von ihnen ergaben deutliche Warnmeldungen, andere einen Verdacht oder schlagen gar nicht an.

4. Hypothese *Statistische Auswertungen zu Abweichungen eines Zahlenwertes in einer Grundgesamtheit stellen eine qualifizierte Aussage zu einem möglichen Betrugsfall dar.*

Die σ -Abweichungen des Schneider-Betrugs in Anhang G und des Miami-Betrugs in 5.3 zeigten deutlich, dass die Wahrscheinlichkeiten der zu prüfenden Daten in einer normalverteilten Menge klar dafür sprachen, dass Betrug vorlag.

5. Hypothese *Einzelne Methoden sind ohne Verknüpfung zu weiteren Analysen wenig aussagekräftig, um qualifizierte Warnhinweise aussprechen zu können.*⁶²

⁶²Unverändert entnommen aus [Jo14b] S. 5

Im Miami-Betrug aus Kapitel 5.3 wäre die isolierte Prüfung auf Verletzung des Benfordschen Gesetz zu der Aussage gekommen, dass kein Betrug vorliegt. Weitere Prüfungen zeigten jedoch, dass Arbeitszeiten außerhalb der Norm sind. Somit ergaben sich zwei unterschiedliche Aussagen, deren unbedachte Betrachtung zu Widersprüchen geführt hätte.

Führt das Aggregieren multipler Aussagen zum Betrugspotenzial und Abbildung auf ein Bewertungsmodell zu einer qualifizierteren Aussage über Fraud-Warnhinweise, als die Betrachtung der einzelnen Ergebnisse?

Das erstellte Bewertungsmodell konnte die an das Modell gestellten Kriterien aus Abschnitt 6.1 erfüllen. Es findet eine Gesamtaussage mit Berücksichtigung schwerwiegender Warnungen und Ausreißern und führt eine beliebig große Menge an Warnwerten zu einer einzigen Aussage zusammen.

7.3. Fazit der gewonnenen Erkenntnisse

Diese Thesis konnte aufzeigen, dass die Grenzbetragsanalyse zwar durchaus ihre Herausforderungen hat, diese jedoch bewältigt werden können. Vor allem die Möglichkeit statistischer Untersuchung großer Datenmengen und die Formulierung abstufbarer Grade aussagenlogischer Betrugspotentiale durch die Fuzzylogik, haben neue Analysemöglichkeiten eröffnet und sind maschinell abbildbar. Der χ^2 -Test machte Hypothesenbildungen und -Prüfungen zu vermuteten Funktionen möglich.

Die Grundlage der unscharfen Logik ist, dass es neben der zweiwertigen Logik der Wahrheitswerte *wahr* und *falsch* noch weitere Wahrheitswerte, als Abstufungen, geben kann. Die maschinelle Umsetzbarkeit erfolgte jedoch, auch aufgrund mathematischer Leistungen und dem Wissen um die Computertechnologie, erst in den 1980er Jahren in Form der Fuzzylogik und brachte die Möglichkeit mit sich, neuartige Aussagen zu treffen, und Lösungen finden zu können.

Ferner wurde aufgezeigt, dass sich Bewertungsmodelle aufstellen lassen und so dedizierte Aussagen zu Warnpotenzialen getätigt werden können. Da vor allem große Mengen an Analyseergebnissen nur schwer zu überblicken und einzuschätzen sind, ist das Nutzen eines Modells für das Aggregieren zu einer Gesamtaussage obligatorisch.

Weiter wurden Herausforderungen und Grenzen vieler Analysen und des erstellten Bewertungsmodells diskutiert. Besonders Verfälschungen von allgemeinen Aussagen müssen bedacht werden. Wie ein Autofahrer seinem Navigationssystem nicht blind vertrauen darf, sondern lediglich dessen Weisungen mit Bedacht Beachtung schenken sollte, sind auch automatisierte Analysen mit anschließender Bewertung durch Softwaresysteme stets mit Distanz zu betrachten und nicht unreflektiert darauf zu vertrauen. Sie können lediglich Hinweise geben über mögliche Schadenspotenziale aber nie die Expertise eines Compliance-Experten ersetzen.

7.4. Ausblick und zukünftige Entwicklung

Eines haben vor allem die Kapitel zu Compliance und den Betrugsszenarien deutlich gemacht: Betrüger agieren auf stellenweise enorm kreative Weise. Kreative Analytiker ermöglichen, in Verbindung mit der Geschwindigkeit und Leistung moderner Computersysteme, bislang aufwändige Analysen durch Automatisierung in kurzer Zeit zu weit geringeren Kosten durchzuführen, und erschließen teils neue Vorgehensweisen.

Die moderne Informationstechnologie, gepaart mit den mathematischen Erkenntnissen der bisherigen Forschung, werden auch in Zukunft das Compliance-Umfeld bereichern und Möglichkeiten aufzeigen, betrügerisches Verhalten aufzudecken.

Diese Thesis beschäftigte sich nur mit einem kleinen, ausgewählten Bereich der mathematischen Möglichkeiten und der Problemstellen des Compliance, nämlich der Grenzbetragsanalyse durch ausgewählte mathematische Verfahren. Das Compliance als eigenständiges Fachgebiet befindet sich noch in frühen Phasen. Die Bedeutung wird in der Wirtschaft mehr und mehr zur Kenntnis genommen. Vor allem die Möglichkeiten durch Informatiksysteme, werden als große Chance erkannt. Die Vertiefung der Möglichkeiten bietet, vor allem im Hinblick auf die gewaltigen Schadenssummen durch Betrug, große Chancen.

Es war Ziel, exemplarisch aufzuzeigen, wie das Fachgebiet erschlossen werden kann. Dieses wurde in dieser Thesis durch die theoretische Aufarbeitung fachlicher und mathematischer Grundlagen, anschließender Vorstellung und Untersuchung von Betrugsszenarien, und der Erarbeitung eines Bewertungsmodells getan.

Ferner wären die nächsten Schritte, die gewonnen Erkenntnisse nicht nur für einen kleinen Prototyp, sondern für ein umfangreiches Tool zu nutzen. Dieses könnte in der Lage sein, gegebene Datenbestände auf Grenzbetragsverletzungen zu untersuchen. In Ansätzen ist dies bereits erprobt worden in Form des Prototypen. Dieser Prototyp stellt eine erste, grundlegende, aber noch frühe Fassung erarbeiteter Konzepte zur Grenzbetragsanalyse dar, und diente der Motivation dieser Thesis. Die Oberfläche und Codefragmente sind in Anhang [A](#) zu finden.

Da sich sowohl das Bewertungsmodell, als auch das Vorgehen der einzelnen Analysen, wie aufgezeigt, maschinell abbilden lassen, bietet ein flexibles, erweiterbares Programm die Chance, einen umfangreichen Katalog verschiedener Analysen zu kombinieren. Ein so erstelltes Werkzeug wird fortan effektiv und effizient Betrugshinweise detektieren können.

Der Ausblick geht dahin, Erkenntnisse aus dieser Thesis umzusetzen und diese zu vertiefen. Die Grundsteine dafür sind gelegt.

Literaturverzeichnis

Die genutzten Quellen unterteilen sich in die drei Abschnitte [Literaturverzeichnis](#), [Onlinequellen](#) und [interne Quellen](#).

- [AC12] ASSOCIATION OF CERTIFIED FRAUD EXAMINERS. *Report to the Nations - on Occupational Fraud and Abuse - 2012 Global Fraud Study..* URL: http://www.acfe.com/uploadedFiles/ACFE_WebsiteContent/rtrtn2012-report-to-nations.pdf. Eingesehen am: 20.11.2014.
- [Al11] ALBRECHT, W. Steve, Chad O. Albrecht, and Conan C. Albrecht. *Fraud Examination*. Aufl. 4. Mason, OH: South Western Educ Pub, 2011.
- [Be38] BENFORD, *The Law of Anomalous Numbers*. Proceedings of the American Philosophical Society 78, no. 4 (March 31, 1938): 551–72.
- [Bo11] BOSSEK, Hubert, Karlheinz Weber, Armin Baeger, et al. *Duden Basiswissen Schule: Mathematik Abitur*. Auflage: 3., neu bearbeitete Auflage. Berlin; Mannheim; Zürich: Bibliographisches Institut, 2011.
- [Bo12] BOURIER, Günther. *Beschreibende Statistik: Praxisorientierte Einführung - Mit Aufgaben und Lösungen*. Auflage: 10., aktual. Aufl. 2012. Wiesbaden: Gabler Verlag, 2012.
- [Br13] BRAUER, Johannes. *Abstraktionskonzepte der Informatik. Aufgaben mit Lösungen*. Nordakademie - Hochschule der Wirtschaft. Elmshorn, 2013.
- [Ca88] CARSLAW, Charles. *Anomalies in Income Numbers: Evidence of Goal Oriented Behavior*. The Accounting Review Vol. 63, No. 2 (Apr., 1988). American Accounting Association. 1988
- [Do13] DOIG, Alan. *Fraud*. Routledge, 2013.
- [Fr02] FRANK, Hubert. *Fuzzy Methoden in der Wirtschaftsmathematik. Eine Einführung*. Aufl. Braunschweig; Wiesbaden: Vieweg Verlag, 2002.
- [Ho08] HOFMANN, Stefan. *Handbuch Anti-Fraud-Management: Bilanzbetrug erkennen - vorbeugen - bekämpfen*. Erich Schmidt Verlag GmbH & Co KG, 2008.
- [HS10] HÄFELE, Prof. Dr. Markus und Schmeisky, Jan Helge. *Fraud Red Flags - Warnsignale zur Erkennung wirtschaftskrimineller Handlungen*. Ausgabe: 510. ZRFC. Pub, 2010.
- [JLZ13] JACKMUTH, Hans-Willi, Christian de Lamboy, and Peter Zawilla. *Fraud Management in Kreditinstituten: Praktiken, Verhinderung, Aufdeckung*. Frankfurt am Main: Frankfurt School Verlag, 2013.
- [Kr13] KRUSE, Rudolf. *Fuzzy-Systeme*. 2. Aufl. 1995. Stuttgart: Springer, 2013.
- [Me12] MEYER, Jörg. *Forensische Datenanalyse: Dolose Handlungen im Unternehmen erkennen und aufdecken*. Berlin: Erich Schmidt Verlag GmbH & Co, 2012.
- [Na14] PEEMÖLLER, Prof Dr Volker H., und HOFMANN Dipl KfmStefan. *Bilanzskandale: Delikte und Gegenmaßnahmen*. Berlin: Erich Schmidt Verlag GmbH & Co, 2005.

- [PH05] PEEMÖLLER, Prof Dr Volker H., and Dipl-Kfm Stefan Hofmann. *Bilanzskandale: Delikte und Gegenmaßnahmen*. Berlin: Erich Schmidt Verlag GmbH & Co, 2005.
- [PS09] PLAUE, Matthias, and Mike Scherfner. *Mathematik für das Bachelorstudium I: Grundlagen, lineare Algebra und Analysis*. Auflage: 2009. Heidelberg: Spektrum Akademischer Verlag, 2009.
- [Qe11] QUENTMEIER, Helma. *Praxishandbuch Compliance*. Auflage: 1. Gabler Verlag, 2011.
- [RB07] RAFELD, Hagen. Prof Dr Friedrich R Then Berg *Digitale Ziffernanalyse in deutschen Rechnungsdaten*. Ravensburg, 2007.
- [Sc10] SCHRÄPER, Jörg-Peter. *Benford's law as an instrument for fraud detection in surveys using the data of the Socio-Economic Panel (SOEP)*. Deutsches Institut für Wirtschaftsforschung. SOEPpapers on Multidisciplinary Panel Data Research. Berlin, Februar 2010.
- [St12] STAAB, Frank. *Logik und Algebra: Eine Praxisbezogene Einführung für Informatiker und Wirtschaftsinformatiker*. Auflage: überarbeitete Auflage. München: De Gruyter Oldenbourg, 2012.
- [Zi07] ZIMMERMANN, Prof Dr Frank. *Diskrete Mathematik Vorlesung im 1. Fachsemester Wirtschaftsinformatik*. Nordakademie - Hochschule der Wirtschaft. Elmshorn, Januar 2007.

Onlinequellen-Verzeichnis

Die genutzten Quellen unterteilen sich in die drei Abschnitte [Literaturverzeichnis](#), [Onlinequellen](#) und [interne Quellen](#).

- [AC12] ASSOCIATION OF CERTIFIED FRAUD EXAMINERS. *Report to the Nations - on Occupational Fraud and Abuse - 2012 Global Fraud Study..* URL: http://www.acfe.com/uploadedFiles/ACFE_Website/Content/rtnn/2012-report-to-nations.pdf. Eingesehen am: 20.11.2014.
- [Az14] AZIZ, Shahariz Abdul & Jeyakody Parthiban *What do ya mean fuzzy ???!* URL: http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/sbaa/report.fuzzysets.html. Eingesehen am: 14.01.15
- [Be38] BENFORD, Frank. *The Law of Anomalous Numbers*. URL: <http://www.jstor.org/stable/984802>. Eingesehen am 28.01.2015
- [Bu14] BUSINESS FUNDAS *Fuzzy Logic and Fuzzy Systems* URL: <http://business-fundas.com/tag/fuzzy-set-theory/>. Eingesehen am: 14.01.15
- [Dm14] DMA *Fuzzy Operators*. URL: http://www.dma.fi.upm.es/java/fuzzy/fuzzyinf/fuzzyop_en.htm. Eingesehen am 21.01.15
- [Dr15] DRDOBBS *Lambda Expressions in Java 8*. URL: <http://www.drdobbs.com/jvm/lambda-expressions-in-java-8/240166764>. Eingesehen am 28.01.15
- [Du14] DUDEN ONLINE *Analyse*. URL: <http://www.duden.de/rechtschreibung/Analyse>. Eingesehen am: 12.12.2014.
- [ES14] ELEKTRONISCHE STEUERPRÜFUNG *Finanzgericht Münster Beschluss*. URL: http://www.elektronische-steuerpruefung.de/rechtspr/fg_chi_quadrat.pdf. Eingesehen am 20.01.15
- [He14] HERZOG, F., Achtelik, O., Warius, S. und Nestler, C. *Geldwäschegesetz*. URL: <http://dejure.org/gesetze/GwG/3.html>. Eingesehen am: 12.12.2014. Auflage: 2. München: C.H. Beck. 2013.
- [He14b] HEISE ONLINE *Alles auf 8 - Java SE 8 ist da*. URL: <http://www.heise.de/developer/artikel/Alles-auf-8-Java-SE-8-ist-da-2149753.html>. Eingesehen am: 12.12.2014.
- [IEC14] INTERNATIONAL ELECTRONICAL COMMISSION. <http://www.iec.ch/> Eingesehen am 18.12.14
- [Ma15] MATHEGURU. *Chi-Quadrat-Test*. URL: <http://matheguru.com/stochastik/248-chi-quadrat-test.html>. Eingesehen am: 28.01.2015.
- [Ma15b] UNI-MAGDEBURG. *Chi-Quadrat-Test*. URL: http://fuzzy.cs.uni-magdeburg.de/studium/ida/txt/chi_squared.pdf. Eingesehen am: 29.01.2015.

- [Na14] NACUBO *GUIDELINES TO FTC RED FLAG RULE - INTERAGENCY GUIDELINES ON IDENTITY THEFT DETECTION, PREVENTION, AND MITIGATION*. URL: http://www.nacubo.org/documents/business_topics/Guidelines%20to%20FTC%20Red%20Flag%20APP.pdf.
Eingesehen am: 27.11.2014.
- [Ph14] PHYSIK DER UNI HALLE *Fehlerrechnung*. URL: <http://www.physik.uni-halle.de/Lehre/Grundpraktikum/gpfehler.htm>. Eingesehen am 21.01.15
- [So14] SOURCEFORGE *jFuzzyLogic*. URL: <http://jfuzzylogic.sourceforge.net/html/fcl.html>. Eingesehen am: 18.12.2014.
- [Te14] TESTBENFORDSLAW *Testing Benford's Law*. URL: <http://testingbenfordslaw.com/>. Eingesehen am: 22.12.2014.
- [Wi14a] LIST OF CITIES IN GERMANY BY POPULATION. http://en.wikipedia.org/wiki/List_of_cities_in_Germany_by_population Eingesehen am 22.12.14
- [Wi15a] FUZZY CONTROL. http://commons.wikimedia.org/wiki/File:Fuzzy_control_-_centroid_defuzzification_using_max-min_inferencing.png. Eingesehen am 21.01.15
- [Wi15b] STANDARDABWEICHUNG. <http://de.wikipedia.org/wiki/Standardabweichung>. Eingesehen am 21.01.15
- [Wi15c] CHI-QUADRAT-VERTEILUNG. <http://de.wikipedia.org/wiki/Chi-Quadrat-Verteilung>. Eingesehen am 28.01.15

Interne Quellen

Die genutzten Quellen unterteilen sich in die drei Abschnitte [Literaturverzeichnis](#), [Onlinequellen](#) und [interne Quellen](#).

- [Ca14] CALI, Martina-Susan. *Interview einer Compliance-Expertin der PPI AG*. Hamburg Winterhude, 05.12.2014.
- [Hi14] HINSCH, Laura. *Bachelorarbeit: Erkennung von Betrugspotenzialen auf Basis von Prozessdaten, die mittels Prozesserhebungsmethoden zusammengestellt wurden*. PPI AG & Nordakademie 2014.
- [Jo14a] JOHANNSEN, Dirk. *Vorstudie: Kategorisierung von Red Flags im Bereich Compliance eines Finanzdienstleisters und Evaluierung der Analysetauglichkeit herausgestellter Kategorien*. Nordakademie, PPI AG, 2014. Internes Dokument. einzusehen bei der PPI AG.
- [Jo14b] JOHANNSEN, Dirk. *Exposé zur Bachelorarbeit: Maschinelle Grenzbetragsanalyse durch Softwaretechnologien und Anwendung mathematischer Prinzipien im Umfeld des Compliance eines Finanzdienstleisters*. Nordakademie, PPI AG, 2014.
- [MJ14] MANKEL, Magdalena und Johannßen, Dirk. *PPI-interne Kurzpräsentation: Compliance Cockpit*. PPI AG, 2014. Internes Dokument. einzusehen bei der PPI AG.

Inhalt der CD

Der Thesis ist eine CD mit folgendem Inhalt beigelegt:

Prototyp-Präsentation	Interne Präsentation der PPI AG von Magdalena Mankel und Dirk Johannßen zur Konzeption eines Prototyp
Exposé	Das Exposé zu dieser Bachelorthesis
Vorstudie	Die Vorstudie zu dieser Thesis mit Titel „Kategorisierung und Analysetauglichkeit von Red Flags“
Onlinequellen	Persistierte Online-Quellen aus Abschnitt Onlinequellen mit Abrufdatum aufgeführt
Java Projekt zu Fuzzy-Bsp.: Miami	Das zur Fuzzy-Untersuchung des Miami-Betrugs eingesetzte Java-Projekt inklusive der Bibliothek jFuzzyLogic und allen Konfigurationen
Java Projekt zu Fuzzy-Bsp.: Schneider	Das zur Fuzzy-Untersuchung des Schneider-Betrugs eingesetzte Java-Projekt inklusive der Bibliothek jFuzzyLogic und allen Konfigurationen
Prototyp	Java-Projekt und kausführbare Jar des Prototypen mit Namen Compliance-Cockpit

A. Prototyp

A.1. Motivation

Im Zuge der wirtschaftlich relevanten Fragestellung nach einer Möglichkeit zur maschinellen Betrugsanalyse ergaben sich verschiedene Ansätze, die in Form eines Prototypen, geschrieben in Java, untersucht wurden. Da es angedacht ist, die Erkenntnisse dieser Thesis in ähnlicher Form praktisch zu erproben und zu vertiefen, wird der Prototyp in diesem Teil des Anhangs grob vorgestellt.

A.2. GUI

Die Verwaltungsoberfläche des Prototyp ist in Abbildung A1 dargestellt. Nach Einlesen einer CSV-Datei mit den zu untersuchenden Daten, dargestellt in Abbildung A2, lässt sich auf dieser Oberfläche das gewünschte Limit definieren, auf das untersucht werden soll. Ferner führen Schaltflächen zu drei Konfigurationen zu folgenden Analysearten: Der Normalverteilung, Fuzzylogik und harte Prozentgrenzen. Ausgewählte Analysen werden auf der Hauptoberfläche markiert, wie in Abbildung A3 dargestellt.

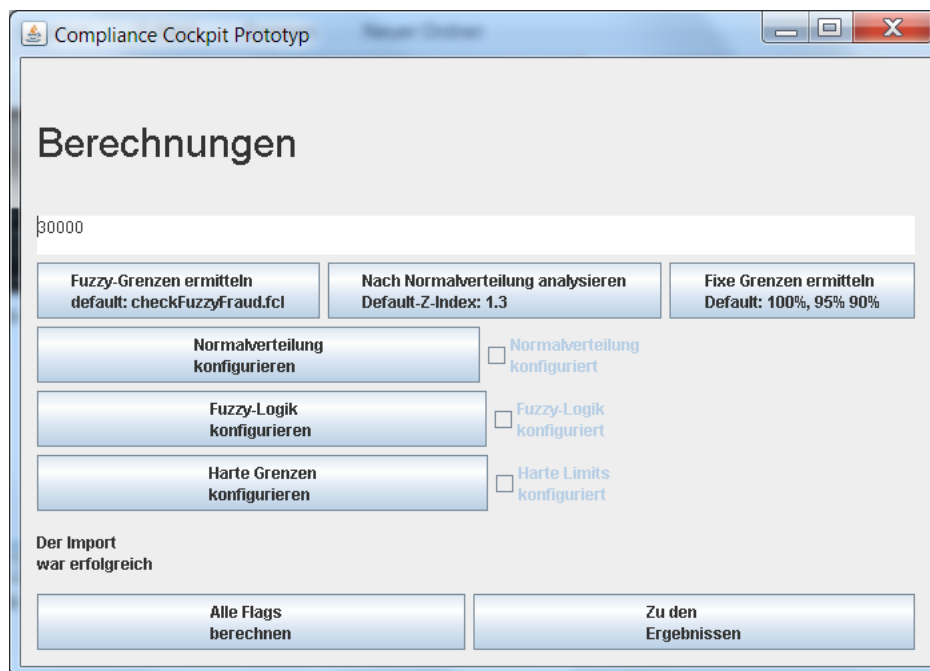


Abbildung A1.: Das Hauptmenü der Oberfläche des Prototyps

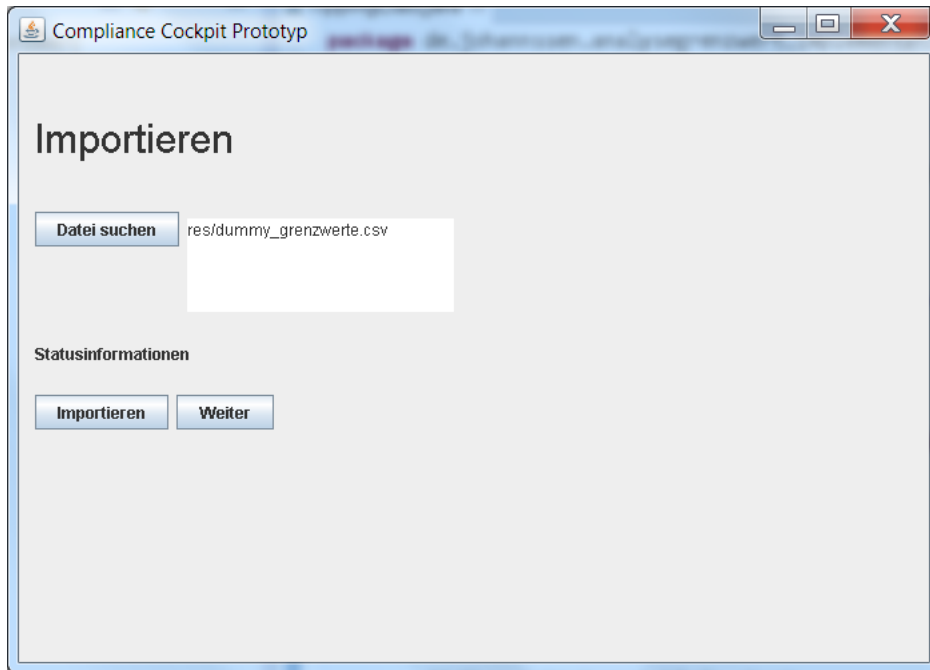


Abbildung A2.: Einführungsdialog zur Auswahl einer CSV-Datei

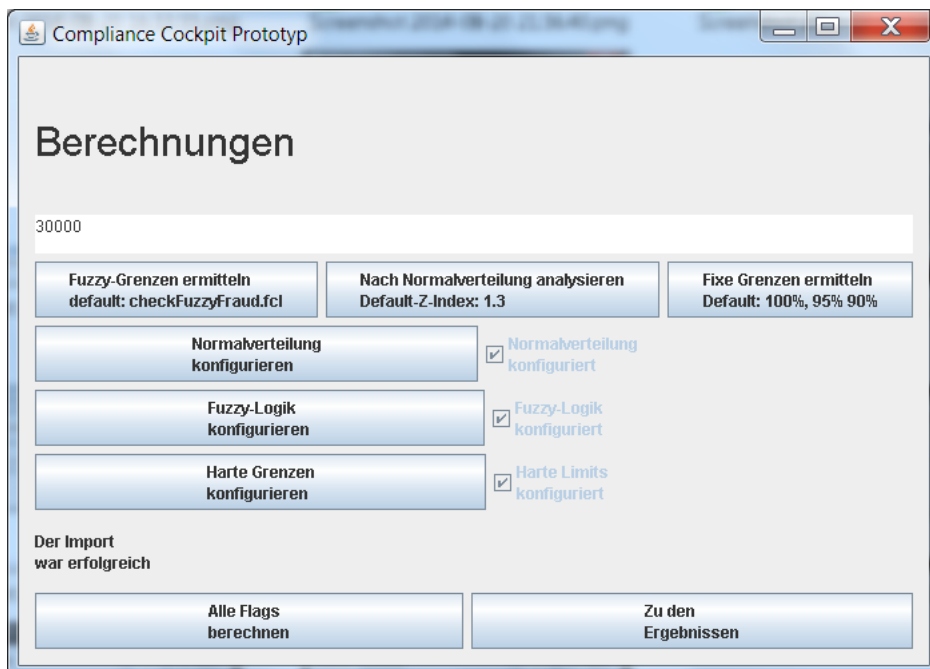


Abbildung A3.: Nutzerdialog für Analysekonfigurationen

Nachdem Analysen konfiguriert wurden, wie exemplarisch die erlaubte Höhe eines Z-Index aus Abbildung A4, markieren Checkboxen die Möglichkeit der Analyse. Eine vereinfachte, schnelle Konfiguration lässt sich durch die horizontalen Schaltflächen erreichen.

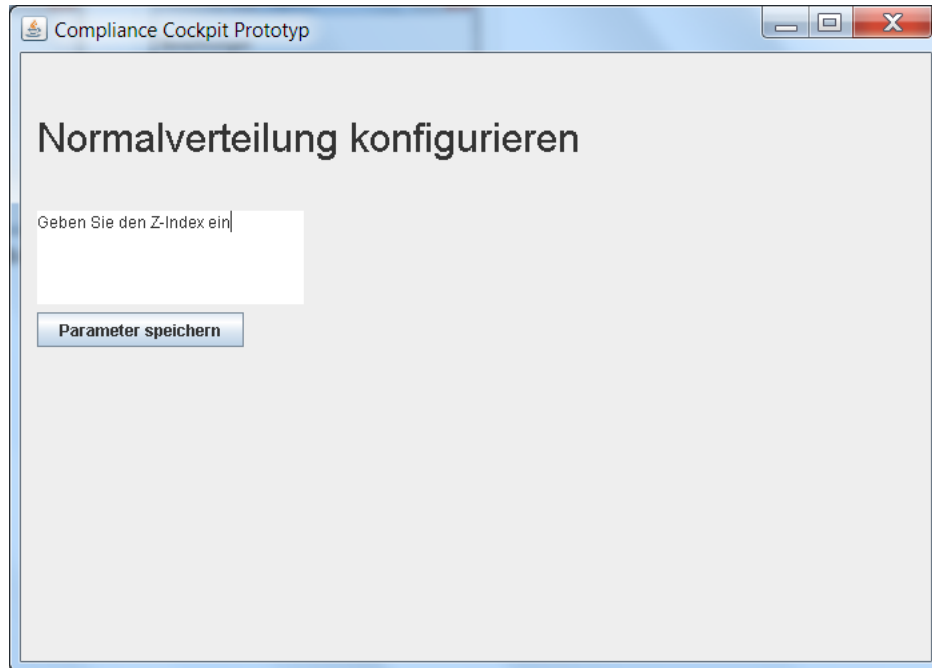


Abbildung A4.: Einfache Eingabe eines zu prüfenden Z-Index des Prototyps

Wurde die Analyse angestoßen, liegen Analyseergebnisse bereit, dargestellt in Abbildung A5. Diese lassen sich durch entsprechende Schaltfläche *Zu den Ergebnissen* aufrufen und betrachten. Die Ergebnisse sind nur grob und bezeichnen mit *true* eine Grenzbetragsverletzung und *false* die Ordnungsmäßigkeit des Wertes. Diese Aussagen entsprechen nicht den Erkenntnissen dieser Thesis, da sich der Prototyp in einem experimentellen, frühen Stadium befindet und die Fragestellung an diesen vereinfacht ist.

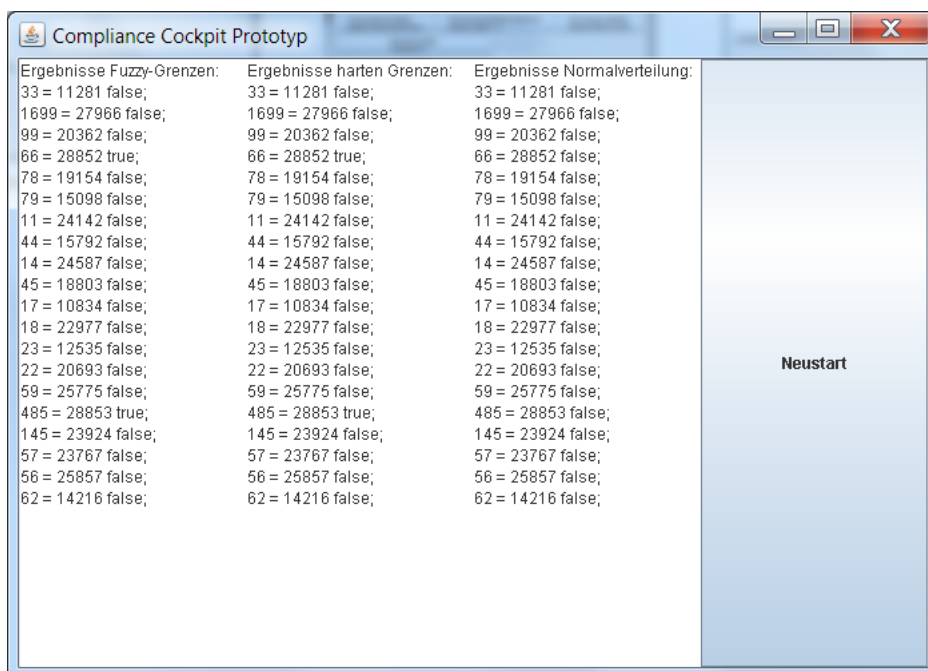


Abbildung A5.: Darstellung der Resultate aus Prüfroutinen

B. Benfordsches Gesetz

Das Benfordsche Gesetz wird an verschiedenen Stellen dieser Thesis behandelt. Hilfreiche Hintergrundinformationen und vertiefende Daten befinden sich in diesem Teil des Anhangs.

B.1. Historische Entwicklung

Der Astronom und Mathematiker Simon Newcomb nutzte Ende des 19. Jahrhunderts für Studien Bücher mit Logarithmentafeln. Dabei fiel ihm auf dass die Seiten aller Tafeln die mit der Ziffer *Eins* begannen wesentlich abgegriffener waren, als etwa die Seiten mit anderen Ziffern, wie zum Beispiel der *Vier*. Daraus folgerte Newcomb, dass logarithmische Tafeln der ersten Ziffer *Eins* häufiger zum Einsatz kamen, da die Ziffer *Eins* in den zu untersuchenden Daten häufiger auftrat. Dies erstaunte den Mathematiker, da die Logarithmentafeln für sehr unterschiedliche Studienfelder genutzt wurden und nicht etwa für ein einziges Gebiet.⁶³

Verwundert über diesen Umstand begann Newcomb zufällig gewählte, empirische Datenbestände auf deren Häufigkeiten erster Ziffern zu untersuchen. Schnell zeigte sich eine außergewöhnliche Häufigkeit der Verteilungen. Newcomb untersuchte mannigfaltige Datenmengen, darunter die Länge von Flüssen, Einwohnerzahlen von Städten, das Gewicht von Atomen und Finanzzahlen. Eine Auswahl der ermittelten Zahlenmenge ist dargestellt in Tabelle B1.

Group	Title	1	2	3	4	5	6	7	8	9	Count
A	Rivers, Area	31,0	16,4	10,7	11,3	7,2	8,6	5,5	4,2	5,1	335
B	Population	33,9	20,4	14,2	8,1	7,2	6,2	4,1	3,7	2,2	3259
C	Constants	41,3	14,4	4,8	8,6	10,6	5,8	1,0	2,9	10,6	104
D	Newspapers	30,0	18,0	12,0	10,0	8,0	6,0	6,0	5,0	5,0	100

Tabelle B1.: Auszug aus Datenerhebungen Newcombs⁶⁴

B.2. Erkenntnisgewinnung aus Newcombs Studien

Nach einer Vielzahl an Auswertungen, weit mehr als in Tabelle B1 dargestellt, ergaben sich für Newcomb die in Tabelle B2 aufgezeigten Wahrscheinlichkeiten für die ersten Ziffern der untersuchten Datenmengen.

⁶³Vgl. [Be38] S. 551 - 572

⁶⁴Vgl. [Sc10] S. 4

Ziffer	Wahrscheinlichkeit			
	Erste Stelle	Zweite Stelle	Dritte Stelle	Vierte Stelle
0		11,97%	10,18%	10,02%
1	30,10%	11,39%	10,14%	10,01%
2	17,61%	10,88%	10,10%	10,01%
3	12,49%	10,43%	10,06%	10,01%
4	9,69%	10,03%	10,02%	10,00%
5	7,92%	9,67%	9,98%	10,00%
6	6,69%	9,34%	9,94%	9,99%
7	5,80%	9,04%	9,90%	9,99%
8	5,12%	8,76%	9,86%	9,99%
9	4,58%	8,50%	9,83%	9,98%
Σ	100%	100%	100%	100%

Tabelle B2.: Wahrscheinlichkeitsverteilung der ersten Ziffer aller untersuchten Zahlen ⁶⁵

Den Namen *Benfordsches Gesetz* erhielt die Gesetzmäßigkeit aufgrund dessen, dass die Veröffentlichung der Forschungsergebnisse Newcombs unbeachtet blieben und in Vergessenheit gerieten, bis der Physiker Frank Benford sie im Jahr 1938 erneut veröffentlichte.⁶⁶

B.3. Testdatenerstellung zu Carslaw, 1988

Die nachfolgende Tabelle stellt die Einwohnerzahlen des Jahres 1950 deutscher Städte in Tausend dar. Die originalen Daten sind überschrieben mit *original*, die manipulierten mit *manipuliert*. Um Platz zu sparen, ist die Tabelle mittig geteilt und nebeneinander dargestellt worden. Der Grund für diese Tabelle ist das Nachstellen der Manipulationen von Daten nach gebrochenen Zahlen, wie sie Carslaw ausfindig machte.

manipuliert	original	manipuliert	original
3336,03	3336,03	122,81	122,81
1605,61	1605,61	118,50	118,50
831,94	831,94	118,18	118,18
617,57	617,57	117,29	117,29
610,30	605,41	116,49	116,49
605,41	594,94	115,39	115,39
532,04	532,04	114,07	114,07
512,71	507,35	112,23	111,59
507,35	500,52	111,59	111,45
504,33	496,49	111,45	109,72

⁶⁵[RB07] S. 27

⁶⁶Vgl. [RB07] S. 26 f.

500,52	494,19	110,23	109,54
444,55	444,55	104,89	104,79
444,30	444,30	104,79	103,28
410,78	410,78	103,28	100,67
363,22	363,22	100,67	99,89
362,46	362,46	94,79	94,79
315,46	315,46	93,27	93,27
302,82	293,37	92,63	89,03
301,11	289,80	88,37	88,37
300,85	289,12	80,31	80,31
260,31	260,31	78,68	78,68
254,45	254,45	78,44	78,44
245,63	245,63	75,53	75,53
238,28	238,28	72,29	72,29
223,76	223,76	71,13	71,13
220,74	220,74	66,44	66,44
202,81	202,81	65,53	65,53
201,45	198,84	64,64	64,64
188,65	188,65	62,50	63,48
185,18	185,18	60,87	60,87
171,88	171,88	59,36	59,87
162,13	162,13	54,14	54,14
153,61	153,61	50,01	50,01
151,17	149,59	45,74	45,74
147,85	147,85	40,27	40,52
146,40	146,40	39,45	40,27
133,11	133,11	38,79	38,79
131,53	129,81	34,83	34,83
124,88	124,88	32,68	32,68
123,87	123,87	25,42	25,42

Tabelle B3.: Manipulierte Einwohnerzahlen Deutscher Städte des Jahres 1950 in Einheiten zu tausend

12	16	17	14	13	14	13
15	19	12	14	16	17	10
18	11	11	10	17	17	19
10	10	11	10	25	28	22
29	20	28	26	21	24	22
23	25	31	36	33	38	39
39	30	37	31	40	43	44
45	42	48	40	55	52	51

53	56	51	62	64	64	65
66	77	71	78	72	86	83
85	89	93	94	97		

Tabelle B4.: Minuten bis zum Feierabend ab 17 Uhr, genutzt in Kapitel 5.3 zur Untersuchung auf Verteilung nach dem Benfordschen Gesetz

B.4. Originale vs. manipulierte Daten Carslaws

Aufbereitet wurden die Daten zum Carslaw-Betrug, indem die einzelnen Ziffern gezählt, die Prozentverteilung ermittelt und die Erwartung aufgestellt wurde. Dies geschah zum einen mit den originalen Daten und den manipulierten Daten, wie in Tabelle B6 aufgeführt. Das Ergebnis stellt für die originalen Daten die Tabelle B5 und für die manipulierten Daten die Tabelle B6 dar.

1. Zif.				2. Zif.				3. Zif.			
	Anz.	%	Erw.%		Anz.	%	Erw.%		Anz.	%	Erw.%
1	28	34,57	30,1	1	12	14,81	11,39	1	6	7,41	10,14
2	11	13,58	17,61	2	8	9,87	10,88	2	8	9,87	10,10
3	7	8,64	12,49	3	7	8,64	10,43	3	10	12,35	10,06
4	9	11,11	9,69	4	9	11,11	10,03	4	12	14,81	10,02
5	7	8,64	7,92	5	7	8,64	9,67	5	8	9,87	9,98
6	7	8,64	6,69	6	6	7,40	9,34	6	6	7,40	9,94
7	5	6,17	5,80	7	1	1,23	9,04	7	7	8,64	9,90
8	4	4,94	5,12	8	8	9,87	8,76	8	9	11,11	9,86
9	3	3,70	4,58	9	9	11,11	8,50	9	7	8,64	9,83
0	0	0	0	0	14	17,28	11,97	0	8	9,87	10,18
Σ	81	100	100	Σ	81	100	100	Σ	81	100	100

Tabelle B5.: Auswertung der Originaldaten

1. Zif.				2. Zif.				3. Zif.			
	Anz.	%	Erw.%		Anz.	%	Erw.%		Anz.	%	Erw.%
1	28	34,57	30,1	1	16	19,75	11,39	1	10	12,35	10,14
2	10	12,35	17,61	2	9	11,11	10,88	2	11	13,58	10,10
3	11	13,58	12,49	3	7	8,64	10,43	3	10	12,35	10,06
4	5	6,17	9,69	4	9	11,11	10,03	4	12	14,81	10,02
5	8	9,88	7,92	5	7	8,64	9,67	5	9	11,11	9,98
6	8	9,88	6,69	6	6	7,40	9,34	6	6	7,40	9,94
7	5	6,17	5,80	7	1	1,23	9,04	7	7	8,64	9,90
8	3	3,70	5,12	8	6	7,40	8,76	8	6	7,4	9,86
9	3	3,70	4,58	9	2	2,47	8,50	9	0	0	9,83
0	0	0	0	0	18	22,22	11,97	0	10	12,35	10,18
Σ	81	100	100	Σ	81	100	100	Σ	81	100	100

Tabelle B6.: Auswertung der manipulierten Daten

C. Octave

C.1. Anbindung einer Java-Bibliothek

Da die maschinelle Umsetzung der theoretischen Konzepte für deren Evaluierung relevant ist, soll in diesem Teil des Anhangs dargestellt werden, wie das Statistik-Programm Octave, dessen Funktionsumfang dem von MATLAB ähnlich ist, als Bibliothek in Java genutzt werden kann.

Die dafür notwendige Bibliothek nennt sich jOctave. Sie unterscheidet sich von anderen Bibliotheken dahingehend, dass sie selbst keine Funktionalität anbietet, sondern lediglich eine Schnittstelle - API - zu Octave bietet. Das bedeutet für die Arbeit mit Octave in Java, dass eine OctaveEngineFactory erstellt und mit Octave-Objekten gearbeitet werden muss. Sie werden in Java an Octave gegeben, von Octave mit Funktionen - definiert als Strings in Java - bearbeitet und als Octave-Objekte wieder zurückgegeben. Anschließend müssen die Octave-Objekte noch in Java-Objekte gewandelt werden.

C.2. Beispiel-Code der Nutzung von jOctave

Der nachfolgende Code berechnet durch den Octave-Kern die Methode *std()*, welche die Standardabweichung zurückgibt. Ergebnisse werden in Sets gespeichert und können anschließend durch Iterieren ausgelesen werden.

```
import dk.ange.octave.OctaveEngine;
import dk.ange.octave.OctaveEngineFactory;
import dk.ange.octave.type.OctaveDouble;

OctaveEngine octave = new OctaveEngineFactory().getScriptEngine();

HashMap<Integer, Integer> tempHash =
    (HashMap<Integer, Integer>) getDataImport().clone();
int length = tempHash.size();
int limit = getLimit();
double[] dataSet = new double[length];

octave.eval("format long g");
OctaveDouble a = null;
try {
    a = new OctaveDouble(dataSet, dataSet.length, 1);
} catch (Exception e) {
    System.out.println(e.getMessage());
}
octave.put("a", a);
```

```
String func = "" //
            + "function res = my_func(a)\n" //
            + " res = std(a)\n" //
            + "endfunction\n" //
            + "";
String func2 = "" //
            + "function res2 = my_func2(a)\n" //
            + " res2 = mean(a)\n" //
            + "endfunction\n" //
            + "";

octave.eval(func);
octave.eval(func2);

octave.eval("[b] = my_func(a);");
octave.eval("[c] = my_func2(a);");

OctaveDouble b = octave.get(OctaveDouble.class, "b");
OctaveDouble c = octave.get(OctaveDouble.class, "c");

double [] aDouble = b.getData();
double [] bDouble = c.getData();

octave.close();
```

D. Fuzzylogik

Für die technische Informatik ist interessant, wie sich AND, OR und NOT in boolesche Algebra abbilden lassen. So ließen sich daraus physische Schaltungen aus Transistoren erstellen. Es werden Mikrocontroller mit physischen Fuzzy-Schaltungen eingesetzt. Da dies nicht von großer Relevanz ist, wird hier nur die Software-Umsetzung der Fuzzylogik betrachtet.

D.1. Anbindung an Java

Eine elegante Art der Software-Umsetzung ist die Bibliothek *jFuzzyLogik* zur Anbindung mit Java. Sie ermöglicht es, mit dem Protokoll *Fuzzy Control Language* (FCL) Fuzzy-Systeme zu beschreiben.⁶⁷ Das Protokoll ist durch die *International Electronical Commission* (IEC) standardisiert (IEC 1331).⁶⁸

FCL ist intuitiv und schnell in der Praxis einsetzbar. Durch die Anbindung an Java mittels der Bibliothek *jFuzzyLogic* lassen sich defuzzifizierte Aussagen gut weiterverarbeiten.

D.2. Grafische Darstellung der Defuzzifizierung

Grafisch dargestellt ist die Defuzzifizierung in Abbildung D1. Die Bedienung einer Regel (beispielsweise IF ... THEN ...) wird angewandt auf eine Fuzzy-Menge und das Ergebnis daraus übertragen auf eine weitere Fuzzy-Menge, definiert in der Antwort der Kondition. Die Fläche ist die Grundlage für die Defuzzifizierung. Alle Flächen aller Regeln zusammen ergeben letztlich den Aussagewert.

⁶⁷Vgl. [So14]

⁶⁸Vgl. [IEC14]

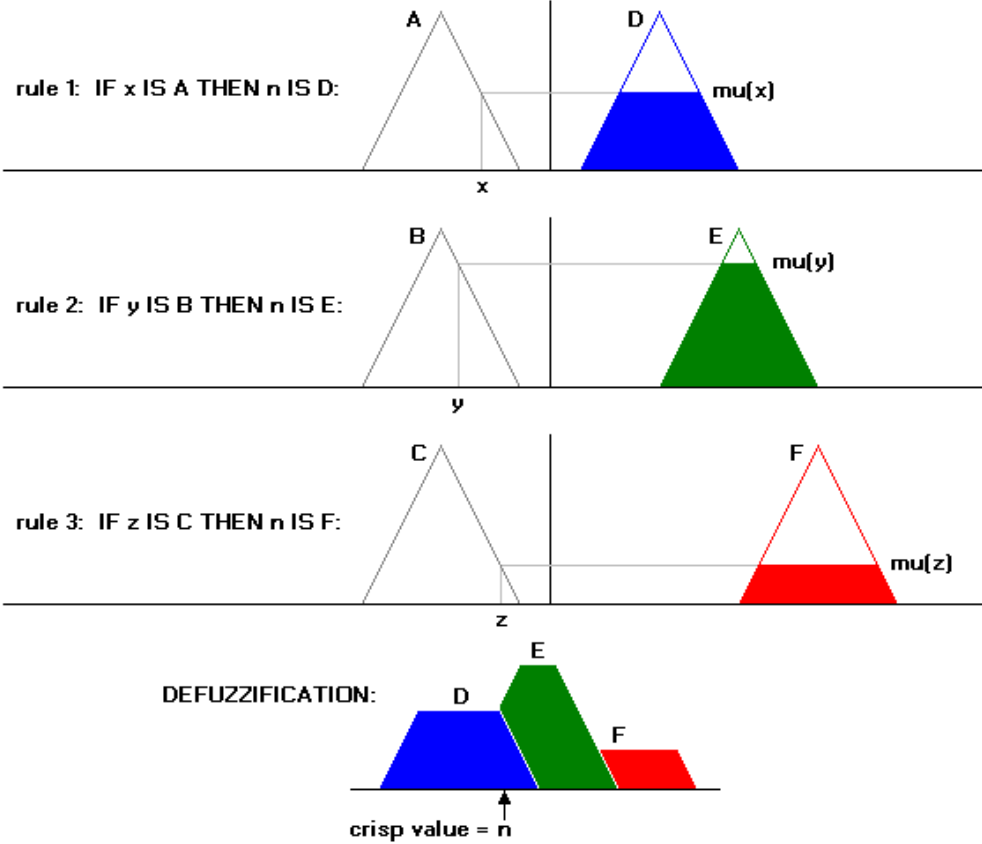


Abbildung D1.: Fuzzifizierung & Defuzzifizierung⁶⁹

⁶⁹[Wi15a]

E. Beispiel-Fuzzy-Code mit grafischen Abbildungen

Die Java-Bibliothek jFuzzyLogic ermöglicht es, in der Auszeichnungssprache Fuzzy-Control-Language (FCL) sowohl Fuzzy-Mengen, Zugehörigkeitsfunktionen und Regeln zu beschreiben, um anschließend anhand von Eingabeparameter Fuzzy-Berechnungen durchführen zu können. In diesem Abschnitt wird ein Beispiel aufgezeigt, welches FCL veranschaulichen soll.

E.1. Mathematische Eigenschaften der t-Norm

In der Thesis wird die t-Norm als zentraler Operand einer Fuzzy-Untersuchung beschrieben. Wie dieser Operand mathematisch beschrieben wird, ist Gegenstand dieses Abschnitts.

Präzise lässt sich die t-Norm anhand bereits bekannter, algebraischer Strukturen definieren, wobei die Fuzzy-Monotonie im Anschluss erläutert wird:

„Eine Funktion $\top : [0, 1]^2 \rightarrow [0, 1]$ heißt *t-Norm*, wenn die Bedingungen

- $\top(a, 1) = a$ (neutrales Element)
- $a \leq b \implies \top(a, c) \leq \top(b, c)$ (Monotonie)
- $\top(a, b) = \top(b, a)$ (Kommutativität)
- $\top(a, \top(b, c)) = \top(\top(a, b), c)$ (Assoziativität)

erfüllt sind.“⁷⁰

Die logische Monotonie, anders als die Monotonie der Analysis, besagt, dass wenn zu einer Menge von Annahmen weitere Annahmen hinzukommen, die bisher vorhandenen Annahmen weiterhin bestehen bleiben.

E.2. FCL-Anweisungen zu einer Fuzzy-Berechnung

Ein anschauliches Beispiel beschreibt das amerikanische Trinkgeld-Verhalten eines Restaurantbesuchers. Dessen Eingabeparameter sind zum einen die Qualität der Speisen, sowie zum anderen die Qualität der Bedienung. Danach berechnet sich das Trinkgeld.

Im Beispiel werden Variablen für die Eingabeparameter *food* und *service* angegeben. Die Relation beider bedingt die Ausgabevariable *tip*. Die Fuzzy-Mengen werden in FCL so beschrieben dass die Kurvenverläufe durch Tupel aufgezeigt werden. Beschrieben werden Relationen durch Zugehörigkeitsfunktionen, in FCL ausgedrückt durch *Rules*.

Nachfolgend sind die exemplarischen Anweisung⁷¹:

⁷⁰[Kr13] S. 23

⁷¹[So14]

```
FUNCTION_BLOCK tipper
VAR_INPUT
    service : REAL;
    food : REAL;
END_VAR

VAR_OUTPUT
    tip : REAL;
END_VAR

FUZZIFY service
    TERM poor := (0, 1) (4, 0) ;
    TERM good := (1, 0) (4,1) (6,1) (9,0);
    TERM excellent := (6, 0) (9, 1);
END_FUZZIFY

FUZZIFY food
    TERM rancid := (0, 1) (1, 1) (3,0) ;
    TERM delicious := (7,0) (9,1);
END_FUZZIFY

DEFUZZIFY tip
    TERM cheap := (0,0) (5,1) (10,0);
    TERM average := (10,0) (15,1) (20,0);
    TERM generous := (20,0) (25,1) (30,0);

METHOD : COG;
DEFAULT := 0;
RULEBLOCK No1
    AND : MIN;
ACT : MIN;
ACCU : MAX;

    RULE 1 : IF service IS poor OR food IS rancid
        THEN tip IS cheap;
    RULE 2 : IF service IS good
        THEN tip IS average;
    RULE 3 : IF service IS excellent AND food IS delicious
        THEN tip is generous;
END_RULEBLOCK
```

Die beschreibenden Zugehörigkeitsfunktionen *service*, *food* und *tip* hätten grafisch aufgezeigt folgende Struktur der Fuzzy-Mengen:



Abbildung E1.: Die Service-Qualität aus dem Beispiel grafisch dargestellt

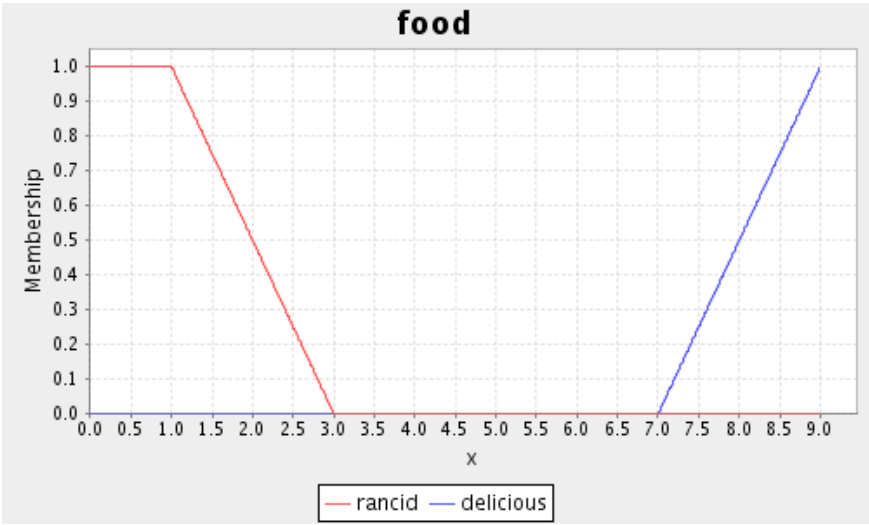


Abbildung E2.: Die Qualität der Speisen aus dem Beispiel grafisch dargestellt

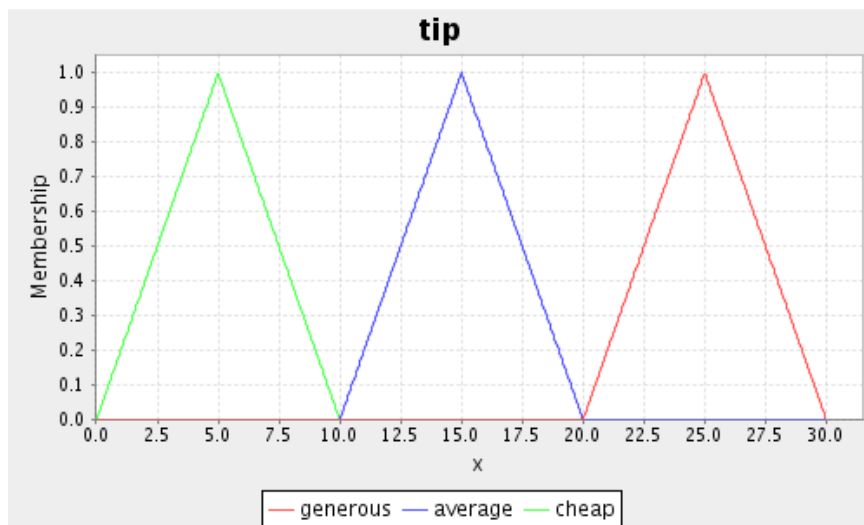


Abbildung E3.: Das resultierende Trinkgeld je nach Service- und Speisequalität aus dem Beispiel

F. FCL der Fuzzy-Analyse des Miami Betrugs

Um die Fuzzy-Untersuchung des Miami-Betrugs nachvollziehen zu können, ist in diesem Teil des Anhangs die FCL-Datei dargestellt. Sie findet sich ebenfalls auf der CD und kann dort als Java-Projekt ausgeführt werden. Von besonderem Interesse sind zum einen die Zugehörigkeitsfunktionen und Regeln. Wie eine FCL-Datei zu lesen ist, wird im Anhang [E.2](#) beschrieben.

```
FUNCTION_BLOCK tipper

VAR_INPUT
    homeTime : REAL;
    amountColleagues : REAL;
END_VAR

VAR_OUTPUT
    warningLevel : REAL;
END_VAR

FUZZIFY homeTime
    TERM early := (6.7, 0) (7,1) (7.9, 0);
    TERM normal := (7.7, 0) (8.2, 1) (8.4, 0);
    TERM late := (8.1, 0) (8.6,1) (8.8,0);
    TERM veryLate := (8.5, 0) (8.8, 1) (9.7, 1);
    TERM strangelyLate := (9.5,0)(15,1);
END_FUZZIFY

FUZZIFY amountColleagues
    TERM many := (80, 0) (90, 1) (100,1);
    TERM relativelyMany := (65, 0)(75, 1)(85, 0);
    TERM few := (35,0) (55, 1) (70, 0);
    TERM veryFew := (12,0)(25,1)(38,0);
    TERM almostNone := (0,1)(10,1)(17,0);
END_FUZZIFY

DEFUZZIFY warningLevel
    TERM none := (0,0)(0.3,0);
    TERM slight := (0.4,0) (0.55,1) (0.6,0);
    TERM moderate := (0.58,0) (0.67,1) (0.73,0);
    TERM severe := (0.7,0) (0.83,1) (0.91,0);
    TERM certain := (0.89, 0)(0.96, 1)(1, 1);
    METHOD : COG;
    DEFAULT := 0;
END_DEFUZZIFY
```

RULEBLOCK No1

AND : MIN;
ACT : MIN;
ACCU : MAX;

RULE 1 : IF homeTime IS late
 AND amountColleagues IS veryFew
 THEN warningLevel IS moderate;
RULE 2: IF homeTime IS veryLate
 AND amountColleagues IS few
 THEN warningLevel IS moderate;
RULE 3: IF homeTime IS late
 AND amountColleagues IS veryFew
 THEN warningLevel IS severe;
RULE 4: IF homeTime IS early
 OR amountColleagues IS many
 THEN warningLevel IS none;
RULE 5: IF homeTime IS early
 AND amountColleagues IS relativelyMany
 THEN warningLevel IS none;
RULE 6: IF homeTime IS normal
 AND amountColleagues IS relativelyMany
 THEN warningLevel IS none;
RULE 7: IF homeTime IS normal
 AND amountColleagues IS few
 THEN warningLevel IS slight;
RULE 8: IF homeTime IS late
 AND amountColleagues IS few
 THEN warningLevel IS moderate;
RULE 9: IF homeTime IS late
 AND amountColleagues IS veryFew
 THEN warningLevel IS severe;
RULE 10: IF homeTime IS veryLate
 AND amountColleagues IS veryFew
 THEN warningLevel IS severe;
RULE 11: IF homeTime IS veryLate
 AND amountColleagues IS almostNone
 THEN warningLevel IS certain;
RULE 12: IF homeTime IS late
 AND amountColleagues IS almostNone
 THEN warningLevel IS severe;
RULE 13: IF homeTime IS late
 AND amountColleagues IS veryFew

```
                THEN warningLevel IS severe;
RULE 14: IF homeTime IS veryLate
        AND amountColleagues IS veryFew
        THEN warningLevel IS severe;
RULE 15: IF homeTime IS late
        AND amountColleagues IS almostNone
        THEN warningLevel IS severe;
RULE 16: IF homeTime IS strangelyLate
        THEN warningLevel IS certain;
END_RULEBLOCK

END_FUNCTION_BLOCK
```

G. Jürgen Schneider Untersuchung

Die Analysen hypothetischer Daten zum Fall Jürgen Schneider sind Gegenstand dieses Abschnitts.

Datenbestand

Den dolosen Handlungen Jürgen Schneiders näher zu kommen und das Schneeballsystem aufzudecken, gestaltet sich dahingehend als herausfordernd, dass ein Schneeballsystem aus Daten nur schwer zu erkennen ist, da ein solches System sich nicht durch eine triviale Kennzahl äußert.

Eine Möglichkeit ist jedoch, die gewöhnliche Höhe von Krediten in Relation zu den zuvor beantragten Krediten anzuschauen. Für ein normales Geschäft würde erwartet werden, dass wenn die Geschäftstätigkeit in einem Jahr gleichbleibend verläuft, die benötigten Gelder ähnlich hoch ausfallen von Antrag zu Antrag.

So wird die benötigte Schrotmenge für die Versorgung von Rindern eines Landwirts nicht sprunghaft ansteigen, da dies keiner sinnvollen Ausführung des Kerngeschäfts entspräche. So wäre auch für eine Sanierung von Gebäuden ähnlicher Kategorien erwartet werden, dass die Höhe beantragter Kredite in zeitlich regelmäßigen Abständen in etwa konstant bleiben sollte.

Um diesen Sachverhalt zu prüfen, sind in Abbildung G1 zum einen die Faktoren (als Operatoren einer Multiplikation) der Höhe von beantragten Krediten in Relation zu den Kredithöhen zuvor auf der Ordinate, und ein Zeitverlauf auf der Abszisse dargestellt. Die rot-gefärbte Linie beschreibt die Kredithöhen der Grundmenge im Durchschnitt, die blaue Linie repräsentiert die Daten zum Betrugsfall Jürgen Schneider.

Ein Schneeballsystem ist nur schwer auszumachen. Es zeichnet sich dadurch aus, dass sich gewöhnlich konstante Datenwerte bei einem Schneeballsystem im Zeitverlauf nicht konstant sind, sondern einen eher logarithmischen Trend aufweisen. Dies ist damit zu erklären dass immer höhere Kreditsummen alte Kredite abdecken und den Schein des Erfolgs nach außen abdecken müssen. Ab einem Punkt bricht ein jedes Schneeballsystem in sich zusammen. Doch besteht die Vermutung dass bereits durch die logarithmische Grundgestalt, erkennbar in der Abbildung auf ein Koordinatensystem und evtl. durch einen χ^2 -Test ein solches System erkannt werden kann. In Abbildung G1 ist dieser Zusammenhang bereits an der blauen Linie erkennbar.

Sowohl die Grundmenge, als auch die Daten zum Betrugsfall Schneider sind hypothetisch und frei erfunden. Sie finden sich in Anhang G.

x	f(x)	$y = 10^x$	$\log_{10}(x)$	G
56	0,002698577	1,006233049	-2,568865153	0,9
56,5	0,008184091	1,019023245	-2,08702956	0,8
57	0,02140727	1,050527124	-1,669438712	0,95
57,5	0,04829558	1,117623643	-1,31609261	1,05
58	0,093974224	1,241578615	-1,026991253	1
58,5	0,157712224	1,43784550	-0,802134643	1,1
59	0,228284919	1,691550309	-0,641522778	1,04
59,5	0,284999659	1,9275234	-0,545155659	0,92
60	0,306878677	2,027116353	-0,513033286	0,87
60,5	0,284999659	1,9275234	-0,545155659	0,94
61	0,228284919	1,691550309	-0,641522778	1,05
61,5	0,157712224	1,43784550	-0,802134643	1,1
62	0,093974224	1,241578615	-1,026991253	1,15
62,5	0,04829558	1,117623643	-1,31609261	1
63	0,02140727	1,050527124	-1,669438712	0,87
63,5	0,008184091	1,019023245	-2,08702956	0,9

Tabelle G1.: Hypothetische Daten des Jürgen-Schneider Betrugs

Klasse	Von	Bis	Dichte	f(x)	n_j	p_{0j}	n_{0j}	χ^2
1	56	56,5	0,005485514	0,010882668	1	0,0025	0,994925928	0,00025876
2	56,5	57	0,013223179	0,029591361	3	0,007	2,785792598	0,0164709
3	57	57,5	0,02688831	0,069702851	7	0,0167	6,646105198	0,018844371
4	57,5	58	0,045678643	0,142269804	14	0,0348	13,84936891	0,0016383
5	58	58,5	0,063738001	0,251686448	25	0,0623	24,79355412	0,0017224
6	58,5	59	0,070572694	0,385997143	39	0,0966	38,44393785	0,008043
7	59	59,5	0,05671474	0,513284578	51	0,1294	51,49736602	0,0048037
8	59,5	60	0,021879018	0,591878336	59	0,1497	59,57616456	0,005572
9	60	60,5	-0,021879018	0,591878336	59	0,1497	59,57616456	0,005572
10	60,5	61	-0,05671474	0,513284578	51	0,1294	51,49736602	0,0048037
11	61	61,5	-0,070572694	0,385997143	39	0,0966	38,44393785	0,008043
12	61,5	62	-0,063738001	0,251686448	25	0,0623	24,79355412	0,0017224
13	62	62,5	-0,045678643	0,142269804	14	0,0348	13,84936891	0,0016383
14	62,5	63	-0,02688831	0,069702851	7	0,0167	6,646105198	0,018844371
15	63	63,5	-0,013223179	0,029591361	3	0,007	2,785792598	0,0164709
SUM					398	0,9955	396,1795044	0,114448103

Tabelle G2.: Hypothetische Daten des Jürgen-Schneider Betrugs

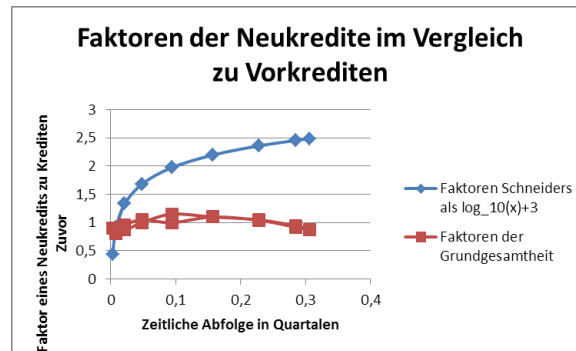


Abbildung G1.: Kredite Schneiders und der Grundgesamtheit

Aufbereitung der Daten und Vermutung

In einem χ^2 -Test der blauen Linie auf eine Normalverteilung musste die Nullhypothese abgelehnt werden, d.h. die blaue Linie scheint nicht Normalverteilt zu sein. Doch springt in Abbildung G1 ins Auge, dass die blaue Linie eine logarithmische Funktion darstellen könnte. Dies wäre für ein Schneeballsystem typisch.

Eine weitere Möglichkeit des χ^2 -Tests ist es, auch auf eine so genannte Lognormalverteilung, also einer logarithmischen Normalverteilung zu prüfen. Wird die Nullhypothese bestätigt, so wäre gewiss, dass die Linie eine logarithmische Funktion darstellt.

Von besonderer Wichtigkeit ist es hierbei zu betonen, dass das optische Erfassen eines solchen Zusammenhangs für geübte Menschen ein Leichtes sein kann, doch die Frage im Raum steht, wie derlei Zusammenhänge mathematisch und automatisiert erkannt werden könnten.

Den Verdacht auf Lognormalverteilung nachgehend, wurden die Daten erst grafisch dargestellt, die vermutete Verteilung erfasst, die Daten in Klassen eingeteilt und letztlich ein χ^2 -Test auf Lognormalverteilung durchgeführt. Dabei ergab sich eine Vermutung zum Erwartungswert μ von 60 (tausend) und eine geschätzte Streuung von σ^2 von 1,3. Das Histogramm ist dargestellt in Abbildung G2. Auffällig an dem Histogramm ist die optische Normalverteilung. Dies deutet darauf hin, dass die Vermutung der Lognormalverteilung zutrifft.

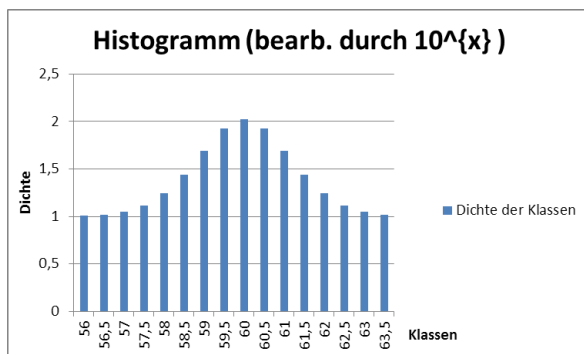


Abbildung G2.: Aufbereitete Daten in logarithmischer Darstellung der Häufigkeiten als Histogramm

Die Tabelle G3 zeigt die χ^2 -Aufbereitung der vorhandenen Daten auf.

Dabei berechnet sich die Wahrscheinlichkeit P_{0j} für ein Histogramm der Klassenlänge von 0,5 mit der ersten Klasse von 56 bis 56,5 exemplarisch folgendermaßen:

$$P_{0j} = P(56 < X < 56,5 | H_0) = P\left(-\frac{56-\mu}{\sigma} < \frac{X-\mu}{\sigma} \leq \frac{56,5-\mu}{\sigma} | H_0\right) = P\left(\frac{56-60}{1,3} < Z \leq \frac{56,5-60}{1,3}\right) = \Phi(-2,6923) - \Phi(-3,0769) = 0,0035 - 0,0001 = 0,0025,$$

mit der Verteilungsfunktion der Standardnormalverteilung Φ , wobei der Grundgedanke die Wahrscheinlichkeit für die Differenz der Ausprägung einer Zufallsvariable X zum Erwartungswert μ über der Standardabweichung σ ist.

Ferner berechnet sich darauffolgend die erwartete Häufigkeit n_{0j} derart:

$$n_{0j} = P_{0j} * n, \text{ angewandt auf die erste Berechnung: } 0,0025 * 398 = 0,994926.$$

Das Ergebnis stellt die Wahrscheinlichkeit der Übereinstimmung von beobachteter zur erwarteten Häufigkeit eines Wertes dar. In der Beispielberechnung also 99,49%.

Nachdem die Tabelle 5.2 aufgestellt wurde, lässt sich χ^2 folgendermaßen bestimmen:

$$\chi^2 = \sum_{j=1}^m \frac{(N_j - n_{0j})^2}{n_{0j}}$$

Angewandt auf das Beispiel heiße dies:

$$\chi^2 = \frac{(1-0,994926)^2}{0,994926} + \dots + \frac{(3-2,785793)^2}{2,785793} = 0,1144$$

Klasse	Intervall	Beob. Häuf.	Wahrschl.	Erwart. Häuf.
j	über bis	n_j	P_{0j}	n_{0j}
1	56 56,5	1	0,0025	0,994926
2	56,5 57	3	0,0413	2,785793
3	57 57,5	7	0,1525	6,646105

4	57,5	58	14	0,3	13,84937
5	58	58,5	25	0,297	24,79355
6	58,5	59	39	0,1525	38,44394
7	59	59,5	51	0,0413	51,49737
8	59,5	60	59	0,0058	59,57616
9	60	60,5	59	0,0058	59,57616
10	60,5	61	51	0,0058	51,49737
11	61	61,5	39	0,0058	38,44394
12	61,5	62	25	0,0058	24,79355
13	62	62,5	14	0,0058	13,84937
14	62,5	63	7	0,0058	6,646105
15	63	63,5	3	0,0058	2,785793
Σ			398	0,9955	396,1795

Tabelle G3.: Aufbereitete Daten der Kreditverteilungen Schneiders

Mit einem gewählten Signifikanzniveau von $\alpha = 0,05$ und einem Freiheitsgrad von $(2 - 1) * (15 - 1) = 14$ ergibt sich für die Testprüfgröße $P(0.1144|14) = 1$. Es gilt ferner für die Untersuchung $\chi^2 = 0,1144 < 1$, was bedeutet dass die Nullhypothese akzeptiert wird und gewiss ist, dass in der Tat die Lognormalverteilung verträglich ist und diese vorliegt.

Maschinell und mathematisch lässt sich an diesem Beispiel aussagen, dass die Lognormalverteilung für den Gegenstand der geforderten Geldmengen in Relation zu vorherigen Krediten unnormal ist und eine deutliche Warnung zur Folge haben muss. Daher wird das Schadenspotential mit 93% eingeschätzt.

Fuzzy-Untersuchung

Da eine Fuzzy-Untersuchung über die Bedenklichkeit der Höhe von Krediten im Kerngeschäft wertvolle Aussagen tätigen kann, bietet diese sich zur Untersuchung des Betrugsfalls Jürgen Schneider an.

Dabei sind zwei Fuzzy-Mengen die Eingabe der Untersuchung: Der Faktor eines Kredits im Vergleich zur Höhe des Kredits davor, entnommen der Daten aus Grafik G1, und die Ratenanzahl pro Jahr, in denen der Kredit getilgt wurde. Ist der Faktor hoch und die Ratenzahl bis zur Tilgung gering, so ist dies ein Hinweis auf ungewöhnliche Finanzbewegungen. Die genutzten Eingaben liegen als Fuzzy-Controlling-Language (FCL) Konfiguration im Anhang D vor.

Die eingesetzten Werte und die daraus resultierenden defuzzifizierten Warnstufen können der folgenden Tabelle G4 entnommen werden.

Nr.	Kred. Faktor.	Raten / Jahr	Fuzzy-Ergebnis
1	0,43	4	0,52
2	0,91	2	0,66
3	1,33	1	0,76
4	1,68	5	0,75
5	1,97	4	0,81
6	2,20	3	0,86
7	2,36	6	0,81
8	2,45	10	0,63
9	2,49	2	0,96
10	2,45	4	0,93
11	2,36	5	0,86
12	2,20	3	0,86
13	1,97	6	0,80
14	1,68	1	0,85
15	1,33	4	0,70
16	0,91	3	0,57
		Σ	12,33
		\bar{x}	0,77

Tabelle G4.: Fuzzy-Untersuchung zum Fall Jürgen Schneider

Die Fuzzy-Untersuchung ergibt folglich für die Eingabedaten einen Warnwert von etwa 77%.

sigma-Abweichungen

Aufgrund dessen, dass die Kreditdaten Jürgen Schneiders im hypothetischen Beispiel lognormal-verteilt sind, ist eine Betrachtung der σ -Abweichungen interessant.

Dafür werden aus den übrigen Durchschnittswerten aus der in Anhang G dargestellten Tabelle G2 die Werte der Grundmenge G und $\log_{10}(x)$ herangezogen. Errechnet wurden $\bar{x} = 1,042666667$, $Var(x) = \frac{(0,9-1,042)^2+(0,8-1,042)^2+\dots+(0,9-1,042)^2}{15} = 0,013496313$ und die Standardabweichung $\sigma_x = \sqrt{Var(x)} = 0,1162$. Somit ergibt sich für die σ -Abweichungen folgende Tabelle:

Nr.	$y = 10^x$	$\log_{10}(x)$	$\mu + / - \sigma$	$P(\mu + / - \sigma)$
1	0,431134847	0,431134847	3,710	99,99
2	0,91297044	0,91297044	7,857	99,99
3	1,330561288	1,330561288	11,45	99,99
4	1,68390739	1,68390737	14,49	99,99
5	1,973008747	1,973008747	16,98	99,99

6	0,157712224	2,197865357	18,91	99,99
7	0,228284919	2,358477222	20,3	99,99
8	0,284999659	2,454844341	21,13	99,99
9	0,306878677	2,486966714	21,40	99,99
10	0,284999659	2,454844341	21,13	99,99
11	0,228284919	2,358477222	20,3	99,99
12	0,157712224	2,197865357	18,91	99,99
13	0,093974224	1,973008747	16,98	99,99
14	0,04829558	1,68390739	14,49	99,99
15	0,02140727	1,330561288	11,45	99,99
16	0,008184091	0,91297044	7,86	99,99

Tabelle G5.: σ -Abweichungen Jürgen Schneider

Aus der σ -Abweichungsuntersuchung wird ersichtlich dass zu 99,99 Prozent eine Auffälligkeit vorliegt. Dies ist daher als eine sehr deutliche Warnung zu interpretieren. Knapp ausgedrückt wird lediglich bestätigt, dass die blaue Linie in ihrer Streuung enorm von der Grundmenge, der roten Linie, abweicht.

Aufgrund dessen, dass der χ^2 -Test ohnehin bereits mit der Lognormalverteilung die Möglichkeit eines Schneeballsystems aufgezeigt hat, macht die erneute Prüfung auf σ -Abweichungen wenig Sinn und wird daher aus der Menge der Warnhinweise herausgelassen. Der Grund dafür ist, dass die σ -Abweichungsuntersuchung für eine Lognormalverteilung nur wenig aussagekräftige Ergebnisse liefert, da sie für die Normalverteilung genutzt wird. Hinter der Lognormalverteilung steht ein logarithmisches System.

Anwendung des Bewertungsmodells

Aus den Analysen Chi-Quadrat und Fuzzy dieses Kapitels geht eine Menge von Warnwerten mit $M = \{93, 77\}$. Dementsprechend stellt dies den Trivialfall des Bewertungsmodells dar.

Das Maximum ergibt sich mit $\max(M) = 93$. Das arithmetische Mittel ergibt $\bar{x} = 85$, die Varianz $Var(M) = 64$ und die Standardabweichung entsprechend $\sigma_X = 8$

Für das Bewertungsmodell im Falle Jürgen Schneiders ergibt sich entsprechend:

$$w_x = \begin{cases} 1 & \text{falls } \exists x \in M : x \geq 0, 95 \\ \frac{(85+8)+93}{2} & \text{sonst} \end{cases}$$

Dies ergibt eine Warnaussage von 93 % und somit ein deutliches Warnpotential.