

# Wissensextraktion zum datengetriebenen Qualitätsmanagement

Zur Erlangung des akademischen Grades eines  
Doktors der Wirtschaftswissenschaften

(Dr. rer. pol.)

von der Fakultät für  
Wirtschaftswissenschaften  
der Universität Karlsruhe (TH)

genehmigte

DISSERTATION

von

**Diplom-Wirtschaftsmathematiker**

**Christian Manuel Strobel**

Tag der mündlichen Prüfung: 10. Februar 2009  
Referent: Prof. Dr. Georg Bol  
Korreferent: Prof. Dr. Karl-Heinz Waldmann  
2008 Karlsruhe



# Vorwort

Die vorliegende Arbeit entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Institut für Statistik und Mathematische Wirtschaftstheorie sowie am Forschungszentrum der Daimler AG in Ulm. An dieser Stelle möchte ich die Gelegenheit nutzen, mich bei all denjenigen zu bedanken, ohne deren Zutun und Unterstützung ein erfolgreiches Gelingen dieser Dissertation nicht möglich gewesen wäre.

Mein ganz herzlicher Dank gilt meinem Doktorvater Herrn Prof. Dr. Georg Bol für die Betreuung, die vielen hilfreichen Diskussionen und Anregungen, sowie für das jederzeit entgegengebrachte Interesse am Fortgang der Arbeit. Ebenfalls möchte ich mich bei Prof. Karl-Heinz Waldmann für die unkomplizierte Zusammenarbeit und die Übernahme des Korreferats bedanken. Ein besonderer Dank gilt Prof. Dr. Gholamreza Nakhaeizadeh und Prof. Dr. J.-F. Luy. Sie ermöglichten es, dass diese Arbeit am Forschungszentrum verfasst werden konnte und trugen somit zum Gelingen maßgeblich bei.

Den Kollegen am Forschungszentrum möchte ich für die fachliche Betreuung und die hilfreiche Diskussionsbereitschaft danken. Einen besonderen Dank möchte ich an dieser Stelle Dr. Tomas Hrycej aussprechen. Seine Unterstützung und Betreuung waren nicht nur fachlich von großer Bedeutung für mich.

Des Weiteren möchte ich mich bei M. Wahrheit für die kritische Korrektur meiner Arbeit und die entgegen gebrachte Freundschaft bedanken.

Ebenso möchte ich meiner Familie, allen voran meinen Eltern, für die Unterstützung während meines Studiums und der Promotion danken. Nicht zuletzt danke ich ganz besonders meiner Freundin Christina. Ihr liebevolles Verständnis und ihre Geduld haben einen großen Anteil am Erfolg dieser Arbeit.



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
1.1	Motivation und Ziele der Arbeit . . . . .	2
1.2	Inhalt der Arbeit . . . . .	4
<b>2</b>	<b>Statistische Prozesslenkung</b>	<b>9</b>
2.1	Qualitätsfähigkeitsindizes . . . . .	14
2.1.1	QFI unter Annahme einer Normalverteilung . . . . .	16
2.1.2	QFI ohne Verteilungsannahme . . . . .	20
2.2	Allgemeiner Ansatz für QFI . . . . .	25
2.2.1	Stetige Prozessparameter . . . . .	26
2.2.2	Nominale Prozessparameter . . . . .	29
2.3	Konfidenzintervalle für QFI . . . . .	31
2.3.1	Konfidenzintervalle für QFI unter Normalverteilungsannahme . . . . .	32
2.3.2	Bootstrapverfahren zur Intervallschätzung für QFI . . . . .	34
2.3.3	Simulationsstudie . . . . .	39
2.4	Zusammenfassung und Bewertung . . . . .	51
<b>3</b>	<b>Fertigungsprozessoptimierung</b>	<b>53</b>
3.1	Varianzanalyse . . . . .	54
3.2	Root Cause Analysis . . . . .	59
3.2.1	Optimierungsproblem . . . . .	67
3.2.2	RCA Algorithmus . . . . .	70
3.2.3	Ergebnisverifikation . . . . .	80
3.3	Konzeptevaluierung . . . . .	81
3.3.1	Simulationsergebnisse . . . . .	82
3.3.2	Beispiel: Prozessoptimierung . . . . .	87
3.4	Zusammenfassung und Bewertung . . . . .	93
<b>4</b>	<b>Wissensextraktion mittels Assoziationsregeln</b>	<b>97</b>
4.1	Assoziationsregeln . . . . .	98
4.2	Interessantheitsmaße zur Bewertung von Regeln . . . . .	105
4.2.1	Kontingenzanalyse . . . . .	106
4.2.2	Statistische Interessantheitsmaße . . . . .	111

4.2.3	Subjektive Interessantheitsmaße . . . . .	129
4.3	Multiples Hypothesenproblem . . . . .	134
4.3.1	Lösungsansätze . . . . .	135
4.3.2	Vergleich der Verfahren . . . . .	144
4.4	Felddaten Analyse . . . . .	150
4.4.1	Modellierung . . . . .	151
4.4.2	Evaluierung und Ergebnisse . . . . .	163
4.5	Zusammenfassung und Bewertung . . . . .	174
<b>5</b>	<b>Zusammenfassung und Ausblick</b>	<b>179</b>
<b>A</b>	<b>Anhang</b>	<b>185</b>
A.1	Sätze und Definitionen . . . . .	185
A.2	Tabellen und Ergebnisse . . . . .	187

# Tabellenverzeichnis

2.1	Grenzwerte für Fähigkeitsindizes . . . . .	20
2.2	$C'_p$ Kategorien für Prozesssimulation . . . . .	41
2.3	Kennwerte der Simulation für $E_{ci}$ . . . . .	43
2.4	Ergebnisse für $E_{ci}$ aufgliedert nach Prozessen . . . . .	44
2.5	Ergebnisse für $E_{ci}$ aufgliedert nach Stichprobenumfang . . . . .	44
2.6	Prozentuale Überdeckung für $E_{ci}$ . . . . .	45
2.7	Ergebnisse der Simulation für $C'_p$ und $C'_{pk}$ . . . . .	47
2.8	Ergebnisse der Simulation für $C'_{pm}$ und $C'_{pmk}$ . . . . .	48
2.9	Prozentuale Überdeckung nach Prozessen . . . . .	49
2.10	Prozentuale Überdeckung nach Fähigkeitsindizes . . . . .	49
2.11	Prozentuale Überdeckung nach Fähigkeitsindizes und Stichprobenumfang . . . . .	50
3.1	Wertemengen eines Beispielprozesses . . . . .	62
3.2	Variablenausprägung Beispielprozess . . . . .	65
3.3	Beispiel Produktionsprozess . . . . .	66
3.4	Binarisierung der Variablen $Y^i$ . . . . .	78
3.5	Mengengerüst analysierter Prozesse . . . . .	83
3.6	Benötigte Rechenzeit für nicht-parametrische Indizes . . . . .	83
3.7	Prozessmerkmale . . . . .	87
3.8	Ergebnisliste Prozessoptimierung . . . . .	88
3.9	Übertragung der Zusammenhänge auf Originalprozess . . . . .	90
4.1	Beispiel Transaktionsdatenbank . . . . .	99
4.2	$2 \times 2$ Kontingenztabelle für die Regel $A \rightarrow B$ . . . . .	107
4.3	$2 \times 2$ Wahrscheinlichkeitsverteilung für die Regel $A \rightarrow B$ . . . . .	109
4.4	Beispiel Kontingenztabelle . . . . .	132
4.5	Kontingenztabelle für Hypothesentests . . . . .	135
4.6	Beispiel: Alternativer Ansatz . . . . .	143
4.7	Auswahlkriterien Assoziationsanalyse . . . . .	145
4.8	Gefundene Zusammenhänge ohne Korrektur . . . . .	146
4.9	Gefundene Zusammenhänge durch Kontrolle der FDR . . . . .	146
4.10	Kontingenztabelle für die Regel $(A_1 \wedge A_2) \rightarrow B$ . . . . .	162
4.11	Kontingenztabelle für die Regel $(A_1 \wedge \neg A_2) \rightarrow B$ . . . . .	162

4.12	Mengengerüst Felddaten Analyse . . . . .	164
4.13	Vorauswahlkriterien Felddaten Analyse . . . . .	165
4.14	Beispielerggebnisse Felddaten Analyse . . . . .	169
4.15	Übersicht Szenarien Felddaten Analyse . . . . .	173
A.1	Simulationsergebnisse $E_{ci}$ . . . . .	187
A.2	Simulationsergebnisse für “guten“ Prozess . . . . .	188
A.3	Simulationsergebnisse für “sehr guten“ Prozess . . . . .	189
A.4	Simulationsergebnisse für “schlechten“ Prozess . . . . .	190
A.5	Beispielevaluation: Gemeinsame Sub-Prozesse für $E_{ci}$ . . . . .	191
A.6	Beispielevaluation: Gemeinsame Sub-Prozesse für $C_{pk}$ . . . . .	191
A.7	Beispielevaluation: Trainings- und Test-Menge für $E_{ci}$ . . . . .	192
A.8	Beispielevaluation: Trainings- und Test-Menge für $C_{pk}$ . . . . .	193

# Abbildungsverzeichnis

2.1	Ursache-Wirkungs-Diagramm nach K. Ishikawa . . . . .	10
2.2	Links: Gut-Schlecht-Betrachtung Rechts: Taguchi-Verlustfunktion . . . . .	11
2.3	Prozessverhalten vs. Verlustfunktion . . . . .	12
3.1	Prozessdichteplot . . . . .	66
3.2	Schematische Darstellung des Optimierungsproblems . . . . .	68
3.3	Baumstruktur für den RCA - Algorithmus für zwei Ebenen . . . . .	71
3.4	Vergleich: Kombinatorisch vs. 1. Branch and Bound Prinzip . . . . .	84
3.5	Vergleich: Binarisierung 1. Branch and Bound Prinzip . . . . .	85
3.6	Vergleich: 1. und 2. Branch and Bound Prinzip . . . . .	86
3.7	Dichteplot optimaler Sub-Prozess vs. Originalprozess für $E_{ci}$ . . . . .	89
3.8	Dichteplot optimaler Sub-Prozesse für $C_{pk}$ . . . . .	90
3.9	Dichteplot optimaler Sub-Prozesse für $E_{ci}$ . . . . .	91
3.10	Dichteplot optimaler Sub-Prozesse für $E_{ci}$ . . . . .	92
4.1	Oszillationsphänomen des Standard- und Wilsonintervalls . . . . .	119
4.2	Clopper-Pearson: Überdeckungswahrscheinlichkeit . . . . .	120
4.3	Jeffrey's-Intervall: Überdeckungswahrscheinlichkeit . . . . .	122
4.4	Agresti-Coull-Intervall: Überdeckungswahrscheinlichkeit . . . . .	123
4.5	Überdeckungswahrscheinlichkeit . . . . .	124
4.6	Intervallfehler . . . . .	125
4.7	Durchschnittliche Überdeckungswahrscheinlichkeit . . . . .	126
4.8	Konfidenzintervall für $n = 20$ . . . . .	127
4.9	Konfidenzintervall für $n = 100$ . . . . .	128
4.10	Interessantheitsfacetten . . . . .	130
4.11	Alternativer Ansatz: Vergleich Datensätze . . . . .	148
4.12	CRISP-DM Prozessmodell . . . . .	153
4.13	Testprozedur Felddaten Analyse . . . . .	156
4.14	Schematische Darstellung der Felddaten Analyse . . . . .	158
4.15	Histogramm Lift und SafeLift . . . . .	166
4.16	Nutzenfunktion der gefundenen Zusammenhänge . . . . .	167
4.17	Anteil zufälliger Regeln Szenario 1 . . . . .	171
4.18	Anteil zufälliger Regeln Szenario 3 . . . . .	173



# 1 Einleitung

Ein Unternehmen, das auf dem globalen Markt mit den unterschiedlichsten internationalen Mitbewerbern langfristig bestehen will, muss konstant qualitativ hochwertige Produkte herstellen und verkaufen. Vor allem für einen Automobilhersteller des Premiumsegments ist es notwendig, seine Produkte und Produktionsprozesse so zu gestalten, dass sie den hohen Qualitätsansprüchen der Kunden gerecht werden. Schlechte Qualität hat zum einen einen nachhaltigen Imageverlust des Unternehmens zur Folge, der zu einer Abwanderung des Kunden zur Konkurrenz führt. Zum anderen kann das Nichterreichen von geforderten Qualitätsstandards zu zukünftigen, unvorhersehbaren, finanziellen Verlusten führen, da es aufgrund von Produktionsausfällen oder minderwertigen, ausgelieferten Produkten zu Regressforderungen seitens der Kunden oder zu kostspieligen Rückrufaktionen kommen kann. Das Ziel eines Unternehmens muss es somit sein, produzierte, qualitativ nicht ausreichende Produkte möglichst früh im Wertschöpfungsprozess zu identifizieren und durch solche Produkte zu ersetzen, die den Qualitätsansprüchen und -standards des Unternehmens und der Kunden genügen. Eine zentrale Funktion zur Überwachung der Qualität nimmt dabei das Qualitätsmanagement eines Unternehmens ein. Das Qualitätsmanagement hat die querschnittliche Aufgabe, Methoden und Konzepte umzusetzen und zu etablieren, welche die Qualität der produzierten Güter sicherstellen oder verbessern.

Damit Qualität ganzheitlich messbar und dokumentierbar ist, werden während der Produktion bzw. im Laufe des Wertschöpfungsprozesses eines Produktes eine Vielzahl unterschiedlicher Prozessparameter und Test- bzw. Messergebnisse erfasst und gespeichert. Diese Daten werden in der Regel unmittelbar nach der Erfassung ausgewertet und darauf aufbauend wird die Qualität bestimmt bzw. werden weitere Maßnahmen, wie Ausschluss minderwertiger Produkte, schlimmstenfalls die Einstellung der Produktion, abgeleitet. Zur Sicherung und Dokumentation der Qualität werden jedoch nicht nur rein produktionsbezogene Daten erfasst, sondern auch solche Daten, die das Verhalten des Produktes im Kundeneinsatz – beispielsweise in Form von Garantie- und Kulanzdaten – beschreiben. Einem Unternehmen wie beispielsweise Automobilherstellern stehen somit im Idealfall eine Vielzahl unterschiedlicher Daten eines Produktes zu den verschiedenen Stufen im Wertschöpfungsprozess zur Verfügung.

Für eine langfristige Wettbewerbsfähigkeit eines Unternehmens ist es jedoch nicht aus-

reichend, Daten, die während des Wertschöpfungsprozesses erfasst werden, lediglich zur Überprüfung der Qualität während der Produktion zu verwenden bzw. zu Dokumentationszwecken abzuspeichern. Vielmehr müssen die erfassten Daten als Rohinformationen betrachtet werden, deren geeignete Analyse und Verwendung es ermöglichen, vorhandenes Optimierungspotential bezüglich der Qualität der gefertigten Produkte zu identifizieren. Dieses extrahierte Potential lässt sich dann auf existierende oder neue Prozesse und Produkte übertragen und führt langfristig zu einer verbesserten Qualität und folglich zu einer Stärkung der Marktposition eines Unternehmens.

Der bis jetzt nicht näher spezifizierte Begriff, die Bestimmung von Optimierungspotential, ist im Kontext des Qualitätsmanagements mit der Identifikation von signifikanten, qualitätsrelevanten Einflüssen bzw. Zusammenhängen gleichzusetzen. Denn gelingt es, Zusammenhänge aufzuzeigen, die einen entscheidenden Einfluss auf die Qualität gefertigter Produkte haben, so kann dieses Wissen durch das Unternehmen genutzt werden um die Qualität eines Produktionsprozesses zu steigern. Die Entwicklung und Modellierung von solchen Verfahren und Methoden, mit deren Hilfe signifikante, qualitätsrelevante Zusammenhänge anhand von erfassten (Qualitäts-)Daten extrahiert werden können, bilden die Kernthemen dieser Abhandlung. Dabei gilt es, die Anforderungen der industriellen Praxis zu beachten, zuallererst die Forderung, dass ökonomische Gesichtspunkte bei der Identifizierung von Zusammenhängen berücksichtigt werden müssen. Denn erst der Einsatz von Verfahren und Modellen, die sowohl die statistische Evidenz eines Zusammenhangs als auch seine wirtschaftliche Bedeutung berücksichtigen, ermöglichen es, extrahiertes Wissen im Sinne eines signifikanten und qualitätsrelevanten Zusammenhangs, nachhaltig in den Qualitätsverbesserungsprozess eines Unternehmens rückzukoppeln.

Im weiteren Verlauf dieses Kapitels werden zunächst die Ziele und die wissenschaftlichen Anforderungen motiviert und formuliert. Im Anschluss wird eine Zusammenfassung dieser Arbeit gegeben, wobei spezielles Augenmerk auf die Neuerungen, die die hier vorgestellten Verfahren implizieren, gelegt wird.

## 1.1 Motivation und Ziele der Arbeit

Betrachtet man die Informationen, die entlang der Wertschöpfungskette eines Unternehmens für die produzierten Güter erfasst werden, so bietet die Kombination dieser Daten unterschiedlichste Möglichkeiten, Optimierungspotential für die Qualität der gefertigten Produkte zu identifizieren. Auf der einen Seite können die zu einem gewissen Zeitpunkt des Wertschöpfungsprozesses erfassten Daten eines Produktes in einen vorgelagerten Zeitpunkt des Produktentstehungsprozesses gespiegelt und gemeinsam analysiert werden. Zum Beispiel können Informationen, die in Form von Garantie- und Kulanzdaten

gesammelt wurden, in die Produktion rückgekoppelt werden, um so Zusammenhänge zwischen Produktion und Gebrauch des Produktes zu identifizieren. Das so extrahierte Wissen kann anschließend dazu genutzt werden, Fehler, die erst beim Einsatz des Produktes beim Kunden auftreten, bereits in der Produktion zu beheben. Auf der anderen Seite können Informationen aus der gleichen Stufe des Produktentstehungsprozesses genutzt werden, um die Qualität in diesem Schritt der Wertschöpfungskette zu verbessern. Beispielsweise können Stichproben aus der laufenden Produktion gezogen, näher analysiert und das so gewonnene Wissen zur Steigerung der Qualität genutzt werden. Beiden beschriebenen Fällen liegt dabei die Idee zugrunde, dass sich die zu analysierenden Informationen in zwei Arten von Daten einteilen lassen, zum einen in die Menge der erklärenden Variablen, die so genannten Einflussgrößen bzw. Prozessmerkmalen. Zum anderen in die Menge der abhängigen Variablen, welche die Qualität näher beschreiben und somit auch als Qualitätsmerkmale bezeichnet werden können.

Legt man diese Betrachtungsweise zugrunde, so beschreibt sie eine Grundaufgabe des Qualitätsmanagements: Mittels geeigneter Methoden sind qualitätsmessende Größen mit Hilfe erfasster Prozessmerkmale zu erklären, um darauf aufbauend Maßnahmen zur Steigerung der Qualität abzuleiten. Um nun einen nachhaltigen Mehrwert zur Qualitätssteigerung zu generieren, reicht es jedoch nicht aus, lediglich die Zusammenhänge zwischen diesen beiden Größen geeignet zu beschreiben. Vielmehr müssen Verfahren und Methoden modelliert und etabliert werden, die es ermöglichen, wenige, jedoch für die Qualität signifikante Zusammenhänge zwischen den Prozessmerkmalen auf der einen und Qualitätsmerkmalen auf der anderen Seite, zu extrahieren. Die Entwicklung von eben solchen mathematisch-statistischen Modellen, unter Beachtung von praktischen Anforderungen, ist das übergeordnete Ziel dieser Arbeit.

Dazu werden in der vorliegenden Ausarbeitung zwei Ansätze zur Rückkoppelung von qualitätsrelevanten, signifikanten Zusammenhängen in einen produktionsnahen Qualitätsverbesserungsprozess unter Berücksichtigung von Umsetzbarkeits-, Wirtschaftlichkeits- und Machbarkeitsaspekten vorgestellt. In beiden Ansätzen werden interessante Zusammenhänge in Form von Regeln generiert, welche eine Aussage über die Qualität bzw. die Qualitätszusammenhänge treffen, da man auf der einen Seite eine Datenquelle mit beschreibenden Produktionsparametern und auf der anderen Seite mit den dazugehörigen Qualitätsergebnissen hat. Dabei ist zu beachten, dass die verwendeten statistisch-mathematischen Verfahren in der Lage sein sollten, große Datenmengen effizient zu analysieren und die darin enthaltenen Informationen mittels geeigneter Maße so zu verdichten und zu aggregieren, dass man letztendlich wenige, jedoch qualitativ hochwertige und verständliche Aussagen über interessante Zusammenhänge erhält.

Das so definierte Ziel dieser Arbeit, die Modellierung von Verfahren zur Extraktion von qualitätsrelevanten Zusammenhängen lässt sich dabei in mehrere, voneinander nicht unabhängige Teilprobleme aufspalten.

Zunächst gilt es, geeignete Maße und Modelle für die Messung von Qualität zu konzeptionieren. Sie sollen es ermöglichen, die Qualität gefertigter Produkte und Prozesse vergleichbar zu machen und den Zusammenhang zwischen qualitätsmessenden und qualitätsbeschreibenden Größen zu bewerten. Bei der Entwicklung muss darauf geachtet werden, dass die Verfahren an wenige Voraussetzungen gebunden sind, um den vielfältigen Ansprüchen der industriellen Praxis zu genügen.

Qualitätsmaße sind an sich nicht in der Lage, signifikante Zusammenhänge zwischen der Qualität eines Produktes oder Prozesses und seinen Einflussgrößen zu identifizieren. Sie erlauben lediglich die Bewertung eines Zusammenhangs. Folglich müssen unter Verwendung von Qualitätsmaßen, mathematisch-statistische Verfahren und Algorithmen zur Identifizierung von qualitätsrelevanten Ursache-Wirkungs-Beziehungen modelliert werden. Auch hierbei ist es erforderlich, dass die genutzten Methoden ein Mindestmaß an Voraussetzungen benötigen, um einen möglichst vielfältigen Einsatz zu gewährleisten.

Um Lösungsansätze für das übergeordnete Ziel dieser Ausarbeitung zu entwickeln, werden Verfahren und Methoden aus den Disziplinen Statistik und Informatik kombiniert. Folglich müssen Implikationen, die sich bei der Kombination der einzelnen Verfahren ergeben, beachtet werden. Vor allem die Datenmenge, die den Qualitätsanalysen zugrunde liegt, muss bei der Konzeptionierung der Lösungsansätze berücksichtigt werden. Darüber hinaus sind die Verfahren stets im Kontext einer praktischen Anwendung zu sehen und umzusetzen. Dies bedeutet, dass neben einer prinzipiellen Umsetzbarkeit der Modelle stets Wirtschaftlichkeitsaspekte berücksichtigt werden müssen.

Abschließend müssen die entwickelten Methoden und Verfahren anhand von realen Daten evaluiert werden.

## 1.2 Inhalt der Arbeit

Die Grundidee der Arbeit ist die Entwicklung von geeigneten Analysemethoden und Techniken, mit deren Hilfe qualitätsrelevante, statistisch signifikante Zusammenhänge aus den unterschiedlichen Datenquellen, die entlang der Wertschöpfungskette eines Produktes entstehen, identifiziert werden können. Die dabei verwendeten Methoden werden durch die jeweilige zugrunde liegende Art der Daten bedingt, so dass sich aufgrund dessen diese Arbeit in zwei Hauptteile gliedert. Im Folgenden wird eine Vorschau auf den Inhalt dieser zwei Hauptteile gegeben, indem die einzelnen Kapitel kurz skizziert und der logische Zusammenhang aufgezeigt werden.

Im ersten Hauptteil der Arbeit werden Verfahren und Methoden betrachtet, mit deren

Hilfen industrielle Fertigungsprozesse bewertet werden können, deren Qualität durch numerische Werte gemessen und durch nominale Werte beschrieben wird. Dazu werden in Kapitel 2 zunächst *Qualitätsfähigkeitsindizes* als statistische Maße für die Qualität von Produktionsprozessen eingeführt. Wie gezeigt wird, hängen die verwendeten Indizes von der jeweiligen Verteilung der zugrunde liegenden (numerischen) Messdaten ab, so dass sowohl parametrische als auch nicht-parametrische Qualitätsfähigkeitsindizes vorgestellt werden. Hierbei wird aufgrund der praktischen Relevanz ein besonderes Augenmerk auf letztere Art von Indizes gelegt und ein darüber hinaus neuartiger, nicht-parametrischer Ansatz zur Qualitätsmessung – der  $E_{ci}$  – motiviert und definiert. Dieser allgemeine Ansatz ist auch die Grundlage für die in Kapitel 3 ebenfalls neu eingeführte *Root-Cause-Analysis*.

Für das Qualitätsmanagement wie auch für die im weiteren Verlauf dieser Ausarbeitung verwendeten Verfahren ist es notwendig, die Aussage über die Qualität eines Prozesses, die mittels eines statistischen Maßes bestimmt wird, abzusichern. Dazu werden im zweiten Teil des Kapitels geeignete Methoden zur Konstruktion von Konfidenzintervallen für die eingeführten Qualitätsfähigkeitsindizes vorgestellt. Vor allem im Falle von nicht-parametrischen Indizes können keine analytisch geschlossenen Konstruktionsverfahren eingesetzt werden, so dass ein Ansatz präsentiert wird, der es erlaubt, Konfidenzintervalle für Fähigkeitsindizes zu konstruieren, deren zugrunde liegende Verteilung der Daten unbekannt ist.

Die in Kapitel 2 vorgestellten Indizes ermöglichen es, die Qualität eines Produktionsprozesses zu bestimmen, sie liefern jedoch keine Aussage über die eigentliche Ursache der (Minder-)Qualität bzw. über die Beziehung zwischen den qualitätsbeschreibenden und qualitätsmessenden Größen, die erfasst werden. Modelle und Verfahren, mit deren Hilfe eine solche Ursachenanalyse unter Berücksichtigung von Qualitätsfähigkeitsindizes durchgeführt werden kann, sind Gegenstand der Betrachtung von Kapitel 3. Dazu werden zunächst klassische Verfahren zur Modellierung von Zusammenhängen zwischen nominalen Einfluss- und numerischen Zielgrößen vorgestellt und deren Schwächen im Hinblick auf den praktischen Einsatz und die Verwendung von Qualitätsfähigkeitsindizes aufgezeigt. Da es nicht möglich ist, diese klassischen Ansätze geeignet zu erweitern, wird ein neues Verfahren motiviert und vorgestellt, welches in der Lage ist, Zusammenhänge in Daten zu modellieren, welche der eingangs beschriebenen Struktur genügen.

Grundlage des neu eingeführten *Root-Cause-Analysis*-Verfahrens ist die Darstellung der Ursachen-Wirkungs-Beziehungen zwischen qualitätsmessenden und qualitätsbeschreibenden Größen in Form von *Sub-Prozessen*, ähnlich den in Kapitel 4 verwendeten Assoziationsregeln. Unter der Verwendung der Sub-Prozess-Darstellung wird die Suche nach signifikanten, qualitätsrelevanten Zusammenhängen in ein Optimierungsproblem überführt. Dabei werden sowohl Wirtschaftlichkeitsaspekte als auch die Qualität im Sinne von Qualitätsfähigkeitsindizes berücksichtigt. Dazu wird auf Basis von Branch and

Bound ein Algorithmus vorgestellt, der eine effiziente Lösung des betrachteten Optimierungsproblems ermöglicht. Darüber hinaus konnte gezeigt werden, dass die konvexe Eigenschaft der neu eingeführten Indizes eine Optimierung des Algorithmus erlaubt. Den Abschluss des Kapitels 3 bildet eine Validierung und Evaluierung des Root-Cause-Analysis-Verfahrens anhand von Produktionsdaten aus der industriellen Praxis. Hierbei werden zur Messung und Bewertung der Qualität die in Kapitel 2 eingeführten Indizes und die jeweils geeigneten Konfidenzintervalle verwendet.

Im Mittelpunkt des zweiten Hauptteils dieser Arbeit stehen Verfahren und Methoden der Assoziationsanalyse, mit deren Hilfe die Extraktion von interessantem Wissen für Produktionsprozesse möglich ist, deren Qualität durch nominale Daten sowohl bewertet als auch beschrieben wird. Dazu werden zunächst Assoziationsregeln als geeignete Methodik zur Darstellung von Zusammenhängen zwischen zwei und mehreren nominalen Größen eingeführt. Um nun die Relevanz von Assoziationsregeln für die Anwendung zu bestimmen, wird zunächst mittels eines statistischen Unabhängigkeitstests bestimmt, ob ein Zusammenhang zwischen Ziel- und Einflussgröße besteht. Existiert ein Zusammenhang, so kann dieser mittels eines *Interessantheitsmaßes* bewertet werden. Dabei lassen sich letztere prinzipiell in zwei Kategorien einteilen: Die bekannteste Gruppe bilden die so genannten *statistischen Maße*, wie beispielsweise *Lift* oder *Support*. Diese können, ähnlich zu den Qualitätsfähigkeitsindizes bzw. Sub-Prozessen, als statistische Schätzer aufgefasst werden, deren Aussage mittels eines Konfidenzintervalls abgesichert werden sollte. Die Wahl einer geeigneten Methode zur Konstruktion eines Intervalls ist Gegenstand der anschließenden durchgeführten Betrachtung, welche zu einer Definition eines neuen statistischen Maßes – dem *SafeLift* – führt.

In praktischen Anwendungen ist es nicht immer ausreichend, relevante Zusammenhänge einzig mittels statistischer Interessantheitsmaße zu begründen. Vielmehr ist es oftmals notwendig, Assoziationsregeln aus der Sicht des Anwenders bzw. der Anwendung zu bewerten. Interessantheitsmaße, die eine solche Betrachtung erlauben, werden unter dem Begriff *subjektive Interessantheitsmaße* zusammengefasst. Exemplarisch für die Gruppe von subjektiven Maßen wird in Abschnitt 4.2.3 ein neues Interessantheitsmaß eingeführt, welches es erlaubt, das wirtschaftliche *Potential* einer Assoziationsregel zu bewerten. Somit ist es, ähnlich zu dem im vorherigen Kapitel vorgestellten Verfahren zur *Root-Cause-Analyse*, auch bei Assoziationsregeln möglich, Wirtschaftlichkeitsbetrachtungen bei der Identifizierung von Zusammenhängen zu berücksichtigen.

Wird mittels eines statistischen Hypothesentests überprüft, ob eine Assoziationsregel einen (statistisch) signifikanten Zusammenhang darstellt, kommt es aufgrund der hohen Anzahl von Regeln, die überprüft werden müssen, zu einer Kumulierung des Fehlers 1. Art bzw. dem *multiplen Hypothesenproblem*. Diese Thematik wird in Abschnitt 4.3 aufgegriffen in dem die in der Literatur gängigen Verfahren und ein neuer, alternativer Ansatz zur Kontrolle des multiplen Hypothesenproblems vorgestellt und anschließend

evaluiert werden.

Den Abschluss des zweiten Hauptteils bildet die Einführung eines Testalgorithmus zur Extraktion von interessanten Zusammenhängen. Dabei kombiniert der Algorithmus diejenigen Methoden und Verfahren, die im Laufe des Kapitels neu eingeführt bzw. vorgestellt und mittels der durchgeführten Simulationsstudien als am Besten geeignet identifiziert wurden. Eine Evaluation und prinzipielle Vorgehensweise dieses Testalgorithmus wird anhand einer Qualitätsanwendung aus dem Automobilssektor in Abschnitt 4.4 durchgeführt. In der Felddaten Analyse werden Informationen, die während des Einsatzes des Produktes beim Endkunden in Form von Garantie- und Kulanzdaten erfasst wurden, den dazugehörigen Produktionsdaten gegenübergestellt und qualitätsrelevante Zusammenhänge mittels Assoziationsregeln extrahiert.

Den Abschluss dieser Ausarbeitung bildet eine kompakte Zusammenfassung der erreichten Ziele der Arbeit und einen Ausblick auf mögliche Verbesserungen, Weiterentwicklungen und weitere Anwendungsgebiete.



## 2 Statistische Prozesslenkung

Das Ziel des Qualitätsmanagements ist es, die Qualitätsansprüche von Kunden und Abnehmern, trotz der ständig zunehmenden Komplexität der zu fertigenden Produkte zu gewährleisten. Dazu werden in der Praxis des Qualitätsmanagements eine Vielzahl von unterschiedlichen statistischen Methoden eingesetzt. Ein Qualitätsmanagementkonzept, das mehrere statistische Methoden vereinigt und darüber hinaus eine Art “Produktionsphilosophie“ beschreibt, ist die so genannte *Statistical Process Control* (SPC) oder zu Deutsch: *Statistische Prozesslenkung*. Unter SPC versteht man zum einen die Anwendung von statistischen Methoden zur Beherrschung von sowohl qualitativen als auch quantitativen Merkmalen eines produzierten Produktes und zum anderen eine kontinuierliche Verbesserung der Produktionsqualität. Der Begriff (Produktions-)Prozess wird in diesem Zusammenhang wie in der DIN 55319 verstanden: “Ein Prozess ist ein Satz von in Wechselbeziehung oder Wechselwirkungen stehenden Elementen, der Eingabe in Ergebnisse umwandelt“ (siehe [DIN Deutsches Institut für Normung e.V. 2002]). Es werden somit alle Vorgänge, die zur Fertigstellung eines Produktes notwendig sind, als Prozess bezeichnet. Je nach Kontext kann dabei ein Prozess aus einem einzigen, aber auch aus mehreren parallel oder seriell ausgeführten Fertigungsschritten bestehen.

Die Ursprünge der SPC gehen auf den amerikanischen Statistiker und Ingenieur W. A. Shewhart (18.3.1891 - 11.3.1967) zurück. Motiviert durch die Forderung seines damaligen Arbeitgebers, der Western Electric Company in Chicago (USA), mittels einheitlicher Produktion die Qualität der Produkte zu verbessern, veröffentlichte er 1931 in seinem Buch *Control of Quality of Manufactured Product* [Shewhart 1931] ein theoretisches Konzept, welches die Grundlage der heutigen Statistischen Prozesslenkung bildet.

Kern seiner Theorie ist die Feststellung, dass die Qualität eines Produktes hauptsächlich von der Streuung seiner Einzelkomponenten abhängt und eine kontinuierliche Prozesssteuerung als Reaktion auf Ausreißer zu einer Erhöhung der Streuung und somit letztendlich zu einer Verschlechterung der Qualität führt ([Shewhart 1931]). Der Begriff Streuung beschreibt dabei die zufällige Abweichung der produzierten Güter von ihrem Optimum (siehe Abschnitt 2.1). Shewhart unterscheidet dabei zwei Arten von Streuung:

1. Zufällige Streuung, die sich als statistisches Rauschen erklären lässt (systematische Streuung)
2. Nicht zufällige Streuung aufgrund besonderer Ursachen (unsystematische Streuung)

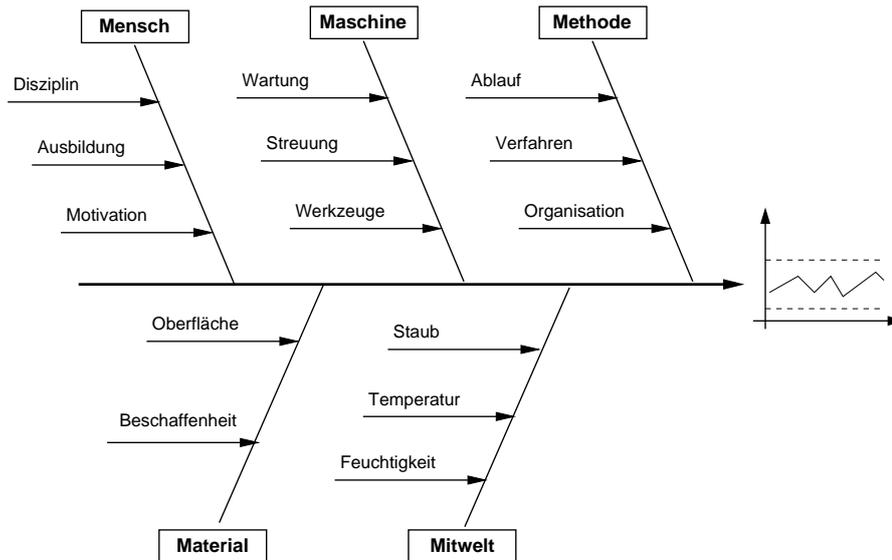


Abbildung 2.1: Ursache-Wirkungs-Diagramm nach K. Ishikawa

Diese Streuung wird durch die unterschiedlichen Faktoren, welche während der Produktion eine Rolle spielen, erzeugt. Denn selbst bei dem heutigen Stand der Technik ist es nicht immer möglich, ein und dasselbe Ergebnis bei mehrmaligem Ausführen eines Produktionsprozesses zu erreichen. Zu viele Faktoren, wie beispielsweise Temperatur oder Materialeigenschaften, spielen eine Rolle bei der Genauigkeit der Produktion. Anfang der 50er Jahre fasste der Chemiker Kaoru Ishikawa diesen Sachverhalt in dem so genannten Ursache-Wirkungs-Diagramm (vgl. Abb. 2.1) zusammen. Die Einflussfaktoren auf einen Produktionsprozess werden oft auch als die *5M* bezeichnet bzw. *7M* wenn man das Management und die Messgenauigkeit als weitere, einflussnehmende Parameter betrachtet.

Shewhart erkannte auch, dass man bei dem Versuch, die Streuung zu minimieren, zwei Arten von Zuweisungsfehlern machen kann:

1. Eine zufällige Abweichung einer nicht zufälligen Streuung zuweisen

## 2. Eine unsystematische Abweichung einer systematischen Streuung zuweisen

Er folgerte auch, dass es nicht möglich ist, gleichzeitig beide Fehler vollständig zu eliminieren. Es ist jedoch möglich, den Anteil der unsystematischen Streuung zu limitieren. Gelingt dies, so spricht man von einem stabilen Prozess (siehe Kapitel 2.1). Eine Reststreuung bleibt in der Praxis jedoch stets erhalten.

Es ist somit in der Regel bestenfalls unter Laborbedingungen möglich, immer genau dasselbe Ergebnis zu produzieren. In der Praxis unterliegt ein Produktionsprozess limitierten Schwankungen. Aus diesem Grund werden für die Produktion eines Merkmals Spezifikationsgrenzen vorgegeben, zwischen denen sich die werkstoffspezifischen Eigenschaften, wie beispielsweise die Länge eines Werkstoffs oder andere messbare Werte, befinden müssen. In der Regel werden zwei Grenzen definiert:

- Untere Spezifikationsgrenze ( $LSL$ )<sup>1</sup>: dieser Wert darf nicht unterschritten werden
- Obere Spezifikationsgrenze ( $USL$ )<sup>2</sup>: dieser Wert darf nicht überschritten werden

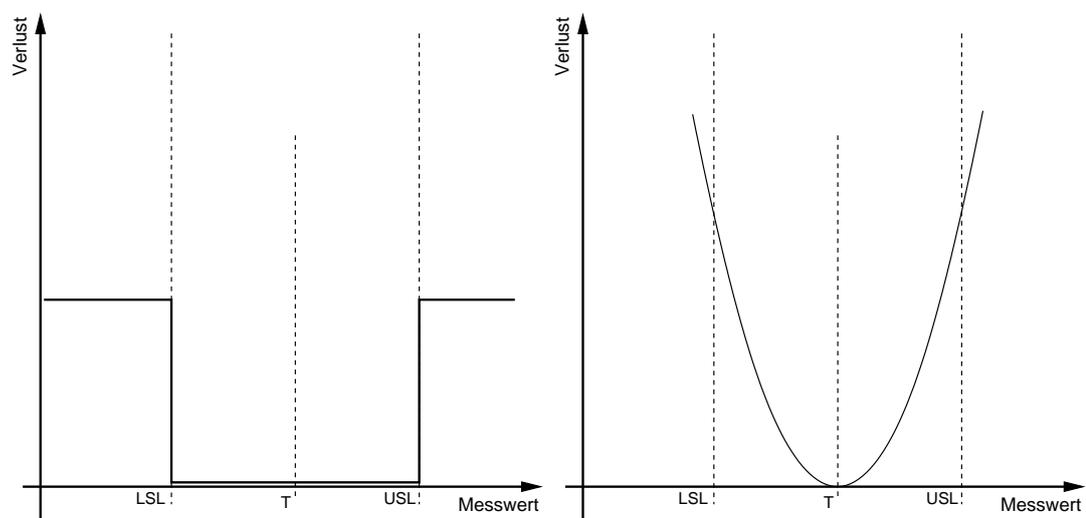


Abbildung 2.2: Links: Gut-Schlecht-Betrachtung Rechts: Taguchi-Verlustfunktion

Man erhält somit ein binäres Qualitätsmerkmal: ist die gemessene Eigenschaft innerhalb der Toleranzgrenzen, so wird das produzierte Teil als gut eingestuft. Liegt der Wert außerhalb der Grenzen, so wird das Teil als qualitativ schlecht eingestuft (vgl. Gut-Schlecht-Betrachtung in Abb. 2.2). Der Nachteil dieses binären Ansatzes ist, dass es nicht ersichtlich ist, aus welchem Grund ein produzierter Werkstoff, der gerade noch

<sup>1</sup>Lower Specification Limit

<sup>2</sup>Upper Specification Limit

innerhalb der Toleranzgrenzen liegt, als gut, hingegen ein Werkstoff, der knapp außerhalb der Grenzen liegt, als schlecht eingestuft wird.

Eine Erweiterung dieses dichotomen Ansatzes baut auf dem Modell der *Verlustfunktion* von G. Taguchi auf. Taguchi beschreibt mit der Verlust-Funktion die Qualitätsminderung eines Produktes durch eine Parabel, die ihr Minimum an der Sollgröße  $T$  erreicht. Die Verlustfunktion erstreckt sich nun, im Gegensatz zur Gut-Schlecht-Betrachtung, über den gesamten Anwendungsbereich (mögliche Ergebnisse der Qualitätsmessung). Je stärker ein Produkt von seinem Sollwert ( $T$ ) abweicht, desto höher ist der durch diese Abweichung verursachte Verlust. Jedoch eben dieser Sollwert  $T$  ist der Punkt, bei dem die Kundenzufriedenheit am größten ist, da die Qualität ihr Optimum erreicht. Eine Produktion möglichst nahe bei diesem Optimum ist somit das Ziel einer zeitgemäßen, kundenorientierten und wettbewerbsfähigen Produktion.

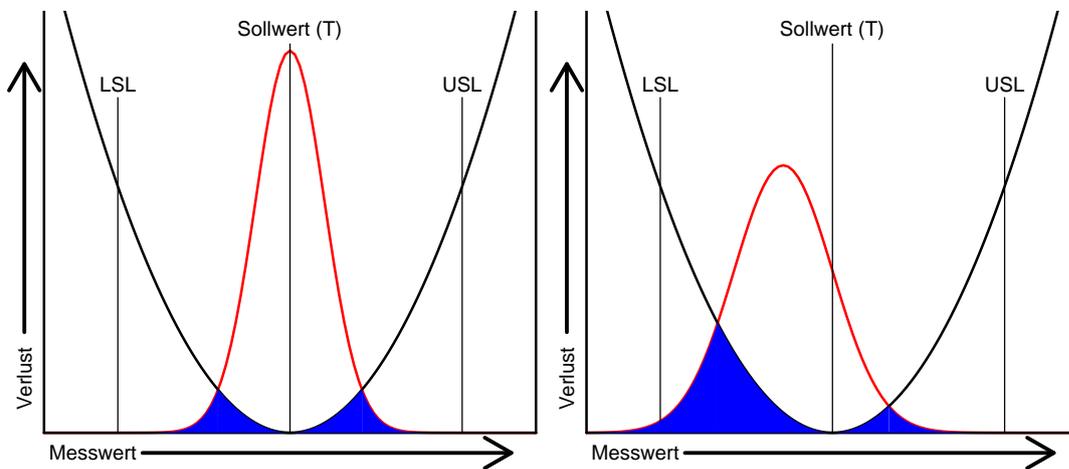


Abbildung 2.3: Prozessverhalten vs. Verlustfunktion

Überlagert man nun das Modell der Verlustfunktion von Taguchi mit einer Normalverteilung, die beispielhaft das Verhalten eines Produktionsprozesses darstellen soll, so ist ersichtlich, dass der Qualitätsverlust minimal ist, wenn die produzierten Werkstoffe sowohl nahe am Sollwert ( $T$ ) sind als auch eine geringe Streuung aufweisen (vergleiche linke Abbildung der Abb. 2.3). Je mehr der Mittelwert des Produktionsprozesses (auch Prozesslage) vom Sollwert abweicht, oder die Streuung des Prozesses zunimmt, umso stärker wird das Ziel der Kundenzufriedenheit durch optimale Qualität verfehlt (rechte Abbildung 2.3).

Eine Schlussfolgerung der vorangegangenen Überlegung ist, dass ein Unternehmen, welches heutzutage im internationalen Wettbewerb bestehen will, die Qualität seiner Produktionsprozesse so optimieren muss, dass die Qualitätsfaktoren sowohl eine möglichst geringe Streuung haben als auch der Mittelwert mit dem Sollwert übereinstimmt. Ende

der 70er Jahre erweiterte das amerikanische Unternehmen Motorola diesen (bekanntesten) statistischen Ansatz um Geschäftsstrategien und Führungsstils zu einem Managementkonzept, welches heute unter dem Namen *Six-Sigma* bekannt ist. Unter Six-Sigma versteht man somit zum einen statistische Methoden, mit deren Hilfe es möglich ist, einen Produktionsprozess bezüglich der oben genannten Parameter zu überwachen und zu steuern. Zum anderen beschreibt Six-Sigma eine Unternehmensphilosophie, die ständig bestrebt ist, die Qualität ihrer Produktionsprozesse zu verbessern (kontinuierliche Qualitätsverbesserung). Dieser Ansatz ist folglich nicht nur für die Produktion nutzbar, sondern auch in anderen Unternehmensbereichen wie Projektsteuerung, Softwareentwicklung und Verwaltung einsetzbar.

Der Name Six-Sigma geht auf die mathematischen Überlegungen, die dem Modell zugrunde liegen, zurück. Betrachtet man einen normalverteilten Prozess, so bezeichnet Six-Sigma die Fähigkeit, dass der Abstand der Toleranzgrenzen LSL und USL von der Prozessmitte 6 Standardabweichungen entfernt ist<sup>3</sup>. Diese Forderung entspricht somit einem Fehleranteil von 0.00197 ppm<sup>4</sup>  $\approx 0\%$ . Mit anderen Worten: bei einem Produktionsprozess, der dieses Kriterium erfüllt, liegen 99.99...% der gefertigten Werkstoffe innerhalb der Toleranzgrenzen. Überträgt man diese Forderung beispielsweise auf den Flugbetrieb am Frankfurter Flughafen, auf dem zur Zeit 1341 Flugzeuge täglich starten und landen<sup>5</sup>, dann entspricht die Six-Sigma Forderung, dass in 1042 Jahren nur ein einziges Flugzeug abstürzen dürfte.

Eine solche Forderung ist offensichtlich sehr stark und kann folglich nicht auf jedes Produkt, Produktmerkmal und zu jedem Zeitpunkt angewandt werden. Aus diesem Grund wird in der praktischen Umsetzung des Six-Sigma Prinzips zwischen zwei Arten von Prozessen unterschieden:

**Kurzzeitstatus** Hier wird gefordert, dass kurzfristig die Prozesslage einen Abstand von  $\pm 6\sigma$  von den vorgegebenen Spezifikationsgrenzen hat.

**Langzeitstatus** Es wird davon ausgegangen, dass die Prozesslage erfahrungsgemäß um  $1.5\sigma$  schwankt (vgl. [Dietrich und Schulze 2005], Seite 9). Dadurch kann sich der Abstand der Prozesslage zu den Toleranzgrenzen auf  $4.5\sigma$  verringern. Unter Annahme einer Normalverteilung resultiert daraus ein Fehleranteil von maximal 3.4 ppm. Obwohl in diesem Fall der Abstand von den Toleranzgrenzen zur Prozesslage  $4.5\sigma$  beträgt, spricht man von  $6\sigma$ .

Das anspruchsvolle Zielniveau, das durch Six-Sigma vorgegeben wird, lässt sich mit

---

<sup>3</sup>An dieser Stelle gehen wir von einem idealisierten, normalverteilten Prozess aus

<sup>4</sup>Parts per Million: Fehleranzahl pro 1 Mio. produzierter Teile

<sup>5</sup>Siehe Betreiberhomepage <http://www.flughafen-frankfurt.de>

statistischen Methoden wie Qualitätsregelkarten überwachen und mit Hilfe von Qualitätsfähigkeitsindizes messen. In den folgenden Abschnitten werden nun Qualitätsfähigkeitsindizes eingeführt und ihre grundlegenden Eigenschaften beschrieben. Auf Qualitätsregelkarten wird an dieser Stelle nicht weiter eingegangen, da sie für diese Arbeit eine untergeordnete Rolle spielen. Eine ausführliche Einführung und Beschreibung von Qualitätsregelkarten findet sich beispielsweise bei [Rinne und Mittag 2002] oder unter [Montgomery 1991].

Ein weiterer Begriff, der im Zuge von unternehmensweiter Qualitätsverbesserung genannt wird, ist das japanische Management-Konzept Kaizen. Kaizen (Kai = Veränderung; ZEN = zum Besseren) verfolgt das Ziel einer kontinuierlichen Verbesserung in allen Unternehmensbereichen, unter Einbeziehung aller Mitarbeiter des Unternehmens. Dabei integriert es im Wesentlichen die folgenden zwei Aspekte:

**Qualitätsverbesserung** Wie bei der SPC, werden verschiedene Instrumente zur ständigen Verbesserung der Qualität eingesetzt. Dabei werden jedoch nicht nur die Faktoren, die einen unmittelbaren Einfluss auf die Qualität eines Produktes haben berücksichtigt, sondern alle Faktoren (Bereiche) eines Unternehmens.

**Vorschlagswesen** Im Gegensatz zu einem westlichen Management-Konzept, bei dem Fortschritt durch Innovation gekennzeichnet ist, wird bei der japanischen Philosophie Kaizen ein Fortschritt durch eine ständige Verbesserung der bestehenden Produkte erreicht. Dabei stehen Vorschläge, die von den einzelnen Mitarbeitern gemacht werden, im Vordergrund.

Da Kaizen auf allen hierarchischen Ebenen eines Konzerns integriert ist, bei dem sowohl Führungskräfte als auch Mitarbeiter mit einbezogen werden, ist es im Vergleich zu SPC ein eher ganzheitlicher Ansatz. Nichtsdestotrotz liegt der gleiche, integrale Bestandteil wie der Statistischen Prozesslenkung zugrunde: die Bewertung von Qualität und die sich daraus ableitenden Handlungsmaßnahmen.

## 2.1 Qualitätsfähigkeitsindizes

Um die Qualität eines Produktionsprozesses zu messen, werden in der Statistischen Prozesslenkung *Qualitätsfähigkeitsindizes* (im weiteren Verlauf mit QFI abgekürzt) verwendet. Sie erlauben es die Qualität eines Produktionsprozesses mathematisch abzubilden, indem werkstoffspezifische Eigenschaften mit dem tatsächlichen Verhalten des Produktionsprozesses verbunden werden. Fähigkeitsindizes ermöglichen es somit, quantitative Aussagen über die Qualität eines Prozesses zu treffen, um darauf aufbauend Handlungs-

empfehlungen für die Produktion abzuleiten. Nach DIN 55319 wird ein QFI als Quotient aus der Toleranz- (der Differenz zwischen oberer und unterer Spezifikationsgrenze) und der Streubreite des betrachteten Prozesses berechnet. Dabei unterscheiden sich die in der Praxis verwendeten Indizes sowohl in ihrer Berechnungsart als auch in ihrer Bedeutung und Anwendbarkeit.

Im Folgenden werden die gängigsten, in der industriellen Qualitätssicherung verwendeten Indizes, ihre Anwendung und Bedeutung erläutert und miteinander verglichen. Darüber hinaus wird ein neuer QFI, basierend auf einer allgemeinen Klasse von Prozesskennzahlen, eingeführt. Da QFI Punktschätzungen sind, ist es, um eine zuverlässige Aussage treffen zu können, notwendig, Konfidenzintervalle für die verschiedenen Indizes zu formulieren. Diese Thematik wird in Kapitel 2.3 für die in den folgenden Abschnitten erläuterten QFI genauer behandelt.

Grundlage für die Berechnung von QFI sind Messergebnisse aus dem Fertigungsprozess eines Produktes. Dabei werden, je nach Anforderung, zu unterschiedlichen oder festgelegten Zeitpunkten, die gefertigten Werkstücke mittels geeigneter Prüfeinrichtungen vermessen. So werden beispielsweise bei der Motorenfertigung die geometrischen Eigenschaften eines Motors, wie die Rauheit der Oberfläche oder der Drall eines Zylinders bei Erreichung eines bestimmten Fertigungsgrades, vermessen und zur anschließenden Auswertung aufgezeichnet. Aufgrund dieser Auswertung werden Handlungsempfehlungen für den betrachteten Produktionsprozess abgeleitet, wie beispielsweise den Prozess neu einzustellen (anhalten), falls die geforderte Qualität nicht erreicht wird.

Prinzipiell kann man zwei Arten von Prüfungen unterscheiden. Zum einen Prüfungen, die für alle produzierten Einheiten gemacht werden und zum anderen stichprobenartige Prüfungen. Im ersten Fall spricht man in diesem Zusammenhang von einer Totalprüfung bzw. einer kontinuierlichen Prozessregelung (vgl. statistische Prozesslenkung). Bei sehr hohen Stückzahlen, sehr kostspieligen oder gar zerstörenden Qualitätsprüfungen ist es nicht möglich, alle produzierten Produkte zu prüfen. Ist dies der Fall, so besteht beispielsweise die Möglichkeit, Stichproben mit einem im Vorfeld definierten Umfang  $n$  aus dem Produktionsprozess zu entnehmen oder sequentielle Prüfverfahren zu verwenden. Im Gegensatz zur Totalprüfung werden in diesem Fall lediglich die Produkte, die in dieser Stichprobe enthalten sind, vermessen. Aufgrund dieser Ergebnisse wird nun eine Aussage über die Gesamtheit des zugrunde liegenden Produktionsprozesses getroffen. Dieses Schließen von einer Stichprobe auf die Grundgesamtheit motiviert auch den Begriff der statistischen Prozessregelung.

Die Berechnung von QFIs hängt somit unmittelbar von der Art der zugrunde liegenden Daten und deren Verteilung ab. Wenn nicht anders angegeben, wird im weiteren Verlauf von stetigen Messdaten ausgegangen. Die Verteilung dieser Messdaten ist jedoch oftmals weder bekannt, noch ist sie zeitinvariant. Um trotzdem auf Basis der gesammelten Daten

eine Aussage über das zukünftige Verhalten des Produktionsprozesses treffen zu können, wird von einem Prozess, der mittels QFI bzw. SPC gelenkt wird, erwartet, dass er *stabil* ist. Unter einem stabilen Prozess versteht man einen, bezüglich seiner Streuung und seiner Lage beherrschbaren Prozess. Dies bedeutet, dass sich die Varianz und die Lage des Prozesses innerhalb vordefinierter Grenzen bewegen muss. Qualitätsregelkarten, wie beispielsweise Shewhart-Karten, ermöglichen es, den zeitlichen Verlauf der Größen Varianz und Lage grafisch zu überwachen. Dazu werden so genannte Eingriffsgrenzen definiert. Bewegt sich der Prozess innerhalb dieser Grenzen, so spricht man von einem stabilen Prozess (vgl. [Dietrich und Schulze 2005], Kapitel 3.12). Erst diese Stabilität macht es möglich, von dem aktuellen Verhalten des Prozesses auf zukünftiges Verhalten zu schließen und somit etwaigen Handlungsbedarf zu identifizieren.

Wird ein Prozess aufgrund seines Verhaltens in der Vergangenheit als stabil eingestuft, kann die Qualität nun mittels QFI bestimmt werden. Dazu muss nach DIN 55319 das Verteilungszeitmodell der zugrunde liegenden Daten bestimmt werden, um damit die korrekte Berechnungsmethode zu identifizieren. Unter Verteilungszeitmodell wird in diesem Zusammenhang die aus zeitlichem Verhalten von Streuung und Lage resultierende vermutete Werteverteilung (Verteilung der Messergebnisse) bezeichnet. Für einige in der Praxis vorkommende Spezialfälle, sind in DIN 55319, in Abhängigkeit von Verteilungszeitmodellen, Berechnungsvorschriften angegeben, auf die an dieser Stelle verwiesen wird.

Im Verlauf dieser Arbeit werden zwei allgemeinere Fälle betrachtet: Zum einen der Fall, dass die zugrunde liegenden Daten einer Normalverteilung folgen und zum anderen der nicht-parametrische Fall, in dem keine bekannte Verteilung für die Messwerte vorliegt. Auch sei an dieser Stelle auf die Arbeiten von Dietrich und Schulze [Dietrich und Schulze 2005] und Kotz [Kotz und Johanson 2002] verwiesen, die verschiedene Eigenschaften und die mathematische Beziehung der Indizes näher erläutern. Die zur Zeit gängigen Verfahren zur Bestimmung der Indizes bei Nichtnormalverteilung können beispielsweise bei Dietrich und Schulze ([Dietrich und Schulze 2005]) oder bei Montgomery ([Montgomery 1991]) nachgelesen werden. Die Berechnungsmethoden unterscheiden sich im Wesentlichen durch eine unterschiedliche Berechnung der *Prozesslage* und *Streuung* in Abhängigkeit von verschiedenen Verteilungszeitmodellen, die einen Produktionsprozess approximativ beschreiben.

### 2.1.1 QFI unter Annahme einer Normalverteilung

Sei  $\mathbf{X} = X_1, \dots, X_n$  eine Stichprobe mit Umfang  $n$  aus einem Produktionsprozess, dessen Messdaten einer Normalverteilung folgen mit Erwartungswert  $\mu$  und Standardab-

weichung  $\sigma$ . Des Weiteren wird

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i \quad (2.1)$$

als ein Schätzer für den Stichprobenmittelwert verwendet und

$$\hat{s} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \hat{\mu})^2} \quad (2.2)$$

beschreibt einen Schätzer für die Standardabweichung der Stichprobe.

Wie auf Seite 11 eingeführt bezeichne  $USL$  und  $LSL$  die produktionsspezifischen Toleranzgrenzen eines Produktes mit Toleranzbreite  $d = (USL - LSL)/2$  und  $M = (USL + LSL)/2$  die Mitte des Toleranzbereiches. Bei symmetrischen Spezifikationsgrenzen ist  $M$  somit der Punkt, an dem die Qualität im Sinne der SPC optimal ist.

Die Bezeichnungen für die im Folgenden eingeführten QFIs variieren in der Literatur. Vor allem in praxisorientierten Büchern wie beispielsweise [Dietrich und Schulze 2005] wird eine unterschiedliche Notation verwendet. Die in der vorliegenden Arbeit genutzten Abkürzungen für die QFI richtete sich nach den Bezeichnung in dem Übersichtsartikel von S. Kotz und N. Johnson [Kotz und Johanson 2002].

Ein erster, 1974 von J.M. Juran eingeführter Index, der auf der Annahme einer Normalverteilung der zugrunde liegenden Daten basiert, ist

$$\begin{aligned} C_p &= \frac{USL - LSL}{6\sigma} \\ &= \frac{d}{3\sigma} \end{aligned} \quad (2.3)$$

Der  $C_p$  vergleicht die Streuung eines Prozesses mit den spezifizierten Toleranzgrenzen (siehe Abbildung 2.4) und kann somit als ein Maß für die Leistung, die bei optimaler Einstellung des Prozesses möglich ist, interpretiert werden (vergleiche [Dietrich und Schulze 2005]). Aus diesem Grund spricht man beim  $C_p$  auch vom Potential eines Prozesses. Ein Schätzer für den  $C_p$  ergibt sich durch:

$$\hat{C}_p = \frac{d}{3\hat{s}}$$

Die in (2.3) vorgestellte Definition entspricht dabei der Berechnungsmethode M1 nach DIN 55319.

Wenn nicht anders angegeben, ergeben sich die Schätzer für QFI durch das jeweilige Einsetzen der Schätzer für Lage (2.1) und Streuung (2.2). Wie in der Literatur üblich, werden Schätzer durch ein ‘‘Dach‘‘ gekennzeichnet.

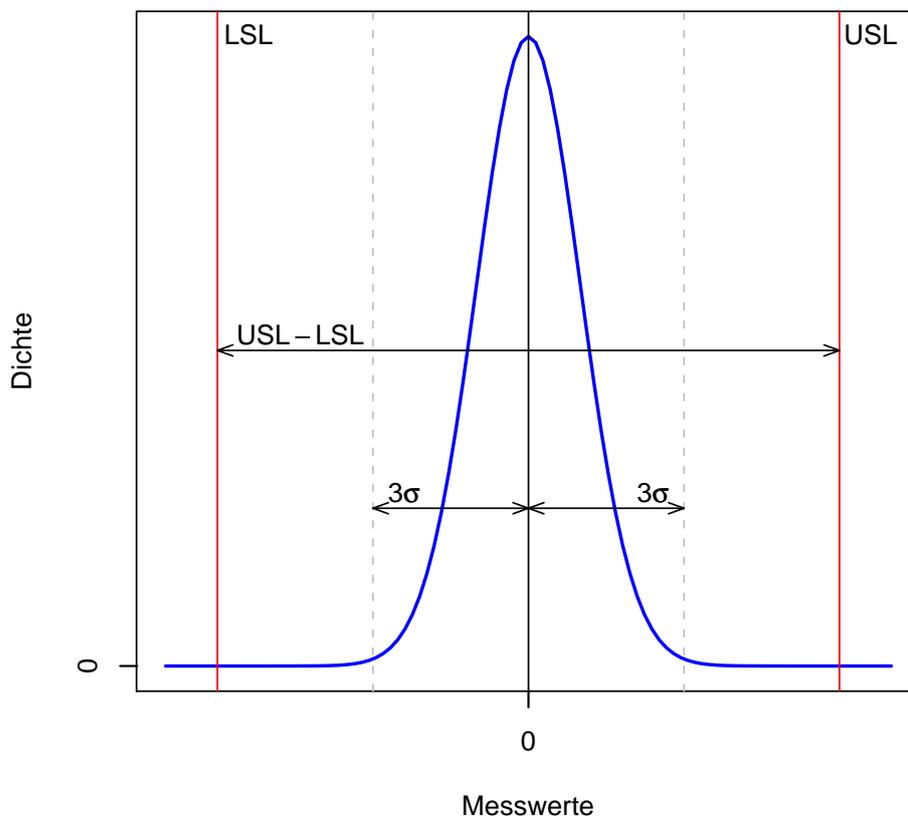


Abbildung 2.4: Standardnormalverteilung mit Toleranzen LSL-USL und einer Prozessstreuung von  $3\sigma$

Der Nachteil des Prozesspotentials ist jedoch, dass es nicht die Lage des Prozesses berücksichtigt und somit in vielen Fällen ungenaue oder ungenügende Aussagen liefert. Ein Index, der beide Größen, Lage und Streuung berücksichtigt ist  $C_{pk}$ . Er misst das Verhältnis zwischen der Differenz der Toleranzbreite und des Abstandes von  $\mu$  zum Mittelpunkt  $M$  und  $3\sigma$  und ist definiert durch:

$$\begin{aligned}
C_{pk} &= \min\left\{\frac{USL-\mu}{3\sigma}, \frac{\mu-LSL}{3\sigma}\right\} \\
&= \frac{d-|\mu-M|}{3\sigma}
\end{aligned}
\tag{2.4}$$

Aus diesem Grund spricht man beim  $C_{pk}$  auch von einem Prozessfähigkeitsindex (vgl. [Dietrich und Schulze 2005], Kap. 5.2).

Offensichtlich gilt

$$C_{pk} = C_p - \frac{1}{3} \left| \frac{\mu - M}{\sigma} \right|$$

In vielen Anwendungen liegt das gewünschte Qualitätsoptimum  $T$  nicht in der Mitte der Toleranzgrenzen. Gilt  $T \neq M$ , so wird als QFI der so genannte  $C_{pm}$  verwendet:

$$\begin{aligned}
C_{pm} &= \frac{d}{3\sqrt{\sigma^2 + (\mu - T)^2}} \\
&= \frac{d}{3\sqrt{E(X - T)^2}}
\end{aligned}
\tag{2.5}$$

Der  $C_{pm}$  wurde 1985 von den Japanern Hsiang und Taguchi eingeführt und ermöglicht die Berechnung von Fähigkeitsindizes mit asymmetrischen Toleranzgrenzen<sup>6</sup> ([Kotz und Johanson 2002]). Wie beim Prozesspotential (2.3) wird beim  $C_{pm}$  die Prozesslage nicht explizit berücksichtigt. Sie wird lediglich zur Korrektur der Standardabweichung benutzt. Eine Möglichkeit, auch die Lage für asymmetrische Toleranzgrenzen zu berücksichtigen, ist eine Hybridlösung aus dem  $C_{pm}$  (siehe 2.5) und  $C_{pk}$  (2.4):

$$C_{pmk} = \frac{d - |\mu - M|}{3\sqrt{E(X - T)^2}} \tag{2.6}$$

Nach Vännman [Vännman 1995] ergibt sich für die QFI folgende einheitliche Darstellung:

$$C_p(u, v) = \frac{d - u|\mu - M|}{3\sqrt{\sigma^2 + v(\mu - T)^2}} \tag{2.7}$$

Dann ist:

$$\begin{aligned}
C_p &\equiv C_p(0, 0); & C_{pk} &\equiv C_p(1, 0) \\
C_{pm} &\equiv C_p(0, 1); & C_{pmk} &\equiv C_p(1, 1)
\end{aligned}
\tag{2.8}$$

Legt man einen nach der SPC idealen Produktionsprozess zugrunde, der einer Normalverteilung folgt mit  $\mu = M = T$  und  $\sigma = \frac{M-LSL}{6}$ , so erhält man für die oben genannten

<sup>6</sup>Im Falle von asymmetrischen Toleranzgrenzen gilt:  $T \neq M$

QFI jeweils den Wert 2.00. Denn für  $\mathbf{X} \sim N(T, (\frac{M-LSL}{6})^2)$  gilt:  $C_p = C_{pk} = C_{pm}$ . Folglich muss ein Prozess, der dem Kriterium des Kurzzeitprozesses nach SPC genügen will (vergleiche Seite 13), einen QFI von mindestens 2.00 haben. Dieser Grenzwert ist jedoch in den meisten Anwendungen unrealistisch. Selbst die Forderung, dass ein Anteil von 0.0034% (entspricht dem Fehleranteil für das Langzeitverhalten) an produzierten Teilen innerhalb der Toleranzgrenzen liegt, ist in der Praxis unrealistisch. Aus diesem Grund wird in der praktischen Umsetzung der SPC für den Kurzzeitstatus ein QFI von mindestens 1.67 und für den Langzeitstatus eines Prozesses ein QFI von 1.33 gefordert (vgl. Tabelle 2.1).

Bezeichnung	Prozesspotential	Prozessfähigkeit
Maschinenabnahme	$C_p \geq 2.3$	$C_{pk} \geq 2.0$
Kurzzeitstatus	$C_p \geq 2.0$	$C_{pk} \geq 1.67$
Langzeitstatus	$C_p \geq 1.33$	$C_{pk} \geq 1.33$

Tabelle 2.1: Grenzwerte für Fähigkeitsindizes nach [Dietrich und Schulze 2005]

Ein weiterer Faktor bei der Wahl der Grenzwerte ist die jeweilige Situation, in der die Daten ermittelt wurden. Basieren die Daten aus einer Maschinenabnahme (Laborbedingungen), so fordert man, wie in Tabelle 2.1 beschrieben, einen höheren Mindestwert für den QFI. Wird anhand von Qualitätsregelkarten ein instabiler Prozess nachgewiesen, wird der Grenzwert ebenfalls um eine Stufe erhöht. So verlangt man beispielsweise für das Prozesspotential in einem solchen Fall anstatt 1.33 einen Wert von 1.67 (vgl. [Dietrich und Schulze 2005]). Diese Werte für die Qualität sind jedoch nicht allgemein gültig. Sie hängen von firmenspezifischen Regelungen ab, sind jedoch oftmals bindend für etwaige Zulieferer.

### 2.1.2 QFI ohne Verteilungsannahme

Im Allgemeinen sind Produktionsprozesse nicht normalverteilt. Liegt eine andere, bestimmbare Verteilung der Daten zugrunde, so lassen sich diese in eine Normalverteilung transformieren. Ein bekanntes Verfahren dafür ist beispielsweise die Box-Cox-Transformation (vergleiche [Albrecht 2005]). Bei diesem Verfahren muss für die Transformation ein Parameter  $\lambda$  bestimmt werden. Die Bestimmung von  $\lambda$  ist nicht immer analytisch möglich, sondern erfolgt oftmals numerisch. Ein weiterer Nachteil ist, dass eine Transformation von nicht-normalverteilten Daten nur im univariaten Fall sinnvoll ist. Im multivariaten Fall ist es in der Regel nicht möglich, die Originaldaten ohne Informationsverlust in eine Normalverteilung zu transformieren. In der Praxis, wie beispielsweise bei der

Motorenfertigung der Daimler AG, wird von einer 95% *normalverteilungsfreien* Produktion ausgegangen<sup>7</sup>. Das bedeutet, dass nur bei 5 Prozent aller Prüf- bzw. Messstände die gemessenen Werte einer Normalverteilung folgen. In den restlichen Fällen ist die Verteilung der Messwerte nicht eindeutig bestimmbar. Somit ist es nicht möglich, die Daten zu transformieren, was zu einer bedingten Aussagekraft der im vorherigen Kapitel eingeführten Indizes führt.

Einen Ausweg bieten nicht-parametrische Indizes. Diese ermöglichen es, die Qualität eines Prozesses, unabhängig von der Verteilung der zugrunde liegenden Daten, zu bestimmen. In dieser Arbeit werden dazu zwei Klassen von nicht-parametrischen QFI vorgestellt: Zum einen Fähigkeitsindizes, die eine Verallgemeinerung der oben eingeführten Indizes darstellen. Zum anderen wird ein neuer Index, der so genannte  $E_{ci}$ , basierend auf einer Klasse von empirischen Prozesskennzahlen (siehe Kapitel 2.2), vorgestellt.

Zur Konstruktion von nicht-parametrischen QFI werden im Folgenden die Begriffe empirische Verteilungsfunktion und empirische Quantile eingeführt.

**Definition 2.1** (Empirische Verteilungsfunktion)

Sei  $(\Omega, \Sigma, P)$  ein Wahrscheinlichkeitsraum und  $(X_1 \dots X_n)$  eine Zufallsstichprobe. Dann heißt die Abbildung  $\hat{F}_n : \Omega \times \mathbb{R} \rightarrow [0, 1]$  mit

$$\hat{F}_n(x, \omega) = \frac{|\{X_i(\omega) : 1 \leq i \leq n, X_i(\omega) \leq x\}|}{n}$$

empirische Verteilungsfunktion der Zufallsstichprobe  $X = (X_1, \dots, X_n)$ .

Die elementaren Eigenschaften der empirischen Verteilungsfunktion sind im folgenden Satz zusammengefasst (vgl. [Schmidt 2004], Kapitel 2)

**Satz 2.1**

Sei  $\hat{F}_n(x) = \hat{F}_n(x, \omega)$  die empirische Verteilungsfunktion einer Zufallsstichprobe  $X_i, i = 1 \dots, n$  iid mit Verteilungsfunktion  $F$ . Dann gilt für alle  $x \in \mathbb{R}$ :

1. Die Zufallsvariable  $n\hat{F}_n(x)$  ist binomialverteilt mit den Parametern  $n$  und  $p = F(x)$  d.h., es gilt  $\forall k \in \{0, 1, \dots, n\}$ :

$$P(n\hat{F}_n(x) = k) = \binom{n}{k} F(x)^k (1 - F(x))^{n-k} \quad (2.9)$$

2. Für den Erwartungswert und die Varianz von  $\hat{F}_n(x)$  gilt

$$\mathbb{E}(\hat{F}_n(x)) = F(x) \quad (2.10)$$

---

<sup>7</sup>Erfahrungswert der Qualitätssicherungsabteilung bei der Daimler AG

$$\text{Var} \left( \hat{F}_n(x) \right) = \frac{F(x)(1 - F(x))}{n} \quad (2.11)$$

3. Mit Wahrscheinlichkeit 1 gilt:  $\lim_{n \rightarrow \infty} \hat{F}_n(x) = F(x)$

4. Falls  $F(x) \in (0, 1)$ , dann gilt für alle  $y \in \mathbb{R}$

$$\lim_{n \rightarrow \infty} P \left( \sqrt{n} \frac{\hat{F}_n(x) - F(x)}{\sqrt{F(x)(1 - F(x))}} \leq y \right) = \Phi(y) \quad (2.12)$$

wobei  $\Phi(y)$  die Verteilungsfunktion der Standardnormalverteilung ist.

Ein Beweis der angeführten Eigenschaften ist in [Schmidt 2004], Kapitel 2 aufgeführt.

**Definition 2.2** (Quantilfunktion)

Sei  $F : \mathbb{R} \rightarrow [0, 1]$  eine Verteilungsfunktion. Die Funktion  $F^{-1} : [0, 1] \rightarrow \mathbb{R}$  mit

$$\mu_\alpha = F^{-1}(\alpha) = \begin{cases} \sup x : F(x) = 0, & \alpha = 0 \\ \inf x : F(x) \geq \alpha, & 0 < \alpha \leq 1 \end{cases}$$

heißt Quantilfunktion der Verteilung  $F$ . Mit  $\alpha \in (0, 1)$  wird  $F^{-1}(\alpha)$  das  $\alpha$ -Quantil von  $F$  genannt.

**Definition 2.3** (Empirisches Quantil)

Sei  $\hat{F}_n$  eine empirische Verteilungsfunktion. Dann ist das empirische Quantil  $\hat{\mu}_\alpha$  definiert durch:

$$\hat{\mu}_\alpha = \hat{F}_n^{-1}(\alpha) = \begin{cases} \sup x : \hat{F}_n(x) = 0, & \alpha = 0 \\ \inf x : \hat{F}_n(x) \geq \alpha, & 0 < \alpha \leq 1 \end{cases}$$

Um die Voraussetzung einer Normalverteilung zur Berechnung von QFI zu umgehen, schlug J.A. Clemens 1989 in [Clemens 1989] vor,  $6\sigma$  in der Formel durch die Differenz aus dem oberen und unteren 0.135% (empirischen) Quantil des Produktionsprozesses zu ersetzen. Denn legt man einen idealen Prozess zugrunde, der einer Normalverteilung folgt, so entspricht das Intervall  $[\mu_\alpha, \mu_{1-\alpha}]$  mit  $\alpha = 0.00135$  wiederum  $[\mu - 3\sigma, \mu + 3\sigma]$ .

Die im März 2007 erschienene DIN ISO 21747 (sie löst die DIN 55319 ab) [DIN Deutsches Institut für Normung e.V. 2007], schlägt zum ersten Mal eine einheitliche Basis zur Berechnung von Kenngrößen zur Schätzung der Qualitätsfähigkeit von Produkt- und Prozessmerkmalen vor. Im Gegensatz zur bisherigen DIN 55319 wird eine einheitliche Methode zur Berechnung des Streuungsschätzers vorgeschlagen. Im Falle, dass keine zugrunde liegende Verteilung bekannt ist, wird im neuen Standard die Varianz des Prozesses anhand der Quantile  $F_{0.00135}$  und  $F_{0.99865}$  ermittelt.

Allgemein erhält man wie in (2.7), basierend auf den empirischen Quantilen des Prozesses, eine Klasse von Fähigkeitsindizes:

**Definition 2.4** (Nicht-parametrische Fähigkeitsindizes)

Bezeichne  $d \in \mathbb{R}$  die Toleranzbreite und  $M \in \mathbb{R}$  die Mitte des Toleranzbereiches. Des Weiteren sei  $0 < \alpha < 0.5$ . Dann ist die Klasse von nicht-parametrischen Fähigkeitsindizes definiert durch:

$$C'_p(u, v, \alpha) = \frac{d - u|\mu_{0.5} - M|}{3\sqrt{\frac{(\mu_{1-\alpha} - \mu_\alpha)^2}{36} + v(\mu_{0.5} - T)^2}} \quad (2.13)$$

Zur Berechnung der nicht parametrischen Qualitätsindizes  $C'_p$ ,  $C'_{pk}$ ,  $C'_{pm}$  und  $C'_{pmk}$  wird  $(u, v)$  wie in (2.8) gewählt. Einen Schätzer für  $C'_p(u, v, \alpha)$  erhält man durch das Ersetzen der Quantile  $\mu_{1-\alpha}$  und  $\mu_\alpha$  in (2.13) durch die zugehörigen empirischen Quantile. Nach Satz 2.2 ist  $\hat{C}'_p(u, v, \alpha)$  asymptotisch normalverteilt.

**Satz 2.2**

Sei  $F$  eine Verteilungsfunktion, die in einer Umgebung von  $\mu_\alpha$ ,  $\mu_{1-\alpha}$ ,  $\mu_{0.5}$  mit  $0 < \alpha < 0.5$  eine Dichte  $f$  besitzt. Des Weiteren sei  $f$  an den Stellen  $\mu_\alpha$ ,  $\mu_{1-\alpha}$  und  $\mu_{0.5}$  stetig. Dann gilt mit obiger Notation für die Klasse nicht-parametrischer Indizes (vgl: Def. 2.4):

$$\sqrt{n} \left( \hat{C}'_p(u, v, \alpha) - C'_p(u, v, \alpha) \right) \xrightarrow{d} N(0, A\Sigma A^T) \quad (2.14)$$

mit

$$\Sigma = \begin{pmatrix} \frac{\alpha(1-\alpha)}{f(\mu_\alpha)^2} & \frac{0.5\alpha}{f(\mu_\alpha)f(\mu_{0.5})} & \frac{\alpha^2}{f(\mu_\alpha)f(\mu_{1-\alpha})} \\ \frac{0.5\alpha}{f(\mu_\alpha)f(\mu_{0.5})} & \frac{0.25}{f(\mu_{0.5})^2} & \frac{0.5(1-\alpha)}{f(\mu_{1-\alpha})f(\mu_{0.5})} \\ \frac{\alpha^2}{f(\mu_\alpha)f(\mu_{1-\alpha})} & \frac{0.5(1-\alpha)}{f(\mu_{1-\alpha})f(\mu_{0.5})} & \frac{\alpha(1-\alpha)}{f(\mu_{1-\alpha})^2} \end{pmatrix}$$

Falls  $(\mu_{0.5} - M) \geq 0$ , dann gilt für  $A$ :

$$A = \begin{pmatrix} \frac{2(\mu_{1-\alpha} - \mu_\alpha)(d - u(\mu_{0.5} - M))}{\sqrt{((\mu_{1-\alpha} - \mu_\alpha)^2 + 36v(\mu_{0.5} - T)^2)^3}} & & \\ & \frac{-2u}{\sqrt{(\mu_{1-\alpha} - \mu_\alpha)^2 + 36v(\mu_{0.5} - T)^2}} + \frac{72((u\mu_{0.5}T - uT^2 - d\mu_{0.5} + dT - uM\mu_{0.5} + uMT))}{\sqrt{((\mu_{1-\alpha} - \mu_\alpha)^2 + 36v(\mu_{0.5} - T)^2)^3}} & \\ & & \frac{2(\mu_\alpha - \mu_{1-\alpha})(d - u(\mu_{0.5} - M))}{\sqrt{((\mu_{1-\alpha} - \mu_\alpha)^2 + 36v(\mu_{0.5} - T)^2)^3}} \end{pmatrix}^T$$

Andernfalls gilt:

$$A = \begin{pmatrix} \frac{2(\mu_{1-\alpha} - \mu_\alpha)(d + u(\mu_{0.5} - M))}{\sqrt{((\mu_{1-\alpha} - \mu_\alpha)^2 + 36v(\mu_{0.5} - T)^2)^3}} & & \\ & \frac{-2u}{\sqrt{(\mu_{1-\alpha} - \mu_\alpha)^2 + 36v(\mu_{0.5} - T)^2}} + \frac{72(-u\mu_{0.5}T + uT^2 - d\mu_{0.5} + dT + uM\mu_{0.5} - uMT)}{\sqrt{((\mu_{1-\alpha} - \mu_\alpha)^2 + 36v(\mu_{0.5} - T)^2)^3}} & \\ & & \frac{2(\mu_\alpha - \mu_{1-\alpha})(d + u(\mu_{0.5} - M))}{\sqrt{((\mu_{1-\alpha} - \mu_\alpha)^2 + 36v(\mu_{0.5} - T)^2)^3}} \end{pmatrix}^T$$

Die asymptotische Varianz von  $\hat{C}'_p(u, v, \alpha)$  ist definiert durch  $\frac{A\Sigma A^T}{n}$ .

**Beweis:** Fallunterscheidung:

1.  $(\mu_{0.5} - M) \geq 0$ : Sei

$$h(x, y, z) = \frac{d - (y - M)}{3\sqrt{(z - x)^2/t^2 + v(y - T)^2}}$$

differenzierbar an der Stelle  $(x, y, z) = (\mu_\alpha, \mu_{0.5}, \mu_{1-\alpha})$ . Dann gilt:

$$\sqrt{n} \left( \hat{C}'_p(u, v, \alpha) - C'_p(u, v, \alpha) \right) = \sqrt{n} \left( h(\hat{\mu}_\alpha, \hat{\mu}_{0.5}, \hat{\mu}_{1-\alpha}) - h(\mu_\alpha, \mu_{0.5}, \mu_{1-\alpha}) \right)$$

Dann folgt aus dem Mittelwertsatz mit der Zwischenstelle

$$\mathbf{x}_0 = (\hat{\mu}_\alpha + s(\mu_\alpha - \hat{\mu}_\alpha), \hat{\mu}_{0.5} + s(\mu_{0.5} - \hat{\mu}_{0.5}), \hat{\mu}_{1-\alpha} + s(\mu_{1-\alpha} - \hat{\mu}_{1-\alpha}))$$

und geeignetem  $s \in \mathbb{R}$

$$\begin{aligned} \sqrt{n} \left( \hat{C}'_p(u, v, \alpha) - C'_p(u, v, \alpha) \right) &= \sqrt{n} (\nabla h(\mathbf{x}_0)) \begin{pmatrix} \hat{\mu}_\alpha - \mu_\alpha \\ \hat{\mu}_{0.5} - \mu_{0.5} \\ \hat{\mu}_{1-\alpha} - \mu_{1-\alpha} \end{pmatrix} \\ &= \sqrt{n} \left( \frac{\partial h}{\partial x}(\mathbf{x}_0), \frac{\partial h}{\partial y}(\mathbf{x}_0), \frac{\partial h}{\partial z}(\mathbf{x}_0) \right) \begin{pmatrix} \hat{\mu}_\alpha - \mu_\alpha \\ \hat{\mu}_{0.5} - \mu_{0.5} \\ \hat{\mu}_{1-\alpha} - \mu_{1-\alpha} \end{pmatrix} \\ &= \sqrt{n} \hat{A} X \end{aligned}$$

Da  $\hat{\mu}_\alpha$ ,  $\hat{\mu}_{1-\alpha}$  und  $\hat{\mu}_{0.5}$  erwartungstreu und konsistent sind, gilt für  $\hat{A}$ :

$$\hat{A} \xrightarrow{p} A = \begin{pmatrix} \frac{2(\mu_{1-\alpha} - \mu_\alpha)(d - u(\mu_{0.5} - M))}{\sqrt{((\mu_{1-\alpha} - \mu_\alpha)^2 + t^2 v(\mu_{0.5} - T)^2)^3}} \\ \frac{-2u}{\sqrt{(\mu_{1-\alpha} - \mu_\alpha)^2 + t^2 v(\mu_{0.5} - T)^2}} + \frac{2t^2 v(u\mu_{0.5}T - T^2 - d\mu^2 + uM\mu^2 + uMT)}{\sqrt{((\mu_{1-\alpha} - \mu_\alpha)^2 + t^2 v(\mu_{0.5} - T)^2)^3}} \\ \frac{2(\mu_\alpha - \mu_{1-\alpha})(d - u(\mu_{0.5} - M))}{\sqrt{((\mu_{1-\alpha} - \mu_\alpha)^2 + t^2 v(\mu_{0.5} - T)^2)^3}} \end{pmatrix}^T$$

Nach Satz A.2 gilt für  $X$ :

$$\sqrt{n} X \xrightarrow{d} N(0, \Sigma)$$

mit

$$\Sigma = \begin{pmatrix} \frac{\alpha(1-\alpha)}{f(\mu_\alpha)^2} & \frac{0.5\alpha}{f(\mu_\alpha)f(\mu_{0.5})} & \frac{\alpha^2}{f(\mu_\alpha)f(\mu_{1-\alpha})} \\ \frac{0.5\alpha}{f(\mu_\alpha)f(\mu_{0.5})} & \frac{0.25}{f(\mu_{0.5})^2} & \frac{0.5(1-\alpha)}{f(\mu_{1-\alpha})f(\mu_{0.5})} \\ \frac{\alpha^2}{f(\mu_\alpha)f(\mu_{1-\alpha})} & \frac{0.5(1-\alpha)}{f(\mu_{1-\alpha})f(\mu_{0.5})} & \frac{\alpha(1-\alpha)}{f(\mu_{1-\alpha})^2} \end{pmatrix}$$

Somit folgt nach Satz von Slutsky (A.3) und Satz A.4 mit  $t = 6$ :

$$\sqrt{n} \hat{A} X \xrightarrow{d} N(0, A\Sigma A^T)$$

2.  $(\mu_{0.5} - M) < 0$ : Analog Fall 1.

□

### Bemerkung 2.1

Für die asymptotische Varianz des Prozesspotentials  $C'_p = C'_p(0, 0, \alpha)$  mit  $\alpha \in (0, 0.5)$  gilt mit Hilfe von (2.14):

$$\text{Var}_n(\hat{C}'_p) = \sqrt{\frac{4\alpha d^2 [(1 - \alpha)(f(\mu_\alpha)^2 + f(\mu_{1-\alpha})^2) - 2\alpha f(\mu_\alpha)f(\mu_{1-\alpha})]}{f(\mu_\alpha)^2 f(\mu_{1-\alpha})^2 (\mu_\alpha - \mu_{1-\alpha})^4}}$$

## 2.2 Allgemeiner Ansatz für QFI

Im Folgenden wird ein neuer, nicht parametrischer Qualitätsfähigkeitsindex, basierend auf der Klasse von Prozesskennzahlen (Definition 2.6) eingeführt. Dabei wird zunächst davon ausgegangen, dass die Prozessparameter bzw. Einflussgrößen stetig sind. Im Allgemeinen kann jedoch nicht von einer Stetigkeit der Einflussgrößen ausgegangen werden, so dass in Abschnitt 2.2.2 ein nicht-parametrischer QFI für diskrete Einflussgrößen hergeleitet wird.

Der in diesem Abschnitt eingeführte und im weiteren Verlauf dieser Arbeit genutzte Begriff eines *Produktionsprozesses* bezieht sich auf die für die Qualitätssicherung notwendigen Daten eines solchen Produktionsprozesses. Dies sind in dem vorliegenden Fall die erfassten Produktionsparameter (Maschinen, Arbeiter, verwendetes Material u.ä.) und die dazugehörigen Messergebnisse eines produzierten Merkmals oder Produktes.

Um einen Produktionsprozess formal zu beschreiben gelte folgende Annahme: sei  $\Omega$  die Menge aller möglichen Produktionsabläufe mit Einflussgrößen und den dazugehörigen Messergebnissen. Der Zufallsprozess des Eintretens einzelner Produktionsabläufe sei beschrieben durch einen Wahrscheinlichkeitsraum  $(\Omega, \Sigma, P)$ .

### Definition 2.5 (Produktionsprozess)

Sei  $(\Omega, \Sigma, P)$  ein Wahrscheinlichkeitsraum und das Tupel  $(X, \vec{Y}) : (\Omega, \Sigma, P) \rightarrow \mathbb{R}^{m+1}$  ein Zufallsvektor, wobei  $X(\omega) \in \mathbb{R}$  das Messergebnis und  $\vec{Y}(\omega) \in B = \{(Y^1, \dots, Y^m)\}$  mit  $Y^i \in \mathbb{R}, i = 1, \dots, m$  die Einflussgrößen beim Produktionsablauf  $\omega$  beschreiben. Dann definiert das Tupel  $\mathcal{X} = (X, \vec{Y})$  einen Produktionsprozess.

Die Zufallsvariable  $\vec{Y}$  beschreibt somit die für den Produktionsprozess relevanten Einflussgrößen (Produktionsparameter) und  $X$  den dazugehörigen Messwert. Damit lässt sich eine Klasse von Prozesskennzahlen wie folgt definieren:

**Definition 2.6** (Klasse von Prozesskennzahlen)

Sei  $(\Omega, \Sigma, P)$  ein Wahrscheinlichkeitsraum und  $\mathcal{X} = (X, \vec{Y})$  ein Produktionsprozess wie in Definition 2.5 gegeben. Des Weiteren sei  $w : \mathbb{R} \rightarrow \mathbb{R}$ ,  $w \in L_1$ , borel-messbar eine Bewertungsfunktion des Messergebnisses. Bei entsprechender Variation von  $w$  erhält man somit eine Klasse von Prozesskennzahlen:

$$Q_{w, \mathcal{X}} = E(w(X)) \quad (2.15)$$

**2.2.1 Stetige Prozessparameter****Bemerkung 2.2**

Besitzt  $\mathcal{X} = (X, \vec{Y})$  eine Dichtefunktion  $f : \mathbb{R}^{m+1} \rightarrow \mathbb{R}$ . Dann gilt für die Qualitätskennzahl:

$$\begin{aligned} Q_{w, \mathcal{X}} &= E(w(x)) \\ &= \int_{\mathbb{R}^m} \int_{\mathbb{R}} w(x) f(x, y) dx dy \\ &= \int_{\mathbb{R}} w(x) f_X(x) dx \end{aligned} \quad (2.16)$$

mit  $f_X(x) = \int_{\mathbb{R}^m} f(x, y) dy$  der Dichtefunktion von  $X$ .

Setzt man in (2.16) beispielsweise  $w(x) = x$  oder  $w(x) = x^2$ , so erhält man das erste bzw. das zweite Moment eines Prozesses. Um jedoch die Qualität eines Prozesses zu messen, reicht es in der Regel nicht aus, die Momente isoliert zu betrachten. Man ist daran interessiert, das Zusammenspiel zwischen den Toleranzgrenzen und dem Verhalten des Prozesses mathematisch zu messen, indem beispielsweise die Bewertungsfunktion  $w(x)$  geeignet gewählt wird.

**Definition 2.7** (Gutanteil)

Sei  $(\Omega, \Sigma, P)$  ein Wahrscheinlichkeitsraum und ein Produktionsprozess  $\mathcal{X} = (X, \vec{Y})$  mit Dichtefunktion  $f(x, y)$  gegeben. Des Weiteren seien  $U, L \in \mathbb{R}$  geeignet gewählte Grenzen. Dann ist der Gutanteil (*usable fraction*) eines Produktionsprozesses  $\mathcal{X}$  gegeben durch

$$Q_{uf, \mathcal{X}} = \int_{\mathbb{R}} \mathbf{1}_{[L, U]}(x) f_X(x) dx \quad (2.17)$$

mit  $f_X(x) = \int_{\mathbb{R}^m} f(x, y) dy$  die Randdichte von  $X$ .

Wird als Bewertungsfunktion die Identitätsfunktion  $w(x) = \mathbf{1}_{[L, U]}(x)$  gewählt, so gibt der  $Q_{uf, \mathcal{X}}$  an, wie viel Prozent der Messwerte aus  $X$  sich zwischen den Grenzen  $U, L$  befinden.

Ein  $Q_{uf,\mathcal{X}}$  von 1 bedeutet folglich, dass 100 Prozent der Messwerte sich innerhalb der Grenzen befinden und ein Wert von 0, dass alle Messwerte außerhalb der Grenzen  $L$  und  $U$  liegen. Eine detailliertere Untersuchung für eine geeignete Wahl der Grenzen  $L$  und  $U$  wird im nächsten Abschnitt vorgestellt.

In vielen Anwendungen der Qualitätssicherung ist es sinnvoll, die Variationsmöglichkeiten bei den Produktionsabläufen zu reduzieren, indem die Produktionsparameter dadurch kontrolliert werden, dass lediglich eine Teilmenge des bisherigen Wertebereichs der Produktionsparameter zugelassen wird. Beispielsweise kann es sinnvoll sein, die Luftfeuchtigkeit (Temperatur etc.) innerhalb vorgegebener Grenzen zu halten. Denn aus der Sicht der Produktionsleitung (im Sinne des Qualitätsmanagements) kann dies wünschenswert sein, wenn es zu einer Qualitätsverbesserung der Produktion führt<sup>8</sup>.

Sei  $B$  der Wertebereich des Zufallsvektors  $\vec{Y}$  und  $B_0 \subseteq B$ , dann lässt sich ein so genannter Sub-Prozess wie folgt definieren:

**Definition 2.8** (Sub-Prozess)

Sei  $(\Omega, \Sigma, P)$  ein Wahrscheinlichkeitsraum und ein Produktionsprozess  $\mathcal{X} = (X, \vec{Y})$  mit  $X(\omega) \in \mathbb{R}$  und  $\vec{Y}(\omega) \in B$  gegeben. Des Weiteren sei  $B_0 \subset B$  eine borelsche Menge mit  $P(\vec{Y} \in B_0) \neq 0$ . Dann definiert das Tupel  $\mathcal{X}_{B_0} = (X, \vec{Y}_0) = (X, \vec{Y} | \vec{Y} \in B_0)$  einen Sub-Prozess von  $\mathcal{X}$ .

**Bemerkung 2.3**

Es gilt:

1. Die Klasse von (stetigen) Prozesskennzahlen (Def. 2.15) lässt sich auch auf das Konzept von stetigen Sub-Prozessen anwenden: Sei ein Sub-Prozess  $\mathcal{X}_{B_0}$  wie in Definition 2.8 gegeben. Dann ist

$$Q_{w,\mathcal{X}_{B_0}} = E\left(w(x) | \vec{Y} \in B_0\right) \quad (2.18)$$

die Prozesskennzahl des Sub-Prozesses. Dabei misst  $Q_{w,\mathcal{X}_{B_0}} - Q_{w,\mathcal{X}}$  die Qualitätsverbesserung (-Verschlechterung) durch Einschränkung auf den Sub-Prozess.

2. Sei  $f(x, y)$  die gemeinsame Dichte von  $\mathcal{X} = (X, \vec{Y})$ . Des Weiteren sei die Dichte von  $\mathcal{X}_{B_0}$  definiert durch

$$f_{B_0}(x, y) = \begin{cases} 0 & \text{für } y \notin B_0 \\ \frac{f(x,y)}{\int_{B_0} f_Y(y) dy} & \text{für } y \in B_0 \end{cases}$$

---

<sup>8</sup>Die Feststellung der Auswirkung von Einschränkungen des Wertebereiches erlaubt es, diejenigen Einflussparameter zu identifizieren, die einen besonders starken Einfluss auf die Qualität der Produktion haben. In Kapitel 3 wird ein Verfahren, welches optimale Einflussfaktoren für die Qualität einer Produktion identifiziert, motiviert und eingeführt

und der Wertebereich von  $\mathcal{X}_0$  sei gegeben durch

$$W_{\mathcal{X}_0} = \left\{ (x, \vec{y}) \in \mathbb{R}^{m+1} : \exists \omega \in \Omega : X(\omega) = x, \vec{Y}(\omega) \in B_0 \right\}$$

Dann gilt für Gleichung (2.18):

$$Q_{w, \mathcal{X}_0} = \frac{\int_{W_{\mathcal{X}_{B_0}}} w(x) f(x, y) dx dy}{\int_{B_0} f_Y(y) dy} \quad (2.19)$$

3. Die ‘‘Vereinigung‘‘ bzw ‘‘Durchschnitt‘‘ zweier Sub-Prozesse erhält man durch Restriktion auf die Vereinigung (Durchschnitt) der beiden Restriktionsmengen:

$$\mathcal{X}_{B_1} \cup \mathcal{X}_{B_2} = \mathcal{X}_{B_1 \cup B_2}$$

und

$$\mathcal{X}_{B_1} \cap \mathcal{X}_{B_2} = \mathcal{X}_{B_1 \cap B_2}$$

Für die Vereinigung zweier Sub-Prozesse  $\mathcal{X}_{B_1, B_2} = \mathcal{X}_{B_1} \cup \mathcal{X}_{B_2}$  mit  $\vec{Y}_1(\omega) \in B_1 \subset B$ ,  $\vec{Y}_2(\omega) \in B_2 \subset B$  und  $B_1 \cap B_2 = \emptyset$  gilt

$$f_{B_1 \cup B_2}(x, y) = \begin{cases} 0 & \text{für } y \notin B_1 \cup B_2 \\ \frac{f(x, y)}{\int_{B_1 \cup B_2} f_Y(y) dy} & \text{für } y \in B_1 \cup B_2 \end{cases}$$

Der Wertebereich für  $\mathcal{X}_{1,2}$  ergibt sich zu

$$\begin{aligned} W_{\mathcal{X}_{B_1, B_2}} &= \left\{ (x, \vec{y}) \in \mathbb{R}^{m+1} : \exists \omega \in \Omega : X(\omega) = x, \vec{Y}(\omega) \in B_1 \cup B_2 \right\} \\ &= \left\{ (x, \vec{y}) \in \mathbb{R}^{m+1} : \exists \omega \in \Omega : X(\omega) = x, \vec{Y}(\omega) \in B_1 \right\} \\ &\cup \left\{ (x, \vec{y}) \in \mathbb{R}^{m+1} : \exists \omega \in \Omega : X(\omega) = x, \vec{Y}(\omega) \in B_2 \right\} \\ &= W_{\mathcal{X}_{B_1}} \cup W_{\mathcal{X}_{B_2}} \end{aligned}$$

mit

$$Q_{w, \mathcal{X}_{B_1} \cup \mathcal{X}_{B_2}} = Q_{w, \mathcal{X}_{B_1, B_2}} = \frac{\int_{W_{\mathcal{X}_{B_1, B_2}}} w(x) f(x, y) dy dx}{\int_{B_1 \cup B_2} f_Y(y) dy} \quad (2.20)$$

Für die auf Seite 26 eingeführte Klasse von QFIs mit stetigen Einflussgrößen (Definition 2.6) lassen sich konvexe Eigenschaften zeigen, welche vor allem in Hinblick auf eine effiziente Bestimmung der Prozesskennzahl von Vorteil sind (siehe Root-Cause-Analysis-Verfahren in Kapitel 3.2).

**Satz 2.3** (Konvexe Eigenschaft von  $Q_{w,\mathcal{X}}$ )

Seien zwei Sub-Prozesse  $\mathcal{X}_{B_1} = (X, \vec{Y}_{B_1})$ ,  $\mathcal{X}_{B_2} = (X, \vec{Y}_{B_2})$  eines Prozesses  $\mathcal{X} = (X, \vec{Y})$  gegeben. Des Weiteren gelte  $B_1 \cap B_2 = \emptyset$ . Dann gilt für die Klasse stetiger Prozessindizes folgende Ungleichung:

$$\min_{\mathcal{Z} \in \{\mathcal{X}_{B_1}, \mathcal{X}_{B_2}\}} Q_{w,\mathcal{Z}} \leq Q_{w,\mathcal{X}_{B_1 \cup B_2}} \leq \max_{\mathcal{Z} \in \{\mathcal{X}_{B_1}, \mathcal{X}_{B_2}\}} Q_{w,\mathcal{Z}}$$

**Beweis:** Mit

$$p = \frac{\int_{B_1} f_Y(y) dy}{\int_{B_1 \cup B_2} f_Y(y) dy}$$

gilt die folgende konvexe Eigenschaft:

$$\begin{aligned} Q_{w,\mathcal{X}_{B_1 \cup B_2}} &= \frac{\int_{W^{\mathcal{X}_{B_1, B_2}}} w(x) f(x, y) dy dx}{\int_{B_1 \cup B_2} f_Y(y) dy} \\ &= \frac{\int_{W^{\mathcal{X}_{B_1}}} w(x) f(x, y) dy dx + \int_{W^{\mathcal{X}_{B_2}}} w(x) f(x, y) dy dx}{\int_{B_1 \cup B_2} f_Y(y) dy} \\ &= p Q_{w,\mathcal{X}_{B_1}} + (1 - p) Q_{w,\mathcal{X}_{B_2}} \end{aligned}$$

Da  $p \in (0, 1)$  gilt obige Behauptung. □

## 2.2.2 Nominale Prozessparameter

In den meisten Anwendungen sind die Produktionsparameter nicht stetig, das heißt es liegen diskrete bzw. nominelle Einflussgrößen vor. In diesem Fall ist der Wertebereich von  $\vec{Y}$  eine diskrete Teilmenge des  $\mathbb{R}^m$  wobei unter Umständen eine Skalierung bzw. Codierung einzelner Merkmale durchgeführt wurde. Auch kann zur Vereinheitlichung bei einzelnen stetigen Merkmalen eine Klassierung erfolgt sein.

Bei der Berechnung einer Prozesskennzahl zu gegebener Bewertungsfunktion  $w$  wirkt sich dies, da das Messergebnis als stetig angenommen wird, nicht aus:

$$Q_{w,\mathcal{X}} = E(w(x)) = \int_{-\infty}^{+\infty} w(x) f_X(x) dx$$

wobei  $f_X(x)$  die Dichtefunktion von  $X$  ist. Die Dichtefunktion kann jetzt aber nicht mehr durch Integration aus einer gemeinsamen Dichtefunktion gewonnen werden.

Liegen  $k$  konkrete Realisierungen eines Produktionsprozesses vor, so können die Realisationen  $\mathbf{x} = (x_i, \vec{y}_i)_{i=1, \dots, k}$  des Produktionsprozesses  $\mathcal{X} = (X, \vec{Y})$  als Stichprobe aufgefasst werden.

Einen Schätzwert für die Prozesskennzahl  $Q_{w, \mathcal{X}}$  erhält man mit dem Stichprobenmittel:

$$\hat{Q}_{w, \mathcal{X}} = Q_{w, \mathbf{x}} = \frac{1}{k} \sum_{i=1}^k w(x_i)$$

Unter der Verwendung der Bewertungsfunktion

$$w(x) = \mathbf{1}_{[L, U]}(x)$$

ist die Prozesskennzahl gerade der Gutanteil der Produktion und wird geschätzt durch den Stichprobenanteil

$$\hat{Q}_{w, \mathcal{X}} = \frac{1}{k} \sum_{i=1}^k \mathbf{1}_{[L, U]}(x_i)$$

Diese Kennzahl, bzw. ihren Schätzwert, bezeichnen wir auch als empirischen Fähigkeitsindex  $E_{ci}$  bzw.  $\hat{E}_{ci}$ .

Zur Bestimmung der Eigenschaften bzw. der Berechnung von Konfidenzintervallen werden im Folgenden  $\mathbf{x}_i, i = 1, \dots, n$  als iid Zufallsvariablen zu  $\mathcal{X}$  aufgefasst. Um die Notation nicht zu unübersichtlich werden zu lassen, wird jedoch dieselbe Bezeichnung verwendet. Aus dem Kontext wird klar, welche Interpretation zu verwenden ist.

Der *empirische Fähigkeitsindex* (**Empirical Capability Index**) lässt sich als Gutanteil eines Produktionsprozesses auffassen, welcher die Stichprobe in Relation zu geeignet gewählten Grenzen  $U, L$  setzt. Ein intuitiver Ansatz bei der Wahl der Grenzen  $U$  und  $L$  wäre, sie den Toleranzgrenzen gleich zu setzen ( $U = USL$  und  $L = LSL$ ). Bei der Analyse von Sub-Prozessen, wie in Kapitel 3.2 vorgestellt, kann es passieren, dass sich alle Messwerte der restringierten Sub-Prozesse innerhalb der technisch vorgegebenen Toleranzgrenzen befinden. In einem solchen Fall würde der  $E_{ci}$  stets einen Wert von 1.0 ergeben. Ein Vergleich bezüglich der Qualität von Sub-Prozessen mittels des  $E_{ci}$  ist unter diesen Voraussetzungen nicht sinnvoll, was eine Bestimmung von  $U$  und  $L$  verschieden zu den Toleranzgrenzen motiviert. Eine Möglichkeit ist, die Grenzen in Abhängigkeit der Prozesslage zu wählen. Mittels empirischer Quantile können die Grenzen  $U$  und  $L$  mit

$$[L, U] = [\hat{\mu}_\alpha, \hat{\mu}_{1-\alpha}]$$

bestimmt werden. In diesem Fall sind jedoch die Grenzen unabhängig von den eigentlichen Toleranzgrenzen. Aus diesem Grund kann alternativ  $[L, U]$  folgendermaßen gewählt werden:

$$[L, U] = \left[ \hat{\mu}_{0.5} - \frac{\hat{\mu}_{0.5} - LSL}{t}, \hat{\mu}_{0.5} + \frac{USL - \hat{\mu}_{0.5}}{t} \right] \quad (2.21)$$

mit  $t \in \mathbb{R}$ . Für  $t = 1$  ist  $[L, U] = [LSL, USL]$ . Mittels  $t$  kann man den  $E_{ci}$  an die tatsächliche Streubreite des Prozesses anpassen. Die Grenzen können mit  $t > 1$  so weit verschoben werden, dass der Abstand zwischen Prozessmitte und den Grenzen gleich  $6\sigma$  ist und somit dem Six-Sigma Prinzip entspricht.

#### Bemerkung 2.4

Einen alternativen, erwartungstreuen Schätzer für den  $E_{ci}$  erhält man durch die Verwendung von empirischen Quantilen. Sei  $\hat{F}_n$  die empirische Verteilungsfunktion von zu  $(x_i, \vec{y}_i)_{i=1, \dots, k}$  und  $U, L \in \mathbb{R}$  geeignete Toleranzgrenzen. Dann gilt:

$$\hat{E}_{ci} = \hat{F}_n(U) - \hat{F}_n(L) \quad (2.22)$$

Der empirische Fähigkeitsindex beinhaltet nur die empirische Verteilungsfunktion. Somit folgt aus den Eigenschaften empirischer Verteilungsfunktionen (Theorem 2.1), dass  $n\hat{E}_{ci} \sim B(n, p)$  mit  $p = \mathbb{E}(\hat{E}_{ci}) = F(U) - F(L)$  und dass der Schätzer für  $E_{ci}$  asymptotisch konsistent ist. Darüber hinaus gilt für die Varianz des Schätzers:

$$\text{Var}(\hat{E}_{ci}) = \frac{p(1-p)}{n} \quad (2.23)$$

Ist die zugrunde liegende Verteilung der Messwerte jedoch nicht bekannt, so kann die Varianz mittels

$$\sigma^2 = \frac{\hat{p}(1-\hat{p})}{n} \quad (2.24)$$

mit  $\hat{p} = \hat{E}_{ci}$  geschätzt werden.

## 2.3 Konfidenzintervalle für QFI

Die im vorigen Kapitel vorgestellten QFI sind Punktschätzer aufgrund einer Stichprobe und somit mit Unsicherheit behaftet. Um verlässlichere Aussagen zu treffen, werden Konfidenzintervalle für die Schätzer der Fähigkeitsindizes bestimmt. Ist die Verteilung der zugrunde liegenden Daten bekannt und stetig, so lassen sich diese mittels geeigneter Methoden in eine Normalverteilung transformieren<sup>9</sup>. Zunächst werden für den Spezialfall, dass eine Normalverteilung der Daten vorliegt, Konfidenzintervalle hergeleitet und anschließend ein allgemeines Verfahren, welches keine bekannte Verteilung voraussetzt, vorgestellt.

<sup>9</sup>Siehe Box-Cox-Transformation [Albrecht 2005]

### 2.3.1 Konfidenzintervalle für QFI unter Normalverteilungsannahme

**Definition 2.9** (Konfidenzintervall)

Die Verteilung einer Stichprobe  $X$  gehöre einer parametrischen Familie von Verteilungen  $\{P_\theta, \theta \in \Theta\}$ ,  $\Theta \subset \mathbb{R}^m$  an. Sei  $\alpha \in (0, 1)$  eine beliebige, jedoch fest vorgegebene Zahl. Dann heißt das Intervall

$$C_{l,u}(X) = [C_l(X), C_u(X)]$$

Konfidenzintervall für  $\theta$  zum Niveau  $\alpha$ , falls

$$P_\theta(C_l(X) \leq \theta \leq C_u(X)) \geq 1 - \alpha$$

für beliebige  $\theta \in \Theta$  gilt.

Sei  $X \sim N(\mu, \sigma^2)$ , dann gilt für die Schätzer (2.1) und (2.2):

$$\hat{\mu} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad (2.25)$$

sowie

$$(n-1) \frac{\hat{s}^2}{\sigma^2} \sim \chi_{n-1}^2 \quad (2.26)$$

Dabei sei  $\chi_{n-1}^2$  die Chi-Quadrat-Verteilung mit  $n-1$  Freiheitsgraden und das  $\frac{\alpha}{2}$ -Quantil von  $\chi_{n-1}^2$  sei gegeben durch  $\chi_{n-1; \frac{\alpha}{2}}^2$ . Daraus folgt für den Quotienten  $\frac{C_p}{\hat{C}_p}$  eines normalverteilten Prozesses:

$$(n-1) \left(\frac{C_p}{\hat{C}_p}\right)^2 = (n-1) \frac{\hat{s}^2}{\sigma^2} \sim \chi_{n-1}^2 \quad (2.27)$$

Damit lässt sich für das Prozesspotential  $C_p$  ein Konfidenzintervall zum Niveau  $\alpha$  wie folgt bestimmen<sup>10</sup>:

$$C_{l,u}^{C_p} = \left[ \hat{C}_p \sqrt{\frac{\chi_{n-1; 1-\frac{\alpha}{2}}^2}{n-1}}, \hat{C}_p \sqrt{\frac{\chi_{n-1; \frac{\alpha}{2}}^2}{n-1}} \right] \quad (2.28)$$

Die genaue Bestimmung eines Konfidenzintervalls für die Prozessfähigkeit  $C_{pk}$  ist sehr aufwändig, da  $\hat{C}_{pkl} = \frac{USL-\mu}{3\sigma}$  und  $\hat{C}_{pku} = \frac{\mu-LSL}{3\sigma}$  (vgl. (2.4)) einer nicht-zentralen  $t$ -Verteilung folgen. Aus diesem Grund werden in der Literatur zwei Vorschläge zur Approximation des Konfidenzintervalls gemacht:

<sup>10</sup>Im Folgenden wird stets die gleiche Notation für Konfidenzintervalle verwendet: der Superskript bezeichnet den zu schätzenden Parameter und der Subskript die Intervallgrenzen

1. Der erste Ansatz erfolgt mit Hilfe einer Normalverteilungsapproximation. [Bissel 1990] schlägt folgendes Konfidenzintervall vor:

$$C_{l,u}^{C_{pk}} = [\hat{C}_{pk} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{1}{9n} + \frac{\hat{C}_{pk}^2}{2(n-1)}}; \hat{C}_{pk} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{1}{9n} + \frac{\hat{C}_{pk}^2}{2(n-1)}}] \quad (2.29)$$

Dabei ist  $z_{1-\frac{\alpha}{2}}$  das  $(1 - \frac{\alpha}{2})$ -Quantil der Standardnormalverteilung.

Da sich das Konfidenzintervall (2.29) leicht berechnen lässt, wird es in der Praxis oft verwendet, jedoch unter der Einschränkung, dass für den Stichprobenumfang  $n$  gilt:  $n \geq 50$ .

2. Eine weitere Möglichkeit, ein Konfidenzintervall für die Prozessfähigkeit zu konstruieren, haben Zhang, Stenbeck und Wardrop in [Zhang et al. 1990] vorgestellt. In ihrer Ausarbeitung leiten sie sowohl eine genaue Berechnungsmöglichkeit als auch eine Approximation für die Varianz  $\text{Var}(\hat{C}_{pk})$  her (vgl. [Zhang et al. 1990]):

Es gelte:

$$\begin{aligned} c &= \sqrt{n}[\mu - M]\sigma \\ t &= \frac{2d}{\sigma} \end{aligned}$$

Dann lässt sich die genaue Varianz bestimmen durch:

$$\begin{aligned} \text{Var}(\hat{C}_{pk}) &= \frac{t^2}{36} \frac{n-1}{n-3} \\ &- \frac{t}{9\sqrt{n}} \frac{n-1}{n-3} \left\{ \sqrt{\frac{2}{\pi}} \exp\left(\frac{-c^2}{2}\right) \right\} \\ &- \frac{t}{9\sqrt{n}} \frac{n-1}{n-3} c [1 - 2\Phi(-c)] \\ &+ \frac{[ \frac{n-1}{9n(n-3)} ] (1 + c^2)}{[ \frac{n-1}{72n} ] \left\{ \frac{\Gamma((n-2)/2)}{\Gamma((n-1)/2)} \right\}^2} \\ &\cdot \left\{ \frac{t}{\sqrt{n}} - 2\sqrt{\frac{2}{\pi}} \exp\left(\frac{-c^2}{2}\right) - 2c[1 - 2\Phi(-c)] \right\}^2 \end{aligned} \quad (2.30)$$

wobei  $\Phi$  die Verteilungsfunktion der Standard-Normalverteilung bezeichnet.

Gelten folgende Einschränkungen:

$$n \geq 25, \hat{C}_{pk} \in (0.75; 4), |c| \leq 100, t \leq 24$$

dann kann (2.30) durch

$$\text{Var}(\hat{C}_{pk}) \approx (n-1) \left( \frac{1}{n-3} - \frac{1}{2} \frac{\Gamma((n-2)/2)^2}{\Gamma((n-1)/2)^2} \right) \hat{C}_{pk}^2 \quad (2.31)$$

approximiert werden.

Mittels der Varianzberechnung aus (2.30) bzw. (2.31) lässt sich nun ein Konfidenzintervall für den  $C_{pk}$  zum Niveau  $(1 - \alpha)$  konstruieren:

$$C_{l,u}^{C_{pk}} = \hat{C}_{pk} \pm z_{1-\frac{\alpha}{2}} \sqrt{\text{Var}(\hat{C}_{pk})}$$

Für die Bestimmung eines Konfidenzintervalls für  $C_{pm}$  gibt es in der Literatur eine Vielzahl unterschiedlicher Verfahren. [Perakis und Xakalaki 2004] gibt eine Übersicht über die verwendeten Verfahren zur Bestimmung eines Vertrauensintervalls, die sich vor allem in der Approximation der nicht-zentralen  $\chi^2$ -Verteilung unterscheiden. Exemplarisch seien an dieser Stelle zwei Verfahren von R.A. Boyles ([Boyles 1991]) zur Bestimmung eines Konfidenzintervalls für den  $C_{pm}$  herausgegriffen. Dabei gelte folgende Notation:

$$\begin{aligned} \hat{\sigma} &= \hat{s} \frac{n-1}{n} \\ \hat{\delta} &= \frac{\hat{\mu}-T}{\hat{\sigma}} \\ \hat{f} &= \frac{n(1+\hat{\delta})^2}{1+2\hat{\delta}} \end{aligned}$$

Dann kann für  $n \leq 100$  ein Vertrauensintervall wie folgt konstruiert werden (siehe [Perakis und Xakalaki 2004]):

$$C_{l,u}^{C_{pm}} = [\hat{C}_{pm} \sqrt{\chi_{\hat{f}; \frac{\alpha}{2}}^2}, \hat{C}_{pm} \sqrt{\chi_{\hat{f}; 1-\frac{\alpha}{2}}^2}]$$

Falls  $n > 100$ , schlägt Boyles folgendes Vertrauensintervall vor:

$$C_{l,u}^{C_{pm}} = \left[ \hat{C}_{pm} \left( 1 - \frac{\Phi^{-1}(1 - \frac{\alpha}{2})}{\sqrt{2\hat{f}}} \right), \hat{C}_{pm} \left( 1 + \frac{\Phi^{-1}(1 - \frac{\alpha}{2})}{\sqrt{2\hat{f}}} \right) \right]$$

### 2.3.2 Bootstrapverfahren zur Intervallschätzung für QFI

Wie in Kapitel 2.1 gezeigt, ist es nicht immer möglich, die Wahrscheinlichkeitsverteilung der zugrunde liegenden Stichprobe zu bestimmen. Um jedoch auch in solchen Situationen Vertrauensintervalle für Fähigkeitsindizes bestimmen zu können, werden computer-gestützte, statistische Auswertungsmethoden benutzt. Die bekannteste dieser so genannten Resampling-Verfahren sind das Bootstrap- und das Jackknife-Verfahren. Bei diesen Verfahren werden auf Basis der vorhandenen Stichprobe mit Hilfe von Simulationsverfahren eine große Anzahl von weiteren (Pseudo-Zufalls) Datensätzen erzeugt (re-sampling). Diese werden dazu verwendet, die Stichprobenverteilung und insbesondere die für die Bildung von Konfidenzintervallen notwendigen Streuungsparameter zu schätzen. Damit umgehen Resampling-Verfahren die klassischen Probleme der Schätzfehlerbestimmung,

da sie weder eine Annahme über das zugrunde liegende Wahrscheinlichkeitsmodell benötigen noch einen minimalen Stichprobenumfang erfüllen müssen.

Die Grundidee der häufigsten in der heutigen Praxis eingesetzten Resampling-Methode, dem Bootstrap-Verfahren, geht auf die Jackknife-Methode zurück. Das Jackknife-Verfahren wurde 1949 von M.H. Quenouille zur Reduzierung der Verzerrung eines Schätzparameters  $\theta_n(X)$  einer Stichprobe  $X$  mit kleinem  $n$  eingeführt. 1958 erkannte Tukey, dass sich dieses Verfahren verallgemeinern lässt und zur Bestimmung von Konfidenzintervallen einer Vielzahl von unterschiedlichen Schätzern  $\theta_n(X)$  herangezogen werden kann.

Beim Jackknife-Verfahren bzw. dem *delete-one-Jackknife*-Verfahren werden  $i = 1, \dots, n$  Stichproben mit Umfang  $n - 1$  aus der vorliegenden Stichprobe  $X = (X_1, \dots, X_n)$  gezogen<sup>11</sup>. Für jede dieser  $n$  Stichproben wird ein Pseudo-Wert  $\hat{\theta}_i$  ermittelt. Diese Werte bilden die Jackknife-Statistik, welche die Grundlage für die Berechnung der Konfidenzintervalle sind (vgl. [Efron und Tibshirani 1993], Kapitel 11.4):

**Definition 2.10** (Pseudo-Werte)

Sei  $\hat{\theta}_n = \hat{\theta}_n(X) = \hat{\theta}_n(X_1, \dots, X_n)$  ein Schätzer für den unbekannt Parameter  $\theta$  der Stichprobe  $X = (X_1, \dots, X_n)$ . Des Weiteren sei  $\hat{\theta}_{(i)} = \hat{\theta}_{(i)}(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$  ein Schätzer für  $\theta$  auf der Stichprobe  $X$  ohne den  $i$ -ten Wert. Dann ist der  $i$ -te Pseudo-Wert von  $\hat{\theta}_n(X)$  definiert als

$$\tilde{\theta}_i(X) = n\hat{\theta}_n - (n-1)\hat{\theta}_{(i)}$$

Unter der Annahme, dass  $\tilde{\theta}_i$  unabhängige Zufallsvariablen mit Erwartungswert  $\theta$  sind kann mit

$$\hat{\mu}_\theta = \frac{1}{n} \sum_{i=1}^n \tilde{\theta}_i(X)$$

und

$$\hat{\sigma}_\theta^2 = \frac{1}{n-1} \sum_{i=1}^n (\hat{\mu}_\theta - \tilde{\theta}_i(X))^2$$

ein Konfidenzintervall zum Niveau  $\alpha$  für  $\theta_n$  wie folgt bestimmt werden (vgl. [Efron und Tibshirani 1993], Kapitel 11.4):

$$C_{l,u}^{\text{Jackknife}} = \left[ \hat{\mu}_\theta - t_{n-1}^{1-\alpha} \sqrt{\hat{\sigma}_\theta^2}, \hat{\mu}_\theta + t_{n-1}^{1-\alpha} \sqrt{\hat{\sigma}_\theta^2} \right]$$

<sup>11</sup>Alternativ können Stichproben mit Umfang  $n - k$  gezogen werden. In einem solchen Fall spricht man vom delete-k-Jackknife

Dabei ist  $t_{n-1}^{1-\alpha}$  das  $(1 - \alpha)$ -Quantil der  $t$ -Verteilung mit  $n - 1$  Freiheitsgraden.

Das Jackknife-Verfahren wurde in der Vergangenheit oft eingesetzt, da es einen geringen Rechenaufwand benötigt. Die stetige Weiterentwicklung der Computer und ihrer Rechenleistung hat jedoch bewirkt, dass sich in den letzten Jahren, eine Erweiterung des Jackknifes, das Bootstrap-Verfahren in der Praxis etabliert hat.

Wie beim Jackknife-Verfahren, soll der Parameter  $\theta$  einer Verteilung unbekanntem Typs mit einer unabhängigen und identisch verteilten Zufallsstichprobe geschätzt sowie die Genauigkeit (Varianz) der ermittelten Schätzfunktion bestimmt werden. Parametrische Verfahren setzen dazu die Kenntnis der Verteilungsklasse voraus. Dieses Problem wird im Bootstrap-Verfahren umgangen, da die zugrunde liegende Stichprobe als die beste Approximation der (unbekannten) Verteilung betrachtet wird.

Sei eine Stichprobe  $X = (X_1, \dots, X_n)$  einer unbekanntem Verteilung  $F$  gegeben. Des Weiteren sei  $\theta = t(F)$  der gesuchte Parameter und  $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n) = s(X)$  ein Schätzer für  $\theta$ . Beim Bootstrap-Verfahren wird das Problem, den unbekanntem Wert der Verteilung  $F$  zu bestimmen, imitiert, indem  $F$  durch die empirische Verteilung  $\hat{F}_n$  ersetzt wird. Dazu werden zunächst, im Gegensatz zum Jackknife-Verfahren,  $B$  Bootstrap-Stichproben

$$X^{*1}, \dots, X^{*B}$$

mit Umfang  $n$  aus der Stichprobe  $X = (X_1, \dots, X_n)$  mit Zurücklegen gezogen. Auf jeder der  $B$  Bootstrap-Stichproben  $X^{*i}$  wird die Statistik

$$\hat{\theta}^*(b) = s(X^{*b}), \text{ mit } b = 1, \dots, B$$

berechnet.

Aus diesen  $B$  Werten, die als Bootstrap-Statistik bezeichnet werden, lässt sich der Mittelwert durch

$$\hat{\theta}^*(\cdot) = \frac{1}{B} \sum_{b=1}^B \hat{\theta}^*(b) \quad (2.32)$$

und die Standardabweichung mit

$$\hat{\sigma}_B = \sqrt{\frac{1}{B-1} \left\{ \sum_{b=1}^B [\hat{\theta}^*(b) - \hat{\theta}^*(\cdot)]^2 \right\}} \quad (2.33)$$

berechnen.

Zur Bildung von Bootstrap-Konfidenzintervallen werden, abhängig von der jeweiligen Anwendung, eine Vielzahl unterschiedlicher Methoden vorgeschlagen. Für die Bildung

von Konfidenzintervallen für QFI werden in der Literatur in der Regel drei Methoden verwendet ([Balamurali und Kalyanasundaram 2002]):

**Bootstrap-Standardabweichung** Ein erstes, intuitives Konfidenzintervall kann mittels der Standardabweichung  $\hat{\sigma}_B$  der Bootstrap-Statistik berechnet werden. Die Standardabweichung wird dabei mit Hilfe von (2.33) berechnet und ein Konfidenzintervall zum Niveau  $1 - \alpha$  kann wie folgt definiert werden:

$$C_{l,u}^{SBBias}(X) = [\hat{\theta} - \hat{\text{bias}}_B - z_{1-\frac{\alpha}{2}} \hat{\sigma}_B, \hat{\theta} - \hat{\text{bias}}_B + z_{1-\frac{\alpha}{2}} \hat{\sigma}_B] \quad (2.34)$$

mit  $\hat{\text{bias}}_B = \hat{\theta}^*(\cdot) - \hat{\theta}$  (vgl. [Efron und Tibshirani 1993], Kapitel 10.2)<sup>12</sup>. In praktischen Anwendungen wird jedoch der Bias bei der Konstruktion des Konfidenzintervalls (2.34) ignoriert. Ist die Voraussetzung einer asymptotisch normalverteilten Statistik erfüllt, so kann  $k$  über die Quantile einer Normalverteilung bestimmt werden. Im Folgenden wird die Bootstrap-Standard-Methode ohne Berücksichtigung des Bias mit *SB* abgekürzt. Wird der Bias zur Berechnung des Konfidenzintervalls hinzugenommen, so wird dieses mit *SBBias* abgekürzt.

**Perzentil-Methode** Eine Methode zur Bestimmung eines Konfidenzintervalls, die ohne die Annahme einer Verteilung auskommt, ist die Perzentil-Methode. Das Konfidenzintervall wird mit Hilfe der empirischen Quantile  $\hat{\mu}_\alpha$  der Bootstrap-Statistik geschätzt:

$$C_{l,u}^{Perzentil}(X) = [\hat{\mu}_{\frac{\alpha}{2}}, \hat{\mu}_{1-\frac{\alpha}{2}}] \quad (2.35)$$

Ein Nachteil der Perzentil-Methode ist jedoch, dass sie nicht invariant gegenüber Verzerrungen (Bias) ist. Denn es ist möglich, dass Bootstrap-Verteilungen, die aus einer Teilmenge der gesamten Bootstrap-Verteilung stammen, verzerrt gegenüber dem erwarteten Wert sind (vgl. [Balamurali und Kalyanasundaram 2002]).

---

<sup>12</sup>Der *Bias* von  $\hat{\theta} = s(X)$  ist definiert als die Differenz zwischen dem Erwartungswert von  $\theta$  und dem wahren Wert des Parameters  $\theta$ . Folglich gilt

$$\text{bias}_F = E_F[s(X)] - t(F)$$

Durch Ersetzen der unbekanntenen Verteilung  $F$  mittels der empirischen Verteilung  $\hat{F}_n$  kann das Bootstrap-Verfahren dazu verwendet werden, den Bias von  $\hat{\theta} = s(X)$  zu bestimmen. Somit gilt

$$\text{bias}_{\hat{F}_n} = E_{\hat{F}_n}[s(X^*)] - t(\hat{F}_n)$$

mit  $X^* = (X^{*1}, \dots, X^{*B})$ . Die Bootstrap-Schätzung des Bias mittels der Bootstrap-Statistik ergibt sich aufgrund von  $\hat{\theta} = t(\hat{F}_n)$  zu

$$\hat{\text{bias}}_B = \hat{\theta}^*(\cdot) - t(\hat{F}_n) = \hat{\theta}^*(\cdot) - \hat{\theta}$$

**BCa-Methode** Die *Bias-corrected-accelerated*-Methode (BCa) ist eine Erweiterung der Perzentil-Methode. Ist die Voraussetzung  $\text{bias} = \hat{\theta}^*(\cdot) - \theta = 0$  nicht erfüllt, müssen die Grenzen des Konfidenzintervalls angepasst werden. Geht man von einer Normalverteilung der Verzerrung von  $\hat{\theta}^*(\cdot)$  und  $\hat{\theta}$  aus, so lässt sich nach Efron und Tibshirani ein “entzerrtes“ Konfidenzintervall zum Niveau  $1 - \alpha$  wie folgt bestimmen:

$$C_{l,u}^{BCa}(X) = [\hat{\mu}_{\alpha_1}, \hat{\mu}_{\alpha_2}] \quad (2.36)$$

Wie bei der Perzentil-Methode, bezeichnen  $\hat{\mu}_{\alpha_1}, \hat{\mu}_{\alpha_2}$  die empirischen Quantile der Bootstrap-Statistik. Die Korrektur der Grenzen erfolgt über die Quantile  $\alpha_1, \alpha_2$  die sich ergeben durch:

$$\begin{aligned} \alpha_1 &= \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z_{\frac{\alpha}{2}}}{1 - \hat{a}(\hat{z}_0 + z_{\frac{\alpha}{2}})}\right) \\ \alpha_2 &= \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z_{1 - \frac{\alpha}{2}}}{1 - \hat{a}(\hat{z}_0 + z_{1 - \frac{\alpha}{2}})}\right) \end{aligned} \quad (2.37)$$

Der Entzerrungsfaktor  $\hat{z}_0$  und der Beschleunigungsfaktor  $\hat{a}$  werden nach [Efron und Tibshirani 1993], Kapitel 14.3 wie folgt bestimmt:

$$\begin{aligned} \hat{z}_0 &= \Phi^{-1}\left(\frac{\mathbf{1}(\hat{\theta}(\cdot) < \hat{\theta})}{B}\right) \\ \hat{a} &= \frac{\sum_{i=1}^B (\hat{\theta}(\cdot) - \hat{\theta}_{(i)})^3}{6[\sum_{i=1}^B (\hat{\theta}(\cdot) - \hat{\theta}_{(i)})^2]^{\frac{3}{2}}} \end{aligned} \quad (2.38)$$

Dabei gilt für  $i = 1, \dots, n$ :

$$\hat{\theta}_{(i)} = s(X_{(i)})$$

und

$$\hat{\theta}_{(\cdot)} = \frac{1}{n} \sum_{i=1}^n \hat{\theta}_{(i)}$$

wobei  $X_{(i)}$  die Originalstichprobe ohne den  $i$ -ten Wert beschreibt (vgl. Jackknife-Verfahren).

In [Efron und Tibshirani 1993] und [Balamurali und Kalyanasundaram 2002] wird eine Vereinfachung der BCa-Methode vorgestellt. Wie auch bei der BCa-Methode kann der Bias durch eine Verschiebung der Grenzen wie folgt korrigiert werden:

$$C_{l,u}^{BCPB}(X) = [\hat{\mu}_{\alpha_1}, \hat{\mu}_{\alpha_2}] \quad (2.39)$$

mit

$$\begin{aligned} \alpha_1 &= \Phi(2z_{\tilde{\alpha}} - z_{\alpha}) \\ \alpha_2 &= \Phi(2z_{\tilde{\alpha}} + z_{\alpha}) \end{aligned} \quad (2.40)$$

und  $\tilde{\alpha} = P(\hat{T}_b^* \leq \hat{\theta})$  (vgl. [Hjorth 1994], Kapitel 6.6). Letzteres Verfahren wird in der Literatur (siehe z.B. [Balamurali und Kalyanasundaram 2002]) auch als

*bias corrected percentile* (BCPB-Verfahren) bezeichnet und entspricht der BCa-Methode, wenn der Beschleunigungsfaktor  $\hat{a} = 0$  gesetzt wird. Gilt  $\hat{a} = \hat{z}_0 = 0$  dann reduziert sich die BCa-Methode zur Perzentil-Methode.

Das erläuterte Bootstrap-Verfahren gehört zu der Klasse der nicht-parametrischen Resampling-Verfahren, da es keine Annahme über die Verteilung  $F$  macht. Ist die parametrische Klasse der Verteilung  $F$  jedoch bekannt, so verwendet man das parametrische Bootstrap-Verfahren. Bei diesem Verfahren werden die Bootstrap-Stichproben nicht aus der vorliegenden Stichprobe gezogen, sondern anhand von Zufallszahlen mit der Verteilung  $F$  berechnet. Die notwendigen Parameter  $\theta \in \Theta$  für die Verteilung  $F$  werden dabei aus der vorliegenden Stichprobe geschätzt.

### 2.3.3 Simulationsstudie

Eine erste Studie, die sich mit der Qualität von Bootstrap-Konfidenzintervallen für Fähigkeitsindizes auseinandersetzt, ist die Arbeit von S. Balamurali und M. Kalyanasundaram. In ihrer Studie “Bootstrap lower confidence limits for the process capability indices  $C_p$ ,  $C_{pk}$  and  $C_{pm}$ “ [Balamurali und Kalyanasundaram 2002] vergleichen sie die Güte von exakten Konfidenzintervallen, wie sie in Kapitel (2.3.1) beschrieben wurden, mit Bootstrap-Konfidenzintervallen. In Anlehnung an eine Studie von L. Franklin [Franklin und Wasserman 1992] wurden in [Balamurali und Kalyanasundaram 2002] drei verschiedene QFI betrachtet:  $C_p$ ,  $C_{pk}$  und  $C_{pm}$ . Darüber hinaus wählten die Autoren drei unterschiedliche Verteilungen ( $F = \mathbf{N}(\mu, \sigma)$ ,  $\chi_4^2$ ,  $\log\text{Normal}(\mu, \sigma)$ ) für die Simulation der zugrunde liegenden Prozesse und verschiedene Stichprobenumfänge ( $n = 20, 40, 70$ ) um ein möglichst realistisches Verhalten zu simulieren. Die einzelnen Schritte der Simulation von Balamurali und Kalyanasundaram sind im Folgenden aufgelistet:

1. **Bestimmung von Lage und Streuung:** Um Prozesse mit unterschiedlicher Qualität zu simulieren, wurden Kombinationen von  $\mu$  und  $\sigma$  so bestimmt, dass sich Prozesse von *ungeeignet* (z.B.  $C_p < 1.0$ ) bis *geeignet* (z.B.  $C_p > 1.5$ ) ergeben. Die daraus resultierenden QFI werden im weiteren Verlauf als “wahre“ Fähigkeitswerte bezeichnet.
2. **Prozesssimulation:** Mit den unterschiedlichen Kombinationen der Parameter  $\mu$  und  $\sigma$  aus dem 1. Schritt werden Zufallszahlen mit der Verteilung  $F$  und einem Umfang  $n$  erzeugt.
3. **Konfidenzintervalle:** Für den im 2. Schritt erzeugten Prozess wird zum einen die Bootstrap-Statistik mit  $B = 1000$  erzeugt. Zur Berechnung der Bootstrap-Konfidenzintervalle wählten die Autoren die oben beschriebenen SB-, Perzentil-

oder PCPB-Methode. Zum Vergleich wird in der Studie ebenfalls ein analytisches Konfidenzintervall in Anlehnung an die in Kapitel 2.3.1 vorgestellten Methoden bestimmt. In [Balamurali und Kalyanasundaram 2002] wird für alle Intervalle das Konfidenzniveau  $\alpha = 0.05$  verwendet.

4. **Gütemessung:** Die Schritte 2. und 3. wurden in der Studie von Balamurali und Kalyanasundaram  $\tilde{N} = 1000$  Mal wiederholt um die Überdeckungshäufigkeit zu berechnen.

Die Überdeckungshäufigkeit (auch Abdeckungshäufigkeit) ergibt sich dabei aus dem Anteil, wie oft  $C_l(X)$  kleiner ist als der "wahre" Fähigkeitsindex bezogen auf die Anzahl der Wiederholungen  $\tilde{N}$ . Auch für die Abdeckungshäufigkeit lässt sich mittels folgender Überlegung ein Konfidenzintervall bestimmen: Der Abdeckungsanteil für die linke Intervallgrenze entspricht einer binomialverteilten Zufallsvariable  $X$  mit  $p = 0.95$  und  $n = \tilde{N} = 1000$ . Daraus ergibt sich für die Abdeckungshäufigkeit ein Konfidenzintervall zum Niveau  $\alpha = 0.01$  mit

$$\begin{aligned} C_{l,u}^{\text{Abdeckung}}(X) &= p \pm z_{1-\alpha} \sqrt{\frac{p(1-p)}{\tilde{N}}} \\ &= 0.95 \pm z_{0.995} \sqrt{\frac{0.0475}{1000}} \\ &\doteq 0.95 \pm 0.018 \end{aligned} \tag{2.41}$$

Dies bedeutet, dass mit 99% Sicherheit eine linke Konfidenzintervall-Grenze zum Niveau 0.05 einen Abdeckungsanteil zwischen 0.9322 und 0.9678 besitzt.

Das Ergebnis der Studie ist somit eine Auflistung der Abdeckungshäufigkeit und die Indizierung einer signifikanten<sup>13</sup> Abweichung von der erwarteten Abdeckungshäufigkeit. Wie erwartet hängt dabei die Güte der einzelnen Methoden von der jeweiligen zugrunde liegenden Verteilung ab.

**Normalverteilung** Für normalverteilte Prozesse war wie erwartet die SB-Methode am genauesten. Bei allen simulierten Indizes war die Abdeckungshäufigkeit in dem vorgegebenen Konfidenzintervall (2.41).

**log-Normalverteilung** Bei der Simulation von log-normalverteilten Prozessen war die Güte der einzelnen Bootstrap-Verfahren abhängig von dem Stichprobenumfang  $n$ . Schnitt bei einem geringen Stichprobenumfang ( $n = 20$ ) die SB-Methode am besten ab, so war bei größeren Umfängen ( $n \geq 40$ ) die BCPB-Methode die geeignetste.

**$\chi^2$ -Verteilung** Die BCPB ist, nach der Studie, die Methode der Wahl bei chi-Quadrat verteilten Prozessen. Bei der Perzentil- und SB-Methode war in vielen Fällen eine signifikante Abweichung vom erwarteten Wert zu beobachten.

<sup>13</sup>In der Studie wird signifikant mit einer Abweichung vom Konfidenzintervall gleichgesetzt

Ein Kritikpunkt an der Studie von Balamurali und Kalyanasundaram ist, dass nur solche Indizes eingesetzt werden, die eine Normalverteilung voraussetzen. Dies ist jedoch nur für einen der drei Fälle gerechtfertigt, denn in den beiden anderen Fällen werden Nicht-Normalverteilungen (log-Normal- und  $\chi^2$ -Verteilung) zur Simulation verwendet. Darüber hinaus sind die Stichprobenumfänge nicht realistisch. In der Praxis wird an dieser Stelle ein Mindestumfang von 50 bei kurzfristiger Prozessfähigkeit und 120 für langfristige Prozessfähigkeit gefordert.

Aus diesem Grund wird im Folgenden eine Simulation zur Identifizierung von geeigneten Methoden zur Berechnung von Konfidenzintervallen für nicht-parametrische Qualitätsfähigkeitsindizes vorgestellt. Dabei werden die im Verlauf dieses Kapitels eingeführten Fähigkeitsindizes und Bootstrap-Konfidenzintervalle verwendet. Um Prozesse mit unterschiedlicher Qualität zu simulieren, wurden drei Qualitätskategorien wie in Tabelle 2.2 beschrieben, gewählt.

Prozessqualität	$C'_p$	$E_{ci}$
Schlecht	$C'_p < 1.0$	$E_{ci} < 0.78$
Gut	$1.0 \leq C'_p < 1.6$	$0.78 \leq E_{ci} < 0.86$
Sehr Gut	$C'_p \geq 1.6$	$0.86 \leq E_{ci} < 0.98$

Tabelle 2.2:  $C'_p$  Kategorien für Prozesssimulation

Um ein Verhalten wie in Tabelle 2.2 zu simulieren, konnten beispielsweise die folgenden Verteilungen gewählt werden, da sich mit ihnen Prozesse der jeweiligen Qualität (schlecht, gut, sehr gut) darstellen lassen:

**Schlechter Prozess** Für die simulierte Stichprobe  $X = (X_1, \dots, X_N)$  gilt:  $X_i$  ist  $\chi^2$ -verteilt mit 9 Freiheitsgraden mit  $USL = 20$ ,  $LSL = 0$  und  $T = 9$ .

**Guter Prozess** Die Messdaten  $X = (X_1, \dots, X_N)$  folgen der Verteilung  $F$  mit einer Dichte  $f = \frac{1}{5}f_1 + \frac{4}{5}f_2$ , wobei  $f_1$  die Dichte von  $N(5.5, 1)$  ist und  $f_2$  die Dichte von  $N(11, 3)$ . Die Toleranzgrenzen wurden auf  $USL = 21$ ,  $LSL = 0$  gesetzt und die Nenngröße des Prozesses auf  $T = 10$  festgelegt.

**Sehr guter Prozess**  $X = (X_1, \dots, X_N)$  mit  $X_i \sim \text{logNormal}(1, 0.45)$ . Ähnlich zu den anderen Prozessen wurde  $USL = 17$ ,  $LSL = 0$  und  $T = 9$  gewählt.

Im Gegensatz zu Balamurali und Kalyanasundaram wurden keine Prozesse mit sehr geringem Umfang simuliert<sup>14</sup>, so dass vier verschiedene Umfänge  $N$  gewählt wurden:

<sup>14</sup>In den meisten Praxisanwendungen sind die Stichprobenumfänge zur Bestimmung von QFIs größer als 20

$N \in \{50, 120, 400, 900\}$ .

In Anlehnung an die Studie von Balamurali und Kalyanasundaram wurde die Simulation mittels folgender Prozedur aufgebaut:

1. **Analytische Bestimmung der Qualitätskennzahl:** Da die "Originalverteilung" des Prozesses bekannt ist, können die Quantile  $F_{0.00135}$  und  $F_{0.99865}$  analytisch bestimmt und die Qualitätskennzahl berechnet werden. Im Falle des Prozessfähigkeitsindex  $E_{ci}$  wurden entsprechend die Quantile  $F_U$  und  $F_L$  ermittelt. Die Grenzen  $U$  und  $L$  wurden dabei mittels (2.21) angepasst, wobei  $t$  auf den Wert 2.0 gesetzt wurde.
2. **Prozesssimulation:** In Abhängigkeit von der Qualität und dem Stichprobenumfang  $N$  werden Zufallszahlen bezüglich der Verteilung  $F$  erzeugt.
3. **Konfidenzintervalle:** Für den im 2. Schritt erzeugten Prozess wird eine Bootstrap-Statistik mit Umfang  $B = 1500$  erzeugt. Mit Hilfe dieser Statistik werden anschließend die verschiedenen Bootstrap-Konfidenzintervalle zum Niveau  $\alpha = 0.1$  berechnet. Die Schritte 2 und 3 werden für jede Prozess-Fähigkeitsindexkombination 1000 Mal wiederholt.
4. **Gütemessung:** Um eine Aussage über die Qualität der ermittelten Konfidenzintervalle treffen zu können, wurde wie in der eingangs vorgestellten Studie von Balamurali und Kalyanasundaram die Häufigkeit gemessen, wie oft die untere Grenze des Konfidenzintervalls in die entsprechende Qualitätskategorie fällt (vgl. Tabelle 2.2).

Analog zu der Studie von Balamurali und Kalyanasundaram lässt sich ein Konfidenzintervall für die Abdeckungshäufigkeit der linken Konfidenzintervallgrenze konstruieren. Mit (2.41) und dem Niveau  $\alpha = 0.1$  ist mit 99% Sicherheit die Abdeckungshäufigkeit in dem Intervall :

$$C_{l,u}^{\text{Abdeckung}} = 0.9 \pm 0.027 \quad (2.42)$$

Somit wird als Gütekriterium für die im Folgenden analysierten Methoden eine Abdeckungshäufigkeit von mindestens 87.3% verlangt.

## Ergebnisse

Für den Vergleich der verschiedenen Methoden zur Berechnung der Konfidenzintervalle wurde neben den eingangs erläuterten Größen ebenfalls der Bias, die Bootstrap-Standardabweichung  $\hat{\sigma}_B$  und die geschätzte Standardabweichung  $\hat{\sigma}$  für die einzelnen Indizes ermittelt.  $\hat{\sigma}$  wird dabei auf den im zweiten Schritt der Simulationsprozedur gezogenen Zufallsvariablen mit Verteilung  $F$ , mittels (2.2) bestimmt.

Die Präsentation der Ergebnisse ist gegliedert nach den beiden Klassen von Fähigkeitsindizes  $E_{ci}$  und  $C'_p$ . Zunächst werden die Simulationsergebnisse für den empirischen Fähigkeitsindex  $E_{ci}$  vorgestellt und zusammengefasst. Anschließend werden die Ergebnisse für die Klasse nicht-parametrischer Fähigkeitsindizes, die in (2.8) definiert sind, analysiert. Zur Erläuterung werden Auszüge aus den Ergebnissen in verschiedenen Tabellen zusammengefasst. Für eine vollständige Auflistung der Ergebnisse dieser Studie sei auf die im Anhang befindlichen Tabellen verwiesen.

### Simulationsergebnisse $E_{ci}$

In Tabelle 2.3 sind die wichtigen statistischen Kennzahlen der Simulation für den  $E_{ci}$  aufgelistet. Da man die genaue Verteilung der zugrunde liegenden Daten kennt, kann die exakte Standardabweichung  $\sigma$  mittels (2.23) berechnet werden.

Qualität	$N$	$E_{ci}$	$\hat{E}_{ci}$	Bias	$\sigma$	$\hat{\sigma}$	$\hat{\sigma}_B$
Schlecht	50	0.80	0.8	0.0025	0.0566	0.0568	0.0596
Schlecht	120	0.79	0.8	0.0015	0.0365	0.0371	0.0390
Schlecht	400	0.79	0.8	0.0004	0.0200	0.0203	0.0213
Schlecht	900	0.79	0.8	0.0001	0.0133	0.0135	0.0142
Gut	50	0.87	0.86	0.0027	0.0491	0.0478	0.0505
Gut	120	0.87	0.86	0.0013	0.0317	0.0311	0.0334
Gut	400	0.87	0.86	0.0004	0.0173	0.0170	0.0181
Gut	900	0.87	0.86	0.0002	0.0116	0.0114	0.0120
Sehr Gut	50	0.97	0.98	0.0003	0.0198	0.0229	0.0217
Sehr Gut	120	0.98	0.98	0.0001	0.0128	0.0130	0.0124
Sehr Gut	400	0.98	0.98	0.0000	0.0070	0.0070	0.0069
Sehr Gut	900	0.98	0.98	0.0000	0.0047	0.0046	0.0046

Tabelle 2.3: Kennwerte der Simulation für  $E_{ci}$  mit Bias =  $\hat{E}_{ci} - E_{ci}$

Bei der Betrachtung der Standardabweichung fällt auf, dass es zwischen der exakten

Standardabweichung  $\sigma$  und der Bootstrap-Standardabweichung  $\hat{\sigma}_B$  kaum Unterschiede gibt, beziehungsweise dass beide ein ähnliches Niveau haben. Noch geringer ist die Differenz zwischen der exakten Standardabweichung und  $\hat{\sigma}$ . Offensichtlich sind die Bootstrap-Schätzungen für die Standardabweichungen sehr genau, hängen jedoch von dem jeweiligen Stichprobenumfang und dem gewählten Prozess bzw. der Verteilung ab. Dieses stabile Verhalten wird durch den Bias bestätigt. Auch in diesem Fall hängt das Niveau des Bias von der Stichprobengröße und dem zugrunde liegendem Prozess ab.

Ein weiterer Faktor, der für die Qualität des  $E_{ci}$  spricht, ist die Genauigkeit der Schätzung. Betrachtet man die Abweichung des  $\hat{E}_{ci}$  von dem jeweiligen analytischen Wert  $E_{ci}$  (Tabelle 2.3), so stellt man fest, dass die Differenz sehr gering ist. Nur in wenigen Fällen ist die prozentuale Abweichung des Schätzers größer als 1 Prozent, jedoch nie größer als 2 Prozent. Folglich kann der  $\hat{E}_{ci}$  als ein stabiler Schätzer betrachtet werden, der auch in Fällen mit geringem Stichprobenumfang eine verlässliche Aussage über die gemessene Qualität ermöglicht.

Qualität	BCa	Perzentil	SBBias	BCPB	SB
Sehr Gut	88.00%	91.00%	90.00%	69.00%	91.00%
Gut	65.00%	72.00%	71.00%	63.00%	73.00%
Schlecht	91.00%	86.00%	86.00%	71.00%	86.00%

Tabelle 2.4: Ergebnisse für  $E_{ci}$  aufgliedert nach Prozessen

Aufgrund dieser Eigenschaft liegt die Vermutung nahe, dass für die Konstruktion von Konfidenzintervallen auf aufwändige Resampling-Verfahren verzichtet werden kann. Betrachtet man jedoch die Ergebnisse aufgliedert nach den einzelnen Prozessqualitäten (Verteilungen), so ist in Tabelle 2.4 zu beobachten, dass es keine durchgängig beste Wahl für die Konstruktion eines Konfidenzintervalls gibt. In Tabelle 2.4 ist für ‘gute’ bzw. ‘sehr gute’ Prozesse zu sehen, dass die SB-Methode am geeignetsten ist. Sie hat in diesen Fällen die höchste prozentuale Überdeckung (91% bzw. 73%). Auf der anderen Seite ist für den simulierten ‘schlechten’ Prozess die BCa-Methode das Mittel der Wahl. In diesem Fall hat die BCa-Methode eine Überdeckung von 91%, wohingegen die SB-Methode lediglich 86% überdeckt.

$N$	BCa	Perzentil	SBBias	BCPB	SB
50	58%	70%	71%	51%	74%
120	85%	87%	84%	74%	85%
400	94%	89%	89%	74%	89%
900	89%	85%	85%	71%	85%

Tabelle 2.5: Ergebnisse für  $E_{ci}$  aufgliedert nach Stichprobenumfang

Ein ähnliches Bild bietet sich, wenn man die Ergebnisse nach dem Stichprobenumfang aufschlüsselt (vgl. Tabelle 2.5). Auch hier ist wiederum keine global beste Methode zu finden und bei einem Stichprobenumfang mit  $N = 120$  ist eine weitere Methode – die Perzentil-Methode – besser als die beiden bisher favorisierten Methoden. Jedoch bleibt festzustellen, dass die meisten untersuchten Methoden eine ähnlich hohe Überdeckung liefern. Eine Ausnahme bildet dabei die BCPB-Methode. Sie hat durchgängig die schlechteste Überdeckung und ist aufgrund der hier durchgeführten Studie nicht zur Konstruktion eines Konfidenzintervalls für den empirischen Fähigkeitsindex  $E_{ci}$  geeignet.

Ein weiterer auffallender Punkt ist die Tatsache, dass für die Prozesse mit der Eigenschaft “gut“ und Stichprobenumfänge  $N \in \{50, 120\}$  die prozentualen Überdeckungsanteile weit unterhalb des in (2.42) konstruierten Intervalls liegen – ein maximaler Anteil wird von der BCPB-Methode mit lediglich 70% erreicht (vgl. Tabelle 2.6). Gleiches erkennt man in Tabelle 2.6 für die Kombination: “schlecht“ und  $N = 900$ . Auch hier wird der Wert für den QFI anhand der Stichprobe zu niedrig geschätzt, so dass entsprechend die Intervallgrenze nicht mehr in die jeweilige Kategorie fällt. Betrachtet man die restlichen Fälle, so kann man beobachten, dass nur in einem Fall (für SBBias in zwei Fällen) der prozentuale Überdeckungsanteil von mindestens 87.3% nicht erreicht wird. Trotzdem ist in allen fünf Fällen die durchschnittliche Überdeckung geringer als die geforderten 87.3% (vgl. unterste Zeile Tabelle 2.6).

Qualität	$N$	BCa	Perzentil	SBBias	BCPB	SB
Schlecht	50	95.00%	88.00%	87.00%	69.00%	90.00%
Schlecht	120	94.00%	90.00%	88.00%	77.00%	88.00%
Schlecht	400	91.00%	86.00%	87.00%	72.00%	87.00%
Schlecht	900	83.00%	79.00%	80.00%	64.00%	80.00%
Gut	50	17.00%	36.00%	41.00%	31.00%	44.00%
Gut	120	61.00%	76.00%	67.00%	70.00%	72.00%
Gut	400	91.00%	90.00%	89.00%	77.00%	90.00%
Gut	900	88.00%	88.00%	86.00%	74.00%	86.00%
Sehr gut	50	60.00%	88.00%	85.00%	52.00%	88.00%
Sehr gut	120	99.00%	96.00%	96.00%	74.00%	96.00%
Sehr gut	400	99.00%	90.00%	91.00%	74.00%	91.00%
Sehr gut	900	95.00%	89.00%	88.00%	76.00%	88.00%
Aggregiert		81.08%	83.00%	82.08%	67.50%	83.33%

Tabelle 2.6: Prozentuale Überdeckung für  $E_{ci}$

Wie gezeigt wurde, scheint es zunächst schwierig zu sein, für die Konstruktion eines Konfidenzintervalls für den  $E_{ci}$  eine durchgängig beste Methode zu identifizieren. Betrachtet man jedoch die aggregierte Überdeckung über alle Prozesse, wie in Tabelle 2.5

dargestellt, so ist zu beobachten, dass die SB-Methode den höchsten Überdeckungsanteil aller simulierter Methoden aufweist. Dies bestätigt die Eingangsvermutung, dass es für die Konstruktion eines Konfidenzintervalls für den empirischen Fähigkeitsindex nicht notwendig ist, eine Bootstrap-Methode zu verwenden. Darüber hinaus kann ein Konfidenzintervall mittels der geschätzten Standardabweichung  $\hat{\sigma}$  anstatt der Bootstrap-Standardabweichung  $\hat{\sigma}_B$  eingesetzt in (2.34) berechnet werden.

### Simulationsergebnisse $C'_p$

Für die Präsentation der Simulationsergebnisse für die Klasse nicht-parametrischer Fähigkeitsindizes wurde zunächst mit Satz 2.2 die approximative Varianz berechnet. Im weiteren Verlauf sind die wichtigsten Kennzahlen der Simulation in den Tabellen 2.7 und 2.8 dargestellt.

Eine erste Auffälligkeit ist das asymptotische Verhalten der betrachteten Indizes. Die Schätzer  $\hat{C}'_p$  für die analysierten Indizes weichen, im Vergleich zu den Ergebnissen für den  $E_{ci}$ , sehr stark von den jeweiligen analytischen Werten  $C'_p$  ab. Dabei wird der Wert, bis auf eine Ausnahme, stets zu hoch geschätzt. Erst bei Zunahme des Stichprobenumfangs nähert sich der Wert des Indizes dem wahren Wert, so dass Aussagen bei einer Stichprobengröße von  $N \geq 400$  erst als zuverlässig gewertet werden können. In diesem Fall beträgt die durchschnittliche Abweichung 12%. Für  $N \leq 120$  hingegen 44%. Dabei weist der  $\hat{C}'_{pk}$  unter allen Indizes die geringste Abweichung vom dem jeweiligen analytischen Wert auf.

Dieses asymptotische Verhalten überträgt sich folglich auf den Bias der betrachteten Fähigkeitsindizes. Durch die Wahl von  $\alpha = 0.00135$  zur Berechnung der  $F_\alpha$  und  $F_{1-\alpha}$  ist der Bias verhältnismäßig hoch – im Vergleich zu den Ergebnissen des  $E_{ci}$  um den Faktor 100 größer. Ein Ausweg wäre die Wahl eines größeren Wertes für  $\alpha$ , was jedoch dem Six-Sigma Prinzip widersprechen würde<sup>15</sup>. Wie auch schon bei der Betrachtung der Abweichung des Schätzers vom analytischen Wert, ist auffällig, dass der  $C'_{pk}$  einen geringeren Bias aufweist als die anderen Indizes.

Betrachtet man die Standardabweichung der Schätzer in den Tabellen 2.7 und 2.8, so ist zunächst zu beobachten, dass die geschätzte asymptotische Standardabweichung  $\hat{\sigma}$  und die Bootstrap-Standardabweichung  $\hat{\sigma}_B$  ein ähnlich hohes Niveau haben. Da man die zugrunde liegende Verteilung kennt, lässt sich die asymptotische Standardabweichung  $\sigma_A$  bestimmen. Vergleicht man diese mit  $\hat{\sigma}$ , so ist zu beobachten, dass sich beide nicht auf demselben Niveau befinden, wie es zwischen  $\hat{\sigma}_B$  und  $\hat{\sigma}$  der Fall ist. Darüber

<sup>15</sup>Eine Studie, die das Verhalten des Bias nicht-parametrischer Indizes  $C'_p$  und  $C'_{pk}$  untersucht ist bei [Nonnemacher 2006], Kapitel 4.3 zu finden

Qualität	QFI	$N$	$\hat{C}'_p$	$C'_p$	Bias	$\sigma_A$	$\hat{\sigma}_B$	$\hat{\sigma}$
Sehr gut	$C_p$	50	3.12	1.73	0.3324	0.0816	0.4810	0.2892
Sehr gut	$C_p$	120	2.51	1.73	0.2100	0.0526	0.3195	0.3004
Sehr gut	$C_p$	400	2.05	1.73	0.0955	0.0288	0.2077	0.2903
Sehr gut	$C_p$	900	1.89	1.73	0.0271	0.0192	0.1999	0.2390
Gut	$C_p$	50	1.66	1.19	0.1068	0.3206	0.1358	0.0615
Gut	$C_p$	120	1.47	1.19	0.0714	0.2070	0.0939	0.0661
Gut	$C_p$	400	1.31	1.19	0.0340	0.1134	0.0662	0.0737
Gut	$C_p$	900	1.24	1.19	0.0068	0.0756	0.0654	0.0588
Schlecht	$C_p$	50	1.22	0.77	0.1121	0.3192	0.1548	0.0744
Schlecht	$C_p$	120	1.02	0.77	0.0664	0.2060	0.0974	0.0779
Schlecht	$C_p$	400	0.87	0.77	0.0308	0.1128	0.0655	0.0720
Schlecht	$C_p$	900	0.82	0.77	0.0086	0.0752	0.0609	0.0585
Sehr gut	$C_{pk}$	50	3.01	1.71	0.3049	0.1037	0.4766	0.3468
Sehr gut	$C_{pk}$	120	2.43	1.71	0.1990	0.0670	0.3163	0.3543
Sehr gut	$C_{pk}$	400	2.00	1.71	0.0924	0.0367	0.2044	0.3406
Sehr gut	$C_{pk}$	900	1.84	1.71	0.0263	0.0245	0.1957	0.2801
Gut	$C_{pk}$	50	1.53	1.19	0.0666	0.3441	0.1529	0.1211
Gut	$C_{pk}$	120	1.38	1.19	0.0520	0.2221	0.1038	0.0916
Gut	$C_{pk}$	400	1.25	1.19	0.0294	0.1216	0.0697	0.0825
Gut	$C_{pk}$	900	1.19	1.19	0.0061	0.0811	0.0657	0.0644
Schlecht	$C_{pk}$	50	1.07	0.77	0.0862	0.3755	0.1525	0.1169
Schlecht	$C_{pk}$	120	0.90	0.77	0.0570	0.2424	0.0967	0.1011
Schlecht	$C_{pk}$	400	0.77	0.77	0.0274	0.1328	0.0620	0.0867
Schlecht	$C_{pk}$	900	0.73	0.77	0.0078	0.0885	0.0555	0.0695

Tabelle 2.7: Ergebnisse der Simulation für  $C'_p$  und  $C'_{pk}$ 

hinaus würde man erwarten, dass die Differenz (Abweichung) zwischen  $\sigma_A$  und  $\hat{\sigma}$  mit steigendem Stichprobenumfang  $N$  abnimmt. Dies konnte jedoch mittels der Studie nicht bestätigt werden. Zwar ist zu beobachten, dass  $\hat{\sigma}$  mit Zunahme des Stichprobenumfangs abnimmt, jedoch nicht in dem Maße wie  $\sigma_A$ . Vielmehr nimmt die Abweichung zwischen der analytischen Varianz und der geschätzten Varianz zu. Im Gegensatz zum empirischen Fähigkeitsindex  $E_{ci}$  kann man somit nicht auf Resampling-Verfahren verzichten, um exakte Konfidenzintervalle zu bestimmen.

Wie in der Studie von Balamurali und Kalyanasundaram hat sich auch hier herausgestellt, dass die BCPB -Methode am besten geeignet ist, um Konfidenzintervalle für die Klasse nicht-parametrischer Indizes (2.8) zu konstruieren.

Qualität	QFI	$N$	$\hat{C}'_p$	$C'_p$	Bias	$\sigma_A$	$\hat{\sigma}_B$	$\hat{\sigma}$
Sehr gut	$C'_{pm}$	50	2.85	1.69	0.2208	0.1130	0.4275	0.3021
Sehr gut	$C'_{pm}$	120	2.39	1.69	0.1722	0.0729	0.2883	0.2307
Sehr gut	$C'_{pm}$	400	2.00	1.69	0.0860	0.0399	0.1942	0.2176
Sehr gut	$C'_{pm}$	900	1.85	1.69	0.0237	0.0266	0.1891	0.1818
Gut	$C'_{pm}$	50	1.54	1.12	0.0377	0.2086	0.1695	0.1752
Gut	$C'_{pm}$	120	1.43	1.12	0.0458	0.1346	0.1013	0.0994
Gut	$C'_{pm}$	400	1.30	1.12	0.0284	0.0737	0.0661	0.0577
Gut	$C'_{pm}$	900	1.24	1.12	0.0046	0.0492	0.0650	0.0418
Schlecht	$C'_{pm}$	50	1.15	0.68	0.0716	0.3035	0.1527	0.0929
Schlecht	$C'_{pm}$	120	1.00	0.68	0.0548	0.1959	0.0952	0.0777
Schlecht	$C'_{pm}$	400	0.86	0.68	0.0285	0.1073	0.0646	0.0687
Schlecht	$C'_{pm}$	900	0.82	0.68	0.0078	0.0715	0.0604	0.0556
Sehr gut	$C'_{pmk}$	50	2.74	1.67	0.1976	0.0816	0.4126	0.2892
Sehr gut	$C'_{pmk}$	120	2.32	1.67	0.1617	0.0526	0.2771	0.3004
Sehr gut	$C'_{pmk}$	400	1.95	1.67	0.0829	0.0288	0.1883	0.2903
Sehr gut	$C'_{pmk}$	900	1.81	1.67	0.0229	0.0192	0.1842	0.2390
Gut	$C'_{pmk}$	50	1.43	1.12	0.0086	0.3206	0.2036	0.0615
Gut	$C'_{pmk}$	120	1.34	1.12	0.0292	0.2070	0.1186	0.0661
Gut	$C'_{pmk}$	400	1.24	1.12	0.0240	0.1134	0.0700	0.0737
Gut	$C'_{pmk}$	900	1.19	1.12	0.0041	0.0756	0.0651	0.0588
Schlecht	$C'_{pmk}$	50	1.01	0.68	0.0536	0.3192	0.1604	0.0744
Schlecht	$C'_{pmk}$	120	0.88	0.68	0.0472	0.2060	0.0977	0.0779
Schlecht	$C'_{pmk}$	400	0.76	0.68	0.0255	0.1128	0.0621	0.0720
Schlecht	$C'_{pmk}$	900	0.72	0.68	0.0071	0.0752	0.0555	0.0585

Tabelle 2.8: Ergebnisse der Simulation für  $C'_{pm}$  und  $C'_{pmk}$

Prozessqualität	BCa	Perzentil	SBBias	BCPB	SB
Sehr Gut	69.98%	71.28%	45.18%	71.44%	53.16%
Gut	83.58%	90.46%	78.64%	89.52%	85.26%
Schlecht	90.63%	87.61%	95.29%	88.49%	94.05%

Tabelle 2.9: Prozentuale Überdeckung der Konfidenzintervalle aufgegliedert nach verschiedenen Prozessen

Dabei wurden, um die Güte der Konfidenzintervalle zu analysieren, zwei verschiedene Sichtweisen gewählt. Zunächst wurde die prozentuale Überdeckung über die verschiedenen Prozesse (sehr gut, gut, schlecht) aggregiert (Tabelle 2.9). Dabei fällt auf, dass sowohl die BCPB als auch die Perzentilmethode ähnlich gute Ergebnisse liefern, jedoch die SB-Methode bzw. SBBias eine verhältnismäßig geringe Überdeckung hat. Durchschnittlich hat die BCPB-Methode eine Überdeckung von 83.15%, hingegen die SB-Methode 77.49% und die SBBias-Methode lediglich 73.04%. Jedoch wurde mittels keiner Methode die geforderte Mindestüberdeckung von 87.3% erreicht. Für den Fall eines “sehr guten“ Prozesses waren sogar alle Überdeckungen zu gering. Die BCPB-Methode und die Perzentil-Methode erfüllen das geforderte Gütekriterium lediglich im Fall eines “schlechten“ bzw. “guten“ Prozesses.

Aggregiert man die Ergebnisse der Simulation sowohl nach dem Fähigkeitsindex und nach dem Stichprobenumfang, so bestätigen sich die Ergebnisse. Aus Tabelle 2.11 ist ersichtlich, dass die SB- bzw. die SB-Bias-Methode – bis auf eine Ausnahme – die schlechtesten Ergebnisse bezüglich des hier betrachteten Gütemaßes liefern, hingegen die BCPB- bzw. die Perzentil-Methode die besten Ergebnisse erzielen (ausgenommen der Fall:  $N = 50$  und  $QFI = C_p$ ). In dieser Darstellung erreicht keine der Methoden die geforderte Mindestüberdeckung.

Qualität	Bca	Perzentil	SBBias	BCPB	SB
$C_p$	77.13%	77.40%	73.80%	77.59%	77.54%
$C_{pk}$	83.27%	84.88%	71.67%	85.13%	76.75%
$C_{pm}$	85.75%	85.53%	75.13%	85.82%	79.35%
$C_{pmk}$	79.43%	84.65%	71.55%	84.08%	76.33%

Tabelle 2.10: Prozentuale Überdeckung der Konfidenzintervalle aufgegliedert nach unterschiedlichen Fähigkeitsindizes

Wird die Überdeckung über die einzelnen Indizes aggregiert (vgl. Tabelle 2.10), so bestätigt sich das Ergebnis aus den vorherigen Beobachtungen: Die BCPB-Methode ist am geeignetsten, um ein Konfidenzintervall für die Klasse nicht-parametrischer Indizes  $C'_p$  zu konstruieren. Sie hat die höchste prozentuale Gesamtüberdeckung von 83.15%.

QFI	$N$	Bca	Perzentil	SBBias	BCPB	SB
$C_p$	50	52.23%	52.23%	66.63%	52.23%	68.70%
$C_p$	120	84.37%	84.37%	74.40%	84.37%	80.37%
$C_p$	400	90.77%	90.77%	77.17%	90.77%	82.27%
$C_p$	900	81.13%	82.23%	77.00%	83.00%	78.83%
$C_{pk}$	50	82.17%	80.07%	69.10%	80.93%	74.80%
$C_{pk}$	120	91.10%	94.30%	73.90%	93.80%	80.53%
$C_{pk}$	400	84.50%	87.90%	72.20%	87.03%	78.10%
$C_{pk}$	900	75.30%	77.23%	71.47%	78.73%	73.57%
$C_{pm}$	50	83.27%	80.90%	71.40%	80.93%	76.33%
$C_{pm}$	120	91.20%	89.93%	75.40%	90.67%	81.50%
$C_{pm}$	400	88.40%	90.00%	77.10%	89.60%	81.73%
$C_{pm}$	900	80.13%	81.30%	76.60%	82.07%	77.83%
$C_{pmk}$	50	76.93%	82.93%	69.73%	82.37%	75.20%
$C_{pmk}$	120	84.47%	92.80%	73.33%	90.50%	79.80%
$C_{pmk}$	400	82.27%	86.93%	72.20%	85.93%	77.53%
$C_{pmk}$	900	74.07%	75.93%	70.93%	77.50%	72.77%

Tabelle 2.11: Prozentuale Überdeckung der Konfidenzintervalle aufgedgliedert nach unterschiedlichen Fähigkeitsindizes und Stichprobenumfang

Jedoch kann bei weitem nicht die Güteeigenschaft erreicht werden, wie es beim  $E_{ci}$  der Fall ist.

## 2.4 Zusammenfassung und Bewertung

Die zentrale Aufgabe des Qualitätsmanagements eines Unternehmens ist die Sicherstellung und die Verbesserung der Qualität gefertigter Produkte. Dazu werden unterschiedliche Konzepte wie beispielsweise die statistische Prozesslenkung (SPC) verwendet, deren Methoden es ermöglichen, minderwertige Produkte und Prozesse schon früh im Wertschöpfungsprozess geeignet zu identifizieren. Die Idee der dabei zugrunde liegenden Methodik ist, die Qualität des jeweiligen Fertigungsprozesses mittels Qualitätsfähigkeitsindizes zu messen und darauf aufbauend, geeignete Maßnahmen für den Fertigungsprozess abzuleiten.

In dem vorliegenden Kapitel wurden zunächst die in der Praxis verwendeten parametrischen und nicht-parametrischen Qualitätsfähigkeitsindizes vorgestellt und um einen neuen, nicht-parametrischen Index erweitert. Um den  $E_{ci}$  zu motivieren, wurde der Begriff *Produktionsprozess*, der im vorliegenden Fall die für die Qualitätsbestimmung notwendigen Produktionsparameter und die dazugehörigen Messergebnisse beschreibt (vgl. Definition 2.5), eingeführt. Mittels des Konzeptes eines Produktionsprozesses und des zugrunde liegenden Wahrscheinlichkeitsraumes lassen sich QFIs als Schätzer für die Qualität eines Produktionsprozesses auffassen, wobei die so getroffenen Qualitätsaussagen mit Unsicherheit behaftet sind. Eine Möglichkeit, diese statistische Unsicherheit zu kontrollieren, ist die Verwendung von Konfidenzintervallen. Wie in Abschnitt 2.3 dargestellt, ist das Problem bei der Bestimmung von Konfidenzintervallen für Qualitätsindizes, dass zum einen für viele QFIs keine analytisch konstruierbaren Konfidenzintervalle existieren und zum anderen die Verteilung der zugrunde liegenden (Mess)Daten im Allgemeinen unbekannt ist.

Um trotz der Unbekanntheit der Verteilung der Messdaten Konfidenzintervalle bestimmen zu können, wird das Bootstrapverfahren verwendet. Die Grundidee des Verfahrens ist, die vorliegende Stichprobe des Produktionsprozesses als die beste Approximation der "echten", unbekanntem Verteilung zu betrachten. Mittels wiederholtem Ziehen mit Zurücklegen aus dieser Stichprobe, welche in diesem Falle als Grundgesamtheit aufgefasst wird, entsteht die so genannte Bootstrap-Statistik, mit deren Hilfe Konfidenzintervalle konstruiert werden. Zur konkreten Bestimmung von Konfidenzintervallen mittels des Bootstrap-Verfahrens können unterschiedliche Methoden eingesetzt werden, die je nach Art des zugrunde liegenden (Qualitäts-)Schätzers unterschiedlich geeignet sind.

Zur Bestimmung der jeweils besten Bootstrap-Konfidenzintervall für die hier vorgestellten QFIs, wurde in Abschnitt 2.3.3 eine Simulationsstudie durchgeführt. Dabei konnte gezeigt werden, dass das *Bias Corrected Percentile Bootstrap*-Verfahren (BCPB) für die in der Praxis häufig verwendete Klasse von nicht-parametrischen QFIs  $C'_p$  (vergleiche Definition 2.4) die beste Wahl ist. Im Gegensatz dazu ergab die Studie, dass es für den

neu eingeführten empirischen Fähigkeitsindex  $E_{ci}$  nicht notwendig ist, ein Konfidenzintervall mittels Bootstrap-Verfahren zu konstruieren. Vielmehr kann ein Intervall mittels der geschätzten Standardabweichung bestimmt werden.

Die in diesem Kapitel vorgestellten Ansätze und Verfahren zur Qualitätssicherung und -lenkung ermöglichen es, die Qualität eines Produktionsprozesses mathematisch zu messen und zu bewerten. Allerdings ist man mittels Qualitätsfähigkeitsindizes im Allgemeinen nicht in der Lage, diejenigen Einflussgrößen (Produktionsparameter) zu bestimmen, welche einen signifikanten Einfluss auf die Qualität eines Fertigungsprozesses haben. Jedoch sind die Bestimmung solcher Einflussfaktoren und die daraus abzuleitenden Maßnahmen entscheidend für eine nachhaltige Qualitätsverbesserung und Wettbewerbsfähigkeit eines Unternehmens. Ein solches Verfahren, mit dessen Hilfe es möglich ist, signifikante Einflussgrößen auf die mittels QFIs gemessene Qualität eines Produktionsprozesses zu bestimmen und somit letztendlich die Qualität des Prozesses zu optimieren, ist Gegenstand des nächsten Kapitels.

# 3 Fertigungsprozessoptimierung

Die Qualität eines Fertigungsprozesses kann als die Fähigkeit betrachtet werden, ein bestimmtes Produkt in gleich bleibender Art und Weise innerhalb der vorgegebenen Toleranzgrenzen ( $USL$  und  $LSL$ ) und möglichst nahe an dem Optimalitätspunkt  $T$  zu produzieren. Da eine Abweichung von diesem optimalen Punkt  $T$  eine Minderung der Qualität zur Folge hat, ist es die Aufgabe des Qualitätsmanagements diese Abweichung (Streuung) zu minimieren. Im vorangegangenen Kapitel wurden verschiedene Qualitätsfähigkeitsindizes (QFI) vorgestellt, die es ermöglichen, die Qualität eines Prozesses, unter Berücksichtigung seiner Lage zu den Spezifikationsgrenzen und seiner Streuung mathematisch abzubilden und skaleninvariant zu bewerten. Aus diesen Bewertungen lassen sich für einen Fertigungsprozess Handlungsmaßnahmen ableiten, ob er beispielsweise noch den Qualitätsansprüchen eines Unternehmens genügt oder nicht.

Der große Nachteil der im vorigen Kapitel vorgestellten Fähigkeitsindizes ist jedoch, dass sie ein aggregiertes Maß für die Qualität liefern, jedoch keine Aussage machen, wie der Prozess verbessert werden kann bzw. ob es Sub-Prozesse (vgl. Definition 3.1) des Fertigungsprozesses gibt, die schlechter oder besser sind als der Gesamtprozess. Ein Fähigkeitsindex kann somit nicht die Frage beantworten, welche Aspekte (Einflüsse) eines Prozesses für eine mindere oder bessere Qualität verantwortlich sind und somit ist er nicht in der Lage, eventuelles Optimierungspotential zu identifizieren. Doch eben dieses Wissen über mögliches Optimierungspotential ist notwendig, um einen Prozess, der mittels SPC gesteuert wird, langfristig und nachhaltig zu verbessern. Fasst man einen Prozess wie in Abschnitt 2.2 auf, so ist folgende Fragestellung zu beantworten: gibt es eine Teilmenge des Wertebereiches der Prozessparameter, welche die Qualität eines Produktionsprozesses verbessert, wenn man den Prozess auf lediglich diese Parameterwerte beschränkt. Mit anderen Worten ist ein Verfahren zu finden, mit dessen Hilfe nominale Einflussgrößen gefunden werden können, die einen signifikanten Einfluss auf stetige Messergebnisse haben.

Im Verlauf dieses Kapitels werden verschiedene Verfahren, zur Identifizierung signifikanter Einflussgrößen motiviert und analysiert. Es wird jedoch gezeigt, dass die in der Literatur bekannten Verfahren auf einem parametrischen Ansatz beruhen und Voraussetzungen benötigen, die in der Praxis nur in wenigen Fällen erfüllt sind. Aus diesem Grund wird im zweiten Teil ein neuer, nicht-parametrischer Ansatz, der wenige Voraussetzun-

gen benötigt, hergeleitet. Dabei wird die Identifizierung von Optimierungspotential auf ein Optimierungsproblem abgebildet, welches mit Hilfe von Branch and Bound Techniken effizient gelöst wird. Im letzten Abschnitt dieses Kapitels wird das vorgestellte Optimierungsmodell anhand von Daten einer Gießerei der Daimler AG evaluiert.

## 3.1 Varianzanalyse

Die am weitesten verbreitete Methode zur Identifizierung von signifikanten, nominalen Parametern, die einen Einfluss auf stetige Größen wie beispielsweise Messergebnisse haben, ist die Varianzanalyse. Die Varianzanalyse (engl.: analysis of variance, ANOVA), ist eine Klasse statistischer, strukturprüfender Verfahren zur Analyse des Einflusses des Niveaus von einem oder mehreren Faktoren ("Einflussgröße") auf den Mittelwert einer oder mehrerer abhängiger Variablen.

Prinzipiell lassen sich zwei Arten der Varianzanalyse unterscheiden:

**Univariate Varianzanalyse** Werden Mittelwertanalysen für eine abhängige Variable durchgeführt, so spricht man von einer univariaten Varianzanalyse. Wird dabei nur eine Einflussvariable zur Analyse genutzt, so spricht man von einer einfaktoriellen Varianzanalyse, werden mehrere Einflussvariablen einbezogen von einer mehrfaktoriellen Analyse.

**Multivariate Varianzanalyse** Hier wird im Gegensatz zur Univariaten Varianzanalyse mehr als eine abhängige Variable betrachtet. Analog zum univariaten Fall unterscheidet man wiederum ein- und mehrfaktorielle multivariate Varianzanalysen. Die multivariate Analyse wird an dieser Stelle nicht weiter verfolgt, da in dieser Arbeit der Fokus auf solchen Prozessen liegt, in denen bei jeder produzierten Einheit lediglich eine Messreihe erhoben wird bzw von Relevanz ist.

Der Begriff "Varianzanalyse" bedeutet dabei nicht, dass die Varianzen der Messergebnisse untersucht werden, sondern es wird der Einfluss der Klassen auf den Erwartungswert analysiert. Die Grundidee dabei ist, die Gesamtvarianz der Zielvariable in die Varianz zwischen den Klassen und die Varianz innerhalb der Klassen aufzuteilen.

Der Begriff Klasse ist in diesem Zusammenhang wie folgt definiert: die gemessenen Daten (Zielgrößen)  $X = (x_1, \dots, x_n)$  werden als Realisierungen bezeichnet. Diese hängen im einfaktoriellen Fall von einer Variablen (oder Faktor)  $Y$  mit  $k$  Klassen – auch Faktorstu-

fen oder Ausprägungen – ab. Bei dem einfachsten Fall der einfaktoriellen Varianzanalyse<sup>1</sup> wird folgendes lineares Modell zu Grunde gelegt:

$$X_{ij} = \theta_i + \epsilon_{ij}, i = 1, \dots, k; j = 1, \dots, n \quad (3.1)$$

wobei  $\theta_i$  die unbekannt Parameter (Mittelwerte) und  $\epsilon_{ij}$  die Störvariablen sind. Dabei wird  $\epsilon_{ij}$  als normalverteilt mit Varianz  $\sigma_i^2$  angenommen. Schematisch lassen sich die beobachteten Werte der  $k$  Faktorstufen wie folgt strukturieren:

Stufe	1	2	3	...	$k$
	$x_{11}$	$x_{21}$	$x_{31}$	...	$x_{k1}$
	$x_{12}$	$x_{22}$	$x_{32}$	...	$x_{k2}$
	$\vdots$	$\vdots$	$\vdots$	...	$x_{k3}$
	$\vdots$	$x_{2n_2}$	$\vdots$		$\vdots$
	$\vdots$		$\vdots$		$x_{kn_k}$
	$x_{1n_1}$		$x_{3n_3}$		

Zur Durchführung der einfaktoriellen Varianzanalyse werden für das Modell (3.1) folgende Voraussetzungen gemacht:

1.  $E(\epsilon_{ij}) = 0, Var(\epsilon_{ij}) = \sigma_i^2 < \infty, \forall i, j. Cov(\epsilon_{ij}, \epsilon_{i'j'}) = 0, \forall i, i', j, j'$  falls  $j' \neq j, i' \neq i$
2.  $\{\epsilon_{ij}\}$  i.i.d. mit  $\epsilon_{ij} \sim N(0, \sigma_i^2)$  für  $i = 1 \dots, k, j = 1, \dots, n_i$  mit  $n_i$  Anzahl der Werte in der  $i$ -ten Faktorstufe
3.  $\sigma_i^2 = \sigma^2$  für  $i = 1 \dots, k$

Mit anderen Worten: die Messergebnisse ergeben sich aus einem klassenspezifischen Parameter  $\theta_i$  und einer Störvariable  $\epsilon_{ij}$ . Eine weitere starke Voraussetzung ist die Homoskedastizität, denn es wird angenommen, dass über alle Faktorstufen die Messergebnisse die gleiche Varianz haben. Ziel der Varianzanalyse ist, die folgende ANOVA-Nullhypothese  $H_0 : \theta_1 = \dots = \theta_k$  zu testen. Wie in [Casella und Berger 2002] gezeigt wird, ist diese Hypothese äquivalent mit der Hypothese  $H_0 : \sum_{i=1}^k a_i \theta_i = 0 \forall a \in \mathcal{A} = \{\mathbf{a} = (a_1, \dots, a_k)^T : \mathbf{a} \neq 0, \sum_{i=1}^k a_i = 0\}$ .

<sup>1</sup>In diesem Kapitel wird nur auf die einfaktorielle Varianzanalyse eingegangen, um das generelle Prinzip zu verdeutlichen. Für mehrfaktorielle oder multivariate Varianzanalysen sei an dieser Stelle auf die Literatur verwiesen (vgl. z.B. [Schmidt 2006])

Damit kann gezeigt werden, dass mit  $\bar{X}_i = \sum_{j=1}^{n_i} \frac{X_{ij}}{n_i}$  gilt:

$$\sum_{i=1}^k a_i \bar{X}_i \sim N \left( \sum_{i=1}^k a_i \theta_i, \sigma^2 \sum_{i=1}^k \frac{a_i^2}{n_i} \right)$$

und somit

$$\frac{\sum_{i=1}^k a_i \bar{X}_i - \sum_{i=1}^k a_i \theta_i}{\sqrt{\sigma^2 \sum_{i=1}^k \frac{a_i^2}{n_i}}} \sim N(0, 1) \quad (3.2)$$

In der Regel kennt man  $\sigma^2$  jedoch nicht. Um trotzdem eine Aussage über  $\theta_i$  treffen zu können, muss  $\sigma^2$  geschätzt werden: die Varianz jeder Faktorstufe lässt sich schätzen durch

$$S_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2 \quad (3.3)$$

mit

$$\frac{(n_i - 1)S_i^2}{\sigma^2} \sim \chi_{n_i-1}^2 \quad (3.4)$$

Da unter Homoskedastizität alle  $S_i^2$  das gleiche  $\sigma^2$  schätzen, kann man einen verbesserten Schätzer – die so genannte *gepoolte* Stichprobenvarianz – bestimmen (vgl. [Schmidt 2006], Kapitel 3.1):

$$S_p^2 = \frac{1}{n - k} \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2 \quad (3.5)$$

Aufgrund der Unabhängigkeit aller  $S_i^2$ , folgt dass

$$\frac{(n - k)S_p^2}{\sigma^2} \sim \chi_{n-k}^2 \quad (3.6)$$

Darüber hinaus ist  $S_p^2$  unabhängig von  $\bar{X}_i$ , so dass gilt:

$$\frac{\sum_{i=1}^k a_i \bar{X}_i - \sum_{i=1}^k a_i \theta_i}{\sqrt{S_p^2 \sum_{i=1}^k \frac{a_i^2}{n_i}}} \sim t_{n-k} \quad (3.7)$$

mit  $t_{n-k}$  einer Student-t-Verteilung mit  $n - k$  Freiheitsgraden.

Um nun die Hypothese  $H_0 : \sum_{i=1}^k a_i \theta_i = 0, \forall a \in \mathcal{A}$  vs.  $H_1 : \exists a \in \mathcal{A}$  mit  $\sum_{i=1}^k a_i \theta_i \neq 0$  (vgl. [Casella und Berger 2002], Kap. 11.2) mit Hilfe von

$$T_a = \left| \frac{\sum_{i=1}^k a_i \bar{X}_i - \sum_{i=1}^k a_i \theta_i}{\sqrt{S_p^2 \sum_{i=1}^k \frac{a_i^2}{n_i}}} \right| \quad (3.8)$$

zu testen, ist es naheliegend,  $H_0$  abzulehnen, falls  $\sup_{a \in \mathcal{A}} T_a > c$  ist, wobei  $c$  so gewählt wird, dass  $P(\sup_{a \in \mathcal{A}} T_a > c) = \alpha$ . Die Bestimmung von  $\sup_{a \in \mathcal{A}} T_a$  ist jedoch äquivalent mit der Bestimmung von  $\sup_{a \in \mathcal{A}} T_a^2$ . Damit kann gezeigt werden (vgl. Schmitt, Kapitel 3, [Casella und Berger 2002], Kap. 11.2), dass unter  $H_0$  mit  $\bar{X}_{..} = \sum_{i=1}^k n_i \bar{X}_i / \sum_{i=1}^k n_i$  gilt:

$$\sup_{a \in \mathcal{A}} T_a^2 = \frac{\sum_{i=1}^k n_i \left( (\bar{X}_i - \bar{X}_{..}) \right)^2}{(k-1)S_p^2} \sim F_{k-1, n-k} \quad (3.9)$$

Somit wird die Nullhypothese  $H_0 : \sum_{i=1}^k a_i \theta_i = 0, \forall a \in \mathcal{A}$  zum Niveau  $\alpha$  abgelehnt falls

$$\frac{\sum_{i=1}^k n_i (\bar{X}_i - \bar{X}_{..})^2}{S_p^2} > (k-1)F_{k-1, n-k, \alpha} \quad (3.10)$$

Eine anschauliche Deutung von Zähler und Nenner der Teststatistik (3.10) liefert die folgende Zerlegung in Quadratsummen<sup>2</sup> ( $SS$ ), die auch die grundsätzliche Idee der Varianzanalyse widerspiegelt.

Bezeichne  $SS_T$  die Gesamtvariabilität,  $SS_B$  die Unterschiedlichkeit zwischen den Klassen<sup>3</sup> und  $SS_W$  die Unterschiedlichkeit innerhalb einer Klasse<sup>4</sup>. Somit gilt:

$$\sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_{..})^2 = \sum_{i=1}^k n_i (\bar{X}_i - \bar{X}_{..})^2 + \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2 \quad (3.11)$$

Die Varianzen innerhalb der Klassen werden im Kontext der Varianzanalyse als mittlere Quadratsummen ( $MS_B$ ) bezeichnet, mit  $MS_B = \frac{SS_B}{k-1}$ . Analog ergibt sich die mittlere Quadratsumme zwischen den Klassen zu  $MS_W = \frac{SS_W}{n-k}$ . Dadurch ist (3.10) äquivalent zu

$$F = \frac{MS_B}{MS_W} > F_{k-1, n-k, \alpha}$$

Mit Hilfe der Varianzanalyse lässt sich somit die Frage beantworten, ob Variablen einen signifikanten Einfluss auf den Mittelwert haben oder nicht. Dabei wird vorausgesetzt, dass die Streuung innerhalb der Klassen der Variablen konstant ist (Homoskedastizität). Darüber hinaus ist die Varianzanalyse ein parametrisches Modell, das eine multivarierte Normalverteilung der zugrunde liegenden Daten fordert. Eine weitere Einschränkung ist die Anzahl der Hypothesentests, die durchgeführt werden müssen, um signifikante

<sup>2</sup>Englisch: **S**um of **S**quare (SS)

<sup>3</sup>**B**etween groups

<sup>4</sup>**W**ithin groups

Haupteffekte und Wechselwirkungen zu identifizieren. Bei der Anwendung von Hypothesentests muss a priori ein Signifikanzniveau  $\alpha$  definiert werden. Es besteht folglich die Gefahr, bei jedem Test die Nullhypothese zu verwerfen, obwohl sie wahr ist (multiples Hypothesenproblem). Dies impliziert, dass beispielsweise für  $\alpha = 0.05$  im Mittel einer von 20 Tests zu einem falschen Ergebnis führt. Für  $m$  Einflussvariablen müssen stets  $m^{\frac{m-1}{2}}$  Hypothesentests für die Haupteffekte durchgeführt werden, so dass mit hoher Wahrscheinlichkeit mindestens ein Test zu einem nicht richtigen Resultat führt. Somit eignet sich ANOVA nur für Analysen mit wenigen Einflussvariablen.

Eine nicht-parametrische Alternative zur Varianzanalyse ist der nach dem amerikanischen Ökonomen Milton Friedman benannte *Friedman-Test* zum Vergleich von mehreren, verbundenen Stichproben. Sei eine Matrix  $Y \in \mathbb{R}^{n \times k}$  gegeben, wobei  $n$  Zeilenanzahl (Blocks),  $k$  die Anzahl der Gruppen ist und jeder Eintrag der Matrix ein Messergebnis darstellt. Zunächst wird für jedes Element von  $Y$  der Rang innerhalb des Blocks (Zeilenrang) bestimmt und anschließend der Wert durch den Rang ersetzt ( $Y \rightsquigarrow R$ ). Getestet wird nun die Hypothese  $H_0$ , dass die Rangsummen aller Gruppen ungefähr gleich groß sind, das heißt, sie weichen nur zufällig voneinander ab vs. die Alternativhypothese  $H_1$ , dass ein nicht zufälliger, systematischer Unterschied besteht. Dies geschieht mit Hilfe der Teststatistik  $F_R$ :

$$F_R = \frac{12}{nk(k+1)} \sum_{i=1}^k R_{.i}^2 - 3n(k+1) \quad (3.12)$$

Dabei ist  $R_{.i}^2$  die Rangsumme der  $i$ -ten Gruppe und die Teststatistik (3.12) folgt asymptotisch einer  $\chi^2$ -Verteilung mit  $k$  Freiheitsgraden. Im Gegensatz zur Varianzanalyse untersucht der Friedman-Test jedoch nicht die Wechselwirkungen zwischen den unabhängigen Variablen und ist somit nur bedingt aussagekräftig.

Angewendet auf das Problem der Identifizierung von Optimierungspotential für Produktionsprozesse ist die Varianzanalyse und der Friedman-Test nur eingeschränkt einsetzbar. Wie im vorherigen Kapitel gezeigt wurde, wird zur Messung der Qualität eines Produktionsprozesses sowohl die Lage als auch die Streuung des Prozesses in das Verhältnis zu den Spezifikationsgrenzen gesetzt. Somit braucht man ein Modell, welches sowohl die Lage als auch die Streuung der Faktorstufen analysiert. Die Möglichkeit, eine Veränderung der Prozessvariabilität mittels quadrierter Messergebnisse zu untersuchen scheitert zum einen an der geforderten Homoskedastizität. Zum anderen birgt die somit erforderliche zusammengesetzte Hypothese ein Hindernis: ist sie verletzt, d.h. Mittelwert und Varianz sind beide unterschiedlich, können sie sich beide gegenseitig in Bezug auf die Verletzung der Toleranzgrenzen entweder verstärken oder aufheben. Zusätzlich würden bei einer solchen Varianzanalyse die (quadrierten) Messergebnisse keiner Normalverteilung mehr folgen.

Eine weitere Restriktion für den Einsatz der vorgestellten Verfahren ist die Voraussetzung, dass sich die Einflussmenge aus wenigen, gut balancierten Variablen zusammensetzt. Mit anderen Worten: es gibt wenige Faktorstufen und die Anzahl der Messergebnisse in allen Faktorstufen ist ungefähr gleich. Der Vorteil des Friedman-Tests, keine Verteilung vorauszusetzen, wird durch die Tatsache, dass Wechselwirkungen nicht beachtet werden, egalisiert. Somit bilden beide Ansätze keine stabilen Verfahren für die Optimierung von SPC gestützten Produktionsprozessen. Um dennoch einen Produktionsprozess auf Basis der statistischen Prozesslenkung optimieren zu können, wird in den folgenden Abschnitten ein neues, nicht parametrisches Verfahren – *Root Cause Analysis* – vorgestellt, mit dessen Hilfe *optimale* Variablen und deren Werte (Klassen) bestimmt werden können.

## 3.2 Root Cause Analysis

Die Verwendung von Qualitätsindizes wie beispielsweise  $C_p$ ,  $C_{pk}$  oder  $E_{ci}$  ermöglicht die Bewertung der Qualität eines Produktionsprozesses mit einer einheitslosen Größe. Aufgrund dieser Größe werden für einen Prozess vom Qualitätsmanagement Maßnahmen abgeleitet, wie Weiterführung der Produktion oder ihrer kurzzeitige Einstellung. Letzteres würde dann zu einer Neukalibrierung des Prozesses genutzt werden können. Dabei versteht man unter einer Kalibrierung die Identifikation und Eliminierung von Ursachen, die dazu führen, dass ein Prozess eine schlechtere Qualität liefert als erwartet bzw. gefordert. Doch eben diese Information, welche Prozessparameter die Ursache für eine mindere Qualität sind, kann durch die Bestimmung eines Indizes nicht identifiziert werden.

Aus diesem Grund wird in der Praxis, falls ein Prozess mit minderer Qualität vorliegt, mit Hilfe von Versuchsaufbauten und Bewertung dieser Versuche die Ursache für die Verschlechterung des Produktionsprozesses ermittelt. Unter Umständen kann es dabei dazu kommen, dass Teile der Produktion angehalten werden müssen, da die Auswertung und Ursachenanalyse oftmals viel Zeit in Anspruch nehmen. Dies kann zu finanziellen Verlusten für ein Unternehmen führen, da die Versuchsaufbauten kostspielig sind und der Stillstand der Produktion zu Ausfällen führt oder gar Regressforderungen aufgrund nicht eingehaltener Liefertermine nach sich zieht. Was man folglich benötigt ist eine Methode, welche Ursachen für eine Qualitätsverschlechterung zeitnah und kostengünstig identifiziert, so dass die Produktion möglichst uneingeschränkt fortgeführt werden kann.

Doch nicht nur in den Extremfällen, in denen die Qualität deutlich unter ein für den jeweiligen Prozess vorgegebenes Mindestniveau sinkt, ist man daran interessiert, mög-

liche Ursachen für Minderqualität zu finden. Auch in einer Situation, in der ein QFI einen (geforderten) Mindestqualitätswert überschreitet (z.B.  $C_{pk} > 1.67$  vgl. Tabelle 2.1), ist es im Zuge des Six-Sigma Prinzips wünschenswert, Verbesserungspotential zu identifizieren. Das im Folgenden vorgestellte *Root-Cause-Analysis*-Verfahren (RCA) ist ein neuer, nicht-parametrischer Ansatz um Ursachen für eine veränderte Qualität eines Produktionsprozesses, der mittels SPC überwacht wird, ohne kostspielige Versuche und Analysen zu identifizieren.

Wie in Kapitel 2.2.2 eingeführt, lässt sich ein Produktionsprozess durch das Tupel von Zufallsvariablen  $\mathcal{X} = (X, \vec{Y})$  beschreiben (vergleiche Def. 2.5) und ein Sub-Prozess wie folgt definieren:

**Definition 3.1** (Sub-Prozess)

Sei  $(\Omega, \Sigma, P)$  ein Wahrscheinlichkeitsraum und  $\mathcal{X} = (X, \vec{Y}) : (\Omega, \Sigma, P) \rightarrow R^{m+1}$  ein Produktionsprozess. Sei  $\mathcal{S} = \{\vec{Y}(\omega) | \omega \in \Omega\}$  der Wertebereich von  $\vec{Y}$  und  $\mathcal{S}_0 \subset \mathcal{S}$  mit  $\mathcal{S}_0 \neq \emptyset$ . Zum eingeschränkten Wertebereich  $\mathcal{S}_0$  (auch als Restriktionsmenge bezeichnet) gibt es eine Teilmenge  $\Omega_0 = \{\omega | \vec{Y}(\omega) \in \mathcal{S}_0\} \subset \Omega$ . Sei  $(\Omega_0, \Sigma_0, P_0)$  der sich aus  $(\Omega, \Sigma, P)$  durch Einschränkung auf  $\Omega_0$  ergebende Wahrscheinlichkeitsraum, so ist  $\mathcal{X}_{\mathcal{S}_0} = (X, \vec{Y}_{\Omega_0}) = (X, \vec{Y} | \vec{Y} \in \mathcal{S}_0)$  ein Produktionsprozess, der als Sub-Prozess von  $\mathcal{X} = (X, \vec{Y})$  bezeichnet wird.

Es wird im weiteren Verlauf die Notation  $\mathcal{X}_{\mathcal{S}_0} \subseteq \mathcal{X}$  verwendet, um einen Sub-Prozess eines gegebenen diskreten Produktionsprozesses zu beschreiben.

Ein Spezialfall der Definition 3.1 ist das Fixieren einiger der Einflussgrößen auf einen speziellen Wert bei gleichzeitigem Nichtfixieren (Variabelhalten) der restlichen Einflussgrößen und deren Werte.

Sei  $S^i = \{s_1^i, \dots, s_{m_i}^i\}$  die Menge möglicher Werte der Einflussgrößen  $Y_i$  für  $i = 1, \dots, m$ . Sei ferner  $\mathcal{S} = S^1 \times \dots \times S^m$ , d.h. alle Kombinationsmöglichkeiten der Werte der Einflussgrößen sind möglich. Dann sind folgende, spezielle Sub-Prozesse von Interesse:

**Definition 3.2** (Singuläre Fixierung)

Sei  $\tilde{s}_i$  ein Element des Wertebereichs der Einflussgröße  $Y_i$ . Dann wird ein Sub-Prozess, der durch das Fixieren der  $i$ -ten Einflussgröße auf den Wert  $\tilde{s}_i$  gegeben ist, mit

$$\mathcal{X}_{\tilde{s}_i} = (X, \vec{Y}_{\tilde{s}_i}) = (X, \vec{Y} | Y_i = \tilde{s}_i)$$

bezeichnet. Die dazugehörige Menge  $\mathcal{S}_0$  ist dann

$$\mathcal{S}_0 = S^1 \times \dots \times \{\tilde{s}_i\} \times \dots \times S^m$$

Der Sub-Prozess  $\mathcal{X}_{\tilde{s}_i}$  wird somit durch die Fixierung genau einer Variablen auf genau einen Wert beschrieben.

Erweitert man das Konzept, indem für genau eine Variable eine Einschränkung auf eine oder mehrere Werte festgelegt wird, so bezeichnen wir dies als einfache Fixierung:

**Definition 3.3** (Einfache Fixierung)

Sei  $\tilde{S}_i$  eine nicht leere Teilmenge des Wertebereiches  $S^i$  der Einflussgröße  $Y_i$  mit<sup>5</sup>  $\tilde{S}_i \subsetneq S^i$ . Dann wird ein Sub-Prozess, der durch das Fixieren der  $i$ -ten Einflussgröße auf die Werte  $\tilde{S}_i$  gegeben ist, mit  $\mathcal{X}_{\tilde{s}_i}$  bezeichnet. Dies wird auch als einfache Fixierung bezeichnet und die Restriktionsmenge  $\mathcal{S}_0$  ist gegeben durch:

$$\mathcal{S}_0 = S^1 \times \cdots \times \tilde{S}_i \times \cdots \times S^m$$

Fordert man hingegen, dass ein Sub-Prozess durch die Fixierung einer oder mehrerer Variablen auf jeweils genau einen Wert entsteht, so bezeichnen wir dies als mehrfache Fixierung:

**Definition 3.4** (Mehrfache Fixierung)

Sei  $J \subseteq \{1, \dots, m\}$ ,  $\tilde{s}_j \in S^j$  für  $j \in J$ . Dann wird der Sub-Prozess, der durch das gleichzeitige Einschränken der Variablen  $Y_j$  auf die Werte  $\tilde{s}_j$  entsteht, mit  $\mathcal{X}_{\tilde{s}_j, j \in J}$  bezeichnet

$$\mathcal{X}_{\tilde{s}_j, j \in J} = \bigcap_{j \in J} \mathcal{X}_{\tilde{s}_j}$$

Wird mindestens eine Variable auf mindestens einen Wert fixiert, so lässt sich dies formal wie folgt definieren:

**Definition 3.5** (Allgemeine Fixierung)

Sei  $J \subset \{1, \dots, m\}$ ,  $\tilde{S}_j \subsetneq S^j$ ,  $\tilde{S}_j \neq \emptyset$  für  $j \in J$ . Dann wird der Sub-Prozess  $\mathcal{X}_{\tilde{S}_j, j \in J}$ , der durch das simultane Einschränken der Variablen  $Y_j$  auf  $\tilde{S}_j$  für alle  $j \in J$  entsteht, als allgemeine Fixierung bezeichnet und die dazugehörige Restriktionsmenge  $\mathcal{S}_0$  ist gegeben durch

$$\mathcal{S}_0 = \bigcap_{j \in J} \bigcup_{\tilde{s}_j \in \tilde{S}_j} S^1 \times \cdots \times \{\tilde{s}_j\} \times \cdots \times S^m \quad (3.13)$$

Die Einschränkung des Wertebereiches, wie in den Definitionen 3.2 bis 3.5 aufgezeigt, ist eine Sichtweise, die in der Informatik, speziell in der Datenbanktechnologie, weit verbreitet ist. In Anlehnung daran kann ein Sub-Prozess, anstatt durch Restriktion aller

<sup>5</sup>Der explizite Ausschluss der Menge  $\tilde{S}_i = S^i$  ist notwendig, da es sich sonst nicht um eine Fixierung im eigentlichen Sinne handelt und alle Werte zugelassen werden würden

Einflussgrößen auf bestimmte Werte, mittels Fixierung bestimmter Variablen und deren Werte, bei gleichzeitigem Variabelhalten der übrigen Variablen beschrieben werden. In einem solchen Fall wird die Einschränkung des Wertebereiches durch die *Fixierungsmenge* beschrieben - es werden nur Variablen angegeben, deren Werte fixiert sind. Werden im Gegensatz dazu die zulässigen Einflussparameter eines Sub-Prozesses wie in Definition 3.1 beschrieben, so bezeichnen wir dies als Restriktion des Wertebereiches bzw. *Restriktionsmenge*.

Mittels (3.13) kann eine Fixierungsmenge, die einen Sub-Prozess durch eine allgemeine Fixierung beschreibt, in eine Restriktionsmenge überführt werden. Der umgekehrte Fall, die Darstellung eines Sub-Prozesses, der durch eine Restriktionsmenge beschrieben wird, ist mit Hilfe einer Fixierungsmenge nicht immer möglich. Für die hier vorliegende Arbeit ist es jedoch ausreichend, nur solche Sub-Prozesse zu betrachten, die sich mittels einer allgemeinen Fixierung darstellen lassen. Eben für diese Menge an Sub-Prozessen gilt:

$$\mathcal{X}_{S_0} = \mathcal{X}_{\tilde{S}_j, j \in J}$$

und folglich lassen sich in diesem Fall Eigenschaften und Ergebnisse für Restriktionsmengen auch auf Fixierungsmengen übertragen.

Variable	Wertemenge
A	$\{a_1, a_2, a_3\}$
B	$\{b_1, b_2\}$

Tabelle 3.1: Wertemengen eines Beispielprozesses

**Beispiel 3.1** (Fixierung eines Prozesses)

Sei ein Produktionsprozess mit Einflussgrößen wie in Tabelle 3.1 gegeben. So sind

$$\mathcal{X}_{\{a_1\}}, \mathcal{X}_{\{a_2\}}, \mathcal{X}_{\{a_3\}}, \mathcal{X}_{\{b_1\}}, \mathcal{X}_{\{b_2\}}$$

singulär fixierte Sub-Prozesse,

$$\mathcal{X}_{\{a_1, a_2\}}, \mathcal{X}_{\{a_1, a_3\}}, \mathcal{X}_{\{a_2, a_3\}}$$

einfach fixierte Sub-Prozesse

$$\mathcal{X}_{\{a_1, b_1\}}, \mathcal{X}_{\{a_2, b_1\}}, \mathcal{X}_{\{a_3, b_1\}}, \mathcal{X}_{\{a_1, b_2\}}, \mathcal{X}_{\{a_2, b_2\}}, \mathcal{X}_{\{a_3, b_2\}}$$

mehrfach fixierte Sub-Prozesse und

$$\mathcal{X}_{\{a_1, a_2, b_1\}}, \mathcal{X}_{\{a_2, a_3, b_1\}}, \mathcal{X}_{\{a_2, a_3, b_2\}}$$

allgemein fixierte Sub-Prozesse.

Der Subskript der Sub-Prozesse beschreibt dabei die jeweiligen Variablen, die fixiert werden, so ist  $\mathcal{X}_{\{a_1\}}$  der Sub-Prozess, der durch Fixieren der Variable  $A$  auf den Wert  $a_1$  bei gleichzeitigem Variabelhalten der Variable  $B$ . Im Falle von  $\mathcal{X}_{\{a_1, b_1\}}$  wird ein Sub-Prozess beschrieben, der durch das Fixieren der Variable  $A$  auf den Wert  $a_1$  und der Variable  $B$  auf den Wert  $b_1$  entsteht.

Für einen Sub-Prozess lassen sich ebenfalls die im vorherigen Kapitel vorgestellten Qualitätsfähigkeitsindizes wie  $C_p$ ,  $C_{pk}$ ,  $C_{pm}$  und  $C_{pmk}$  berechnen. Sei  $\mathbf{x} = (x_i, \dots, x_k)$  mit  $x_i = (x_i, \vec{y}_i)$  für  $i = 1, \dots, k$  eine konkrete Realisierung eines Produktionsprozesses vom Umfang  $k$ . Wird nun ein Sub-Prozess mit Restriktionsmenge  $\mathcal{S}_0$  betrachtet, so kann eine Qualitätskennzahl für den Sub-Prozess auf der Basis der Realisierung geschätzt werden, die der Forderung  $\vec{Y} \in \mathcal{S}_0$  genügt:

$$\mathbf{x} \mid \mathcal{S}_0 = (x_{i_1}, \dots, x_{i_l}) = ((x_{i_1}, \vec{y}_{i_1}), \dots, (x_{i_l}, \vec{y}_{i_l}))$$

mit  $\vec{y}_{i_t} \in \mathcal{S}_0$  für  $t = 1, \dots, l$ ,  $\{i_1, \dots, i_l\} \subset \{1, \dots, k\}$  und  $\vec{y}_i \notin \mathcal{S}_0$  für  $i \notin \{i_1, \dots, i_l\}$ .

Für die oben eingeführten Einschränkungen (Definitionen 3.2 bis 3.5) werden bei der Realisierung die analogen Bezeichnungen  $\mathbf{x}_{\tilde{s}_j}$ ,  $\mathbf{x}_{\tilde{s}_j}$ ,  $\mathbf{x}_{\tilde{s}_j, j \in J}$  und  $\mathbf{x}_{\tilde{s}_j, j \in J}$  für den Teil der Realisierung verwendet, der der entsprechenden Einschränkung genügt. Offensichtlich gilt dann:

$$\mathbf{x}_{\tilde{s}_j} = \bigcup_{\tilde{s}_j \in \tilde{S}_j} \mathbf{x}_{\tilde{s}_j}, \mathbf{x}_{\tilde{s}_j, j \in J} = \bigcap_{j \in J} \mathbf{x}_{\tilde{s}_j}$$

Im Folgenden wird für Sub-Prozesse der Begriff **bedingter Fähigkeitsindex** mit der Notation  $Q_{\mathcal{X}_{\mathcal{S}_0}}$  bzw.  $Q_{\mathbf{x}_{\mathcal{S}_0}}$  verwendet. Ein *unbedingter* QFI wird stets mit  $Q_{\mathcal{X}}$  bzw. der Schätzer auf einer konkreten Realisierung mit  $Q_{\mathbf{x}}$  notiert.

Der **Support** oder die **Mächtigkeit**  $n_{\mathcal{X}_{\mathcal{S}_0}}$  eines Sub-Prozesses  $\mathcal{X}_{\mathcal{S}_0}$  ist definiert als die Wahrscheinlichkeit des Ereignisses, dass die Einflussparameter  $\vec{Y}$  in  $\mathcal{S}_0$  sind. Damit gilt:

$$n_{\mathcal{X}_{\mathcal{S}_0}} = P(\vec{Y} \in \mathcal{S}_0) \quad (3.14)$$

Liegt eine Realisierung  $\mathbf{x} = (x_i, \vec{y}_i), i = 1, \dots, k$  eines Produktionsprozesses vor, ergibt sich ein Schätzer für den Support zu

$$\hat{n}_{\mathcal{X}_{\mathcal{S}_0}} = n_{\mathbf{x} \mid \mathcal{S}_0} = \frac{1}{k} \sum_{i=1}^k \mathbf{1}(\vec{y}_i \in \mathcal{S}_0) \quad (3.15)$$

In der Praxis ist oftmals vorteilhafter, bei einer Realisierung den Support nicht als relativen Anteil, sondern als absolute Häufigkeit zu betrachten. Dementsprechend gilt:

$$\begin{aligned}\hat{N}_{\mathcal{X}_{\mathcal{S}_0}} &= k\hat{n}_{\mathcal{X}_{\mathcal{S}_0}} = kn_{X|\mathcal{S}_0} \\ &= \sum_{i=1}^k \mathbf{1}(\vec{y}_i \in \mathcal{S}_0) \\ &= N_{X|\mathcal{S}_0}\end{aligned}$$

Somit ist  $\hat{N}_{\mathcal{X}_{\mathcal{S}_0}} \in [0, k]$  und für den *unbedingten* Fall  $\hat{N}_{\mathcal{X}} = k$ .

Beschreibt man einen Sub-Prozess wie beispielsweise in Definition 3.5 durch eine Fixierungsmenge, dann ergibt sich der Schätzwert für den Support zu

$$\begin{aligned}\hat{N}_{\mathcal{X}_{\tilde{\mathcal{S}}_j, j \in J}} &= \sum_{i=1}^k \mathbf{1}(\vec{y}_i \in \bigcap_{j \in J} \bigcup_{\tilde{s}_j \in \tilde{\mathcal{S}}_j} S^1 \times \cdots \times \{\tilde{s}_j\} \times \cdots \times S^m) \\ &= \sum_{i=1}^k \mathbf{1}(y_{ij} \in \tilde{\mathcal{S}}_j, j \in J)\end{aligned}$$

Um nun einen gegebenen Produktionsprozess  $\mathcal{X}$  auf Basis der statistischen Prozesslenkung zu optimieren, ist es notwendig, einen geeigneten Sub-Prozess  $\mathcal{X}_{\mathcal{S}_0}$  zu identifizieren, der im Sinne der Statistischen Prozesslenkung bessere Charakteristika aufweist als der Originalprozess (vgl. Kapitel 2.1).

Er sollte die folgenden beiden Kriterien aufweisen:

**Geringere Prozessvariabilität**  $Var(\mathcal{X}_{\mathcal{S}_0}) \leq Var(\mathcal{X})$

**Bessere Zentrierung**  $|\mu(\mathcal{X}_{\mathcal{S}_0}) - M| \leq |\mu(\mathcal{X}) - M|$  bzw.  $|\mu(\mathcal{X}_{\mathcal{S}_0}) - T| \leq |\mu(\mathcal{X}) - T|$ .

wobei  $M$  die Mitte des Toleranzbereiches (vgl. Seite 17) und  $T$  den Sollwert (siehe Seite 12) des Produktionsprozesses beschreibt<sup>6</sup>

Die Suche nach einer geeigneten Restriktionsmenge  $\mathcal{S}_0$  kann nun so erfolgen, dass man die Forderung an die entsprechenden Schätzwerte aufgrund der vorliegenden Realisierung des Produktionsprozesses stellt. Der Nachteil bei dieser Vorgehensweise ist, dass

<sup>6</sup>Bei Verwendung von nicht-parametrischen QFIs werden entsprechend der Erwartungswert durch den Median ersetzt

der für die Schätzung zur Verfügung stehende Stichprobenumfang von der gewählten Restriktionsmenge abhängt. Daher sollte bei der Wahl einer Restriktionsmenge  $\mathcal{S}_0$  auch die Zuverlässigkeit der Schätzung berücksichtigt werden. Dies kann durch eine Forderung an den resultierenden Stichprobenumfang erfolgen. Erfüllen viele Abläufe der konkreten Realisierung die Restriktion, so spricht dies auch dafür, dass der Aufwand zur Umsetzung des Sub-Prozesses in der Realität vergleichsweise gering ist.

Variable	Variablenwerte
BAZ	{1, 2, 3, 4}
Spindel	{1, 2}
Center	{links, rechts}
Person	{Gast, Müller, Rixinger etc.}

Tabelle 3.2: Variablenausprägung des Beispielprozesses

Abbildung 3.1 veranschaulicht eine solche Verbesserung eines geeigneten Sub-Prozesses gegenüber seinem Originalprozess. Die dafür verwendeten Daten stammen aus der in Abschnitt 3.3 vorgestellten Konzeptevaluierung bzw. sind Messergebnisse aus einer Gießerei und haben die folgenden Eigenschaften:

**Messgrößen** Die gemessenen Werte beschreiben die geometrische Beschaffenheit (Unebenheit) des verwendeten Materials mit 0.0 als untere und 0.15 als obere technische Toleranzgrenze.

**Produktionsparameter** Die aufgezeichneten Einflussgrößen setzen sich aus vier unterschiedlichen Variablen zusammen (vgl. Tabelle 3.2):

1. Bearbeitungszentrum (BAZ) mit vier Ausprägungen
2. Zwei Arten von Spindeln, gekennzeichnet mit Werten 1 und 2
3. Verwendete Maschine im BAZ (Center), gekennzeichnet jeweils durch “links“ und “rechts“
4. Zusätzlich wurde der Name des Arbeiters, der dieses Werkstück bearbeitet hat, aufgezeichnet (insgesamt neun verschiedene Namen)

Messgröße	BAZ	Spindel	Center	Arbeiter
0.0326	1	1	rechts	Gast
0.0711	2	2	rechts	Gast
0.0384	4	2	rechts	Gast
0.0759	1	2	links	Gast
0.0849	2	2	links	Gast
0.0755	3	2	rechts	Gast
0.0763	1	1	rechts	Gast
0.0896*	4	1	rechts	Gast
0.0556	4	2	rechts	Müller
0.0486*	2	1	rechts	Gast
0.0502*	2	1	rechts	Gast
0.0297	1	1	links	Gast
0.0438	4	2	links	Gast
0.0719	3	2	rechts	Rixinger
0.0604	2	2	rechts	Gast
0.0372	1	1	rechts	Gast

Tabelle 3.3: Beispiel eines Produktionsprozesses. Die mit \* gekennzeichneten Messgrößen stellen Einträge dar, die die Bedingungen des Sub-Prozesses *BAZ not in (1)* und *Center in ('rechts')* erfüllen

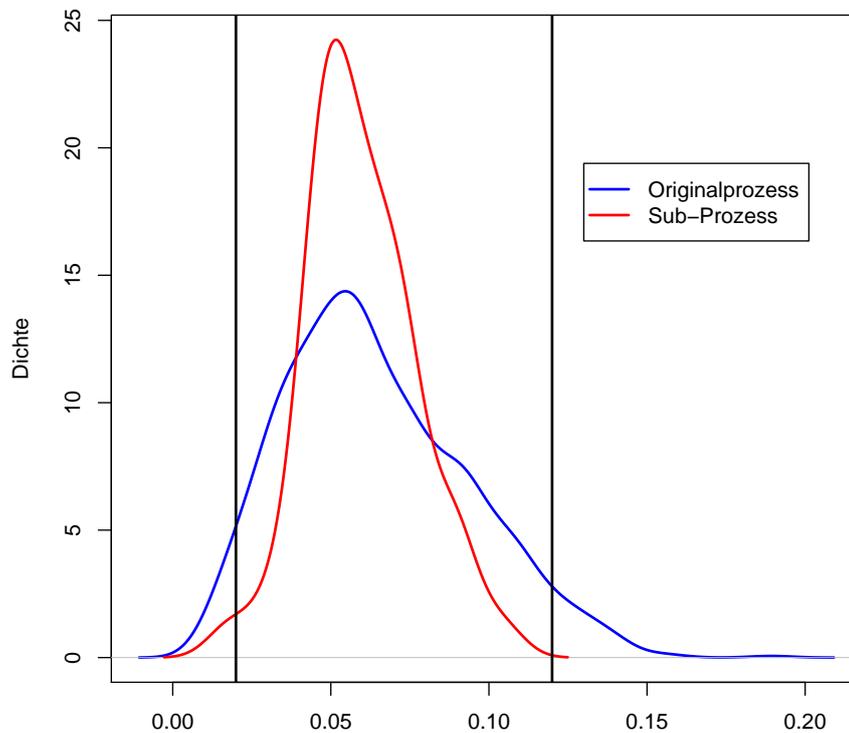


Abbildung 3.1: Prozessdichteplot

In Tabelle 3.3 ist ein Ausschnitt der verwendeten Daten aufgelistet. Zum Beispiel weist der, durch Einschränkung des Wertebereiches (Fixierungsmenge) auf die Werte *BAZ not in (1)* und *Center in ('rechts')*, resultierende Sub-Prozess eine geringere Prozessvariabilität und eine verbesserte Prozesslage auf,

$$|\mu(X, \vec{Y} \mid Y_1 \neq 1 \wedge Y_3 = \text{'rechts'}) - M| \leq |\mu(X, \vec{Y}) - M|$$

so dass sich in diesem speziellen Fall der  $C_{pk}$  von 1.56 um 66% auf 2.59 erhöht. Gleichzeitig verringert sich der Support von  $N_{\mathcal{X}} = 968$  auf  $N_{\mathcal{X}_{\tilde{s}_0}} = 454$ .

Die Problemstellung ist nun, wie man einen solchen optimalen Sub-Prozess identifizieren kann, denn offensichtlich existiert weder eine analytische Lösung, noch können bekannte, statistische Verfahren eingesetzt werden. Einen Ausweg bietet die Transformation des Problems in ein ganzzahliges Optimierungsproblem, welches im Folgenden hergeleitet wird. Die Grundidee des Verfahren ist, einen Ansatz, der aus dem Data Mining bekannt ist mit einem Lösungsverfahren aus dem Operations Research zu kombinieren.

### 3.2.1 Optimierungsproblem

Ein erster, naiver Ansatz einen optimalen Sub-Prozess zu finden ist, über alle Sub-Prozesse  $x_0$  einer Realisierung eines Produktionsprozesses  $x$  zu iterieren und den Sub-Prozess  $\tilde{x}_0$  zu bestimmen, für den folgendes Optimalitätskriterium erfüllt ist:

$$Q_{\tilde{x}_0} \geq Q_{x_0} \text{ für alle Sub-Prozesse} \quad (3.16)$$

Neben dem Nachteil des hohen Rechenaufwandes für die kombinatorische Suche besteht die Gefahr, dass ein Sub-Prozess  $\tilde{x}_0$  als optimale Lösung identifiziert werden würde, der einen sehr geringen Support aufweist:  $N_{\tilde{x}_0} \ll k$ . Eine praktische verwertbare Aussage über Qualitätseinflüsse muss jedoch den Aufwand (Support) für die Beeinflussung, die zur anschließenden Qualitätsverbesserung führt mit berücksichtigen. Idealerweise würde man die auszuschließenden Einflüsse mit den Kosten für die Entfernung bzw. Verbesserung der qualitätsmindernden Einflüsse gewichten.

Da Erfassung und Bestimmung der Kosten praktisch mit vertretbarem Aufwand nicht möglich sind, kann man davon ausgehen, dass seltene Einflusskombinationen "billiger" auszuschließen sind als häufige. Das heist, je größer der Support eines Sub-Prozesses ist, desto einfacher ist es, der Empfehlung, die durch den optimalen Sub-Prozess definierten Prozesseinflüsse auszuschließen, zu folgen. Mit anderen Worten: Sub-Prozesse mit einem hohen Support sind denjenigen Sub-Prozessen vorzuziehen, die einen hohen QFI

besitzen, jedoch einen kleinen Support aufweisen. Dieses Prinzip wird auch bei dem klassischen Data Mining Verfahren, der Assoziationsanalyse verwendet. Bei der Generierung von Assoziationen werden oftmals lediglich solche Kombinationen betrachtet, die einen so genannten *Minimum-Support* aufweisen. Das Ziel ist somit, einen Sub-Prozess zu finden, der zum einen einen “optimalen“ Support hat und gleichzeitig einen “optimalen“ QFI. In Abbildung 3.2 ist zur Verdeutlichung dieses zweiseitigen Problems beispielhaft ein Prozess mit zwei Sub-Prozessen dargestellt. Dabei weist ein Sub-Prozess ( $\tilde{x}_0$ ) einen hohen QFI und niedrigeren Support auf, der andere Prozess ( $x_0$ ) besitzt einen höheren Support, jedoch einen geringeren QFI.

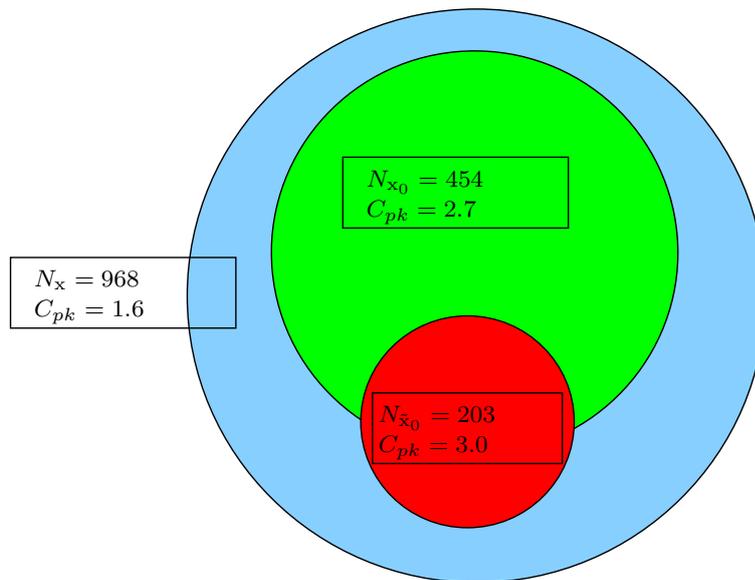


Abbildung 3.2: Schematische Darstellung des Optimierungsproblems

Ein Verfahren, welches das Optimalitätskriterium (3.16) verwendet, würde den “roten“ Sub-Prozess (Abbildung 3.2) als optimale Lösung identifizieren, denn es würde denjenigen Sub-Prozess identifizieren, der den höchsten QFI hat. Eine Anpassung des gesamten Fertigungsprozesses an diesen Sub-Prozess würde zwar auf der einen Seite die Qualität auf ein Maximum erhöhen, auf der anderen Seite aber wäre eine solche “Prozessverbesserung“ sehr kostspielig, da in der Regel viele Änderungen vorgenommen werden müssten. Im Gegensatz dazu liefert der “grüne“ Prozess eine nicht so hohe Qualität, jedoch ist eine Neueinstellung des Prozesses kostengünstiger realisierbar, da wenige Änderungen vorzunehmen sind. Folglich ist es zur Bestimmung eines optimalen Sub-Prozesses notwendig, einen Prozess zu identifizieren, der auf der einen Seite einen “optimalen“ (im Sinne eines möglichst hohen) Support aufweist, auf der anderen Seite jedoch gleichzeitig über einen hohen QFI verfügt.

In der Definition 3.1 wurde der Begriff Sub-Prozess allgemein eingeführt, da für einen

Sub-Prozess eine beliebige Einschränkung des Wertebereichs zugelassen wurde. In der Praxis ist meist eine beliebige Restringierung des Wertebereichs nicht möglich oder nur aufwendig zu realisieren, so dass es sinnvoll ist, die Menge der möglichen Einschränkungen entsprechend zu reduzieren. Fasst man Einschränkungen des Wertebereichs nicht als Restriktionsmengen – für jede Variable wird ein Wertebereich festgelegt – sondern als Fixierungsmengen auf, bieten mehrfach fixierte Sub-Prozesse wie in Definition 3.4 eine erste Möglichkeit, den zulässigen Raum einzuschränken<sup>7</sup>. Sei  $\tilde{\mathcal{A}}$  die Menge dieser Sub-Prozesse;  $\tilde{\mathcal{A}}$  besteht folglich aus allen Sub-Prozessen  $\mathcal{X}_{\tilde{s}_j, j \in J}$  mit  $J \subseteq \{1, \dots, m\}$  und  $\tilde{s}_j \in S^j$  für  $j \in J$ . Das heißt,  $J$  durchläuft alle Teilmengen von  $\{1, \dots, m\}$  und  $\tilde{s}_j$  ist ein Element des Wertebereiches der Variable  $Y^j$ .

Ein Lösungsansatz um Sub-Prozesse auf  $\tilde{\mathcal{A}}$  mit “optimalem“ Support zu identifizieren ist die Verwendung von Algorithmen wie beispielsweise *frequent item sets* (vgl. Abschnitt 4.4.1). Dabei werden solche Sub-Prozesse  $x_i$  aus der Realisierung identifiziert, deren Werte in allen Einflussgrößen übereinstimmen und besonders häufig sind. Folglich werden in einem ersten Schritt, diejenigen Sub-Prozesse mit maximalem Support ermittelt und anschließend auf dieser Menge der Sub-Prozesse bestimmt, der den höchsten QFI aufweist. In den meisten Anwendungen jedoch ist die vorhandene Stichprobe von geringem Umfang. Würde man lediglich solche (mehrfach fixierten) Sub-Prozesse betrachten, würde das zu Prozessen mit sehr kleinem Support führen ( $\tilde{N}_{x_{\tilde{s}_0}} \ll k$ ).

Diese erlauben jedoch keine fundierte Aussage, da sie unter Umständen nur statistisches Rauschen darstellen. Eine verlässliche Aussage, um anhand dieser (kleinen) Sub-Prozesse Handlungsempfehlungen für einen Produktionsprozess abzuleiten, ist nicht möglich. Aus diesem Grund ist es sinnvoll, den Raum der möglichen Sub-Prozesse gegenüber dem oben beschriebenen *frequent-item-sets*-Ansatz zu erweitern: Bezeichne  $\mathcal{A}$  die Menge aller allgemein fixierten Sub-Prozesse wie in Definition 3.5 eingeführt<sup>8</sup>. Die Suche nach dem (den) Sub-Prozess(en) aus der so gebildeten Kandidatenmenge  $\mathcal{A}$ , die auf der einen Seite über einen Mindestwert beim Qualitätsfähigkeitsindex verfügen und auf der anderen Seite einen maximalen Support haben, lässt sich somit als das folgende, ganzzahlige Optimierungsproblem formulieren:

**Definition 3.6** (RCA-Optimierungsproblem)

Sei eine konkrete Realisierung eines Produktionsprozesses  $x$  mit  $(x_i, \vec{y}_i), i = 1, \dots, k$  und  $\vec{y}_i \in \mathcal{S} \subseteq \{S^1 \times \dots \times S^m\}$  gegeben. Des Weiteren sei der Raum zulässiger Lösungen durch die Menge  $\mathcal{A}$  aller allgemein fixierter Sub-Prozesse beschrieben. Dann wird das

<sup>7</sup>Prinzipiell ist es auch möglich, nur einzelne Einflüsse auf einen Produktionsprozess zu untersuchen. Da man aber im Allgemeinen von multifaktoriellen Einflüssen auf die Qualität eines Prozesses ausgehen kann und singuläre Sub-Prozesse eine Teilmenge der mehrfach fixierten Sub-Prozesse darstellen, wird an dieser Stelle auf eine gesonderte Untersuchung verzichtet

<sup>8</sup>Zur Illustration ist auf Seite 71 der Raum möglicher Sub-Prozesse mit zwei Einflussvariablen und drei bzw. zwei Variablen dargestellt

Optimierungsproblem

$$(P_x) = \begin{cases} N_{\tilde{x}} \rightarrow \max \\ Q_{\tilde{x}} \geq q_{\min} \\ \tilde{x} \in \mathcal{A} \end{cases} \quad (3.17)$$

als *Root-Cause-Analysis-Problem (RCA)* bezeichnet.

Die Lösung  $\tilde{x}^*$  des Optimierungsproblem es ist folglich ein Sub-Prozess mit einem maximalen Support und gleichzeitig einer Qualität besser oder gleich dem Schwellwert  $q_{\min}$ . In der Regel wird  $q_{\min}$  gleich einem vorgegebenen Wert wie beispielsweise der langfristigen Prozessfähigkeit oder dem Prozesspotential  $C_{pk}$  gesetzt (vgl. Tabelle 2.1). Ist die Qualität einer vorliegenden Realisierung jedoch wesentlich schlechter als dieser (vorgegebene) Wert, so dass dieser (vorgegebene) Wert – wenn überhaupt – nur bei sehr geringem Support erreicht werden würde, kann der Schwellwert für die Optimierung wie folgt gesetzt werden:  $q_{\min} = Q_x$ .

Die Optimierungsaufgabe besteht nun darin, einen Sub-Prozess mit mindestens gleicher Qualität zu finden. Da der Trivialprozess (nämlich der Gesamtprozess ohne Einschränkungen) ausgeschlossen wurde, wird mit dieser Festlegung des Schwellwertes die Hoffnung verbunden, dass ein Sub-Prozess mit besserer Qualität und geringfügigen Einschränkungen gefunden wird. Diese Vorgehensweise kann dann mit einem zum Beispiel erhöhten Schwellwert wiederholt werden. Vor allem in Hinblick auf die Philosophie des Six-Sigma Prinzips eignet sich diese Vorgehensweise, um die Qualität eines Produktionsprozesses sukzessive zu verbessern.

### 3.2.2 RCA Algorithmus

Bei der Lösung des Optimierungsproblems 3.6, das ja auf der Basis konkreter Prozessdaten (der Stichprobe  $x$ ) formuliert ist, handelt es sich um ein kombinatorisches Optimierungsproblem. Dabei wurde bis jetzt stets von nominalen Einflussgrößen ausgegangen. Um mittels des hier vorgestellten Optimierungsproblems und des Lösungsalgorithmus (RCA) auch Einflüsse numerischer Natur mit einbeziehen zu können, ist es notwendig, die numerischen Einflussvariablen zu klassieren, das heißt, die Einteilung des Wertebereichs numerischer Variablen in Intervalle. Die Intervalle können dabei je nach Erfordernis der Anwendung gewählt werden. In verschiedenen praktischen Anwendungen hat sich eine Einteilung auf die 10%, 20%, 80% und 90% Quantile als geeignet erwiesen.

Eine Möglichkeit, das Optimierungsproblem (3.17) zu lösen, ist die Implementierung eines kombinatorischen Ansatzes, bei dem alle möglichen Lösungen im Sinne von (3.17) evaluiert werden. Dieser kombinatorische Lösungsansatz ist jedoch sehr aufwendig und

rechenintensiv, da die Menge der zu untersuchenden Kombinationen exponentiell mit der Anzahl der Einflussvariablen und ihren Werten wächst. Im speziellen Fall, dass alle Prozessparameter ( $\mathcal{S} = (S^1 \times \dots \times S^m)$ ) die gleiche Anzahl an Ausprägungen haben ( $|S^i| = c, i = 1, \dots, m$ ), müsste eine Menge von

$$\sum_{i=1}^m \binom{m}{i} (2^c - 2)^i \quad (3.18)$$

Sub-Prozessen evaluiert werden. Bereits bei 10 Einflussvariablen ( $m = 10$ ) mit jeweils 4 Werten ( $c = 4$ ) müsste über eine Menge von  $2.8 \cdot 10^{13}$  Sub-Prozessen optimiert werden. Dies würde selbst bei der Rechenkapazität aktueller Computer zu viel Zeit in Anspruch nehmen, so dass ein Einsatz in einer Produktivumgebung nur praktikabel wäre, wenn die Anzahl der Variablen und deren Ausprägungen klein ist.

Das Problem 3.6 fällt jedoch in die Klasse der ganzzahligen Optimierungsprobleme, wobei der zulässige Bereich eine diskrete, endliche Menge ist (vgl. Raum möglicher Sub-Prozesse  $\mathcal{A}$ ). Um nicht über alle möglichen Elemente des zulässigen Bereiches zu iterieren wird eine Methode, die häufig im Operations Research verwendet wird, eingesetzt: **Branch and Bound**.

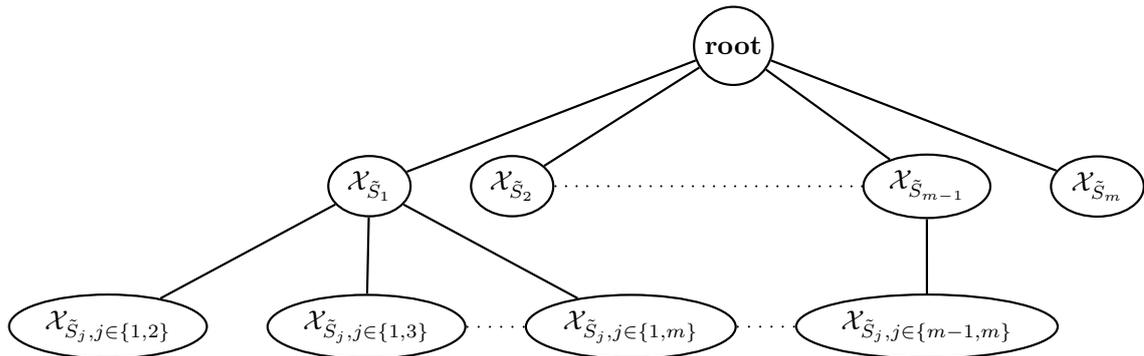


Abbildung 3.3: Baumstruktur für den RCA - Algorithmus für zwei Ebenen

Beim **Branch and Bound** wird der zulässige Bereich in mehrere Teilmengen aufgeteilt, was zu einer Aufspaltung des ursprünglichen Problems führt. Dieser so genannte *Branching*- Schritt wird rekursiv auf die so entstehenden Teilmengen ausgeführt, so dass eine Baum-Struktur (Entscheidungsbaum) entsteht. Im Falle des RCA-Optimierungsproblems entspricht eine so entstehende Teilmenge einem Sub-Prozess und somit stellt jeder Knoten dieses Baumes eine mögliche Lösung des Problems dar (vgl. Abbildung 3.3 und Beispiel 3.2).

### Beispiel 3.2

Sei ein Prozess mit zwei Einflussgrößen und drei bzw. zwei Ausprägungen gegeben (vgl.

Tabelle 3.1 und Beispiel 3.1). Dann würde der Knoten links der ersten Ebene in Abbildung 3.3 folgende Sub-Prozesse darstellen<sup>9</sup>:

$$\mathcal{X}_{\tilde{S}_1} = \{X_{\{a_1\}}, X_{\{a_2\}}, X_{\{a_3\}}, X_{\{a_1, a_2\}}, X_{\{a_1, a_3\}}, X_{\{a_2, a_3\}}\}$$

und für den unteren, linken Knoten gilt:

$$\begin{aligned} \mathcal{X}_{\tilde{S}_j, j \in \{1, 2\}} = & \{X_{\{a_1, b_1\}}, X_{\{a_2, b_1\}}, X_{\{a_3, b_1\}}, X_{\{a_1, a_2, b_1\}}, X_{\{a_1, a_3, b_1\}}, X_{\{a_2, a_3, b_1\}}\} \\ & \cup \{X_{\{a_1, b_2\}}, X_{\{a_2, b_2\}}, X_{\{a_3, b_2\}}, X_{\{a_1, a_2, b_2\}}, X_{\{a_1, a_3, b_2\}}, X_{\{a_2, a_3, b_2\}}\} \end{aligned}$$

Zum systematischen Traversieren eines (Mehrfach)-Baumes können prinzipiell zwei verschiedene Verfahren verwendet werden:

**Tiefensuche** Es wird stets das als letztes eingefügte, noch nicht bearbeitete Teilproblem gewählt (Last In, First Out). Somit arbeitet sich das Verfahren im Baum möglichst schnell in die Tiefe und man erhält relativ schnell eine Lösung. Jedoch kann man nichts über die Qualität der Lösung sagen (es kann einen Knoten geben, der zu einer besseren Lösung führt).

**Breitensuche** Das erste in den Baum eingefügte, noch nicht bearbeitete Teilproblem wird gewählt (First In, First Out). Das heißt, es werden die Knoten im Baum pro Ebene abgearbeitet, bevor tiefer gelegene Knoten (Lösungen) analysiert werden. Dadurch erhält man verhältnismäßig langsam eine zulässige Lösung, deren Qualität tendenziell gut ist.

Der eigentliche Schritt, um die Anzahl der zu analysierenden Knoten (Sub-Prozesse) zu reduzieren, ist das *Bounding*. Es hat die Aufgabe, die Zweige und Äste eines Baumes abzuschneiden, in denen keine Lösung mehr möglich ist und somit den Suchraum zu verkleinern. Im ungünstigsten Fall jedoch werden alle Kombinationen betrachtet. Wie jedoch später gezeigt wird, ist dies nur in ganz speziellen Fällen notwendig.

---

#### Algorithmus 1 Branch & Bound Tiefensuche

---

```

1: procedure TRAVERSE TREE(x)
2:    $\mathcal{A} = \text{GENERATE SUB PROCESSES}(x)$ 
3:   for all  $\tilde{x} \in \mathcal{A}$  do
4:     TraverseTree( $\tilde{x}$ )
5:   end for
6: end procedure

```

---

<sup>9</sup>Die einzelnen Sub-Prozesse  $x_*$  entstehen dabei durch das Fixieren der Einflussparameter  $A$  und  $B$  auf die im Subskript angegebenen Ausprägungen (Fixierungsmenge)

Um eine optimale Lösung für das *Root-Cause-Analysis* Problem zu finden wird eine Kombination von Breiten- und Tiefensuche genutzt, um den Mehrfachbaum zu traversieren (vergleiche Algorithmen 1 und 2)<sup>10</sup>. Dabei werden zwei **Branch and Bound-Prinzipien** verwendet um die Menge der möglichen Lösungen zu reduzieren. Das erste und allgemeinere Prinzip basiert auf den folgenden Überlegungen.

Der zu traversierende Mehrfachbaum (Abbildung 3.3) ist so aufgebaut, dass mit zunehmender Tiefe des Baumes, die Anzahl der Variablen und deren Werte, die fixiert werden, zunimmt. Mit anderen Worten: der Support der Sub-Prozesse nimmt ab (vgl. Bemerkung 3.1, 1. Branch and Bound Prinzip). Daher ist es nicht immer sinnvoll, Sub-Prozesse mit beliebiger Kombinationstiefe zu betrachten, sondern den Raum möglicher Sub-Prozesse wie folgt zu beschränken:

**Definition 3.7** (*l*-stufiger Sub-Prozess)

Sei  $\mathcal{A}_l = \{\mathbf{x}_{\tilde{S}_j, j \in J} : J \subseteq \{1, \dots, m\} \text{ und } |J| = l\}$ . Dann definiert  $\tilde{\mathbf{x}} \in \mathcal{A}_l$  einen *l*-stufigen Sub-Prozess von  $\mathbf{x}$ .

Mit dieser Notation lässt sich das RCA-Optimierungsproblem 3.6 wie folgt verallgemeinern:

**Definition 3.8** (*l*-stufiges RCA-Optimierungsproblem)

Sei  $l \leq m$ . Dann lässt sich das Optimierungsproblem 3.6 schreiben als

$$(P_x^l) = \begin{cases} N_{\tilde{\mathbf{x}}} \rightarrow \max \\ Q_{\tilde{\mathbf{x}}} \geq q_{\min} \\ \tilde{\mathbf{x}} \in \mathcal{A}_i, \forall i = 1, \dots, l \end{cases} \quad (3.19)$$

Mit  $l = m$  entspricht das obige Optimierungsproblem Definition 3.6. Mit Hilfe von (3.19) kann eine Lösung für das Optimierungsproblem auf einer (sinnvollen) Teilmenge der möglichen Sub-Prozesse gefunden werden.

**Bemerkung 3.1** (1. Branch & Bound Prinzip)

Sei  $\mathcal{X} = (X, \vec{Y}(\omega))$  ein Produktionsprozess wie in Definition 2.5. Seien  $S_1, S_2$  zwei Restriktionsmengen mit  $S_1 \subset S_2$  und  $\mathcal{X}_{S_1}, \mathcal{X}_{S_2}$  die zugehörigen Sub-Prozesse, dann gilt:

$$N_{\mathcal{X}_{S_1}} \leq N_{\mathcal{X}_{S_2}}$$

**Beweis:**

$$P(\vec{Y} \in S_1) \leq P(\vec{Y} \in S_2)$$

□

<sup>10</sup>Algorithmus 1 entspricht der Tiefen- und Algorithmus 2 der Breitensuche

Folglich kann bei einer Erweiterung der Restriktionsmenge der Support eines Sub-Prozesses zunehmen und durch Einschränkungen der Restriktionsmenge der Support abnehmen. Betrachtet man stattdessen Fixierungsmengen, so führt eine Erweiterung bezüglich einer bereits fixierten Variablen zu einer Zunahme des Supports. Fixiert man hingegen eine zusätzliche Variable, so hat dies im Allgemeinen eine Reduktion des Support des resultierenden Sub-Prozesses zur Folge. Aus dieser Eigenschaft folgt, dass der Support der Sub-Prozesse beim Absteigen des Baumes nicht größer werden kann (tendenziell abnimmt), da die Anzahl der Einflussvariablen, die eingeschränkt werden (Fixierungsmenge), zunimmt. Demzufolge, wenn ein Knoten einen niedrigeren Support aufweist als eine aktuelle Schranke, existiert kein Knoten in dem Ast unterhalb dieses Knotens, der einen höheren Support als diese Schranke aufweist.

---

**Algorithmus 2** RCA-Algorithmus für das 1. Branch and Bound Prinzip

---

```

1: procedure GENERATESUBPROCESSES( $\mathcal{X}$ )
2:    $\mathcal{A} = \emptyset$ 
3:   for all  $\tilde{x} \subseteq \mathcal{X}$  do
4:     if  $N_{\tilde{x}} > n_{max}$  and  $Q_{\tilde{x}} \geq q_{min}$  then
5:        $n_{max} = N_{\tilde{x}}$ 
6:     end if
7:     if  $N_{\tilde{x}} > n_{max}$  and  $Q_{\tilde{x}} < q_{min}$  then
8:        $\mathcal{A} = \{\mathcal{A} \cup \tilde{x}\}$ 
9:     end if
10:  end for
11:  return  $\mathcal{A}$ 
12: end procedure

```

---

Diese *bounding*-Eigenschaft reduziert die Anzahl der benötigten Schritte um eine optimale Lösung zu finden signifikant, da ein Großteil des Entscheidungsbaumes nicht traversiert werden muss und charakterisiert das **1. Branch and Bound Prinzip** zur Lösung des RCA-Optimierungsproblems. Ein weiterer Aspekt dieses Prinzips ist, wie eingangs erläutert, dass es in Praxisanwendungen oftmals unnötig ist, in allen Ebenen des Lösungsraumes (Entscheidungsbaumes) nach einem "optimalen" Sub-Prozess zu suchen. Denn wie später in der Evaluierung des RCA-Verfahrens gezeigt wird, ist der Support von Sub-Prozessen mit vielen (fixierten) Einflussvariablen in der Regel wesentlich kleiner als der Support für Prozesse mit einer geringeren Anzahl von (fixierten) Prozessparametern und demzufolge kann die Suche beim Erreichen einer bestimmten Ebene abgebrochen werden. Das heißt, dass in der Praxis für die Lösung des Problems 3.8  $l < m$  gewählt wird.

Eine algorithmische Umsetzung des **1. Branch and Bound Prinzips** wird in Algorithmus 2 wiedergegeben. Dabei beschreibt die Funktion GENERATESUBPROCESSES die Breitensuche des Branch and Bound Verfahrens: beim Traversieren des Mehrfachbaumes

(in Abbildung 3.3 von links nach rechts) werden nur solche Knoten bzw. Sub-Prozesse weiter betrachtet (in die Suchmenge  $\mathcal{A}$  aufgenommen), für die gilt, dass ihr Support höher oder gleich dem aktuellen, bis zu diesem Knoten gefundenen maximalen Support  $n_{max}$  ist. Die aktuelle Lösung ist der Knoten (Sub-Prozess)  $\tilde{x}$  für den gilt (Algorithmus 2, Zeile 3 und 4):

1.  $N_{\tilde{x}} > n_{max}$
2.  $Q_{\tilde{x}} \geq q_{min}$

Das **zweite Branch and Bound Prinzip** basiert auf der Betrachtung von disjunkten Mengen. Für die Klasse von konvexen Prozessindizes wie sie in Definition 2.6 dargestellt sind<sup>11</sup>, kann das zweite Branch and Bound Prinzip, analog zum stetigen Fall (Satz 2.3), aus dem folgenden Satz hergeleitet werden:

**Satz 3.1** (Konvexe Eigenschaft von  $Q_{w,\mathcal{X}}$ )

Seien zwei Sub-Prozesse  $\mathcal{X}_1 = (X, \vec{Y}_1)$ ,  $\mathcal{X}_2 = (X, \vec{Y}_2)$  eines Produktionsprozesses  $\mathcal{X} = (X, \vec{Y})$  gegeben (vgl. Definitionen 2.5 und 3.1). Des Weiteren gelte  $\vec{Y}_1 \in \mathcal{S}_1 \subset \mathcal{S}$  und  $\vec{Y}_2 \in \mathcal{S}_2 \subset \mathcal{S}$  mit  $\mathcal{S}_1 \cap \mathcal{S}_2 = \emptyset$ . Dann gilt für die Klasse von Prozesskennzahlen diskreter Einflussgrößen folgende Ungleichung:

$$\min_{\tilde{\mathcal{X}} \in \{\mathcal{X}_1, \mathcal{X}_2\}} Q_{w, \tilde{\mathcal{X}}} \leq Q_{w, \mathcal{X}_1 \cup \mathcal{X}_2} \leq \max_{\tilde{\mathcal{X}} \in \{\mathcal{X}_1, \mathcal{X}_2\}} Q_{w, \tilde{\mathcal{X}}} \quad (3.20)$$

**Beweis:** Mit  $p = \frac{P(\vec{Y} \in \mathcal{S}_1)}{P(\vec{Y} \in \mathcal{S}_1 \cup \mathcal{S}_2)}$  gilt folgende konvexe Eigenschaft:

$$\begin{aligned} Q_{w, \mathcal{X}_1 \cup \mathcal{X}_2} &= E \left( w(x) | \vec{Y}(\omega) \in \mathcal{S}_1 \cup \mathcal{S}_2 \right) \\ &= \frac{E \left( w(x) \mathbf{1}_{\{\vec{Y}(\omega) \in \mathcal{S}_1 \cup \mathcal{S}_2\}} \right)}{P(\vec{Y}(\omega) \in \mathcal{S}_1 \cup \mathcal{S}_2)} \\ &= \frac{E \left( w(x) \mathbf{1}_{\{\vec{Y}(\omega) \in \mathcal{S}_1\}} \right) + E \left( w(x) \mathbf{1}_{\{\vec{Y}(\omega) \in \mathcal{S}_2\}} \right)}{P(\vec{Y}(\omega) \in \mathcal{S}_1 \cup \mathcal{S}_2)} \\ &= p \frac{E \left( w(x) \mathbf{1}_{\{\vec{Y}(\omega) \in \mathcal{S}_1\}} \right)}{P(\vec{Y}(\omega) \in \mathcal{S}_1)} + (1-p) \frac{E \left( w(x) \mathbf{1}_{\{\vec{Y}(\omega) \in \mathcal{S}_2\}} \right)}{P(\vec{Y}(\omega) \in \mathcal{S}_2)} \end{aligned}$$

□

<sup>11</sup>Aufgrund der konvexen Eigenschaft dieser Indizes (vgl. Satz 2.3 und Satz 3.1), werden sie auch als konvexe QFIs bezeichnet

Daraus folgt, dass für die Vereinigung zweier disjunkter Restriktionsmengen der Wert eines konvexen QFIs das Maximum der Kennzahl aus diesen beiden Mengen nicht überschritten und das Minimum nicht unterschritten werden kann. Betrachtet man Fixierungsmengen, so gilt:

**Bemerkung 3.2**

Sei  $j \in \{1, \dots, m\}$  und  $\tilde{S}_j^1, \tilde{S}_j^2 \subsetneq S^j$  zwei disjunkte Fixierungsmengen. Dann gilt:

$$\min_{\tilde{x} \in \{\mathcal{X}_{\tilde{S}_j^1}, \mathcal{X}_{\tilde{S}_j^2}\}} Q_{w, \tilde{x}} \leq Q_{w, \mathcal{X}_{\tilde{S}_j^1} \cup \mathcal{X}_{\tilde{S}_j^2}} \leq \max_{\tilde{x} \in \{\mathcal{X}_{\tilde{S}_j^1}, \mathcal{X}_{\tilde{S}_j^2}\}} Q_{w, \tilde{x}} \quad (3.21)$$

**Beweis:** Die Fixierungsmengen  $\tilde{S}_j^1, \tilde{S}_j^2$  lassen sich wie folgt als Restriktionsmengen darstellen:

$$\begin{aligned} \mathcal{S}_1 &= S^1 \times \dots \times \tilde{S}_j^1 \times \dots \times S^m \\ \mathcal{S}_2 &= S^1 \times \dots \times \tilde{S}_j^2 \times \dots \times S^m \end{aligned}$$

Damit folgt Behauptung aus Satz 3.1. □

Diese Eigenschaft reduziert die Häufigkeit, dass ein konkreter, konvexer QFI wie beispielsweise der  $\hat{E}_{ci}$  zur Lösung des Optimierungsproblems berechnet werden muss. Denn beim Aufbau des Mehrfachbaumes zur Identifizierung einer optimalen Lösung des Problems 3.8 bzw. 3.6, setzten sich die einzelnen Knoten (Sub-Prozesse) oftmals aus der Vereinigung zweier (disjunkter) Fixierungsmengen zusammen (vgl. Beispiel auf Seite 71). In diesen Fällen kann der konvexe QFI anhand des Minimums und des Maximums abgeschätzt und somit entschieden werden, ob die Nebenbedingung von Problem 3.6 erfüllt ist oder nicht.

Konkret muss der Wert eines konvexen Prozessindizes nur für solche Sub-Prozesse berechnet werden für die gilt:

- Der Sub-Prozess lässt sich nicht aus der Vereinigung der Restriktionsmenge zweier bereits “bekanntere“ Sub-Prozesse darstellen
- Der Sub-Prozess lässt sich aus bereits bekannten Sub-Prozessen darstellen, es gilt jedoch:

$$\min_{\tilde{x} \in \{x_1, x_2\}} Q_{w, \tilde{x}} \leq q_{\min} \leq \max_{\tilde{x} \in \{x_1, x_2\}} Q_{w, \tilde{x}}$$

In den anderen Fällen erlaubt die Abschätzung (3.20) es, die “rechenintensive“ Berechnung des konvexen QFI zu vermeiden.

**Beispiel 3.3** (Abschätzung 2. Branch & Bound Prinzip)

Sei der Beispielprozess wie in Tabelle 3.1 gegeben und gelte  $Q_{w,x_{\{a_1\}}} = q_1$ ,  $Q_{w,x_{\{a_2\}}} = q_2$ . Dann lässt sich der QFI des Sub-Prozesses, der durch die Vereinigung von  $x_{\{a_1\}}$  und  $x_{\{a_2\}}$  beschrieben wird, mittels Gleichung (3.21) abschätzen. Konkret gilt dann:

$$Q_{w,x_{\{a_1,a_2\}}} = Q_{w,x_{\{a_1\} \cup \{a_2\}}} \in [\min\{q_1, q_2\}, \max\{q_1, q_2\}]$$

Bei der Ursachenanalyse für Prozessoptimierung ist man in der Praxis meist nicht nur an einer globalen Lösung für das Optimierungsproblem (3.17) interessiert, sondern vielmehr an den  $r$  besten Sub-Prozessen mit einem QFI höher als dem vorgegebenen Schwellwert  $q_{\min}$ . Ein Experte kann somit auf Basis seines Wissens aus dieser Menge einen oder mehrere Prozesse wählen und ihn (sie) als Referenz- bzw. Benchmark-Sub-Prozess(e) verwenden, um den (unbedingten) Produktionsprozess zu verbessern. Um eine solche  $r$ -elementige Menge von "optimalen" Sub-Prozessen zu finden, müssen zusätzlich die Äste des Entscheidungsbaumes traversiert werden, die bereits eine optimale Lösung hervorgebracht haben und abgeschnitten wurden, da sie keine bessere Lösung enthalten können.

Der Grund für das *Wiederaufnehmen* ist, dass ein optimaler Sub-Prozess  $x^*$  mit  $N_{x^*} > n_{\max}$  und  $Q_{x^*} \geq q_{\min}$  einen Knoten in dem Baum unterhalb des Knotens  $x^*$  haben kann, welcher einer zweitbesten (globalen) Lösung des RCA-Optimierungsproblems entsprechen könnte. Dementsprechend muss Zeile 4 in Algorithmus 2 erweitert werden, indem ein Ast mit einer optimalen Lösung nicht abgeschnitten wird, sondern der Menge  $\mathcal{A}$  der zu traversierenden Knoten hinzugefügt wird. Folglich ist der aktuelle, maximale Support nicht länger durch die aktuell beste Lösung definiert sondern durch die (aktuell)  $r$  beste Lösung. In diesem Kontext ist auch die Frage nach einer eindeutigen Lösung hinfällig, da stets mehrere Prozesse als Verbesserungspotential in Frage kommen.

Bei der Bestimmung der  $r$  besten Sub-Prozesse tritt jedoch das Problem von redundanten Lösungen auf. Seien  $x_{S_1}, x_{S_2}$  zwei Sub-Prozesse mit  $S_1 \subset S_2$  und  $N_{x_{S_1}} \equiv N_{x_{S_2}}$ , dann enthält  $x_{S_2}$  keine wesentliche neue Information gegenüber  $x_{S_1}$  und ist somit redundant (vgl. Beispiel 3.4). In solchen Fällen wird  $x_{S_2}$  nicht als Lösung angegeben.

**Beispiel 3.4** (Redundante Lösungen)

Für den Produktionsprozess aus Beispiel 3.1, Tabelle 3.1 seien folgende Restriktionsmengen  $S_1 = \{a_1, a_2\} \times \{b_1, b_2\}$  und  $S_2 = \{a_1, a_2\} \times \{b_1\}$  gegeben. Gilt  $N_{x_{S_1}} \equiv N_{x_{S_2}}$ , so ist  $S_2$  eine Spezialisierung von  $S_1$  und stellt keinen Mehrwert für die Verbesserung des Produktionsprozesses dar.

Wird der Mehrfachbaum wie eingangs beschrieben mittels Fixierungsmengen aufgebaut, werden sukzessive alle Teilmengen<sup>12</sup> der Wertemenge  $S^i$  der Variablen  $Y^i$ ,  $i = 1, \dots, m$

<sup>12</sup>Außer der leeren Menge und der Wertemenge  $S^i$  selber

$\tilde{S}_j^i \in \{\tilde{S}^i \subset S^i\}$	Darstellung als Binärzahl
$\{a_1\}$	001
$\{a_2\}$	010
$\{a_2, a_1\}$	011
$\{a_3\}$	100
$\{a_3, a_1\}$	101
$\{a_3, a_2\}$	110

Tabelle 3.4: Binarisierung der Variablen  $Y^i$  mit Wertebereich  $S^i = \{a_1, a_2, a_3\}$

durchlaufen. Die Umsetzung des ersten und zweiten Branch and Bound Prinzips lässt sich dabei effizient mittels einer Binarisierung der Einflussparameter realisieren. Dies bedeutet, dass jede Teilmenge  $\tilde{S}^i \subset S^i = \{\tilde{s}_1^i, \dots, \tilde{s}_{m_i}^i\}$  als  $m_i$ -stellige Binärzahl analog zu Tabelle 3.4 dargestellt wird.

Gilt zusätzlich die Voraussetzung  $|S^i| = m_i \leq 32$ , so lässt sich jede Teilmenge als eine Integerzahl mit 32 Bit implementieren und die Elemente von  $\{\tilde{S}^i \subsetneq S^i\}$  mittels binären Operationen effizient generieren, was zu einer Reduzierung der Laufzeit führt (vgl. Simulationsergebnisse, Seite 85).

In Algorithmus 3 (Seite 79) ist der Pseudocode für die binarisierte Implementierung dargestellt. Wie in Tabelle 3.4 exemplarisch dargestellt, baut der Algorithmus die Fixierungsmenge  $\{\tilde{S}^i \subset S^i\}$ ,  $i = 1, \dots, m$  sukzessive auf, indem zunächst einzelne Werte der betrachteten “nächsten“ Variablen (Algorithmus 3, Zeile 14 und 2) berechnet werden und dann Kombinationen dieser Werte. Mit Hilfe der Abschätzung (3.20) kann nun der Wert des  $\hat{E}_{ci}$  bestimmt werden (Algorithmus 3, Zeile 17 und 18). Ist eine Abschätzung nicht möglich, so muss der Wert für den konvexen Prozessindex berechnet werden (Algorithmus 3, Zeile 20). Für den Fall, dass ein nicht konvexer Prozessindex verwendet wird<sup>13</sup>, muss für jede Variablenwertkombination der QFI berechnet werden. Ein Vergleich der verschiedenen Optimierungstufen und Branch and Bound Prinzipien wird in der Konzeptevaluierung in Kapitel 3.3 vorgestellt.

<sup>13</sup>Außer dem  $E_{ci}$  sind alle hier vorgestellten Indizes nicht konvex

**Algorithmus 3** Optimiertes 1. und 2. Branch & Bound Prinzip

---

```

1: procedure GENERATESUBPROCESSES(x)
2:    $\mathcal{A} = \emptyset$ 
3:    $i =$  Index of next influence variable
4:   for  $j \in \{2^k\}, k = 1, \dots, m_i$  do
5:     Calculate  $N_{x_{\tilde{s}_j}}$  and  $Q_{x_{\tilde{s}_j}}$ 
6:     if  $N_{x_{\tilde{s}_j}} > n_{max}$  then
7:        $\mathcal{A} = \{\mathcal{A} \cup x_{\tilde{s}_j}\}$ 
8:       if  $Q_{x_{\tilde{s}_j}} > q_{min}$  then
9:          $n_{max} = N_{x_{\tilde{s}_j}}$ 
10:      end if
11:    end if
12:    for  $l = 1, \dots, j - 1$  do
13:      if  $N_{x_{\tilde{s}_j \cup \tilde{s}_l}} \leq n_{max}$  then
14:        Continue
15:      end if
16:      if  $q_{min} \notin [\min_{\tilde{x} \in \{x_{\tilde{s}_j}, x_{\tilde{s}_l}\}} Q_{\tilde{x}}, \max_{\tilde{x} \in \{x_{\tilde{s}_j}, x_{\tilde{s}_l}\}} Q_{\tilde{x}}]$  then
17:        Estimate  $Q_{x_{\tilde{s}_j \cup \tilde{s}_l}}$ 
18:      else
19:        Calculate  $Q_{x_{\tilde{s}_j \cup \tilde{s}_l}}$ 
20:      end if
21:       $\mathcal{A} = \{\mathcal{A} \cup x_{\tilde{s}_j \cup \tilde{s}_l}\}$ 
22:      if  $Q_{x_{\tilde{s}_j \cup \tilde{s}_l}} \geq q_{min}$  then
23:         $n_{max} = N_{x_{\tilde{s}_j}}$ 
24:      end if
25:    end for
26:  end for
27:  return  $\mathcal{A}$ 
28: end procedure

```

---

### 3.2.3 Ergebnisverifikation

Eine Möglichkeit die Ergebnisse des RCA- Algorithmus statistisch zu verifizieren ist, den zu analysierenden Produktionsprozess (zufällig) in einen Test- und in einen Trainingsprozess aufzuteilen. Sei  $\mathbf{x} = (x_i, \vec{y}_i), i = 1, \dots, n$  eine konkrete Realisierung eines Prozesses wie nach Definition 2.5. Des Weiteren sei  $\mathbf{x}_1 = (x_i^1, \vec{y}_i^1), i = 1, \dots, n_1$  der Trainings- und  $\mathbf{x}_2 = (x_i^2, \vec{y}_i^2), i = 1, \dots, n_2$  der Testprozess mit  $n = n_1 + n_2$ . Eine Verifizierung der Ergebnisse des RCA-Algorithmus erfolgt, indem analysiert wird, ob sich die Ergebnisse aus der Trainingsmenge auf die Testmenge übertragen lassen. Konkret werden folgende zwei Schritte durchgeführt:

1. Für den Trainingsprozess  $\mathbf{x}_1$  werden mittels des RCA-Algorithmus die Menge der  $r$  besten Sub-Prozesse  $\mathcal{A}^1$  bestimmt. Da in realen Anwendungen die betrachteten Stichproben, mittels deren Hilfe Handlungsempfehlungen abgeleitet werden, verhältnismäßig klein sind, kann eine Lösung des Optimierungsproblems 3.6 nur in einem stochastischen Zusammenhang definiert werden. Aus diesem Grund wird für die  $r$  besten Prozesse in  $\mathcal{A}^1$  die linke Grenze des Konfidenzintervalls  $C_l^Q(\cdot)$  bestimmt um somit eine verlässlichere Aussage über den wahren Wert des gefundenen Qualitätsniveaus treffen zu können.

Für den speziellen Fall, dass die Messergebnisse eines Produktionsprozesses einer Normalverteilung folgen, können Konfidenzintervalle für QFIs analytisch – im Falle des  $C_p$  – oder numerisch bzw. approximativ bestimmt werden (vergleiche Abschnitt 2.3.1). Folgen die Daten einem anderen, bekannten Verteilungstyp, so müssen die Daten mittels geeigneter Transformationen (z.B. Box-Cox-Transformation) in normalverteilte Daten transformiert werden, um die vorgestellten Verfahren zur Konstruktion eines Konfidenzintervalls verwenden zu können. Ist es nicht möglich eine Annahme über die zugrunde liegende Verteilung des Produktionsprozesse zu treffen, können Konfidenzintervalle mittels der in Kapitel 2.3.2 eingeführten Bootstrap-Methoden konstruiert werden.

2. Analog zum vorherigen Schritt werden nun für die Testmenge  $\mathbf{x}_2$  zunächst die  $r$  besten Sub-Prozesse  $\mathcal{A}^2$  bestimmt. Dabei werden jedoch keine Konfidenzintervalle für die  $r$  besten Sub-Prozesse in  $\mathcal{A}^2$  bestimmt. Anschließend werden diejenigen Sub-Prozesse ausgegeben, für die gilt, dass sie zum einen die gleiche Restriktionsmenge (Fixierungsmenge) besitzen und zum anderen ein mindestens so hohes Qualitätsniveau haben wie der entsprechende Prozess der Trainingsmenge. Mit anderen Worten ist die Menge optimaler Sub-Prozesse gegeben durch

$$\mathcal{A}^* = \left\{ \mathbf{x}_{Test} \in \mathcal{A}^2 : \exists \mathbf{x}_{Train} \in \mathcal{A}^1 \ni \vec{y}_{Train} = \vec{y}_{Test} \text{ und } Q_{\mathbf{x}_{Test}} > C_l^Q(\mathbf{x}_{Train}) \right\}$$

Gilt für  $\mathcal{A}^* = \emptyset$ , so kann davon ausgegangen werden, dass mittels des RCA-Ver-

fahrens kein Optimierungspotential für den Produktionsprozess  $x$  gefunden werden kann.

Im Folgenden wird eine Konzeptevaluierung der in den vorherigen Abschnitten vorgestellten Methoden und Verfahren anhand von Daten aus einer Gießerei der Daimler AG vorgestellt. Da die in dieser Studie vorhandenen Daten keiner bekannten Verteilung folgen, wird der Fokus auf die Verwendung nicht-parametrischen Methoden gelegt.

### 3.3 Konzeptevaluierung

Im Rahmen eines Projektes mit der Daimler AG wurden in einer betriebszugehörigen Gießerei Daten für einen SPC gesteuerten Prozess erhoben. Dabei wurden zunächst verschiedene Messmerkmale für den Prozess erfasst, indem Stichproben der Produktion auf einem Prüfstand vermessen wurden. Diese gemessenen, numerischen Werte beschreiben dabei zum größten Teil geometrische Eigenschaften wie beispielsweise die Position einer Bohrung, die Oberflächenbeschaffenheit oder Dralleigenschaften des zu analysierenden Werkstückes. Parallel zu der Messung auf dem Prüfstand wurden ebenfalls verschiedene, charakteristische Prozessmerkmale aufgenommen, die im Laufe der Produktion nicht konstant gehalten werden konnten. Die erhobenen Prozessmerkmale (Produktionsparameter) sind im Allgemeinen nicht numerisch sondern bis auf wenige Ausnahmen nominal.

Den meisten stichprobenartig erfassten Prozessen war gemein, dass die Qualität des Prozesses unter dem vorgegebenen Mindestqualitätsniveau lag. Das Ziel des Projektes war nun, die Prozessfähigkeit wieder herzustellen, ohne dabei kostspielige Versuche durchführen zu müssen. Für den Fall, dass Prozesse doch die Qualitätsvorgaben erfüllt haben, sollte mögliches Optimierungspotential identifiziert werden.

Eine erste Idee, das Problem zu lösen war eine Verallgemeinerung der Varianzanalyse für den Fall nicht-parametrischer Daten. Wie jedoch in dieser Arbeit gezeigt wurde, ist es nicht möglich, die bisher bekannten Verfahren auf das hier vorliegende Problem anzuwenden. Stattdessen wurde ein neues Verfahren – Root-Cause-Analysis – entwickelt, mit dessen Hilfe eine SPC-gestützte Prozessoptimierung möglich ist.

Die Tatsache, dass in der Literatur noch keine Verfahren zur SPC gestützten Prozessoptimierung bekannt sind hat den Nachteil, dass die RCA-Methode derzeit nicht theoretisch evaluiert werden kann. Um dennoch die Richtigkeit des Verfahrens zu überprüfen wurden die gefundenen Ergebnisse den Verantwortlichen des Qualitätsmanagements zur Umsetzung übergeben. Dazu wurde ein aktuelles Problem der Produktion im Zu-

ge eines Projektes analysiert. Ausgangspunkt des Projektes war, dass Kurbelgehäuse von Sechszylinder-Motoren aus der Gießerei nach der mechanischen Bearbeitung in der Lecktestanlage, Leckraten oberhalb der durch die Konstruktion vorgegebenen Spezifikationsgrenzen aufzeigten. Das Ziel des Projektes war es somit, mittels verschiedener Optimierungsversuche die Ursachen-Wirkungs-Beziehungen<sup>3</sup> zwischen den Gießprozessparametern und den Leckraten zu identifizieren. Mit dem so gewonnenen Wissen über die Wirkungsbeziehungen, sollte die geforderte Prozessfähigkeit wieder hergestellt werden.

Mittels des hier vorgestellten RCA-Verfahrens auf Basis des Fähigkeitsindizes  $E_{ci}$ , wurden zunächst 20 optimale Sub-Prozesse berechnet, deren Ausschussraten dem geforderten Niveau entsprechen. Anschließend wurde vom Experten der Gießerei der Sub-Prozess (Benchmark) ausgewählt, der unter verschiedenen Gesichtspunkten am geeignetsten erschien. Bei der Auswahl des richtigen Sub-Prozesses mussten dabei sowohl wirtschaftliche als auch technische Aspekte berücksichtigt werden, so dass weder der Prozess mit dem höchsten Support noch der mit dem höchsten Qualitätsmesswert ausgewählt wurde. Basierend auf den Einflussparametern des Benchmark-Prozesses wurde ein Bestätigungsexperiment durchgeführt. Dazu wurden die Prozessparameter an die Prozessgrößen des optimalen Prozesses angepasst und anschließend 100 Kurbelgehäuse gegossen. Bei der abschließenden Prüfung stellte sich heraus, dass keines der Gehäusen aus dem Bestätigungsexperiment undicht war. Durch eine Übertragung auf den gesamten Gießprozess konnte die Ausschussrate langfristig um 75% reduziert werden.

Da lediglich eine empirische Evaluation des RCA-Verfahrens möglich ist, wird in der hier beschriebenen Konzeptevaluierung keine Verifikation der RCA-Methode mit anderen Modellen vorgestellt, sondern vielmehr werden zunächst die Ergebnisse der verschiedenen Branch and Bound-Prinzipien und der Optimierungsmöglichkeiten für den RCA-Algorithmus beschrieben. Anschließend wird eine Fertigungsprozessoptimierung exemplarisch präsentiert.

### 3.3.1 Simulationsergebnisse

Für die Evaluierung des RCA-Konzeptes wurden drei verschiedene Gruppen von Fertigungsprozessen betrachtet. Der Stichprobenumfang der hier betrachteten Prozesse reichte dabei von 500 bis 1000 Messergebnissen und bis zu neun verschiedenen Prozessparametern (Einflussgrößen). Die erste zu analysierende Gruppe besteht aus 12 verschiedenen Messgrößen mit jeweils vier verschiedenen Einflussgrößen. Die zweite Gruppe setzt sich aus 20 unterschiedlichen Messgrößen zusammen, wobei jede von bis zu 14 Prozessparametern beschrieben wird mit bis zu 7 verschiedenen Ausprägungen. Zusätzlich wurde eine weitere Messgröße untersucht, welche die Ergebnisse einer Zylinderdrallmessung bein-

Gruppe	Messgrößen	Einflussgrößen ( $m$ )	Ausprägungen
1	12	4	max. 9
2	20	14	max. 7
3	1	76	5

Tabelle 3.5: Mengengerüst analysierter Prozesse

Gruppe	Cpi [s]	$E_{ci}$ [s]	Kombinatorisch [s]
1	19.3	6.7	28.0
2	14.13	5.1	1035.7
3		926	-

Tabelle 3.6: Benötigte Rechenzeit für nicht-parametrische Indizes

haltet. In diesem dritten Datensatz sind die Prozessparameter nicht wie in den anderen beiden Fällen nominaler Natur, sondern bestehen aus 76 numerischen Größen. Um auch diesen Prozess analysieren zu können, wurden die Daten klassiert, in dem die stetigen Prozessparameter auf die 10%, 20%, 80% und 90% Quantile abgebildet wurden.

Für jeden der verwendeten 32 Datensätze der 1. und 2. Gruppe (vergleiche Tabelle 3.5)<sup>14</sup> wurde mittels des RCA-Algorithmus eine Liste der besten 20 Sub-Prozesse nach dem in (3.19) definierten Optimierungsproblem erstellt. Dabei wurde  $q_{min}$  gleich dem unbedingten QFI gesetzt ( $q_{min} = Q_x$ ) und  $l = 4$  gewählt, da eine Wahl von  $l > 4$  keine weiteren (besseren) Sub-Prozesse geliefert hat. In den Fällen, in denen der unbedingte  $E_{ci}$  einen Wert größer als 0.98 hatte<sup>15</sup>, wurden die Grenzen  $U$  und  $L$  nach (2.21) angepasst, wobei  $t = 2$  gewählt wurde, um somit *weit genug in die Stichprobe* zu gelangen.

Um die Effizienz des ersten Branch and Bound Prinzips zu zeigen, wurden in einer ersten Simulation die Lösungen des Optimierungsproblems kombinatorisch berechnet und in einem zweiten Schritt mit Hilfe des 1. Branch and Bound Prinzips verglichen. Dazu wurde die jeweils durchschnittlich benötigte Rechenzeit für die nicht-parametrischen Indizes  $C'_p(u, v, 0.00135)$ <sup>16</sup>,  $u, v \in \{0, 1\}$  und  $E_{ci}$  ermittelt. Wie in Abbildung 3.4 zu sehen ist, steigt die Rechenzeit mit der Anzahl der Einflussvariablen. Diese Tatsache erklärt den Sprung in Abbildung 3.4 für die Rechenzeit des kombinatorischen Ansatzes – die ersten 12 Datensätze entsprechen der ersten Gruppe analysierter Daten mit maximal vier Einflussgrößen.

<sup>14</sup>Der zusätzliche Datensatz mit stetigen Einflussgrößen (3. Gruppe) wird am Ende dieses Abschnittes gesondert betrachtet

<sup>15</sup> $U = USL$  und  $L = LSL$

<sup>16</sup>Abgekürzt mit Cpi

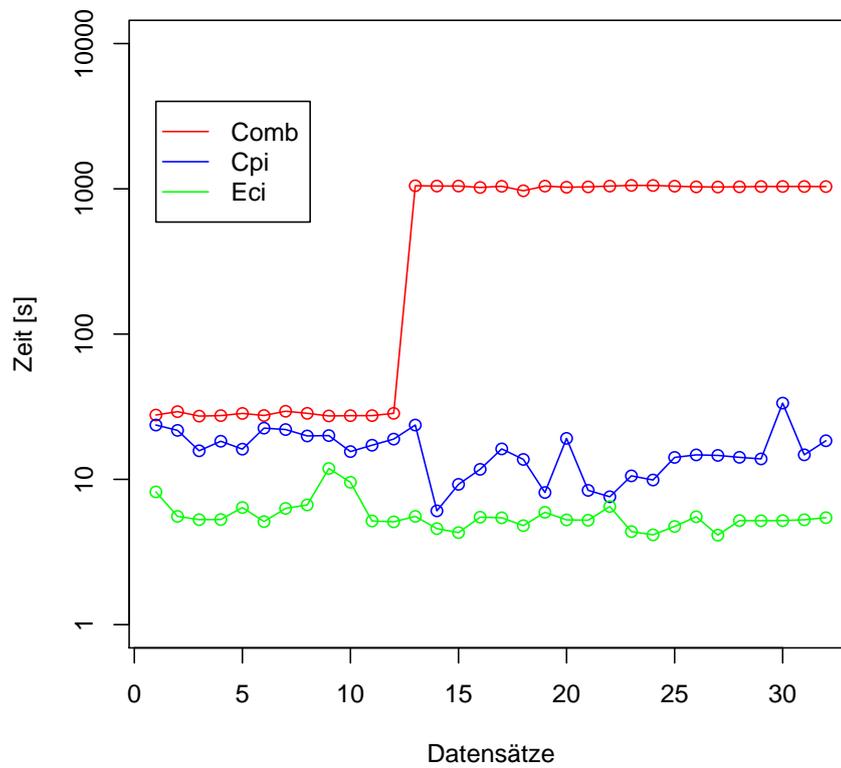


Abbildung 3.4: Vergleich: kombinatorisch vs. 1. Branch and Bound Prinzip (logarithmisch skaliert)

Im Vergleich dazu bleibt die benötigte Rechenzeit unter Verwendung von Algorithmus 2 – Seite 74, 1. Branch and Bound Prinzip – für alle analysierten Datensätze auf dem gleichen Niveau und ist im Durchschnitt vierzig Mal schneller als der kombinatorische Ansatz. Unter Verwendung des kombinatorischen Ansatzes dauert es für die Berechnung des Datensatzes mit wenigen Einflussgrößen (1. Gruppe) lediglich 28 Sekunden, hingegen bei der 2. Gruppe von Prozessdaten rund 17 Minuten. Dahingegen benötigt man mit Hilfe des ersten Branch and Bound Prinzips durchschnittlich 19 Sekunden für die 1. Gruppe und 14 Sekunden für die zweite Gruppe. Anhand dieser Ergebnisse ist deutlich zu erkennen, dass der RCA-Algorithmus vor allem bei der Analyse von Datensätzen mit vielen Einflussgrößen effizient ist (siehe Tabelle 3.6).

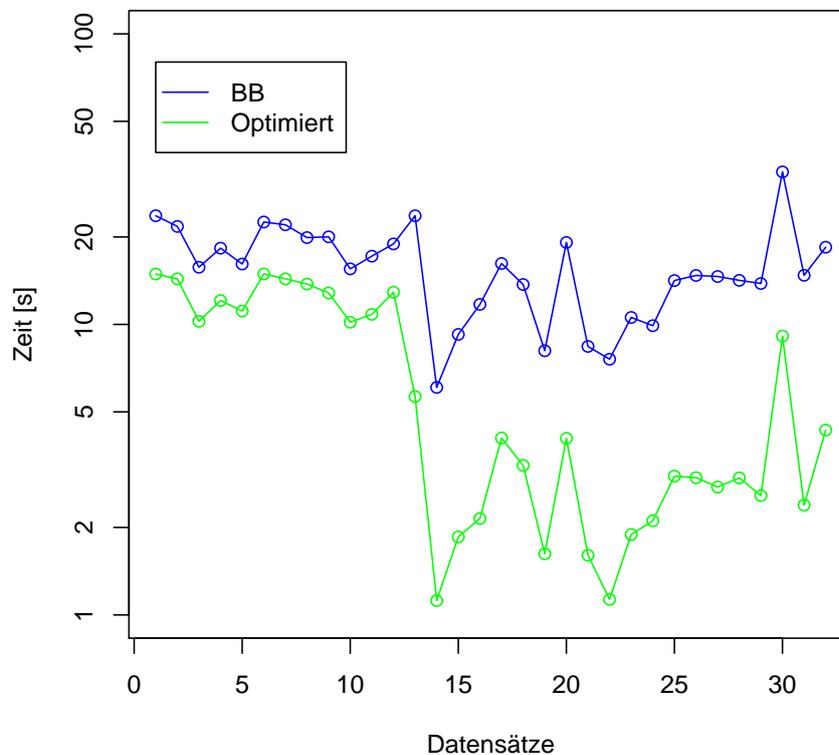


Abbildung 3.5: Vergleich: Binarisierung 1. Branch and Bound Prinzip

Eine zusätzliche Verringerung der Rechenzeit für das erste Branch and Bound Prinzip konnte mit Hilfe der Binarisierung (vgl. Kapitel 3.2.2) erreicht werden. Durch geschickte Anordnung der Werte der Variablen und Abbildung auf binäre Vektoren konnte die

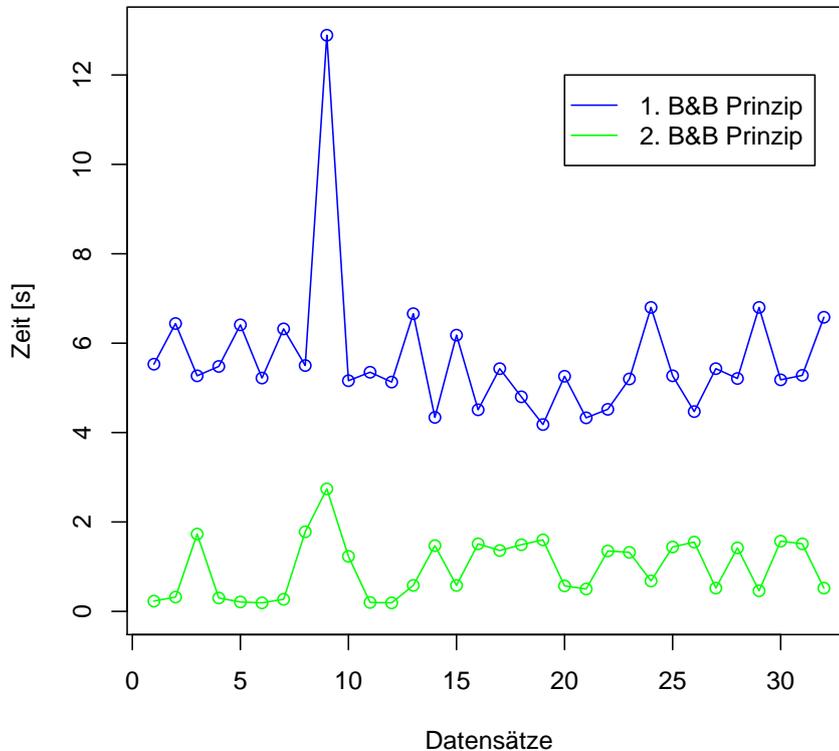


Abbildung 3.6: Vergleich: 1. Branch and Bound Prinzip vs. 2. Branch and Bound Prinzip

benötigte Rechenzeit für die 1. und 2. Gruppe von Daten von ursprünglich 16.6 Sekunden auf 6.6 Sekunden reduziert werden (siehe Abbildung 3.5)

Die Anwendung des zweiten Branch and Bound Prinzips (Satz 3.1) brachte, wie in Abschnitt 3.2.2 erläutert, eine weitere Reduktion der Rechenzeit für den konvexen Fähigkeitsindex  $E_{ci}$ . Wie in Abbildung 3.6 zu sehen ist, wurde die benötigte Zeit für die Ermittlung der 32 *optimalen* Lösungsmengen (jeweils bestehend aus den 20 besten Sub-Prozessen) deutlich gesenkt – im Schnitt um den Faktor 5 von 5.71 Sekunden (1. Branch and Bound Prinzip) auf 0.98 Sekunden. Mit anderen Worten: die benötigte Zeit zur Identifizierung der 20 optimalen Sub-Prozesse über alle, zur Simulation benutzten Datensätze, konnte um 80% verringert werden. Diese Verringerung der Rechenzeit ist zum einen bedingt durch die effiziente, binarisierte Implementierung und zum anderen durch die verminderte Häufigkeit, für die der  $E_{ci}$  konkret berechnet werden musste. Im Durchschnitt konnten 25% der Berechnungen des  $E_{ci}$  durch die Abschätzung (3.20) vermieden werden.

Die Analyse des Datensatzes zur Zylinderdrallmessung (3. Gruppe) konnte für den kombinatorischen Ansatz nicht durchgeführt werden. Denn die benötigte Rechenzeit wäre zu groß, da in diesem Fall rund  $4.9 \cdot 10^{10}$  Kombinationen evaluiert werden müssten. Selbst durch die Verwendung des zweiten Branch and Bound Prinzips wurden die optimalen Sub-Prozesse erst nach 926 Sekunden identifiziert.

### 3.3.2 Beispiel: Prozessoptimierung

Im Folgenden soll anhand eines konkreten Beispiels die Vorgehensweise bei der SPC basierten Prozessoptimierung mit Hilfe des RCA-Algorithmus veranschaulicht werden. Dazu wurde ein Datensatz aus den im vorigen Kapitel beschriebenen Produktionsprozessen ausgewählt. Die geometrische Eigenschaft, die in diesem Falle als Qualitätsmerkmal auf dem Prüfstand erhoben wurde, ist die Position einer Bohrung auf dem Werkstück. Zur Verdeutlichung sind relevante Prozessmerkmale in Tabelle 3.7 dargestellt.

USL	0.15
LSL	0.0
T	0.075
$E_{ci}$	0.49
$C'_{pk}$	0.82
Einflussvariablen	BAZ, Spindel, Center
Stichprobenumfang	968

Tabelle 3.7: Prozessmerkmale

Da der Verteilungstyp des Prozesses nicht eindeutig bestimmbar ist, wurden zur Messung der Qualität nicht-parametrische Indizes verwendet. Im Speziellen wurde eine Analyse mittels des konvexen  $E_{ci}$  und des  $C'_{pk}$  durchgeführt. Letzterer erhielt den Vorzug gegenüber den anderen nicht-parametrischen Indizes, da der Prozess keine asymmetrischen Grenzen aufweist und die Prozessfähigkeit aussagekräftiger ist als das Prozesspotential. Wie aus Tabelle 3.7 ersichtlich ist, ist die Qualität des Prozesses für beide Indizes unter der geforderten Mindestqualität (vergleiche Schwellwerte in Tabelle 2.1). Das Ziel der Prozessoptimierung ist somit, Sub-Prozesse zu identifizieren, mit deren Hilfe der Prozess so eingestellt werden kann, dass sich die Qualität nachhaltig verbessert.

Wie im vorherigen Kapitel erläutert besteht die RCA-Analyse im Wesentlichen aus zwei Schritten, die mit den jeweiligen Ergebnissen kurz dargestellt werden.

Im ersten Schritt wird der vorliegende Prozess in zwei gleich mächtige, disjunkte Men-

gen (Trainings- und Testmenge) aufgeteilt. Für die beiden Indizes  $C'_{pk}$  und  $E_{ci}$  wird im ersten Schritt auf der Trainingsmenge eine Liste der 20 besten Sub-Prozesse ermittelt. Anschließend wird für jeden gefundenen besten Prozess ein Konfidenzintervall bestimmt und die linke Grenze  $C_l^Q$  des entsprechenden Konfidenzintervalls als Maß für die Qualität der "Trainingsprozesse" verwendet. Aufgrund der Studie in Kapitel 2.3.3 wird zur Konstruktion eines Konfidenzintervalls für den  $C_{pk}$  die *BCPB-Methode* und für den  $E_{ci}$  die *SB-Methode* verwendet. Eine detaillierte Auflistung der gefundenen optimalen Sub-Prozesse ist im Anhang in den Tabellen A.7 und A.8 zu finden.

Die im vorherigen Schritt gefundenen Zusammenhänge sollen nun mittels der Testmenge evaluiert werden. Dazu werden wiederum die 20 besten Sub-Prozesse auf der Testmenge identifiziert und wie folgt verglichen:

1. Zunächst werden diejenigen Prozesse der beiden Mengen identifiziert, welche die gleichen Prozessparameter aufweisen. Dabei reduziert sich im vorliegenden Fall die Menge der noch relevanten Sub-Prozesse mit gleichen Einflussgrößen auf 3 für den  $C_{pk}$  bzw. 9 für den  $E_{ci}$  (vgl. Tabellen A.6 und A.5).
2. Anschließend werden diejenigen Sub-Prozesse auf der verbleibenden Schnittmenge identifiziert, für die gilt:  $Q_{Test} > C_l^Q$ . Damit bleiben 2 bzw. 9 (für  $E_{ci}$ ) optimale Sub-Prozesse übrig.

Index	Sub-Prozess	$N_{Test}$	$Q_{Test}$	$N_{Train}$	$C_l^Q$
$C_{pk}$	BAZ in (1,4) and Spindel in (1)	127	1.10	118	1.01
$C_{pk}$	BAZ in (2,3) and Spindel in (2)	124	1.03	127	1.02
$E_{ci}$	BAZ in (1,2)	244	0.81	210	0.81
$E_{ci}$	BAZ in (2,3) and Spindel in (2)	124	0.85	127	0.72
$E_{ci}$	BAZ in (2,4)	225	0.88	236	0.81
$E_{ci}$	BAZ not in (1)	356	0.81	377	0.77
$E_{ci}$	BAZ not in (1) and Center in (rechts)	190	0.85	183	0.75
$E_{ci}$	BAZ not in (1) and Spindel in (2)	178	0.85	194	0.78
$E_{ci}$	BAZ not in (3)	353	0.81	343	0.79
$E_{ci}$	BAZ not in (3) and Center in (rechts)	174	0.87	174	0.8
$E_{ci}$	Center in (rechts)	243	0.83	244	0.77

Tabelle 3.8: Ergebnisliste Prozessoptimierung

Überträgt man nun die gefundenen Zusammenhänge auf den Originalprozess (unrestringierter bzw. unbedingter Prozess), so stellt man fest, dass sich auch dort der jeweilige QFI verbessert (Tabelle 3.9). Wählt man einen optimalen Sub-Prozess aus Tabelle 3.8,

und beschränkt den Originalprozess beispielsweise auf die Menge  $BAZ$  in (1,2) (Abbildung 3.7) so verbessert sich die durch den  $E_{ci}$  gemessene Qualität des Prozesses von ursprünglich 0.49 um 69 Prozent auf 0.83. Eine grafische Übersicht der gefundenen optimalen Sub-Prozesse (blauer Graph) im Vergleich zum Originalprozess (roter Graph) ist in Abbildung 3.8 bis Abbildung 3.10 gegeben.

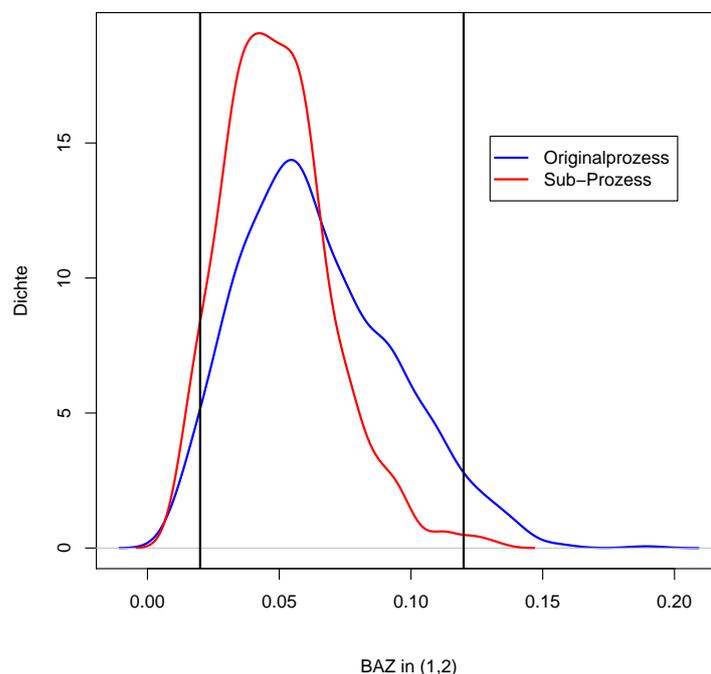


Abbildung 3.7: Dichteplot optimaler Sub-Prozess vs. Originalprozess für  $E_{ci}$  für die Fixierungsmenge  $BAZ$  in (1,2)

Bei der Betrachtung der Grafiken fällt zunächst auf, dass die Sub-Prozesse, deren Qualität mittels des  $C_{pk}$ 's gemessen wurden, nur wenig von dem Originalprozess abweichen (vgl. Abbildung 3.8). Jedoch werden durch die Restriktion auf den Sub-Prozess diejenigen Messergebnisse "ausgelassen", die jenseits der rechten Toleranzgrenze liegen, und somit eine negative (verringemde) Wirkung auf das Qualitätsmaß haben. Ein ähnliches Verhalten ist für die  $E_{ci}$ -basierten, optimalen Sub-Prozesse zu beobachten (Abbildungen 3.7, 3.9 und 3.10). Auch hier werden diejenigen Messpunkte "eliminiert", die rechts von  $U$  liegen.

Index	Sub-Prozess x	$N_x$	$Q_x$	Verbesserung
$C_{pk}$	BAZ in (1,4) and Spindel in (1)	245	1.12	37.07%
$C_{pk}$	BAZ in (2,3) and Spindel in (2)	251	1.10	34.15%
$E_{ci}$	BAZ in (2,3) and Spindel in (2)	251	0.84	71.56%
$E_{ci}$	BAZ in (2,4)	461	0.86	76.19%
$E_{ci}$	BAZ not in (1)	733	0.81	65.38%
$E_{ci}$	BAZ not in (1) and Center in (rechts)	373	0.83	69.06%
$E_{ci}$	BAZ not in (1) and Spindel in (2)	372	0.85	73.91%
$E_{ci}$	BAZ not in (3)	696	0.83	68.60%
$E_{ci}$	BAZ not in (3) and Center in (rechts)	348	0.85	74.17%
$E_{ci}$	Center in (rechts)	487	0.82	67.20%
$E_{ci}$	BAZ in (1,2)	454	0.83	69.02%

Tabelle 3.9: Übertragung der gefundenen, optimalen Zusammenhänge auf den Originalprozess (unbedingte QFIs:  $E_{ci} = 0.49$ ,  $C_{pk} = 0.82$ , vgl. Tabelle 3.7)

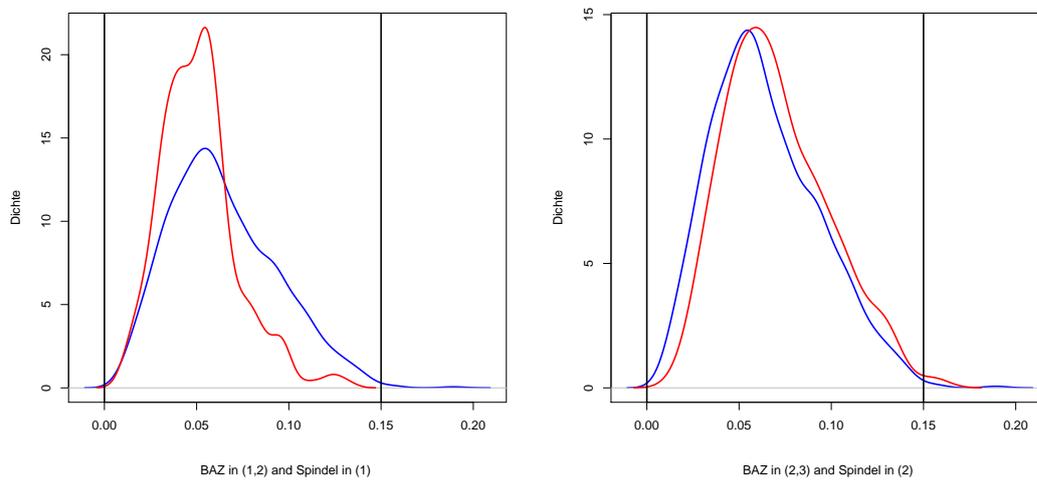


Abbildung 3.8: Dichte optimaler Sub-Prozesse (rot) vs. Originalprozess (blau) für  $C_{pk}$

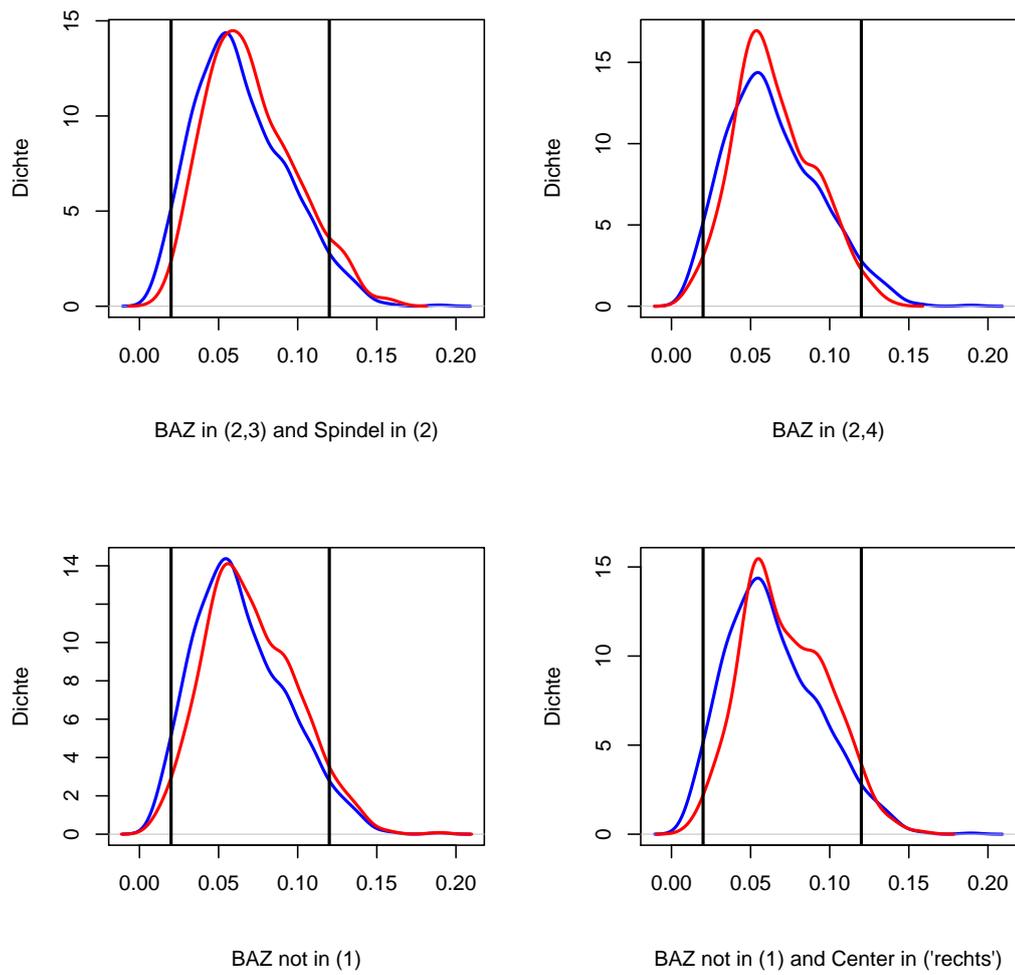


Abbildung 3.9: Dichte optimaler Sub-Prozesse (rot) vs. Originalprozess (blau) für  $E_{ci}$

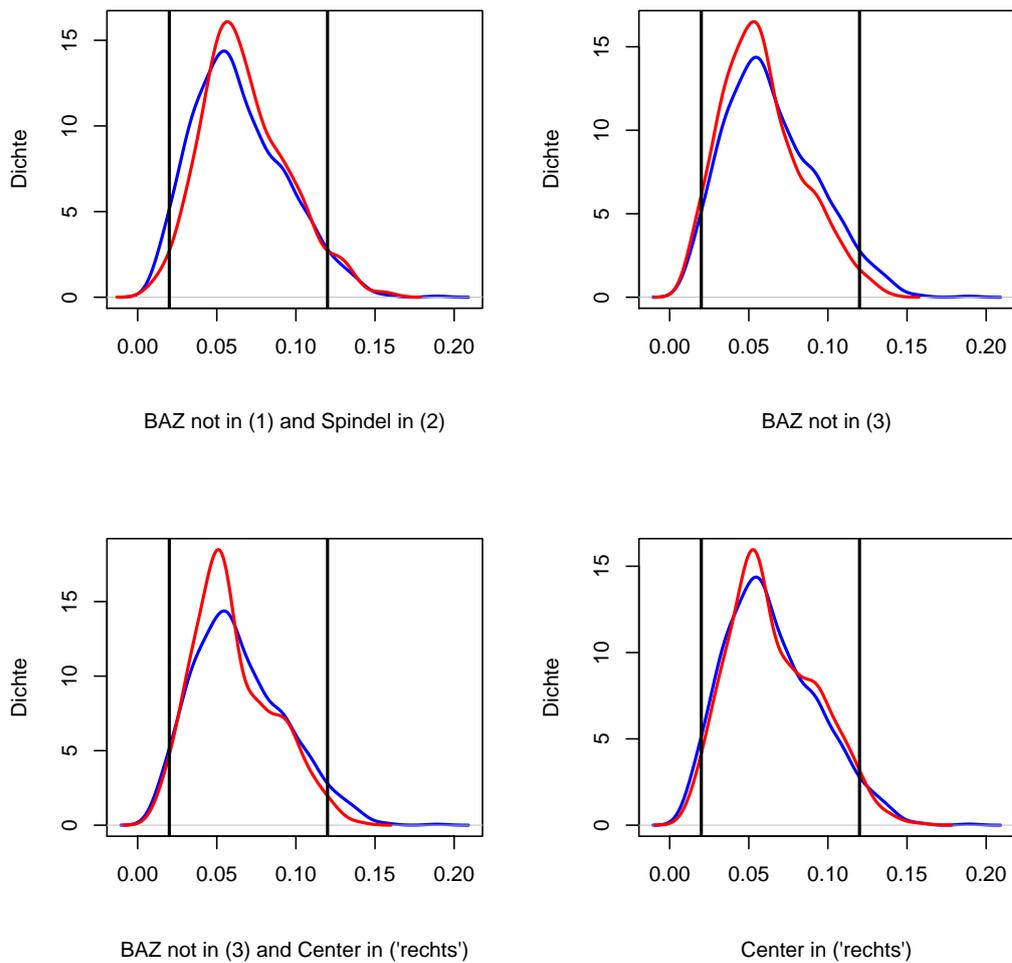


Abbildung 3.10: Dichte optimaler Sub-Prozesse (rot) vs. Originalprozess (blau) für  $E_{ci}$

## 3.4 Zusammenfassung und Bewertung

Qualitätsfähigkeitsindizes wie in Kapitel 2.1 eingeführt, werden in der industriellen Praxis des Qualitätsmanagements eingesetzt, um die (produzierte) Qualität von Gütern mit geforderten Qualitätsstandards zu vergleichen. Die Abbildung der Qualität auf eine einheitslose Größe erlaubt es jedoch nicht, die Ursachen-Wirkungs-Beziehungen zwischen qualitätsbeschreibenden (Einflussvariablen) und qualitätsmessenden Produktionsgrößen (Zielvariablen) ausreichend zu erklären. Denn mittels QFI ist es nicht möglich, diejenigen Produktionsparameter zu bestimmen, welche sich signifikant auf die Qualität der gefertigten Produkte auswirken. Wie eingangs dieses Kapitels beschrieben, scheitern klassische Ansätze wie die Varianzanalyse oder verwandte Verfahren, mit deren Hilfe diese Beziehung zwischen stetigen Zielgrößen und nominalen Einflussgrößen modelliert werden kann, an den jeweiligen restriktiven Vorgaben wie beispielsweise der Homoskedastizität oder der Forderung, dass die Verteilung der zugrunde liegenden (Mess)Daten bekannt sein muss. Um trotzdem in der Lage zu sein, die für eine nachhaltige Qualitätsverbesserung notwendigen Zusammenhänge darstellen zu können, wurde in Kapitel 3.2 das *Root-Cause-Analysis*-Verfahren, ein neues, nicht-parametrisches Verfahren zur Modellierung der Ursachen-Wirkungs-Beziehungen zwischen qualitätsmessenden, numerischen Zielgrößen und nominalen Einflussgrößen motiviert und hergeleitet.

Die Grundlage des RCA-Verfahrens ist die in Kapitel 2.2 eingeführte Darstellung der Ziel- und Einflussgrößen eines Fertigungsprozesses als *Produktionsprozess*, bzw. als ein Tupel aus qualitätsmessenden und qualitätsbeschreibenden Größen (vgl. Definition 2.5). Eine Realisierung eines Produktionsprozesses (ein vorliegender Datensatz) ist folglich eine Stichprobe aus allen möglichen Produktionsabläufen mit Einflussgrößen und den dazugehörigen Messergebnissen. Aufbauend auf dieser Definition lassen sich die Ursachen-Wirkungs-Beziehungen zwischen Produktionsparametern und Messdaten als Sub-Prozess bzw. als Einschränkung der Variationsmöglichkeiten der Produktionsabläufe modellieren. Mit dem RCA-Verfahren ist somit die Suche nach denjenigen Produktionsparametern, die einen signifikanten Einfluss auf die Qualität eines Fertigungsprozesses haben, äquivalent zu der Identifizierung desjenigen Sub-Prozesses, der im Sinne der Anwendung *optimal* ist. Dazu wurde gezeigt, dass es sinnvoll ist, denjenigen Sub-Prozess als optimal zu bezeichnen, der sowohl nach wirtschaftlichen als auch nach qualitativen (statistischen) Gesichtspunkten *besser* ist als alle anderen Sub-Prozesse. Das subjektive Kriterium der Wirtschaftlichkeit wurde als Support und die Qualität als QFI des Sub-Prozesses modelliert. Da die Identifizierung eines optimalen Sub-Prozesses nicht analytisch möglich ist, wurde die Lösungssuche in ein Optimierungsproblem überführt. Ein effizienter Algorithmus zur Lösung des so entstandenen Optimierungsproblems lässt sich mittels Branch and Bound Techniken implementieren: Überführt man den zulässigen Lösungsraum (alle möglichen Sub-Prozesse des Produktionsprozesses) in einen Entscheidungsbaum, dessen Knoten jeweils einen Sub-Prozess repräsentieren und bei dem mit zunehmender Tiefe

des Baumes der Support der Sub-Prozesse abnimmt, so erhält man das erste Branch and Bound Prinzip. Mit Hilfe dieses Prinzips lässt sich die Anzahl der zulässigen Lösungen, die auf Optimalität hin untersucht werden müssen, signifikant reduzieren, da Äste mit zu niedrigem Support nicht weiter betrachtet werden müssen. Für eine spezielle Klasse von Indizes, die so genannten konvexen QFIs wie beispielsweise der  $E_{ci}$ , konnte ein weiteres Branch and Bound Kriterium abgeleitet und implementiert werden: ein geschicktes Kombinieren von bekannten Sub-Prozessen erlaubt die Abschätzung der Qualität eines (unbekannten) Sub-Prozesses. Dies führt zu einer Verminderung der Häufigkeit, für die ein konvexer QFI mittels der Messdaten berechnet werden muss (2. Branch and Bound Prinzip). Aufgrund der verwendeten subjektiven Optimalitätskriterium ist es in der Praxis notwendig, nicht nur die global beste Lösung des Problems anzugeben, sondern eine Liste mit besten Lösungen. Da alle gefundenen Lösungen jedoch mit statistischer Unsicherheit behaftet sind, wird eine bekannte Technik – die Verifizierung der gefundenen Ergebnisse mittels der Übertragbarkeit von Ergebnissen der Trainings- auf die Testmenge – eingesetzt.

In der abschließenden Evaluierung des RCA-Verfahrens wurde zunächst die Aussagekraft des Verfahrens anhand eines realen Problems aus einer Gießerei erfolgreich getestet: die aufgrund des identifizierten, optimalen Sub-Prozesses durchgeführten Änderungen am Fertigungsprozess führten zu der Wiederherstellung der geforderten Fertigungsqualität. Des Weiteren wurde in der anschließend durchgeführten Simulationsstudie gezeigt, dass die Implementierung des Lösungsalgorithmus unter Verwendung des 1. bzw. 2. Branch and Bound Prinzips den Einsatz des Verfahrens in einer Produktivumgebung ermöglicht. Dabei wurde auch deutlich, dass die Laufzeit zur Identifizierung von optimalen Lösungen stark von der Anzahl der zu analysierenden Einflussparameter abhängt, was einerseits ein Nachteil des Verfahrens ist. Andererseits ist es durch die zugrunde liegende Methodik prinzipiell möglich, das Verfahren auch in anderen Bereichen als dem Qualitätsmanagement einzusetzen. Denn die Kombination von Wirtschaftlichkeit (Support) und Qualität (QFI) als Optimalitätskriterium erlaubt die Austauschbarkeit der Funktion, welche im vorliegenden Fall die Qualität misst. Dieser Vorteil ist auch gleichzeitig eine Schwäche des Verfahrens, denn aufgrund der Verwendung von zwei subjektiven Kriterien zur Identifizierung von optimalen Zusammenhängen, bedarf jede Lösung der Einordnung durch einen Experten. Den Abschluss der Evaluierung bildet eine exemplarisch durchgeführte Fertigungsprozessoptimierung mittels des RCA-Verfahrens, bei der die gefundenen Ergebnisse zur Verdeutlichung für den Leser zusätzlich grafisch dargestellt wurden.

In den Kapiteln 2 und 3 wurden verschiedene Möglichkeiten aufgezeigt, wie für das Qualitätsmanagement relevantes Wissen aus Produktionsdaten extrahiert werden kann. Die zugrunde liegenden (Produktions-)Daten ließen sich dabei als ein Paar von Ziel- und Einflussgrößen auffassen, die jeweils eingeschränkten Datentypen genügen mussten: Zielvariablen waren stets numerischer Natur und Einflussgrößen entweder nominaler oder

numerischer Natur (letztere lassen sich durch geeignete Klassierung in nominale Werte überführen). In vielen Fertigungsprozessen der Industrie können Daten, welche die Qualität messen, jedoch nicht immer numerisch dargestellt werden. Beispielsweise kann die Qualität durch eine *Gut-Schlecht*-Einteilung gemessen werden und folglich sind die Daten im Allgemeinen nominal. Verfahren, die es erlauben, auch in Fällen, in denen Produktionsprozesse aus Paaren von nominalen Ziel- und Einflussgrößen bestehen, qualitätsbezogenes Wissen zu extrahieren, sind Gegenstand des nächsten Kapitels.



## 4 Wissensextraktion mittels Assoziationsregeln

In vielen Anwendungen des Qualitätsmanagements beschreiben die zu analysierenden Daten Messergebnisse oder Ereignisse, deren Ausprägungen nicht numerischer (stetiger) sondern nominaler Art sind. Ein typisches Verfahren, mit dessen Hilfe *Wissen* aus nominalen Daten extrahiert und beispielsweise in den Qualitätsverbesserungsprozess rückgespiegelt werden kann, ist die so genannte *Assoziationsanalyse*. Sie beschreibt zum einen die Suche *nach* und zum anderen die Identifizierung *von* starken Zusammenhängen innerhalb nominaler Datensätze.

Das bekannteste Anwendungsgebiet der Assoziationsanalyse ist die *Warenkorbanalyse* (*market basket analysis*). Bei der Warenkorbanalyse wird das Kaufverhalten von Supermarktkunden untersucht, indem die einzelnen Einkäufe in Form von Transaktionen über einen gewissen Zeitraum aufgezeichnet werden. Aufbauend auf der so gewonnenen *Transaktionsdatenbank* lässt sich nun ermitteln, welche Produkte (im weiteren Verlauf *Items* oder *Ereignisse*) von der Kundschaft häufig gemeinsam erworben werden bzw. welche Produkte den Kauf anderer Produkte implizieren.

Die Grundidee der Assoziationsanalyse geht auf die Forschungsgruppe um R. Agrawal zurück. In einer Publikation Anfang der neunziger Jahre veröffentlichte die Gruppe einen Algorithmus – den so genannten *apriori*-Algorithmus – mit dessen Hilfe es auch bei den damaligen Rechenkapazitäten möglich war, große, (nominale) Datensätze effizient auf mögliche Zusammenhänge hin zu untersuchen (vgl. [Agrawal et al. 1993]). Dabei verwendeten die Autoren *Assoziationsregeln* (*association rules*), um Zusammenhänge zwischen Daten zu beschreiben. Im Laufe der vergangenen Jahre hat sich, zum einen durch die steigende Rechenleistung heutiger Computersysteme und zum anderen durch eine stetige Weiterentwicklung der Algorithmen, die Assoziationsanalyse als Standardwerkzeug zur Analyse nominaler Daten etabliert.

Das Ziel der Assoziationsanalyse ist es, diejenigen Itemkombinationen zu ermitteln, deren gemeinsames Auftreten für die jeweilige Anwendung signifikant ist. Eine solche Kombination von signifikanten Zusammenhängen wird im weiteren Verlauf als *interessant* bezeichnet. Die Kernproblematik dabei ist, geeignete *Interessanzmaße* zu finden, mit

deren Hilfe *signifikante* Zusammenhänge von *nicht signifikanten* getrennt werden können. Bei der Wahl eines geeigneten Interessantheitsmaßes können dabei zwei, zum Teil gegensätzliche Schwerpunkte gesetzt werden. Wie im weiteren Verlauf dieses Kapitels näher erläutert wird, kann zum einen der Fokus auf rein *statistische* Interessantheitsmaße gelegt werden und zum anderen auf *subjektive* Maße, welche die Interessantheit aus Sicht des Anwenders widerspiegeln. Dazu werden in Abschnitt 4.1 Assoziationsregeln formal eingeführt und im weiteren Verlauf des Kapitels 4 unterschiedliche Interessantheitsmaße vorgestellt und evaluiert.

## 4.1 Assoziationsregeln

Sei  $I = \{i_1, i_2, \dots, i_m\}$  eine Menge von Items oder eine Ereignismenge. Des Weiteren sei  $T = (T_1, \dots, T_n)$  eine Folge von  $n$  Teilmengen von  $I$  wobei jedes Folgeelement  $T_i \subseteq I, T_i \neq \emptyset, i = 1, \dots, n$ , als eine Transaktion bezeichnet wird. Dann definiert  $T$  eine Transaktionsdatenbank *über* der Ereignismenge  $I$ .

Seien  $A \subseteq I$  und  $B \subseteq I$  nicht leere, disjunkte Teilmengen – genannt Itemsets. Dann beschreibt die Implikation

$$A \rightarrow B$$

eine Assoziationsregel (*association rule*) mit der Prämisse (*antecedent*)  $A$  und Konklusion (*conclusion*)  $B$ <sup>1</sup>, wenn folgende Bedingungen erfüllt sind:

1.  $\text{supp}_T(A \rightarrow B) > 0$  (Support)
2.  $\text{conf}_T(A \rightarrow B) > 0$  (Konfidenz)

Dabei ist der Support definiert durch

$$\text{supp}_T(A \rightarrow B) = \frac{|\{i : A \cup B \subseteq T_i, i = 1, \dots, n\}|}{n} \quad (4.1)$$

und Konfidenz durch:

$$\text{conf}_T(A \rightarrow B) = \frac{|\{i : A \cup B \subseteq T_i, i = 1, \dots, n\}|}{|\{i : A \subseteq T_i, i = 1, \dots, n\}|} \quad (4.2)$$

<sup>1</sup>In der Literatur wird die Prämisse auch oft als Regelrumpf (*rule body*) und die Konklusion als Regelkopf (*rule head*) bezeichnet

Der Support einer Regel  $A \rightarrow B$  ist somit definiert als der (relative) Anteil von Transaktionen in einer Transaktionsdatenbank, die sowohl die Prämisse als auch die Konklusion enthalten. Die Konfidenz hingegen misst den Anteil der Transaktion, welche Prämisse und Konklusion enthalten, unter denjenigen Transaktionen, welche die Prämisse enthalten.

In diesem Zusammenhang ist der Begriff Implikation jedoch nicht wort-wörtlich zu verstehen, da eine Assoziationsregel keine "echte" Implikation darstellt im Sinne von: "wenn  $A$  dann  $B$ ". Vielmehr untersucht man mit einer Assoziationsregel  $A \rightarrow B$  das gemeinsame Auftreten zweier Itemsets in einer Transaktionsdatenbank  $T$  und weist mittels Interessantheitsmaße der Implikation eine Bedeutung im Kontext der Anwendung zu.

Ein Beispiel für eine Transaktionsdatenbank mit Transaktionen  $T = \{T_1, \dots, T_5\}$  ist in Tabelle 4.1 dargestellt. Dabei sind die Transaktionen jeweils Teilmenge der Ereignismenge  $I = \{i_1, \dots, i_4\}$ .

Transaktion	Ereigniskombination
$T_1$	$\{i_1, i_2\}$
$T_2$	$\{i_1, i_3, i_4\}$
$T_3$	$\{i_1, i_2, i_4\}$
$T_4$	$\{i_1, i_2, i_3, i_4\}$
$T_5$	$\{i_2, i_4\}$

Tabelle 4.1: Beispiel Transaktionsdatenbank

Aufbauend auf dem in Tabelle 4.1 gegeben Beispiel lässt sich exemplarisch folgende Assoziationsregel beschreiben:  $\{i_1\} \rightarrow \{i_2\}$  mit

$$\text{supp}_T(\{i_1\} \rightarrow \{i_2\}) = \frac{|\{T_1, T_3, T_4\}|}{|T|} = \frac{3}{5} = 60\%$$

und

$$\text{conf}_T(\{i_1\} \rightarrow \{i_2\}) = \frac{|\{T_1, T_3, T_4\}|}{|\{T_1, T_2, T_3, T_4\}|} = \frac{3}{4} = 75\%$$

Eine Implikation  $A \rightarrow B$ , bei der sowohl Prämisse als auch Konklusion jeweils aus nur einem Element bestehen ( $|A| = |B| = 1$ ), wird im weiteren Verlauf als *einstellige Assoziationsregel* bezeichnet. Eine Regel, deren Prämisse oder Konklusion aus mindestens zwei Elementen besteht, wie beispielsweise  $\{i_1, i_2\} \rightarrow \{i_4\}$ , bezeichnen wir als *mehrstellige Assoziationsregel*.

Für die Regel  $\{i_1, i_2\} \rightarrow \{i_4\}$  gilt:

$$\text{supp}_T(\{i_1, i_2\} \rightarrow \{i_4\}) = \frac{|\{T_3, T_4\}|}{|T|} = \frac{2}{5} = 40\%$$

und

$$\text{conf}_T(\{i_1, i_2\} \rightarrow \{i_4\}) = \frac{|\{T_3, T_4\}|}{|\{T_1, T_3, T_4\}|} = \frac{2}{3} = 66.67\%$$

Ein- und mehrstellige Assoziationsregeln werden unter dem Begriff *einfache Assoziationsregeln* zusammengefasst. Um eine Assoziationsanalyse mit zusätzlichen Informationen anzureichern, können einfache Assoziationsregeln mit *virtuellen Items* kombiniert werden (vgl. [Hettich und Hippner 2001]). Beispielsweise kann im Zuge einer Warenkorbanalyse die Marke oder der Standort des Artikels als virtuelle Variable hinzugefügt werden. Das Konzept von virtuellen Items erlaubt auch die Betrachtung von “negativen“ Items. Negative Items – im Folgenden bezeichnet mit  $\neg i_k$  – sind Items, die nicht in dem jeweiligen Itemset enthalten sind, sprich  $\neg i_k \notin I, i_k \in I, i = 1 \dots, n$ .

Ein Beispiel für eine Assoziationsregel mit einem negativen Item ist die Implikation  $\{\neg i_3\} \rightarrow \{i_1\}$ . Analog zu einfachen Regeln lassen sich Support und Konfidenz wie folgt berechnen:

$$\text{supp}_T(\{\neg i_3\} \rightarrow \{i_1\}) = \frac{|\{T_1, T_3\}|}{|T|} = \frac{2}{5} = 40\%$$

und

$$\text{conf}_T(\{\neg i_3\} \rightarrow \{i_1\}) = \frac{|\{T_1, T_3\}|}{|\{T_1, T_3, T_5\}|} = \frac{2}{3} = 66.67\%$$

Folglich enthält eine Transaktion, welche zufällig aus  $T$  gezogen wird, die *nicht*  $i_3$  enthält, in  $2/3$  aller Fälle das Ereignis (Item)  $i_1$ . Eine formale Definition eines wahrscheinlichkeitstheoretischen Modells für Assoziationsregeln wird auf Seite 106 gegeben.

Die Verwendung von virtuellen Items bzw. negativen Items ist auch auf mehrstellige Assoziationsregeln übertragbar. Dabei lassen sich Prämisse bzw. Konklusion als eine Verknüpfung der Items mittels des logischen UND Operators darstellen. Werden negative Items verwendet, so erfolgt die Verknüpfung von Itemsets mit Hilfe des logischen *UND NICHT* Operators. In der Literatur werden mehrstellige Assoziationsregeln, deren Prämisse bzw. Konklusion mittels des logischen UND NICHT Operators verknüpft sind, auch als *Dissociation Rules*<sup>2</sup> (vgl. [Berry und Linoff 1997]) bezeichnet. Beispielsweise gilt für  $\{i_1, \neg i_3\} \rightarrow \{i_4\}$  analog zu einfachen Regeln:

$$\text{supp}_T(\{i_1, \neg i_3\} \rightarrow \{i_4\}) = \frac{|\{T_3\}|}{|T|} = \frac{1}{5} = 20\%$$

<sup>2</sup>Dissociation: Ausgrenzung, Absonderung

und

$$\text{conf}_T(\{i_1, \neg i_3\} \rightarrow \{i_4\}) = \frac{|\{T_3\}|}{|\{T_1, T_3\}|} = \frac{1}{2} = 50\%$$

Das Prinzip der Verknüpfung von Elementen eines Itemsets ist jedoch nicht beschränkt auf den *UND NICHT* Operator. Prinzipiell lässt sich jeder logische Operator zur Bildung von mehrstelligen Assoziationsregeln nutzen. Im weiteren Verlauf werden Regeln, deren Prämisse oder Konklusion mittels logischer Operatoren gebildet wird, als *komplexe Assoziationsregeln* bezeichnet.

## Klassifikationsregeln

Die in Abschnitt 4.1 eingeführten Assoziationsregeln können als so genannte “ungerichtete“ Regeln interpretiert werden, da eine Teilmenge aus  $I$  sowohl Prämisse als auch Konklusion sein kann. In vielen Fällen ist es jedoch durch die Anwendung vorgegeben, dass bestimmte Items nur als Konklusion(en) sinnvoll interpretiert werden können. Will man mittels Assoziationsregeln Zusammenhänge zwischen qualitätsmessenden Zielgrößen und prozessbeschreibenden Einflussgrößen identifizieren, so genügt es, eine bestimmte Klasse von Regeln zu betrachten: In einem solchen Fall sind lediglich Regeln von Interesse, deren Prämissen durch Einfluss- und deren Konklusion durch Zielgrößen beschrieben werden.

Sei beispielsweise eine Produktionsdatenbank eines Automobilherstellers gegeben, in der zu jedem produzierten Fahrzeug die entsprechenden Produktionsdaten gespeichert sind. Des Weiteren sei für all diese Fahrzeuge bekannt, ob ein Fehler in einem bestimmten elektronischen Steuermodul des Fahrzeugs aufgetreten ist oder nicht. Dann kann die Kombination der Fehlerinformation (Steuermodul) mit der Produktionsdatenbank als eine Transaktionsdatenbank über der Menge der Werte aller Produktionsparameter und der Fehlerinformation aufgefasst werden. Dabei besteht jede Transaktion aus zwei disjunkten Itemsets bzw. einem fahrzeugspezifischem Datensatz, der sich zusammensetzt aus:

1. Produktionsparameter wie beispielsweise verwendetes Material, Maschinen, Produktionsmonat etc. Diese können auch als (unabhängige) Einflussvariablen aufgefasst werden.
2. Informationen bezüglich des Fehlerverhaltens des betrachteten Steuermoduls – ist der Fehler in dem Steuermodul des jeweiligen Fahrzeugs aufgetreten? In dem vorliegenden Kontext beschreibt diese Information die zu erklärende, abhängige Zielvariable.

Will man nun das Fehlerverhalten des Steuermoduls mittels der Produktionsparameter erklären<sup>3</sup>, so kann eine Assoziationsanalyse durchgeführt werden, bei der die Prämisse aus den Produktionsparametern besteht und die Konklusion das Fehlerverhalten beschreibt. Ein Beispiel für eine Implikation, bei der die verwendeten Itemsets in zwei disjunkte Klassen aufgeteilt sind, ist die Assoziationsregel mit der Prämisse *Produktionsmonat = Januar* und Konklusion *Fehler im Dieselmotorsteuergerät*.

Das Konzept der Einteilung von Daten, die zur Qualitätssicherung erfasst werden, in (disjunkte) Ziel- und Einflussgrößen, wurde bereits in den Kapiteln 2 und 3 vorgestellt. In diesem Kontext wurde auch der Begriff Produktionsprozess eingeführt (vgl. Definition 2.5), der als ein Tupel, bestehend aus Ziel- und Einflussgrößen, definiert ist. Ziel- und Einflussvariablen eines Produktionsprozesses sind dabei durch folgende Datenstruktur charakterisiert:

**Einflussgrößen** Die Produktionsparameter werden durch unterschiedliche, in der Regel nominale Variablen beschrieben, die im Lauf des Fertigungsprozesses nicht konstant gehalten werden konnten, wie beispielsweise die Temperatur, die verwendete Maschine oder der Arbeiter. Da eine Variable mehrere Ausprägungen hat, jedoch pro Datensatz (Element einer Realisierung eines Produktionsprozesses) nur einen Wert annehmen kann, sprechen wir von einer komplexen Datenstruktur.

**Zielgrößen** Um die Qualität des Produktionsprozesses zu dokumentieren und mittels Qualitätsfähigkeitsindizes zu bewerten, wird eine stetige Messgröße wie beispielsweise eine Maßzahl für die (resultierende) Oberflächenbeschaffenheit oder andere geometrische Eigenschaften eines Werkstücks erfasst. Konkret wird pro Element einer Realisierung (produziertes Gut) eines Produktionsprozesses genau eine Messgröße erfasst.

Um signifikante Parameter, die einen Einfluss auf die qualitätsmessende Zielgröße haben, zu bestimmen, wird im Zuge der Root-Cause-Analysis (RCA) der Begriff *Sub-Prozess* eingeführt. Mittels eines Sub-Prozesses lässt sich der Wertebereich eines Produktionsprozesses einschränken (Restriktionsmenge) und der bedingte Qualitätsfähigkeitsindex für eine Realisierung eines Produktionsprozesses bestimmen. Die Restriktionsmenge beschreibt dabei alle zugelassenen Variablen eines Sub-Prozesses, indem für jede Einflussvariable vorgegeben wird, welche Werte sie annehmen kann.

In der Praxis werden jedoch nicht immer alle Variablen gleichzeitig einschränken. Prinzipiell lassen sich zwei Arten unterscheiden:

---

<sup>3</sup>Eine ausführliche Motivation und Beschreibung verwendeter Methoden und Daten für eine solche Assoziationsanalyse ist in Kapitel 4.4 gegeben

- Variablen, die auf eine Teilmenge des zulässigen Wertebereichs beschränkt sind.
- Variablen, die *nicht* restringiert sind, sprich sie können jeden Wert des jeweiligen zulässigen Wertebereichs annehmen.

Anstatt nun für alle Variablen eine Restriktionsmenge vorzugeben, kann ein Sub-Prozess durch die Vorgabe einer *Fixierungsmenge* beschrieben werden, welche nur die restringierten Variablen angibt, bei gleichzeitigem Variabelhalten der übrigen Einflussgrößen<sup>4</sup>.

Betrachtet man eine Assoziationsregel  $A \rightarrow B$ , so kann diese ebenfalls als eine Art Fixierungsmenge aufgefasst werden. Denn durch die Implikation werden nur diejenigen Elemente einer Transaktionsdatenbank “erfasst“, welche sowohl die Prämisse als auch die Konklusion enthalten. Somit basieren Sub-Prozesse und Assoziationsregeln auf der gleichen grundlegenden Idee – der Einschränkung des zulässigen Wertebereiches durch die Fixierung von Variablen – jedoch existieren einige wesentliche Unterschiede, die im Folgenden kurz erläutert werden:

**Variablenart** Sub-Prozesse werden durch ein Tupel aus Ziel- und Einflussvariablen beschrieben, wobei die Zielvariable stetig ist. Assoziationsregeln hingegen erlauben es Zielvariablen nominaler Art zu betrachten.

**Datenstruktur** Die Grundidee der Assoziationsanalyse geht auf die Warenkorbanalyse zurück. Dabei wird eine Transaktion als eine Menge von (unterschiedlichen) Produkten aufgefasst. Die Itemsetmenge  $I$  berücksichtigt folglich keine komplexe Struktur, wie die Unterteilung in Variablen mit unterschiedliche Ausprägungen. Bei der RCA liegen jedoch komplexe Datenstrukturen für die Einflussgrößen vor. Für den Fall, dass komplexe Datenstrukturen mittels der Assoziationsanalyse untersucht werden sollen, so müssen die Daten in eine *flache* Struktur überführt werden (vgl. Anwendungsbeschreibung in Kapitel 4.4). Darüber hinaus erlaubt die Assoziationsanalyse die simultane Betrachtung von mehreren, nominalen Zielvariablen. Das RCA-Verfahren hingegen modelliert die Ursachen-Wirkungs-Beziehungen zwischen einer Ziel- und mehreren Einflussgrößen.

**Restriktion bzw. Fixierung** Bei der Root-Cause-Analysis wird die Zielvariable nur implizit durch die Festlegung des Wertebereiches der Einflussgrößen festgelegt (Sub-Prozess). Im Gegensatz dazu werden bei der Assoziationsregel sowohl Ziel- als auch Einflussvariablen durch die Konklusion und Prämisse restringiert.

Schränkt man stets Konklusion und Prämisse auf zwei disjunkte Teilmengen ein, so

---

<sup>4</sup>Auf Seite 62 wurde gezeigt, dass sich jede Fixierungsmenge in eine Restriktionsmenge überführen lässt. Für die in dieser Arbeit betrachteten Fälle ist auch die Umkehrung möglich

spricht man von *Klassifikationsregeln*. Diese lassen sich formal wie folgt beschreiben:

Seien  $I = \{i_1, \dots, i_I\}$ ,  $J = \{i_1, \dots, i_J\}$  zwei disjunkte Mengen von Ereignissen und eine Transaktion bezüglich  $I$  und  $J$  sei eine Kombination  $(T_I, T_J)$  von nicht leeren Teilmengen  $T_I \subset I$ ,  $T_J \subset J$ . Des Weiteren beschreibe die Folge  $\mathbf{T} = (T_1, \dots, T_n)$  eine Transaktionsdatenbank. Dann wird ein Paar  $A, B$  von nicht leeren Teilmengen  $A \subset I$  und  $B \subset J$  als Klassifikationsregel  $A \rightarrow B$  bezeichnet.

Sei wie in obigem Beispiel eine Transaktionsdatenbank gegeben, wobei als Produktionsparameter der Produktionsmonat und die Funktionsfähigkeit eines bestimmten Steuermoduls (0 funktioniert, 1 funktioniert nicht) erfasst sind. Dann ist  $I = \{1, \dots, 12\}$  und  $J = \{0, 1\}$ . Wird ein zusätzliches Produktionsmerkmal, wie beispielsweise der Name des Arbeiters ( $A, B$  oder  $C$ ), der dieses Steuermodul verbaut hat, erfasst, so erweitert sich die Menge der Einflussgrößen zu:

$$I = \{1, \dots, 12\} \cup \{A, B, C\}$$

All diejenigen Transaktionen<sup>5</sup>, die im Juni von Arbeiter  $A$  erstellt wurden und ein funktionierendes Steuermodul besitzen, würden durch folgende Klassifikationsregel beschrieben werden:  $\{6, A\} \rightarrow \{0\}$

Aufgrund der in dieser Arbeit vorgestellten Charakteristik von Qualitätsmanagementanwendungen – der Einteilung in Ziel- und Einflussvariablen – wird im weiteren Verlauf der Begriff Assoziationsregel als Synonym für Klassifikationsregeln verwendet. Dabei gelte für eine gegebene Transaktionsdatenbank

$$\mathbf{T} = \{T_1, \dots, T_n\} = \{(T_{I_1}, T_{J_1}), \dots, (T_{I_n}, T_{J_n})\}$$

über  $I \times J$  und eine Assoziationsregel  $A \rightarrow B$  folgende, vereinfachende Notation:

$$\begin{aligned} n_{A \wedge B} &= |\{i = 1 \dots, n : A \subseteq T_{I_i} \wedge B \subseteq T_{J_i}\}| \\ n_A &= |\{i = 1 \dots, n : A \subseteq T_{I_i}\}| \\ n_B &= |\{i = 1 \dots, n : B \subseteq T_{J_i}\}| \\ n &= |\mathbf{T}| = |\{(T_{I_1}, T_{J_1}), \dots, (T_{I_n}, T_{J_n})\}| \end{aligned} \tag{4.3}$$

Analog gilt für “negative“ Items:

$$\begin{aligned} n_{\neg A \wedge B} &= |\{i = 1 \dots, n : A \not\subseteq T_{I_i} \wedge B \subseteq T_{J_i}\}| \\ n_{\neg A \wedge \neg B} &= |\{i = 1 \dots, n : A \not\subseteq T_{I_i} \wedge B \not\subseteq T_{J_i}\}| \\ n_{\neg A} &= |\{i = 1 \dots, n : A \not\subseteq T_{I_i}\}| \end{aligned} \tag{4.4}$$

<sup>5</sup>Alternativ kann an dieser Stelle der Begriff *Transaktion* mit *Fahrzeug* ersetzt werden

## 4.2 Interessantheitsmaße zur Bewertung von Regeln

Ziel einer Assoziationsanalyse ist es, mit Hilfe der im vorherigen Abschnitt eingeführten Assoziationsregeln, eine vorliegende Datenmenge auf interessante Zusammenhänge hin zu untersuchen. In unserer Anwendungssituation entspricht dabei jeder Datensatz (Transaktion) einem Vektor aus den Werten der Einflussgrößen  $A$ , kombiniert mit einem oder mehreren Werten der Zielgröße  $B$ . Zu jeder Kombination aus einer Teilmenge  $A$  des Wertebereichs der Einflussgrößen und einer Teilmenge  $B$  des Wertebereichs der Zielgröße(n), können nun Assoziationsregeln  $A \rightarrow B$  gebildet werden.

Wird ein Datensatz hinsichtlich sämtlicher Assoziationsregeln analysiert, so ergibt sich schon bei relativ kleinen Wertebereichen für Prämissen und Konklusionen eine unüberschaubare Liste an Zusammenhängen. Diese sind die auf interessante und im Sinne der Aufgabenstellung brauchbaren Assoziationsregeln durchzusehen: ([Dong und Li 1998]):

“When the amount of such high level knowledge is large, which is typically the case for association rules, the selection of interesting patterns becomes a serious problem for the human user“

Um diese große Anzahl an Regeln auf eine sinnvolle Menge zu reduzieren, müsste in einem ersten Schritt der Anwender mittels manueller Durchsicht und Interpretation jede einzelne Regel bewerten und den tatsächliche Aussagegehalt bestimmen. Dies ist sicherlich nicht in einem vertretbaren Zeitaufwand möglich, so dass es für die Anwendung der Assoziationsanalyse notwendig ist, Verfahren und Maße zu entwickeln, welche interessante von uninteressanten Regeln trennen.

Ein erster Schritt, um dem Anwender, angesichts der Informationsfülle eine Übersicht über die generierten Regeln zu geben, ist zunächst mittels verschiedener Darstellungsarten der gefundenen Zusammenhänge möglich. So können Regeln mittels ihrer Merkmale (Prämisse, Konklusion, Support, Konfidenz etc.) geeignet sortiert oder gefiltert werden<sup>6</sup>. Auch eine grafische Darstellung von Regeln ist möglich ([Bollinger 1996], S.259). Diese soll dem Anwender ein schnelles Erkennen von Zusammenhängen ermöglichen. Doch bei der grafischen Darstellung von vielen komplexen Regeln, die sich aus der Kombination mit unterschiedlichen booleschen Operatoren ergeben, ist eine Interpretation schwierig. Somit reichen grafische Darstellungsformen nicht aus, um, aus der “großen“ Menge an generierten Regeln, die *interessanten* Zusammenhänge heraus zu extrahieren. Eine automatische Trennung von interessanten und uninteressanten Zusammenhängen ist notwendig. Hierfür müssen geeignete Interessantheitsmaße verwendet werden. Sie erlauben

---

<sup>6</sup>Zum Beispiel können Regeln absteigend nach Konfidenz sortiert werden oder es können nur solche Regeln angezeigt werden, die einen gewissen Mindestwert für Support oder Konfidenz aufweisen

es, relevante Sachverhalte aufgrund von statistischen *und* bzw. *oder* anwendungsspezifischen Sachverhalten zu identifizieren und bieten sowohl eine quantitative als auch eine qualitative Sichtweise auf gefundene Zusammenhänge.

Die Betrachtung verschiedener Interessantheitsmaße zur Extraktion von signifikanten Zusammenhängen bildet den Kern des Abschnitts 4.2. Dazu werden zunächst einige, der in der Literatur gängigen Maße motiviert und interpretiert. Wie gezeigt wird, sind diese nicht in allen Anwendungen zielführend, so dass sowohl ein entscheidungsorientierter als auch ein statistischer Ansatz zur Bewertung von Regeln vorgestellt wird. Ersterer ermöglicht die Berücksichtigung von wirtschaftlichen Aspekten bei der Identifizierung von interessanten Assoziationsregeln.

### 4.2.1 Kontingenzanalyse

Um Zusammenhänge zwischen nominalen Variablen zu identifizieren, werden im Allgemeinen *Kontingenztabellen* verwendet. Untersuchungen von Abhängigkeitsstrukturen nominaler Variablen nutzen eben diese Tabelle und werde somit unter dem Begriff Kontingenzanalyse zusammengefasst.

Seien  $X$  und  $Y$  zwei nominale Variablen, wobei  $X$  über  $i = 1 \dots I$  und  $Y$  über  $j = 1 \dots J$  Ausprägungen verfügt. Offensichtlich gibt es  $I \cdot J$  unterschiedliche Kombinationen  $(X_i, Y_j)$ . Betrachtet man die Merkmale auf einer statistischen Masse vom Umfang  $n$ , so kann für jede Kombination  $(X_i, Y_j)$  die absolute Häufigkeit  $n_{ij}$  und die relative Häufigkeit  $p_{ij}$  ermittelt und jeweils in einer Kontingenztabelle<sup>7</sup> bestehend aus  $I$  Zeilen und  $J$  Spalten zusammengefasst werden. Diese relative Häufigkeit entspricht dabei auch der Wahrscheinlichkeit, bei zufälligem Ziehen, diese Kombination  $(X_i, Y_j)$  an Merkmalswerten zu beobachten.

Summiert man die Spalten bzw. die Zeilen einer Kontingenztabelle auf, so erhält man die (empirischen) *Randverteilungen* (Marginalverteilungen). Diese werden im Falle von Reihenvariablen mit  $\{p_{i\cdot}\}$  und im Falle von Spaltenvariablen mit  $\{p_{\cdot j}\}$  bezeichnet. Dabei gilt:

$$p_{i\cdot} = \sum_j p_{ij}$$

und

$$p_{\cdot j} = \sum_i p_{ij}$$

<sup>7</sup>Der Begriff Kontingenztabelle geht auf den Statistiker Karl Pearson zurück. Alternativ wird in der Literatur auch der Begriff cross-classification table (vgl. [Agresti 1990], Kapitel 2) verwendet

Analog sind die absoluten Randhäufigkeiten  $n_{i\cdot} = \sum_j n_{ij}$  und  $n_{\cdot j} = \sum_i n_{ij}$  definiert.

Fixiert man eine Variable, so kann man die bedingte Verteilung bestimmen. Bezeichne  $p_{j|i}$  die bedingte Häufigkeit von  $Y_j$  (Spalte  $j$ ) bei einer gegebenen Ausprägung  $X_i$  (Zeile  $i$ ) mit  $\sum_j p_{j|i} = 1$ , dann bildet  $\{p_{1|i}, \dots, p_{J|i}\}$  die bedingte Verteilung von  $Y$  unter  $i$  von  $X$ . Die bedingte Verteilung ergibt sich aus der Randverteilung und der gemeinsamen Verteilung (vgl. [Agresti 1990] Kapitel 2):

$$p_{j|i} = p_{ij}/p_{i\cdot} \quad (4.5)$$

Im vorliegenden Kontext sind zwei Variablen statistisch unabhängig (vgl. [Agresti 1990], Kapitel 2.1.2), wenn die gemeinsame, relative Verteilung dem Produkt der relativen Randverteilungen entspricht. Nutzt man diese Tatsache, so erhält man eine einfache, erste Möglichkeit, die Unabhängigkeit zweier Variablen durch die bedingte Verteilung auszudrücken. Sei  $X$  eine Ziel- und  $Y$  eine Einflussvariable, so sind beide unabhängig wenn gilt:

$$p_{j|1} = \dots = p_{j|I}, \forall j = 1, \dots, J \quad (4.6)$$

Um den Zusammenhang zwischen der Prämisse und der Konklusion einer Assoziationsregel zu analysieren, lässt sich das eben vorgestellte Konzept verwenden. Sei  $T = (T_1, \dots, T_n)$  eine Transaktionsdatenbank und  $A \rightarrow B$  eine Assoziationsregel. Da für eine Assoziationsregel die Prämisse (bzw. Konklusion) entweder in einer Transaktion  $T \in T$  enthalten ist oder nicht, lässt sich der gesuchte Zusammenhang mittels einer  $2 \times 2$  Kontingenztabelle (siehe Tabelle 4.2) darstellen. Häufig wird an dieser Stelle auch von einer Vierfeldertafel gesprochen. Dabei ist

- $n_{AB}$  die Anzahl der Transaktionen  $T_i$  mit  $A \cup B \subseteq T_i$
- $n_{A\bar{B}}$  die Anzahl der Transaktionen  $T_i$  mit  $A \subseteq T_i, B \not\subseteq T_i$
- $n_{\bar{A}B}$  die Anzahl der Transaktionen  $T_i$  mit  $A \not\subseteq T_i, B \subseteq T_i$
- $n_{\bar{A}\bar{B}}$  die Anzahl der Transaktionen  $T_i$  mit  $A \not\subseteq T_i$  und  $B \not\subseteq T_i$

	$B$	$\bar{B}$	
$A$	$n_{AB}$	$n_{A\bar{B}}$	$n_A$
$\bar{A}$	$n_{\bar{A}B}$	$n_{\bar{A}\bar{B}}$	$n_{\bar{A}}$
	$n_B$	$n_{\bar{B}}$	$n$

Tabelle 4.2:  $2 \times 2$  Kontingenztabelle für die Regel  $A \rightarrow B$

Damit kann mit Hilfe einer Kontingenzanalyse untersucht werden, ob bei der Transaktionsdatenbank die Merkmale “ $A$  erfüllt ( $A \subseteq T_i$ ) oder erfüllt nicht ( $A \not\subseteq T_i$ )“ und “ $B$  erfüllt ( $B \subseteq T_i$ ) oder erfüllt nicht ( $B \not\subseteq T_i$ )“ abhängig sind.

Seien  $I$  und  $J$  disjunkte Itemsets und  $\mathcal{T} = \{T | T \text{ ist eine Transaktion}\}$  die Menge aller Transaktionen wie oben beschrieben. Des Weiteren sei  $\Sigma$  die Potenzmenge von  $\mathcal{T}$  und  $P$  ein Wahrscheinlichkeitsmaß. Dann definiert das Tripel  $(\mathcal{T}, \Sigma, P)$  einen Wahrscheinlichkeitsraum mit der Menge aller technisch möglichen Transaktionen als Grundgesamtheit. Der so definierte Wahrscheinlichkeitsraum ist somit ein hypothetischer Raum, aus dem eine konkrete Transaktion  $T$  per Zufall entsprechend dem Maß  $P$  gezogen wird. Mit Hilfe des Wahrscheinlichkeitsraumes lässt sich der im vorherigen Abschnitt eingeführte Begriff der Assoziation- bzw. Klassifikationsregel  $A \rightarrow B$  als das gemeinsame Auftreten der Ereignisse  $A$  und  $B$  auffassen und wie folgt definieren:

**Definition 4.1** (Klassifikationsregel)

Sei das Tripel  $(\mathcal{T}, \Sigma, P)$  ein Wahrscheinlichkeitsraum zu den Itemsets  $I = \{i_1, \dots, i_I\}$ ,  $J = \{i_1, \dots, i_J\}$  und gelte für jede Transaktion  $T \in \mathcal{T}$ :

$$T = (T_I, T_J) \text{ mit } T_I \subseteq I, T_J \subseteq J \text{ und } T_I, T_J \neq \emptyset$$

Des Weiteren seien  $A \subseteq I, B \subseteq J$  zwei Itemsets. Dann definiert die Implikation  $A \rightarrow B$  eine Klassifikationsregel.

**Definition 4.2**

Als Support der in 4.1 definierten Klassifikationsregel definieren wir:

$$\begin{aligned} \text{supp}_{\mathcal{T}}(A \rightarrow B) &= P(A \subseteq T_I, B \subseteq T_J) \\ &= P(\{T \in \mathcal{T} : A \subseteq T_I, B \subseteq T_J\}) \\ &= \sum_{T: A \subseteq T_I, B \subseteq T_J} P(T) \end{aligned} \quad (4.7)$$

und als Konfidenz

$$\begin{aligned} \text{conf}_{\mathcal{T}}(A \rightarrow B) &= \frac{P(A \subseteq T_I, B \subseteq T_J)}{P(A \subseteq T_I)} \\ &= \frac{P(\{T \in \mathcal{T} : A \subseteq T_I, B \subseteq T_J\})}{P(\{T \in \mathcal{T} : A \subseteq T_I\})} \\ &= \frac{\sum_{T: A \subseteq T_I, B \subseteq T_J} P(T)}{\sum_{T: A \subseteq T_I} P(T)} \end{aligned} \quad (4.8)$$

Mit

$$\langle A \cup B \rangle = \{T = (T_I, T_J) \in \mathcal{T} : A \subseteq T_I \wedge B \subseteq T_J\}$$

bzw.

$$\langle A \rangle = \{T = (T_I, T_J) \in \mathcal{T} : A \subseteq T_I\}$$

vereinfacht sich dies zu

$$\begin{aligned} \text{supp}_{\mathcal{T}}(A \rightarrow B) &= P(\langle A \cup B \rangle) \\ \text{conf}_{\mathcal{T}}(A \rightarrow B) &= \frac{P(\langle A \cup B \rangle)}{P(\langle A \rangle)} \end{aligned}$$

Eine Transaktionsdatenbank  $\mathcal{T}$  lässt sich dann als eine Stichprobe mit Zurücklegen aus  $\mathcal{T}$  entsprechend dem Wahrscheinlichkeitsmaß  $P$  auffassen. Damit sind (4.2) und (4.1) Schätzer für die Konfidenz und den Support der Assoziationsregel  $A \rightarrow B$ .

Analog zu Tabelle 4.2 ergibt sich daraus die folgende Tabelle:

	$B$	$\neg B$	
$A$	$P(\langle A \cup B \rangle) = \pi_{11}$	$P(\langle A \cup \neg B \rangle) = \pi_{12}$	$P(\langle A \rangle) = \pi_1.$
$\neg A$	$P(\langle \neg A \cup B \rangle) = \pi_{21}$	$P(\langle \neg A \cup \neg B \rangle) = \pi_{22}$	$P(\langle \neg A \rangle) = \pi_2.$
	$P(\langle B \rangle) = \pi_{.1}$	$P(\langle \neg B \rangle) = \pi_{.2}$	1

Tabelle 4.3:  $2 \times 2$  Wahrscheinlichkeitsverteilung für die Regel  $A \rightarrow B$

Dabei ist  $P(\langle A \cup \neg B \rangle) = P(\{T = (T_I, T_J) \in \mathcal{T} : A \subseteq T_I, B \not\subseteq T_J\})$  und analog für die übrigen Größen.

Sei nun  $\mathcal{T} = (T_1, \dots, T_n)$  eine Stichprobe vom Umfang  $n$  und  $n_{ij}$  ( $p_{ij}$ ) die beobachtete Häufigkeit (relative Häufigkeit) für den Zelleintrag  $i, j$  in Tabelle 4.2. Der bekannteste Test zur Untersuchung der Unabhängigkeit zweier nominalen Variablen ist der  $\chi^2$ -Unabhängigkeitstest. Die Grundidee des Tests ist, die beobachtete Häufigkeit der Zelleinträge mit dem erwarteten Wert bei Unabhängigkeit (4.6) zu vergleichen. Definiere  $\pi_{ij}$ ,  $i, j \in \{1, 2\}$  die Wahrscheinlichkeit des Zelleintrages  $i, j$  und  $\pi_{i.}$ ,  $\pi_{.j}$  die Randverteilungen aus Tabelle 4.2. Dann gilt bei Unabhängigkeit beider Variablen die Nullhypothese

$$H_0 : \pi_{ij} = \pi_{i.} \pi_{.j}$$

Dabei sind die Werte für  $\pi_{ij}$ ,  $\pi_{i.}$  und  $\pi_{.j}$  nicht beobachtbar, jedoch aufgrund der Stichprobe schätzbar. Sei  $\hat{m}_{ij} = np_{i.}p_{.j}$  ein Schätzer für die Zelhäufigkeiten  $\pi_{ij}$ , dann kann die Nullhypothese – beide Variablen sind unabhängig – mittels der folgenden Teststatistik überprüft werden:

$$\chi^2 = \sum_i \sum_j \frac{(n_{ij} - \hat{m}_{ij})^2}{\hat{m}_{ij}} \quad (4.9)$$

Die in (4.9) definierte Teststatistik folgt dabei einer  $\chi^2$ -Verteilung mit nur einem Freiheitsgrad (siehe [Agresti 1990], Kapitel 3.3.1). Sei  $(1-\alpha)$  das geforderte Signifikanzniveau und ist die resultierende Prüfgröße größer als das  $(1-\alpha)$ -Quantil der  $\chi^2$ -Verteilung, so wird  $H_0$  abgelehnt.

Eine Alternative zum  $\chi^2$ -Unabhängigkeitstest ist der *Binomialtest*. Dabei wird implizit überprüft, ob ein Zusammenhang zwischen Regelkopf und -körper besteht, indem getestet wird ob, unter Annahme der Prämisse, die Regel signifikant ist. Sei eine Assoziationsregel  $A \rightarrow B$  gegeben und sei  $\nu_0 = P(B)$  die *unbedingte* Wahrscheinlichkeit des Auftretens des Ereignisses  $B$  in allen Transaktionen und  $\nu = P(B|A)$  die *bedingte* Wahrscheinlichkeit des Ereignisses  $B$  unter allen Transaktionen, die  $A$  enthalten. Dann besteht kein signifikanter statistischer Zusammenhang zwischen dem Regelkopf  $A$  und der Konklusion  $B$ , wenn die bedingte Wahrscheinlichkeit  $\nu$  kleiner oder gleich der unbedingten Wahrscheinlichkeit  $\nu_0$  ist. Mit anderen Worten:  $A$  und  $B$  haben einen signifikanten Zusammenhang, wenn die Wahrscheinlichkeit, dass das Ereignis  $B$  gegeben  $A$  höher ist, als im unbedingten Fall.

Um somit Prämisse und Konklusion auf einen statistischen Zusammenhang mittel des Binomialtests zu untersuchen, muss die Nullhypothese

$$H_0 : \nu \leq \nu_0$$

vs. die Alternativhypothese

$$H_1 : \nu > \nu_0$$

getestet werden. Sei  $n_B/n$  ein Schätzer für die unbedingte und  $n_{A \wedge B}/n_A$  ein Schätzer für die bedingte Wahrscheinlichkeit  $\nu_0$  bzw.  $\nu$  aus einer Stichprobe  $T = (T_1, \dots, T_n)$ . Des Weiteren sei ein Signifikanzniveau  $\alpha$  gegeben, so wählt man einen "kritischen" Bereich der Form

$$\mathcal{K} = \{k, k+1, \dots, n_A\}.$$

$H_0$  wird verworfen, falls  $n_{A \wedge B}$  "zu häufig" ist. Dazu wird  $k$  wie folgt bestimmt:

$$k = \min \left\{ i \in \{0, 1, \dots, n_A\} : \sum_{j=i}^{n_A} \binom{n_A}{j} \nu_0^j (1 - \nu_0)^{n_A-j} \leq \alpha \right\}$$

Ist  $n_{A \wedge B} \geq k$  bzw. gilt für den  $p$ -Wert

$$p = 1 - \sum_{i=0}^{n_{A \wedge B}-1} \binom{n_A}{i} \nu_0^i (1 - \nu_0)^{n_A-i} \leq \alpha$$

wird  $H_0$  verworfen.

Die in diesem Abschnitt eingeführten Tests geben Aufschluss darüber, ob es einen statistisch signifikanten Zusammenhang zwischen der Prämisse und der Konklusion einer

Regel gibt. Jedoch geben sie keinen Hinweis darauf, wie stark der Zusammenhang zwischen beiden ist. Um diesen zu messen, werden in der Literatur Assoziations- bzw. Interessantheitsmaße verwendet. Ziel dieser Maße ist es, die in der Kontingenztabelle abgebildete Abhängigkeitsstruktur auf eine Zahl abzubilden (vgl. [Hartung 1999], S. 442). Trotz intensiver Forschung auf diesem Gebiet ist es jedoch bis heute nicht gelungen ein "allgemeines" Maß zu identifizieren, welches die Interessantheit einer Regel am Besten abbildet, so dass in der Literatur eine Vielzahl unterschiedlicher Interessantheitsmaße zu finden sind (siehe zum Beispiel [Tan et al. 2002], [Hipp 2004]).

Prinzipiell lassen sich Interessantheitsmaße in zwei unterschiedliche Kategorien einteilen: In statistische und in subjektive Interessantheitsmaße. Bei *statistischen* Interessantheitsmaßen werden aufgrund der Werte in der Kontingenztabelle, Maße für die Signifikanz einer Regel erstellt. Liegt das Interessantheitsmaß über einer vorgegebenen Schwelle, so wird die Regel als signifikant gewertet. Eine weitere Art, Regeln zu bewerten ist die Gewichtung der Interessantheit nach subjektiven Gesichtspunkten. So kann beispielsweise *externe Information*<sup>8</sup> zur Bildung der Regel benutzt wird. Dabei steht die Bedeutung der Regel für die Anwendung bzw. dem Anwender im Vordergrund. Diese muss nicht immer der statistischen Relevanz entsprechen muss. Somit kann es zu einem *Zielkonflikt* zwischen einer statistischen und einer anwendungsspezifischen Betrachtung von Zusammenhängen kommen.

In den nachfolgenden Abschnitten werden für beide Kategorien gängige Maße exemplarisch vorgestellt und jeweils um ein neues Maß ergänzt, welche sich in der praktischen Anwendung als besonders geeignet erwiesen haben.

### 4.2.2 Statistische Interessantheitsmaße

Ein bekanntes Verfahren zur Bestimmung der Abhängigkeit zweier Variablen ist die Korrelation. Je nach Wert gibt sie an, ob zwei Variablen positiv oder negativ korreliert sind. Liegt eine konkrete Realisierung vor, so ist die Berechnung des Korrelationskoeffizienten zwischen Prämisse und Konklusion einer Regel mit Hilfe der  $2 \times 2$  Kontingenztabelle möglich (vgl. Tabelle 4.2). Der Koeffizient ist definiert als

$$\varphi = \frac{p_{11}p_{22} - p_{12}p_{21}}{\sqrt{p_{1\cdot} \cdot p_{\cdot 1} p_{2\cdot} \cdot p_{\cdot 2}}} \quad (4.10)$$

wobei gilt:  $\varphi \in [-1, 1]$ .

<sup>8</sup>Information die nicht aus den Zählhäufigkeiten von Prämisse und Konklusion ermittelt werden können wie beispielsweise Kosten

Ein Maß, welches in der deskriptiven Statistik weite Verbreitung findet, ist die so genannte *odds-ratio* – oder auch Quotenverhältnis. Es basiert auf dem Quotient folgender beider *Odds*<sup>9</sup> einer Realisierung:

- Die Odds für das Eintreten der Konklusion unter denjenigen Transaktionen, für welche die Prämisse erfüllt ist:

$$\text{odds}(B|A) = \frac{\frac{p_{11}}{p_{1\cdot}}}{1 - \frac{p_{11}}{p_{1\cdot}}} = \frac{p_{11}}{p_{12}}$$

- Die Odds für das Eintreten der Konklusion unter den Transaktionen, für welche die Prämisse nicht erfüllt ist:

$$\text{odds}(B|\neg A) = \frac{\frac{p_{21}}{p_{2\cdot}}}{1 - \frac{p_{21}}{p_{2\cdot}}} = \frac{p_{21}}{p_{22}}$$

Folglich ergibt sich die *odds-ratio*  $\theta$  zu:

$$\theta = \frac{p_{11}p_{22}}{p_{21}p_{12}}$$

Die Odds-ratio<sup>10</sup> ist ein Maß dafür, um wie viel größer die Chance ist, eine Transaktion zu erhalten für welche die Konklusion und die Prämisse erfüllt sind, gegenüber derjenigen Chance Transaktionen zu erhalten, welche die Konklusion jedoch nicht die Prämisse enthalten. Dabei nimmt die odds-ratio Werte von 0 bis unendlich an<sup>11</sup>. Ein Wert von 1 bedeutet ein gleiches Quotenverhältnis und lässt sich als Unabhängigkeit von Prämisse und Konklusion interpretieren. Dabei gilt, je stärker  $\theta$  von 1 abweicht, um so stärker weicht der Zusammenhang der Variablen von der Unabhängigkeit ab. In manchen praktischen Fällen, in denen alle Zelleinträge von 0 verschieden sind ist es vorteilhafter  $\log(\theta)$  zu betrachten. In diesem Fall ist  $\log(\theta) = 0$  wenn die Variablen unabhängig sind.

Eine Erweiterung der odds-ratio bildet Yule's  $Q$

$$Q = \frac{p_{11}p_{22} - p_{12}p_{21}}{p_{11}p_{22} + p_{12}p_{21}} = \frac{\theta - 1}{\theta + 1}$$

<sup>9</sup>Odds ist der Quotient aus der Wahrscheinlichkeit, dass ein Ereignis eintritt und der Komplementärwahrscheinlichkeit:  $\text{odds}(A) = \frac{P(A)}{1-P(A)}$

<sup>10</sup>Alternativ wird  $\theta$  auch *cross-product-ratio* genannt, da es dem Verhältnis des Produktes der schräg gegenüber liegenden Zellen  $p_{11}p_{22}$  und  $p_{12}p_{21}$  entspricht

<sup>11</sup> $\theta = \infty$  falls der Nenner gleich 0 ist

$Q$  ist eine Transformation der odds-ratio auf das Intervall  $[-1, 1]$  und ähnelt somit stark dem Korrelationskoeffizient (4.10).

Trotz der Vielzahl unterschiedlicherer Maße um die statistische Signifikanz einer Assoziationsregel zu bewerten, hat sich unter den Data Mining Anwendern der Lift, als das bekannteste Maß etabliert. Formal ist der Lift definiert durch:

$$\text{lift}_{\mathcal{T}}(A \rightarrow B) = \frac{P(B|A)}{P(B)} = \frac{P(A \cup B)}{P(A)P(B)} \quad (4.11)$$

$$= \frac{P(\langle A \cup B \rangle)}{P(\langle A \rangle)P(\langle B \rangle)} = \frac{\text{conf}_{\mathcal{T}}}{P(\langle B \rangle)} \quad (4.12)$$

Ähnlich zur Odds-Ratio gibt der Lift an, um wie viel häufiger die Konklusion in den Transaktionen, welche die Prämisse enthalten, vertreten ist, gegenüber der Häufigkeit der Konklusion in der Grundgesamtheit. Gilt  $\text{lift}_{\mathcal{T}}(A \rightarrow B) > 1$  so kann von einer positiven Korrelation des Regelkopfes und des Körpers gesprochen werden. Ist  $\text{lift}_{\mathcal{T}}(A \rightarrow B) = 1$ , so sind diese unabhängig. Im Falle von  $\text{lift}_{\mathcal{T}}(A \rightarrow B) < 1$ , sind Prämisse und Konklusion negativ korreliert bzw. es gilt: wenn eine Transaktion  $A$  enthält, dann sinkt die Wahrscheinlichkeit, dass sie auch  $B$  enthält. Der Lift misst somit die Abweichung von der Unabhängigkeit zweier Variablen. Analog zu Konfidenz und Support ist

$$\text{lift}_{\mathcal{T}}(A \rightarrow B) = \frac{\frac{n_{A \wedge B}}{n}}{\frac{n_A}{n} \frac{n_B}{n}} = \frac{n_{A \wedge B} n}{n_A n_B}$$

ein Schätzer für den Lift.

Sei  $Z_{A \rightarrow B} : \mathcal{T} \rightarrow \{0, 1\}$  eine Zufallsvariable mit

$$Z_{A \rightarrow B}(T) = \begin{cases} 1 & \text{falls } A \cup B \subseteq T \\ 0 & \text{falls } A \cup B \not\subseteq T \end{cases}$$

Dann ist

$$P(Z_{A \rightarrow B} = 1) = P(\langle A \cup B \rangle) = \pi_{11}$$

mit  $Z_{A \rightarrow B} \sim B(1, p = \pi_{11})$ . Seien  $Y_1, \dots, Y_n$  iid wie  $Z_{A \rightarrow B}$ , dann ist  $\sum_{i=1}^n Y_i \sim B(n, \pi_{11})$  verteilt und somit gilt für den Support einer Assoziationsregel bezüglich einer Transaktionsdatenbank  $\mathcal{T} = (T_1, \dots, T_n)$ :

$$n \cdot \text{supp}_{\mathcal{T}}(A \rightarrow B) = \sum_{i=1}^n Y_i \sim B(n, \pi_{11}) \quad (4.13)$$

Folglich lassen sich Maße für die Interessantheit von Assoziationsregeln wie beispielsweise der Support, auf binomial-verteilte Zufallsvariablen zurückführen. Liegt eine konkrete

Stichprobe in Form einer Transaktionsdatenbank vor, so können die eingeführten Maße als Punktschätzer interpretiert werden. Punktschätzer auf Basis einer Stichprobe sind jedoch mit Unsicherheit behaftet, so dass eine Aussage bezüglich des Wertes eines Interessantheitsmaßes mit Hilfe von Konfidenzintervallen abgesichert werden sollte.

Sei  $Z$  eine Bernoulli-verteilte Zufallsvariable. Dann gilt nach dem Zentralen Grenzwertsatz von Lindenberg/Levy (vgl. [Vogel 2005], S. 198) für eine Folge  $(Z_i)_{i \in \mathbb{N}}$  von unabhängigen, identisch wie  $Z$  verteilten Zufallsvariablen, dass

$$\bar{Z}_n = \frac{1}{n} \sum_{k=1}^n Z_k \sim N\left(p, \frac{p(1-p)}{n}\right) \quad (4.14)$$

näherungsweise normalverteilt ist mit Mittelwert  $p = P(Z = 1)$ , Varianz  $\frac{p(1-p)}{n}$  und damit ist

$$\frac{\bar{Z}_n - p}{\sqrt{p(1-p)/n}} \quad (4.15)$$

näherungsweise standardnormalverteilt.

Zur Konstruktion eines Konfidenzintervalls für  $p$  zum Konfidenzniveau  $(1 - \alpha)$  ist  $z_{1-\frac{\alpha}{2}}$  so zu wählen, dass gilt:

$$P\left(-z_{1-\frac{\alpha}{2}} \leq \frac{\bar{Z}_n - p}{\sqrt{p(1-p)}} \sqrt{n} \leq z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha \quad (4.16)$$

Mit (4.15) ist (4.16) äquivalent zu

$$\Phi(z_{1-\frac{\alpha}{2}}) - \Phi(-z_{1-\frac{\alpha}{2}}) = 1 - \alpha$$

bzw.

$$\Phi(z_{1-\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$$

Wird die sich aus (4.16) ergebende Ungleichung

$$|\bar{Z}_n - p| \sqrt{n} \leq z_{1-\frac{\alpha}{2}} \sqrt{p(1-p)}$$

nach  $p$  aufgelöst, so erhält ein erstes Konfidenzintervall:

$$C_{l,u}^{Wilson} = \frac{1}{2(z_{1-\frac{\alpha}{2}} + n)} \left( 2n\bar{Z}_n + z_{1-\frac{\alpha}{2}}^2 \pm 2nz_{1-\frac{\alpha}{2}} \sqrt{\frac{z_{1-\frac{\alpha}{2}}}{4n^2} + \frac{\bar{Z}_n(1-\bar{Z}_n)}{n}} \right) \quad (4.17)$$

In der Literatur wird (4.17) auch als das Wilson Konfidenzintervall (vgl. [Vogel 2005], S. 214) bezeichnet.

Bei der Analyse von Assoziationsregeln ist man im Allgemeinen an einer konservativen Abschätzung der Interessantheit einer Regel interessiert. Ziel ist es somit, eine Aussage über den Zusammenhang zu treffen, die mit möglichst “wenig“ Unsicherheit behaftet ist. Eine Möglichkeit, die vorhandene Unsicherheit zu reduzieren, ist die linke Grenze des Konfidenzintervalls eines Interessantheitsmaßes zu verwenden. Dies ermöglicht eine statistisch abgesicherte Aussage sowohl über die Stärke als auch die Signifikanz einer Abhängigkeit. Für das in (4.17) hergeleitete Konfidenzintervall ergibt sich die untere Intervallgrenze zu

$$C_l^{Wilson} = \frac{n}{n + z_{1-\frac{\alpha}{2}}^2} \left( \bar{Z}_n + \frac{z_{1-\frac{\alpha}{2}}^2}{2n} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\bar{Z}_n(1 - \bar{Z}_n)}{n} + \frac{z_{1-\frac{\alpha}{2}}^2}{4n^2}} \right) \quad (4.18)$$

Für den Support einer Assoziationsregel  $A \rightarrow B$  gilt nach (4.13) für den Schätzwert  $n \cdot \text{supp}_T(A \rightarrow B)$ :

$$n \text{supp}_T(A \rightarrow B) \sim B(n, \pi_{11})$$

wobei  $\pi_{11}$  der “wahre Wert“ des Supports der Assoziationsregel  $A \rightarrow B$  ist.

Sei

$$\hat{\pi}_{11} = \bar{Z}_n = \frac{1}{n} \sum_{k=1}^n Z_k = \frac{n_{A \wedge B}}{n}$$

ein Schätzer für den unbekanntem Parameter  $\pi_{11}$ . Eingesetzt in (4.18) ergibt sich eine linke Konfidenzschranke für den Support einer Regel zu

$$C_l^{Supp} = \frac{n}{n + z_{1-\frac{\alpha}{2}}^2} \left( \frac{n_{A \wedge B}}{n} + \frac{z_{1-\frac{\alpha}{2}}^2}{2n} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\frac{n_{A \wedge B}}{n} (1 - \frac{n_{A \wedge B}}{n})}{n} + \frac{z_{1-\frac{\alpha}{2}}^2}{4n^2}} \right) \quad (4.19)$$

Ähnlich lässt sich eine linke Konfidenzschranke für die Konfidenz einer Regel  $A \rightarrow B$  berechnen. Schränkt man die Grundgesamtheit auf diejenigen Transaktionen  $\mathcal{T}^A \subseteq \mathcal{T}$  ein, welche die Regelprämisse enthalten, dann lässt sich wiederum eine binäre Zufallsvariable  $Z : \mathcal{T}^A \rightarrow \{0, 1\}$  definieren mit

$$Z(T) = \begin{cases} 1 & \text{falls } A \cup B \subseteq T \\ 0 & \text{falls } A \cup B \not\subseteq T \end{cases}$$

Damit gilt, analog zum Support:

$$n \cdot \text{conf}_T(A \rightarrow B) \sim B(n_A, p)$$

wobei jetzt  $n_A$  von der Stichprobe abhängt, also zufällig ist.

Mit dem Schätzer  $\hat{p} = n_{A \wedge B} / n_A$  für den unbekannt Parameter  $p$  erhält man folgende, linke Konfidenzschranke für die Konfidenz einer Assoziationsregel:

$$C_l^{Konf} = \frac{n_A}{n_A + z_{1-\frac{\alpha}{2}}^2} \left( \frac{n_{A \wedge B}}{n_A} + \frac{z_{1-\frac{\alpha}{2}}^2}{2n_A} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\frac{n_{A \wedge B}}{n_A} (1 - \frac{n_{A \wedge B}}{n_A})}{n_A} + \frac{z_{1-\frac{\alpha}{2}}^2}{4n_A^2}} \right) \quad (4.20)$$

Vergleicht man die Konfidenzintervalle für den Support und die Konfidenz, so stellt man fest, dass im ersten Fall das Intervall schmaler ist als im zweiten, da stets  $n_A \leq n$  ist. Diese Tatsache wird bei der Berechnung eines Konfidenzintervalls für den Lift verwendet. Wie in (4.11) eingeführt, ist der Lift der Quotient aus  $P(B|A)$  und  $P(A)$ . Um nun nicht zwei Konfidenzintervalle berechnen zu müssen, wird lediglich die linke Intervallgrenze für die bedingte Wahrscheinlichkeit bestimmt. Somit erhält man wie in (4.20) ein Konfidenzschranke, welche durch die Wahrscheinlichkeit der Konklusion dividiert wird:

$$\begin{aligned} C_l^{Lift} &= \frac{C_l^{Konf}}{n_B/n} \\ &= \frac{n_A n}{n_A n_B + z_{1-\frac{\alpha}{2}}^2 n_B} \left( \frac{n_{A \wedge B}}{n_A} + \frac{z_{1-\frac{\alpha}{2}}^2}{2n_A} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\frac{n_{A \wedge B}}{n_A} (1 - \frac{n_{A \wedge B}}{n_A})}{n_A} + \frac{z_{1-\frac{\alpha}{2}}^2}{4n_A^2}} \right) \end{aligned} \quad (4.21)$$

In der Praxis kann man von einem genügend großen Wert für die Häufigkeit der Regelprämisse  $n_A$  ausgehen (vgl. Anwendung Kapitel 4.4), so dass eine vereinfachte Berechnung eines Konfidenzintervalls für den Lift möglich ist. Sei  $n_A \gg 1$  dann gilt:

$$\frac{n_A}{n_A + z_{1-\frac{\alpha}{2}}^2} \approx 1, \quad \frac{z_{1-\frac{\alpha}{2}}^2}{2n_A} \approx 0 \quad \text{und} \quad \frac{z_{1-\frac{\alpha}{2}}}{4n_A^2} \approx 0$$

Damit lässt sich ein neues Interessantheitsmaß, der so genannte *SafeLift* für eine Assoziationsregel der Form  $A \rightarrow B$  definieren:

$$\text{SafeLift}_T(A \rightarrow B) = \frac{n}{n_B} \left( \frac{n_{A \wedge B}}{n_A} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\frac{n_{A \wedge B}}{n_A} (1 - \frac{n_{A \wedge B}}{n_A})}{n_A}} \right) \quad (4.22)$$

Die Vereinfachung des Wilson-Konfidenzintervalls findet man in der Literatur unter dem Namen *Standard-Intervall*<sup>12</sup>, da es sich, analog zu Konfidenzintervallen einer normal-

<sup>12</sup>Eine alternative Bezeichnung ist Wald-Intervall

verteilten Grundgesamtheit, aus dem Schätzer für den unbekanntem Wert und einem Vielfachen der Standardabweichung zusammensetzt (vgl. 2.3.2).

$$C_{l,u}^{Standard} = \bar{Z}_n \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\bar{Z}_n(1-\bar{Z}_n)}{n}}$$

Bei den eben vorgestellten Verfahren zur Konstruktion von Konfidenzintervallen treten im Wesentlichen zwei Probleme auf, die im Verlauf dieses Kapitels noch nicht ausreichend behandelt wurden:

1. Die Approximation der Binomialverteilung durch die Normalverteilung
2. Die Qualität der eben vorgestellten Intervalle

Um ersteres Problem zu “lösen“, wird in den gängigen Statistiklehrbüchern ein einfacher Ansatz mittels Faustregeln verwendet. In [Brown et al. 1999] findet man einen Auszug der wichtigsten Regeln zur Anwendung des Standard-Intervalls<sup>13</sup>:

- $np \geq 5$  bzw. 10 und  $n(1-p) \geq 5$  bzw. 10
- $np(1-p) \geq 5$  bzw. 10
- $n \geq 50$ , wenn  $p$  nicht sehr klein

Die Grundidee der genannten Faustregeln ist, mittels der Bedingungen zu überprüfen, ob die Verwendung des Zentralen Grenzwertsatzes möglich ist und damit eine Normalverteilungsapproximation gerechtfertigt ist oder nicht (vgl. [Brown et al. 1999]). In vielen praktischen Anwendungen der Assoziationsanalyse sind die oben aufgeführten Regeln nicht bzw. nur in wenige Fälle erfüllt. Vor allem in Anwendungen, mit einer hohen Grundgesamtheit  $n$  jedoch einem seltenen gemeinsamen Auftreten von Regelkonklusion und Prämisse ( $n_{A \wedge B}$ ), sollten Konfidenzintervalle nicht mittels der Standardmethode berechnet werden, da folglich  $p$  sehr klein ist. Um dennoch eine Aussage über die Interessantheit von Zusammenhängen mittels Konfidenzintervallen zu treffen wird im Folgenden die Qualität der betrachteten Konfidenzintervalle untersucht.

Das Messen der Qualität eines Konfidenzintervalls für Binomial-verteilte Zufallsvariablen ist mittels der *Überdeckungswahrscheinlichkeit* (vgl. [Brown et al. 1999]) möglich. Sei  $X \sim B(n, p)$  und  $C = [C_l, C_u]$  ein  $(1-\alpha)$ -Konfidenzintervall des unbekanntem Parameters

<sup>13</sup>Die Angabe in den Faustregeln sind oftmals sehr ungenau und erlauben somit einen großen Spielraum bezüglich der Anwendbarkeit der Regeln

$p$ . Dann ist es das Ziel, ein Konfidenzintervall für  $p$  zu konstruieren, so dass für die Überdeckungswahrscheinlichkeit (coverage probability) gilt:

$$P(p \in C) \approx 1 - \alpha$$

Dabei gelten folgende Annahmen:

- $C_l(X_1) < C_l(X_2)$  und  $C_u(X_1) < C_u(X_2)$  für  $X_1 < X_2$  mit  $X_1, X_2 \in \{0, \dots, n\}$
- $\forall p \in (0, 1) \exists X_0$  so dass  $p \in (C_l(X_0), C_u(X_0))$

Ein Gütekriterium für den unbekannt Parameter  $p$  einer Binomialverteilung lässt sich dann wie folgt definieren:

**Definition 4.3** (Überdeckungswahrscheinlichkeit)

Sei  $C_l(X), C_u(X)$  wie oben eingeführt. Des Weiteren sei

$$k_0(p) = \min_{X=0, \dots, n} \{X : p < C_u(X)\}$$

und

$$k_1(p) = \max_{X=0, \dots, n} \{X : C_l(X) < p\}$$

Dann ist die Überdeckungswahrscheinlichkeit  $CP(n, p)$  eines Konfidenzintervalls für den Parameter  $p$  definiert als:

$$\begin{aligned} CP(n, p) &= P(p \in (C_l, C_u)) \\ &= P(C_l(X) < p < C_u(X)) \\ &= P(k_0(p) \leq X \leq k_1(p)) \\ &= \sum_{i=k_0(p)}^{k_1(p)} \binom{n}{i} p^i (1-p)^{n-i} \end{aligned} \tag{4.23}$$

Betrachtet man die Überdeckungswahrscheinlichkeit der Konfidenzintervalle in Abbildung 4.1, so stellt man zunächst fest, dass beide ein überraschendes Verhalten aufweisen. Fixiert man einen Wert für  $p$ , so oszilliert die Überdeckungswahrscheinlichkeit selbst bei einem steigenden Stichprobenumfang  $n$ . Es existieren somit Tupel  $(n, p)$  für welche  $CP(n, p)$  über dem geforderten Niveau von  $(1 - \alpha)$  liegen. Andererseits gibt es Tupel, für die genau das Gegenteil der Fall ist: Die entsprechenden Werte für  $CP(n, p)$  liegen unterhalb des geforderten Niveaus (siehe [Brown et al. 1999] “lucky“ bzw. “unlucky pairs“). Selbst für *sehr große* Stichprobenumfänge ( $n \gg 1000$ ) ist dieses Verhalten noch zu bemerken. Zwar nimmt die Oszillationsstärke ab, ist jedoch erst bei sehr hohen Werten für  $n$  vernachlässigbar. Des Weiteren ist zu beobachten, dass das Standard-Intervall

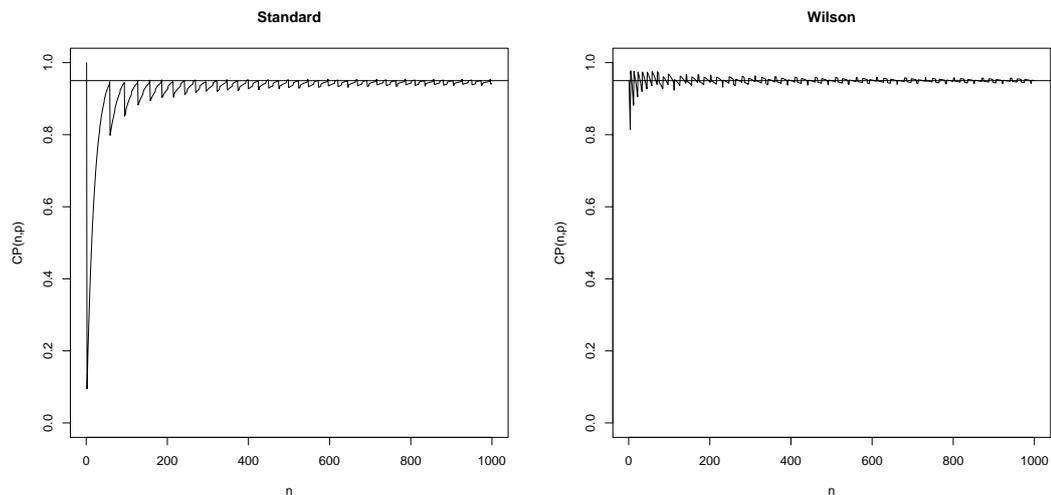


Abbildung 4.1: Oszillationsphänomen des Standard- und Wilsonintervalls zum Niveau  $\alpha = 0.05$  für  $p = 0.05$  und  $n = 0 \dots 1000$

zum größten Teil unter dem geforderten Konfidenzniveau liegt und eine stärkere Oszillation aufweist als das Wilson-Intervall. Für beliebige Kombinationen von  $p$  und  $n$  sind die beiden Intervalle nur bedingt geeignet. Aus diesem Grund sind in den vergangenen Jahrzehnten eine Vielzahl weiterer Verfahren zur Konstruktion von Konfidenzintervallen entwickelt worden. Jedoch ist es trotz intensiver Forschung auf diesem Gebiet bis heute nicht gelungen ein Konfidenzintervall zu identifizieren, welches von allen Seiten als am geeignetsten betrachtet wird (vgl. [Wang 2006]).

Ende der neunziger Jahre wurde von den Autoren L.D. Brown, T. Cai und A. DasGupta (im weiteren Verlauf mit BCD abgekürzt) in [Brown et al. 1999] ein Versuch unternommen, die bekannten Konfidenzintervalle für eine binomialverteilte Grundgesamtheit zu ordnen und die geeignetsten Intervalle in Abhängigkeit von der Stichprobengröße zu benennen. Dazu wurden die Intervalle in zwei Klassen eingeteilt:

1. Exakte Intervalle
2. Intervalle basierend auf einer Normalverteilungsapproximation

In ihrer Studie kamen die Autoren zu dem Ergebnis, dass es drei Konfidenzintervalle gibt, welche im Hinblick auf Berechenbarkeit und Genauigkeit besser sind als die anderen, in ihrer Arbeit vorgestellten Intervalle. Im Folgenden werden diese “besten“ Konfidenzintervalle eingeführt und anschließend ihr Verhalten für die praktischen Gegebenheiten einer Assoziationsanalyse untersucht.

Das bekannteste Verfahren zur Konstruktion von exakten Konfidenzintervallen ist das *Clopper-Pearson*-Intervall. Es basiert auf der Invertierung eines Binomialtests mit der Nullhypothese  $H_0 : \theta = \theta_0$  gegen die Alternativhypothese  $H_1 : \theta \neq \theta_0$ . Damit ist das Clopper-Pearson-Konfidenzintervall zum Niveau  $(1 - \alpha)$  definiert als die Lösung der zwei Gleichungen

$$P_\theta(X \geq x) = \frac{\alpha}{2} \text{ und } P_\theta(X \leq x) = \frac{\alpha}{2} \quad (4.24)$$

Unter Berücksichtigung der Tatsache, dass für  $X \sim B(n, \theta)$  folgender Zusammenhang gilt (siehe z.B. [Casella und Berger 2002], Kapitel 2.5):  $P_\theta(X \geq x) = P(Y \leq \theta)$  mit  $Y \sim \text{Beta}(X, n - X + 1)$ , lässt sich zeigen, dass (4.24) äquivalent ist zu

$$C_{l,u}^{CP} = [\text{Beta}_{\frac{\alpha}{2}; X, n-X+1}, \text{Beta}_{1-\frac{\alpha}{2}; X+1, n-X}] \quad (4.25)$$

Dabei ist  $\text{Beta}(\beta_1, \beta_2)$  eine Beta-Verteilung mit den Parametern  $\beta_1, \beta_2$  und  $\text{Beta}_{\alpha; \beta_1, \beta_2}$  die dazu gehörige Quantilfunktion.

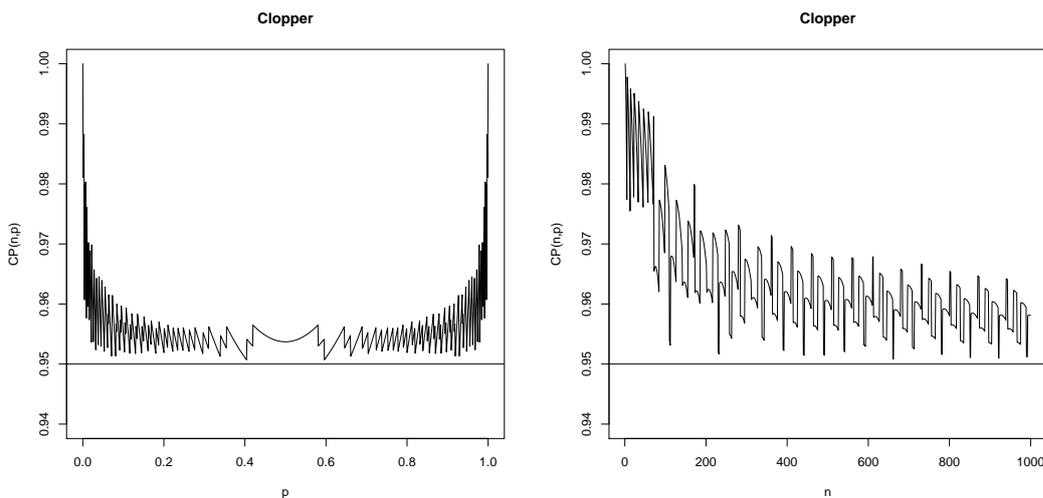


Abbildung 4.2:  $CP(n, p)$  zum Niveau  $\alpha = 0.05$  für Clopper-Pearson-Intervall mit  $n = 1000$  (links) und  $p = 0.05$  (rechts)

Die positive Eigenschaft des *Clopper-Pearson* Intervalls ist, dass es stets eine Überdeckungswahrscheinlichkeit liefert, die mindestens gleich groß ist wie das vorgegebene Signifikanzniveau (siehe Abbildung 4.2). Mit anderen Worten:  $CP_\alpha(n, p) \geq (1 - \alpha)$ . Jedoch kann dabei für ein festes  $p$  die Überdeckungswahrscheinlichkeit wesentlich höher sein als das geforderte Niveau, so dass dieses Verfahren als zu konservativ eingestuft wird. Ein weiterer Nachteil ist die aufwändige Bestimmung der Quantile der Beta-Verteilung.

Ein Konfidenzintervall, welches eine Zwischenstellung zwischen exakten und approximierten Vertrauensintervallen einnimmt, ist das *Jeffrey's* Intervall. In der Bayes-Inferenz-Statistik hat ein unbekannter Parameter  $\theta$  mit A-priori-Verteilung  $f(\theta)$  eine A-Posteriori-Verteilung mit Dichte

$$f(\theta|x) \propto f(x|\theta) \cdot f(\theta) \quad (4.26)$$

wobei  $f(x|\theta)$  die Stichprobenverteilung bezeichnet.

**Definition 4.4** (Konjugierte Verteilung)

Sei  $\mathcal{F} = \{f(x|\theta), \theta \in \Theta\}$  eine Familie von Wahrscheinlichkeiten bzw. Dichtefunktionen der Daten  $X = x$ . Man nennt nun eine Klasse  $\mathcal{G}$  von A-priori-Verteilungen konjugiert bezüglich  $\mathcal{F}$ , falls für alle A-priori-Verteilungen  $f(\theta) \in \mathcal{G}$  gilt:  $f(\theta|x) \in \mathcal{G}$

Sei nun  $X \sim B(n, p)$  und sei  $Beta(\beta_1, \beta_2)$  die dazugehörige konjugierte A-priori-Verteilung mit  $\beta_1, \beta_2 > 0$ . Mit

$$f(x|\theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}$$

und

$$f(\theta) = \frac{1}{Beta(\beta_1, \beta_2)} \theta^{\beta_1-1} (1 - \theta)^{\beta_2-1}$$

ist nach (4.26) die A-posteriori-Verteilung gegeben durch

$$\begin{aligned} f(\theta|x) &\propto \binom{n}{x} \theta^x (1 - \theta)^{n-x} \frac{1}{Beta(\beta_1, \beta_2)} \theta^{\beta_1-1} (1 - \theta)^{\beta_2-1} \\ &\propto \theta^{x+\beta_1-1} (1 - \theta)^{n-x+\beta_2-1} \end{aligned}$$

und folglich gilt:  $\theta|x \sim Beta(x + \beta_1, n - x + \beta_2)$ .

Vergleicht man den Bayes'schen Ansatz mit der klassischen, frequentistischen Inferenzstatistik, so wird im ersteren Fall keine Asymptotik verwendet, um ein Konfidenzintervall zu berechnen. Hingegen benötigt man in der Bayes'schen Statistik eine A-priori-Verteilungsannahme für den unbekannt Parameter  $\theta$ . Um möglichst wenig Vorwissen in die A-priori-Verteilung einfließen zu lassen, wird das so genannte *Jeffrey's A-Priori* verwendet. Im Falle eines Binomialexperimentes wie im vorliegenden Fall, ergibt sich für das Jeffrey's A-Priori eine Betaverteilung mit  $\beta_1 = \beta_2 = 0.5$  (siehe z.B. [Held 2007], Kapitel 4.2). Somit lässt sich das *Jeffrey's Konfidenzintervall* zum Niveau  $(1 - \alpha)$  mittels der Quantilfunktion der Betaverteilung wie folgt definieren:

$$C_{u,l}^{Jeffrey} = [Beta_{\frac{\alpha}{2}; x+1/2, n-x+1/2}, Beta_{1-\frac{\alpha}{2}; x+1/2, n-x+1/2}] \quad (4.27)$$

Trotz der engen Verwandtschaft mit dem Clopper-Person-Intervall hat das Jeffrey's Vertrauensintervall keine so hohe Überdeckungswahrscheinlichkeit – es ist nicht zu konservativ (vgl. Abbildung 4.3). Wie jedoch auch im Falle des Wilson- bzw. Standard-Intervalls oszilliert  $CP(n, p)$  um das vorgegebene Signifikanzniveau.

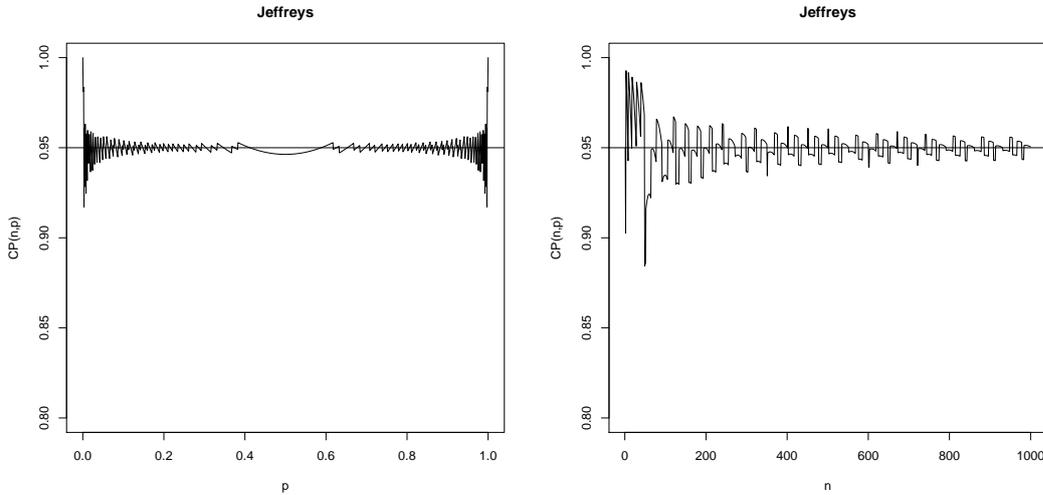


Abbildung 4.3:  $CP(n, p)$  zum Niveau  $\alpha = 0.05$  für Jeffrey's-Intervall mit  $n = 1000$  (links) und  $p = 0.05$  (rechts)

Ein nicht exaktes Konfidenzintervall, welches wie das Wilsonintervall eine Normalverteilungsapproximation verwendet, ist ein von Agresti und Coull eingeführtes Intervall. Basierend auf der Idee des Standard-Intervalls zum Niveau  $\alpha = 0.05$  wird die zugrunde liegenden Grundgesamtheit so verändert, dass zu den tatsächlich beobachteten Fällen 4 Pseudo-Beobachtungen hinzu gefügt werden, mit 2 *Erfolgen* und 2 *Misserfolgen*. In diesem Fall erhält man  $\tilde{n} = n + 4$  und  $\tilde{p} = (X + 2)/\tilde{n}$ . Da  $z_{1-\alpha/2} = z_{0.975} = 1.959964 \approx 2$ , entsprechen die addierten Erfolge dem entsprechenden  $(1 - \alpha/2)$ -Quantil der Normalverteilung. Dieses Prinzip lässt sich auf beliebige Konfidenzniveaus mit  $\tilde{n} = n + z_{1-\alpha/2}^2$  und  $\tilde{X} = X + z_{1-\alpha/2}^2/2$  übertragen, so dass das *Agresti-Coull*-Intervall zum Niveau  $(1 - \alpha)$  mit  $\tilde{p} = \tilde{X}/\tilde{n}$  wie folgt definiert ist:

$$C_{l,u}^{AC} = \tilde{p} \pm z_{1-\alpha/2} \sqrt{\frac{\tilde{p}(1 - \tilde{p})}{\tilde{n}}} \quad (4.28)$$

Die "neue" Wahrscheinlichkeit  $\tilde{p}$  entspricht somit dem gewichteten Wert der ursprünglichen Wahrscheinlichkeit  $p$  und der Wahrscheinlichkeit  $1/2$ , da gilt:

$$\tilde{p} = p \left( \frac{n}{n + z_{1-\alpha/2}^2} \right) + \frac{1}{2} \left( \frac{z_{1-\alpha/2}^2}{n + z_{1-\alpha/2}^2} \right)$$

Trotz der Approximation durch eine Normalverteilung hat das Agresti-Coull-Intervall bessere Eigenschaften als das Standard-Intervall (vgl. Abbildung 4.4). Vergleicht man

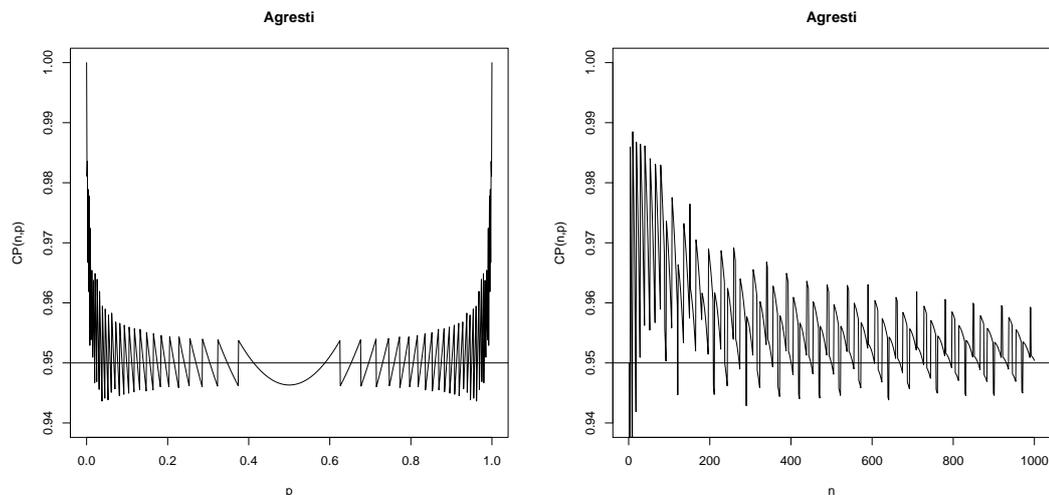


Abbildung 4.4:  $CP(n, p)$  zum Niveau  $\alpha = 0.05$  für Agresti-Coull-Intervall mit  $n = 1000$  (links) und  $p = 0.05$  (rechts)

die beiden Überdeckungswahrscheinlichkeiten, so stellt man zunächst fest, dass  $CP(n, p)$  zwar ebenfalls oszilliert, jedoch ist sie in der Nähe des Konfidenzniveaus  $1 - \alpha$ . Darüber hinaus besitzt es ähnlich gute Eigenschaften wie das eingeführte, exakte Vertrauensintervall.

In ihrer Studie haben BCD verschiedene Kriterien zur Messung der Qualität von Konfidenzintervallen verwendet. Neben der bereits erwähnten Überdeckungswahrscheinlichkeit  $CP(n, p)$  wurde auch die erwartete Länge der jeweiligen Konfidenzintervalle verglichen: Ein Konfidenzintervall für eine binomialverteilte Zufallsvariable hat die erwartete Länge

$$E_{n,p}(\text{Länge}(C)) = \sum_{X=0}^n (C_u(X) - C_l(X)) \binom{n}{X} p^X (1-p)^{n-X}$$

wobei  $C_u$  und  $C_l$  die obere bzw. die untere Intervallgrenze bezeichnen. Zusätzlich wurde die durchschnittliche Überdeckungswahrscheinlichkeit

$$\bar{CP}(n, p) = \int_0^1 CP(n, p) dp$$

untersucht.

In der Zusammenfassung ihrer Untersuchung kamen die Autoren zu dem Ergebnis, dass für geringere Stichprobenumfänge ( $n \leq 40$ ) das Jeffrey's und das Wilson-Konfidenzintervall am geeignetsten sind. Für  $n > 40$  sind sie zu dem Schluss gekommen, dass sowohl das Jeffrey's und Wilson- als auch das Agresti-Coull-Intervall geeignet sind. Dabei wurde das

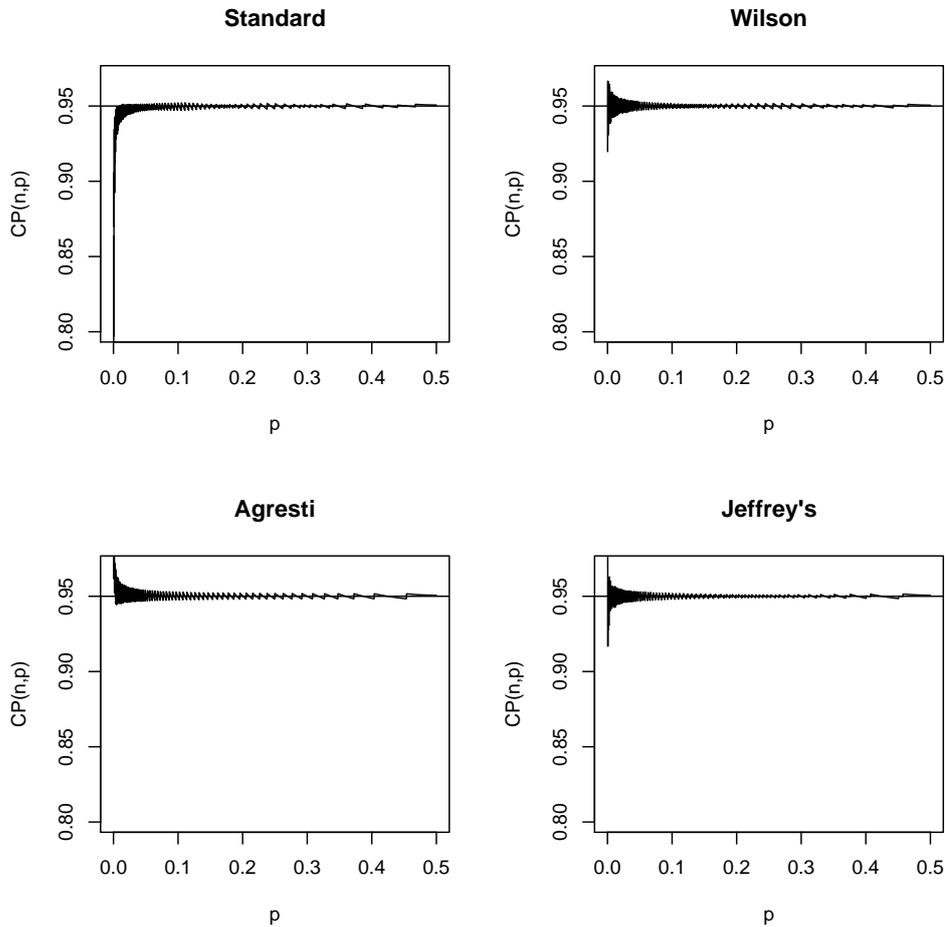


Abbildung 4.5: Überdeckungswahrscheinlichkeit für  $n = 5000$

exakte Clopper-Pearson-Intervall aufgrund seiner hohen Überdeckung  $CP(n, p) > (1 - \alpha)$  nicht weiter beachtet.

Bei ihrer Studie hatten die Autoren sich jedoch lediglich auf Untersuchungen mit einem Stichprobenumfang von weniger als  $n = 200$  gestützt. In vielen praktischen Anwendungen der Assoziationsanalyse ist  $n$  jedoch wesentlich höher<sup>14</sup>, so dass für die eingeführten Intervalle eine Analyse der Überdeckungswahrscheinlichkeit mit einem großen Stichprobenumfang und gleichzeitig sehr kleinem  $p$  ( $p < 0.05$ ) durchgeführt wurde. Als Konfidenzniveau wurde stetes  $\alpha = 0.05$  gewählt. Da in den vorherigen Abschnitten gezeigt wurde, dass das Clopper-Pearson-Intervall zu konservativ ist, werden in der folgenden Analyse lediglich das Standard-, Wilson-, Jeffrey's- und Agresti-Coull-Intervall unter-

<sup>14</sup>Auch die Regelprämisse  $n_A$  ist oftmals größer als 200

sucht.

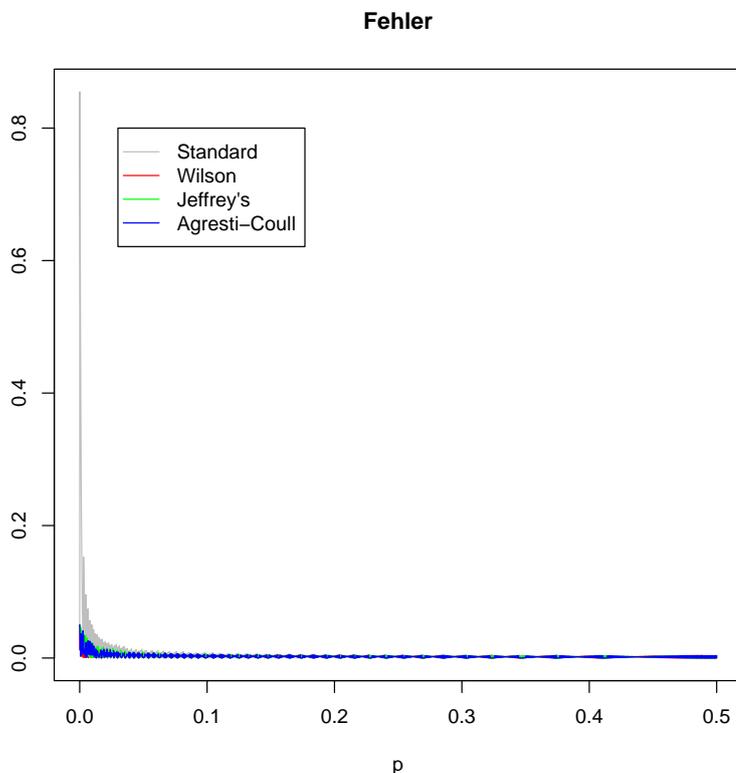


Abbildung 4.6:  $|CP(n, p) - (1 - \alpha)|$  für  $n = 1000$  und  $p \leq 0.05$

In Abbildung 4.5 ist die Überdeckungswahrscheinlichkeit für  $n = 5000$  und  $p \leq 0.05$  aufgezeigt. Dabei ist zu beobachten, dass trotz des hohen Stichprobenumfangs,  $CP(n, p)$  um das nominale Signifikanzniveau oszilliert. Wie erwartet, hat das Standard-Intervall selbst bei hohen Stichprobenumfängen eine geringere Überdeckungswahrscheinlichkeit als die anderen Konfidenzintervalle. Vor allem für  $p < 0.01$  ist die absolute Abweichung der Überdeckungswahrscheinlichkeit vom nominalen Signifikanzniveau verhältnismäßig hoch und für  $p \rightarrow 0$  nicht kontrollierbar (siehe Abbildung 4.6). Für die anderen Intervalle nimmt der Abweichungsfehler  $Err_{\alpha, n}(p) = |CP(n, p) - (1 - \alpha)|$  mit fallendem  $p$  ebenfalls zu, jedoch ist das Verhalten nicht so stark ausgeprägt.

Betrachtet man die durchschnittliche Überdeckungswahrscheinlichkeit  $\bar{CP}(n, p)$ , so kann man anhand der Abbildung 4.7 erkennen, dass für einen steigenden Stichprobenumfang  $\bar{CP}(n, p) \approx (1 - \alpha)$  gilt. Isoliert betrachtet spricht dies für den Einsatz des Standard-Intervalls in Anwendungen mit  $n \gg 1$ . Wie jedoch in [Brown et al. 1999] gezeigt wurde,

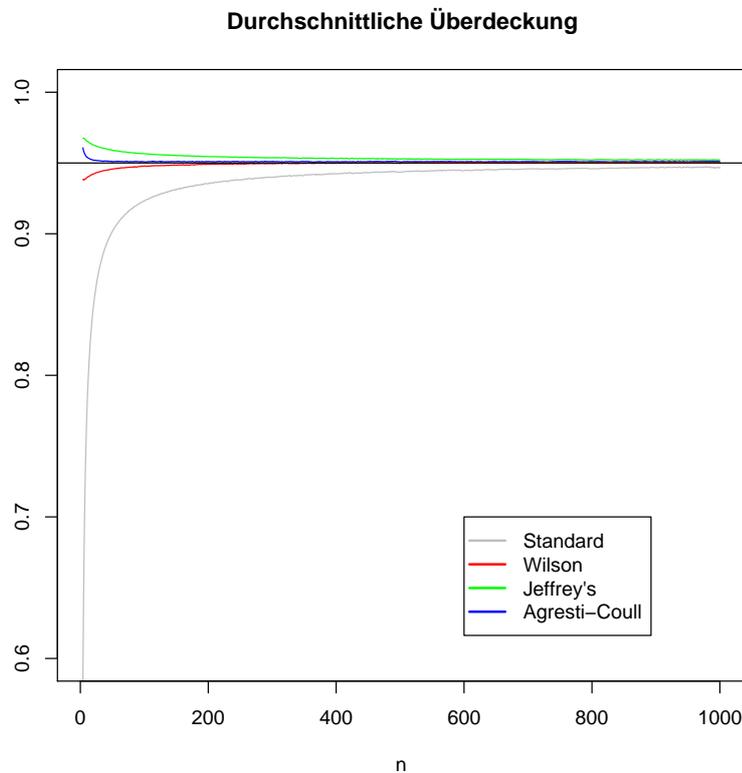


Abbildung 4.7:  $\bar{C}P(n, p)$  für  $n = 0 \dots 1000$

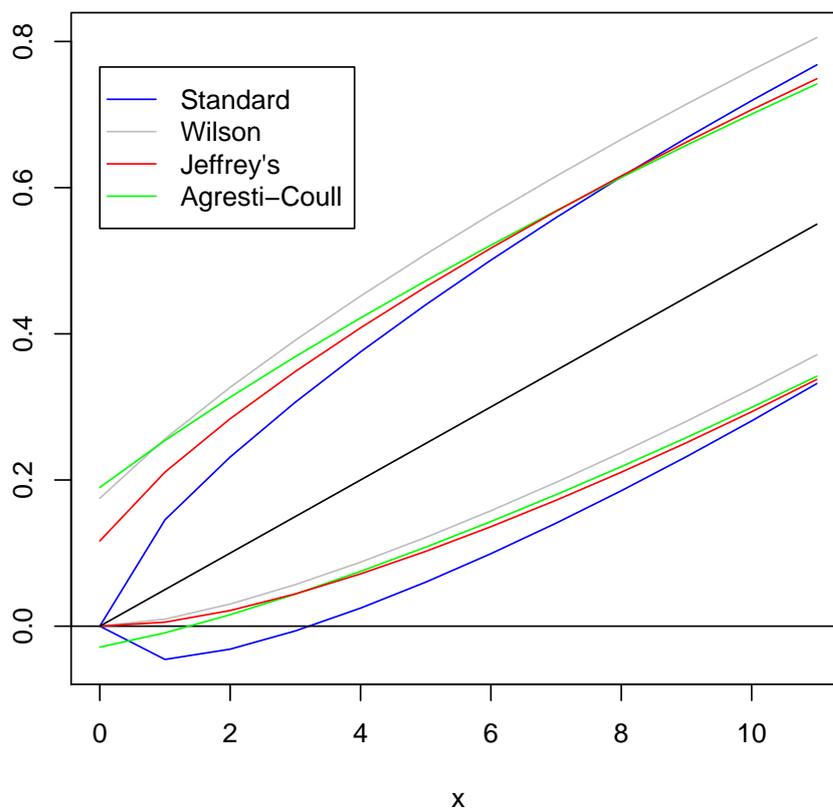
gilt speziell für die Überdeckungswahrscheinlichkeit des Standard-Intervalls:

$$\liminf_{n \rightarrow \infty} C(n, \frac{X}{n}) = 0.860$$

$$\liminf_{n \rightarrow \infty} C(n, \frac{X}{n}) = 0.870$$

$$\liminf_{n \rightarrow \infty} C(n, \frac{X}{n}) = 0.905$$

Somit nähert sich auf der einen Seite die durchschnittliche Überdeckungswahrscheinlichkeit des Standard-Intervalls asymptotisch dem nominalen Signifikanzniveau, auf der anderen Seite gibt es jedoch Fälle, in denen selbst bei großem Stichprobenumfang  $n$  gilt:  $CP(n, p) < (1 - \alpha)$ . Folglich ist es mittels der Standardmethode nur bedingt möglich, ein zuverlässiges Konfidenzintervall für alle Kombinationen von  $n$  und  $p$  zu konstruieren. Dieses Verhalten ist auch in den anderen drei Fällen nachweisbar, jedoch mit geringerem Ausmaß (vgl. [Brown et al. 1999]).

Abbildung 4.8: Konfidenzintervall für  $n = 20$ 

Überträgt man diese Ergebnisse auf das in den vorherigen Abschnitten entwickelte Konzept der Konfidenzintervalle für Interessantheitsmaße, so müssen die verwendeten Intervalle modifiziert werden. Für praktische Anwendungen, in denen eine sehr große Anzahl<sup>15</sup> von Assoziationsregeln bewertet werden muss, das Jeffrey's-Intervall nur eingeschränkt geeignet, da die Quantile einer Beta-Verteilung bestimmt werden müssen. Analysiert man die beiden verbleibenden Konfidenzintervalle – Wilson und Agresti-Coull – so ist zunächst festzuhalten, dass  $CP(n, p)$  für beide einen ähnlich hohen Wert liefert und das Fehlerverhalten  $Err_{\alpha, n}(p)$  vergleichbar ist. Das Agresti-Coull-Intervall hat jedoch den Vorteil einer leichteren Berechenbarkeit gegenüber dem Wilson-Intervall, da es weniger Rechenoperationen benötigt. Auch für kleinere  $n$  liefert das Agresti-Coull-Intervall akzeptable Ergebnisse, so dass im Folgenden das in (4.22) eingeführte Interessantheitsmaß modifiziert wird:

**Definition 4.5** (SafeLift)

Sei die Implikation  $A \rightarrow B$  eine wie in Definition 4.1 eingeführte Assoziationsregel und

<sup>15</sup>In vielen Anwendungen sind oftmals viele tausende Assoziationsregeln zu bewerten (vgl. Kapitel 4.4)

gelte die Notation aus (4.3). Des Weiteren sei

$$\tilde{n}_A = n_A + z_{1-\frac{\alpha}{2}}^2 \quad \text{und} \quad \tilde{n}_{A \wedge B} = n_{A \wedge B} + \frac{z_{1-\frac{\alpha}{2}}^2}{2}$$

mit  $\alpha \in (0, 1)$  und  $z_{1-\frac{\alpha}{2}}$  dem  $(1 - \alpha/2)$ -Quantil der Normalverteilung. Dann ist der SafeLift einer Assoziationsregel gegeben durch:

$$\text{SafeLift}_T(A \rightarrow B) = \frac{n}{n_B} \left( \frac{\tilde{n}_{A \wedge B}}{\tilde{n}_A} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\tilde{n}_{A \wedge B} (1 - \frac{\tilde{n}_{A \wedge B}}{\tilde{n}_A})}{\tilde{n}_A}} \right) \quad (4.29)$$

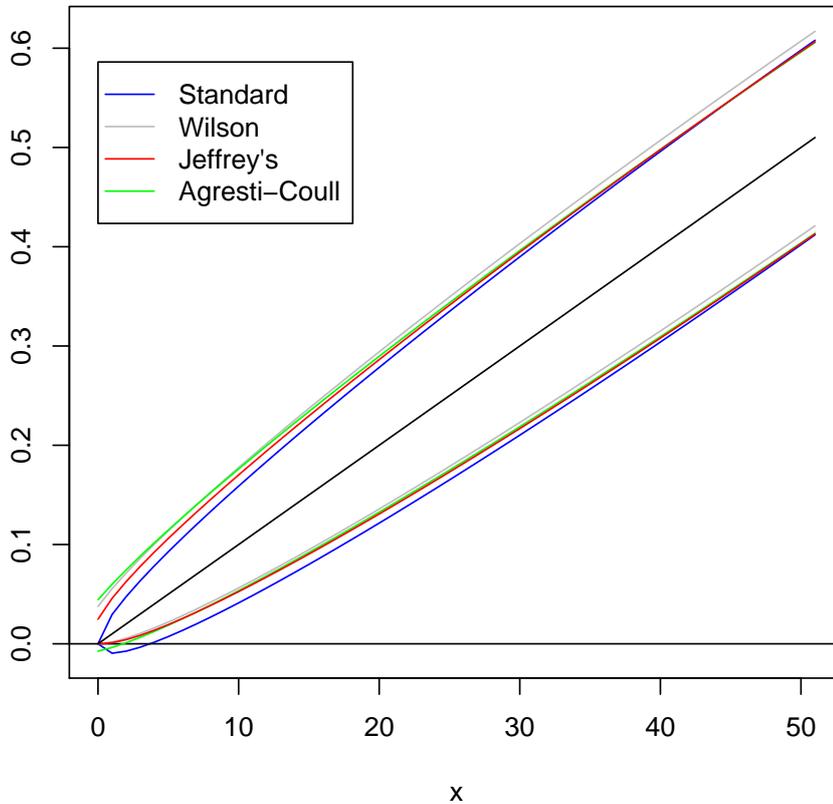


Abbildung 4.9: Konfidenzintervall für  $n = 100$

Trotz der guten Eigenschaften des Agresti-Coull-Intervalls ist für kleine Werte von  $n$  und  $p$  das Intervall ungenau. Wie man in Abbildung 4.8 und 4.9 sehen kann, liefert das Agresti-Coull-Intervall für  $np < 5$  negative Werte. Lediglich das Jeffrey's und das Wilson-Intervall liefert als linke Intervallgrenze einen positiven Wert. Folglich kann die Verwendung des SafeLifts auf Basis von Agresti-Coull in einen negativen Wert resultieren. Da dies jedoch der eigentlichen Definition des Lifts widersprechen würde, ist es

sinnvoll, diesen so zu definieren, dass er minimal einen Wert von 0 annehmen kann. Alternativ kann für solche Werte auch das Wilson-Konfidenzintervall zugrunde gelegt werden oder es ist darauf zu achten, dass Regeln mit  $n_{A \wedge B} < 5$ , nicht beachtet werden.

### 4.2.3 Subjektive Interessantheitsmaße

Die eben eingeführten Interessantheitsmaße messen den Grad des Zusammenhangs zwischen Regelprämisse und Konklusion mittels statistischer Methoden. In vielen praktischen Anwendungen spielen jedoch subjektive Maße, welche die Interessantheit einer Regel aus der Sicht des Anwenders widerspiegeln eine ebenso wichtige Rolle. Dabei gibt es unterschiedliche Aspekte, nach denen ein Anwender die Interessantheit einer Assoziationsregel bewerten kann, die so genannten *Interessantheitsfacetten* (vgl. [Müller et al. 1998]). Abbildung 4.10 stellt die verwendeten Facetten schematisch dar. Im Folgenden wird eine kurze Erläuterung und Zuordnung der Interessantheitsfacetten, wie sie in [Müller et al. 1998] eingeführt wurden, gegeben.

1. Die erste und intuitive Interessantheitsfacette ist nach [Müller et al. 1998] die *Validität* eines Zusammenhangs. Sie gibt an, wie gültig eine Aussage (in dem vorliegenden Kontext eine Regel) in Bezug auf neue Daten ist. Mit anderen Worten, die Validität einer Regel ergibt sich aus den Aspekten *Sicherheit*, *Allgemeingültigkeit* und *Signifikanz*. Dieses Konzept lässt sich am besten durch statistische Interessantheitsmaße, wie sie im vorherigen Abschnitt eingeführt wurden, darstellen.
2. Ein weitere Facette, die bei der Bewertung von Assoziationsregeln relevant ist, wird durch die *Neuheit* einer Regel beschrieben. Dabei kann zunächst als Maß für die Abweichung vom Erwarteten die *Unerwartetheit* betrachtet werden. Des Weiteren kann nach [Müller et al. 1998] gemessen werden, ob eine Aussage das bisherige Wissen über eine Domäne ergänzt oder ob es im Widerspruch dazu steht, spricht der *Unbekanntheitsgrad* einer Assoziationsregel kann bewertet werden. Die Identifikation von ergänzendem Wissen ist jedoch im Allgemeinen nicht trivial, so dass ein weiterer Aspekt bei der Facette *Neuheit* betrachtet werden muss: die Bewertung von Regeln unter Einbeziehung von Hintergrundwissen<sup>16</sup> (*Redundanzfreiheit*). Dabei wird explizit bekanntes Domänenwissen (Expertenwissen) bei der Erzeugung von Regeln berücksichtigt, so dass lediglich *neue* (ergänzende) Informationen als signifikant angesehen werden. Eine ausführliche Einführung und Erläuterung dieser Art von Bewertung von Zusammenhängen ist in [Blumenstock et al. 2007] zu finden.

---

<sup>16</sup>In diesem Zusammenhang spricht man auch von Domänenwissen

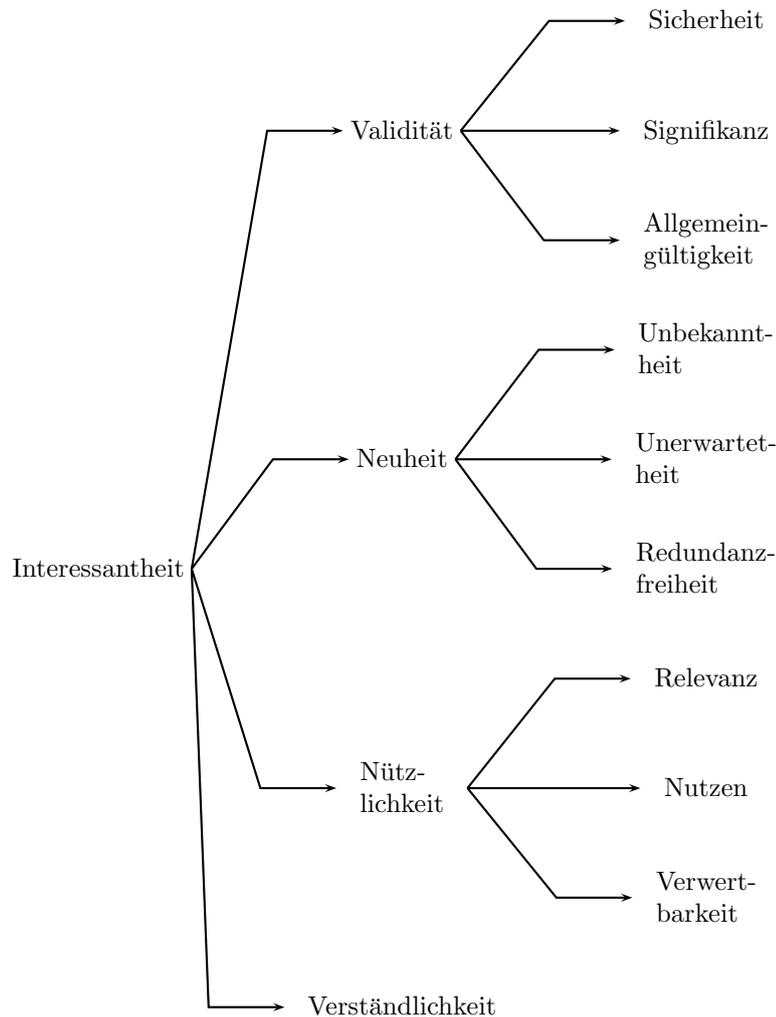


Abbildung 4.10: Interessantheitsfacetten nach [Müller et al. 1998]

3. Eine oftmals unterschätzte Facette bei der Bewertung von Assoziationsregeln ist die *Verständlichkeit* des zugrunde liegenden Interessantheitsmaßes. In vielen praktischen Anwendungen ist es nützlich, dass der Anwender die verwendete Methodik (Motivation) eines Maßes versteht. Denn dies führt zu einer höheren Akzeptanz von Seiten des Anwenders für die gefundenen Zusammenhänge.
4. Sind die bisherigen Interessantheitsfacetten allgemeingültig erklärt und modelliert worden, sprich sie sind unabhängig von der jeweils zugrunde liegenden Anwendung, so ist dies bei der *Nützlichkeit* einer Assoziationsregel nicht möglich. Hier steht die Interaktion mit dem Anwender bzw. der Einfluss externer Information im Vordergrund. Nach [Müller et al. 1998] werden dazu drei verschiedene Subfacetten unterschieden. Die *Relevanz* gibt an, "...inwieweit sich der Bereich, über den etwas

ausgesagt wird, auf den vom Benutzer spezifizierten Fokus bezieht“ ([Müller et al. 1998]). Mittels *Verwertbarkeit* kann bewertet werden, wie oft eine Aussage sich verwenden lässt. Dabei gilt, je öfter eine Regel anwendbar ist, desto “besser“ ist sie. Eng damit verbunden ist der *Nutzen* einer Regel, der sich aus dem Wissen um einen interessanten Zusammenhang ergibt.

Das zentrale Problem, das in [Müller et al. 1998] betrachtet wird, ist die Auswahl der richtigen Regeln anhand der unterschiedlichen Interessantheitsmaße. Dazu haben die Autoren einen Algorithmus entwickelt, der Aussagen (Regeln) anhand der verschiedenen Facetten sortiert. Dies soll dem Anwender einen schnellen Zugriff auf die, für ihn interessanten Zusammenhänge ermöglichen. In praktischen Anwendungen ist es jedoch nicht immer zielführend, dem Anwender die Auswahl der Regel anhand aller Interessantheitsfacetten zu überlassen. Denn oftmals enthält jede Facette einen, für den Anwender interessanten Zusammenhang, so dass das eigentliche Ziel, eine Reduzierung der Information, nicht erreicht wird. Aus diesem Grund hat es sich in der Praxis als sinnvoll erwiesen, neben statistischen Maßen, die eine Aussage über signifikante Zusammenhänge liefern, auch ein subjektives Maß zu verwenden, welches das Ziel der (Gesamt)Analyse berücksichtigt. Dieses ist in den meisten Data-Mining-Anwendungen die Reduzierung von Kosten im weitesten Sinne. Eine intuitive Möglichkeit dies zu tun, ist die monetäre Bewertung des Nutzens einer Regel. Letzteres spielt vor allem bei der Qualitätssicherung eine entscheidende Rolle, da es dort das Ziel ist, mit möglichst wenig Veränderungen (Eingriffen in die Fertigung) die Qualität nachhaltig und umfangreich zu verbessern.

Jede Klassifikations- bzw. Assoziationsregel stellt zunächst einen Zusammenhang zwischen der Menge der Einflussitems  $A$  und der Zielitems  $B$  dar, obgleich man nicht weiß, ob dieser signifikant ist. Eine Teilmenge dieser Assoziationsregeln kann beispielsweise nach einem Hypothesentest wie dem Binomialtest als signifikant eingestuft werden. Legt man das Modell zugrunde, dass die Zielitems (-variablen) Fehler und die Einflussitems mögliche Fehlerursachen darstellen, dann kann die gefundene, signifikante Klassifikationsregel dazu verwendet werden, um die vorhandene Fehlerrate zu reduzieren. Unter idealen Umständen ist es nun möglich, diese Reduzierung monetär zu quantifizieren.

Im Allgemeinen jedoch kann diese Quantifizierung der Einsparung nur teilweise geschehen, so dass der Erfolg (Benefit) einer Regel in finanziellen Dimensionen nicht exakt gemessen werden kann. Eine realistische und intuitive Annahme ist jedoch, dass mittels des Wissens aus einer Assoziationsregel die bedingte Fehlerrate (bedingte Wahrscheinlichkeit) auf die unbedingte Fehlerrate (unbedingte Wahrscheinlichkeit) reduziert werden kann. Denn ist die bedingte Fehlerrate höher als die unbedingte<sup>17</sup> ( $P(B|A) > P(B)$ ), dann müsste, im Zuge der Qualitätssicherung, der zugrunde liegende Fertigungsprozess dahingehend verbessert werden, dass zumindest die durchschnittliche (unbedingte) Feh-

<sup>17</sup>Eine erhöhte bedingte Fehlerrate liegt stets vor, wenn gilt :  $\text{Lift}_T(A \rightarrow B) > 1$

lerrate erreicht wird. In diesem Falle würden sich die Einsparungen durch die Reduzierung der Fehlerrate (bei gleicher Anzahl produzierter Einheiten) wie folgt berechnen:

$$n_A \left( \frac{n_{A \wedge B}}{n_A} - \frac{n_B}{n} \right) c_B = \left( n_{A \wedge B} - \frac{n_A n_B}{n} \right) c_B \quad (4.30)$$

Dabei beschreibt  $c_B$  die nicht immer bestimmbareren Folgekosten des Fehlers  $B$ . Formel (4.30) gibt somit an, wie stark die Reduzierungsmöglichkeit bzw. das **Potential** einer Assoziationsregel ist.

Warum es vor allem in der Qualitätssicherung nicht immer zielführend ist, sich einzig auf die statistischen Interessantheitsmaße zu verlassen, zeigt folgendes Beispiel. Sei in Tabelle 4.4 die Kontingenztabelle zu zwei Assoziationsregeln der Form  $A \rightarrow B$  gegeben.

	$n_{A \wedge B}$	$n_B$	$n_A$	$n$
Beispiel 1	4	43	22	10777
Beispiel 2	218	257	1089	12145

Tabelle 4.4: Beispiel Kontingenztabelle

Für das erste Beispiel erhält man mittels (4.11) einen *Lift* von 45.52 und nach (4.29) einen *SafeLift*<sup>18</sup> von 16.42, jedoch ist das *Potential* dieser Regel lediglich  $3.9c_B$ . Somit ist das ein Beispiel für eine statistisch signifikante Regel, die jedoch kaum Auswirkungen hat, da das Potential sehr gering ist. Mit anderen Worten: würde man nur den Lift bzw. den *SafeLift* als Interessantheitsmaß verwenden und diese Assoziationsregel in der Praxis umsetzen, so würde man die durch den Fehler  $B$  entstehenden Aufwendungen, um lediglich 4 Kosteneinheiten<sup>19</sup> reduzieren, was hinsichtlich der hohen Grundgesamtheit von 10777 eine verschwindend geringe Zahl ist. Betrachtet man stattdessen das zweite Beispiel in Tabelle 4.4, so erhält man einen *Lift* von 9.46 und einen *SafeLift* von 8.37. In diesem Fall sind die statistischen Interessantheitsmaße nicht so hoch wie im ersten Fall, jedoch liefert die Potentialberechnung einen Wert von  $196.96c_B$ . Somit würde in diesem speziellen Fall eine Umsetzung der Regel eine Kostenersparnis von rund 197 Einheiten bringen.

Aufgrund der obigen Ausführungen stellt sich nun die Frage, wie man am geeignetsten die Menge der interessanten Zusammenhänge identifizieren und reduzieren kann. Denn wie gezeigt wurde, gibt es einerseits Fälle, in denen statistische Maße nicht in der Lage sind die Ziele der Analyse vollständig darzustellen und zu quantifizieren. Andererseits reicht eine Potentialanalyse ebenfalls nicht aus, da in einem solchen Fall auch nicht signifikante

<sup>18</sup>Zum Niveau  $\alpha = 0.05$

<sup>19</sup>Z.B. Kosten zum Austausch des, durch die Konklusion  $B$  bezeichneten, defekten Steuergerätes. Da diese Kosten oftmals unbekannt sind, werden sie im weiteren Verlauf mit eins gleich gesetzt

Zusammenhänge (Regeln mit (zu) kleinem SafeLift) identifiziert werden würden. Aus diesem Grund hat sich in der Praxis nachfolgende Vorgehensweise (Testalgorithmus) als geeignet herausgestellt:

1. Bestimmung, ob ein signifikanter Zusammenhang zwischen der Regelprämisse und der Regelkonklusion besteht. Dazu wird der in Kapitel 4.2.1 vorgestellte Binomialtest verwendet.
2. Der im vorherigen Schritt durchgeführte Test liefert eine Aussage, ob ein statistischer Zusammenhang zwischen dem Regelkopf und -Körper besteht. Er sagt jedoch nichts über die Stärke des Zusammenhangs aus, so dass in einem nächsten Schritt mittels eines statistischen Maßes die Interessantheit quantifiziert wird. Dazu kann beispielsweise der Lift oder der SafeLift (4.29) zum Niveau  $\alpha = 0.05$  verwendet. Um die Menge der gefundenen Zusammenhänge weiter einzuschränken, kann an dieser Stelle ein Schwellwert  $c$  verwendet werden, so dass für alle gefundenen Regeln gilt:  $\text{SafeLift}_T(A \rightarrow B) \geq c$ . Dieser Wert ist von der jeweiligen Anwendung bzw. Anwender abhängig, es hat sich jedoch in der Praxis ein Wert von 1.5 als geeignet herausgestellt<sup>20</sup> (vgl. Kapitel 4.4).
3. Um zusätzlich die gefundenen Ergebnisse mit einem subjektiven Interessantheitsmaß zu analysieren, können die Regeln nun mittels des Potentials nach (4.30) bewertet werden. Eine Auswahl der umzusetzenden Regeln geschieht mit Hilfe eines Experten und seinem Domänenwissen.

---

<sup>20</sup>Aufgrund der Beziehung zwischen Konfidenzintervall und Hypothesentest, ist es, bei der Verwendung von hohen Schwellwerten für den SafeLift, unnötig, den 1. Schritt auszuführen

### 4.3 Multiples Hypothesenproblem

Führt man eine Assoziationsanalyse mittels des im vorherigen Kapitel eingeführten Testalgorithmus<sup>21</sup> durch, so werden im ersten Schritt für jede mögliche Kombination von Ziel- und Einflussgrößen Hypothesentests durchgeführt mit der Nullhypothese, dass kein Zusammenhang besteht. Diese sind notwendig, um *echte* Zusammenhänge in den Daten von statistischem Rauschen zu unterscheiden. Denn habe beispielsweise eine Regel einen Lift größer als 1, kann erst durch einen Hypothesentest (zum Beispiel ein Binomialtest) entschieden werden, ob auch wirklich ein signifikanter Zusammenhang besteht.

Sei beispielsweise eine Transaktionsdatenbank  $D$  gegeben, in der qualitätsbeschreibende Faktoren ( $I$ ) und qualitätsmessende Informationen ( $J$ ) über produzierte Fahrzeuge gespeichert sind (vgl. Felddaten Analyse, Kapitel 4.4). Sei also  $D$  eine Transaktionsdatenbank über  $I \times J$  mit deren Hilfe interessante Zusammenhänge mittels einer Assoziationsanalyse identifiziert und in den Qualitätsverbesserungsprozess eines Automobilunternehmens rückgekoppelt werden sollen. Üblicherweise ist in solchen Anwendungen die Anzahl der Einflussfaktoren  $I$  und Zielvariablen  $J$  sehr groß – oft mehrere tausend Werte. Will man nun, wie in Kapitel 4.1 eingeführt, mittels einstelliger Assoziationsregeln signifikante Zusammenhänge zwischen Ziel- und Einflussfaktoren beschreiben, so müssen im schlimmsten Fall<sup>22</sup>  $N = |I \times J|$  Regeln mittels eines Signifikanztests auf Unabhängigkeit überprüft werden. Mit anderen Worten, es werden  $N$  Tests durchgeführt, mit deren Hilfe bestimmt wird, ob zwischen der Prämisse und der Konklusion einer Regel eine signifikante Abhängigkeit existiert.

Liefere nun ein solches Verfahren zur Identifizierung von signifikanten Zusammenhängen  $R$  signifikante Regeln, das heißt es wurden  $R$  Hypothesen zum Niveau  $\alpha$  abgelehnt. Dann stellt sich nun die Frage, ob die gefundene Anzahl signifikanter Zusammenhänge zu *groß* oder zu *klein* ist, denn man würde  $N\alpha$  irrtümlich abgelehnte Hypothesen unter den  $R$  gefundenen Zusammenhängen erwarten, jedoch weiß man nicht welche es sind: Sei eine Transaktionsdatenbank über  $I \times J$  mit  $I = \{i_1, \dots, i_{n_I}\}$  der Menge der Einflussgrößen und  $J = \{j_1, \dots, j_{n_J}\}$  der Menge der Schadensschlüssel gegeben. Des Weiteren sei  $n_I = 6260$  und  $n_J = 3789$  und fordere man für jede Assoziationsregel  $n_{A \wedge B} \geq 4$ . Dann müssten in diesem Beispiel  $N = 261198$  einfache Assoziationsregeln auf einen signifikanten Zusammenhang zwischen Prämisse und Konklusion hin analysiert werden. Wird als Signifikanzniveau  $\alpha = 0.05$  festgelegt, so würde man nach Durchführung der  $N$  Hypothesentests rund 13060 fälschlicherweise abgelehnte Nullhypothesen erwarten. Bei einer solchen hohen Anzahl durchgeführter Hypothesentests kommt es demnach zu

<sup>21</sup>Unter einem Testalgorithmus wird im weiteren Verlauf stets eine Folge von Operationen verstanden, die aus einer Menge möglicher Zusammenhänge, diejenigen identifiziert, welche signifikant sind

<sup>22</sup>Viele Algorithmen zur Berechnung von Assoziationsregeln nutzen Schwellwerte wie *minimum Konfidenz* oder *minimum Support*, um den Raum möglicher Kombinationen sinnvoll einzuschränken

einer *Kumulierung* des Fehlers 1. Art bzw. zu dem so genannten *Multiplen Hypothesenproblem*.

### 4.3.1 Lösungsansätze

Zur Formalisierung des multiplen Hypothesenproblems wurde von Y. Benjamini und Y. Hochberg in [Benjamini und Hochberg 1995] eine Kontingenztabelle für die Ereignisse multipler Hypothesentests aufgestellt (vgl. Tabelle 4.5). Basierend auf dieser Tabelle, lassen sich Indikatoren für die Verlässlichkeit der gemachten (multiplen) Hypothesen ableiten.

	Nicht abgelehnt	abgelehnt	
Wahre Nullhypothesen	$U$	$V$	$m_0$
Nicht-wahre Nullhypothesen	$T$	$S$	$m_1$
	$N - R$	$R$	$N$

Tabelle 4.5: Kontingenztabelle für Hypothesentests nach [Benjamini und Hochberg 1995]

Prinzipiell gibt es mehrere Ansätze, dieses fundamentale Problem der Assoziationsanalyse und anderer Data Mining Verfahren, welche zur Wissensextraktion statistische Tests nutzen, zu “lösen“, bzw. zu umgehen<sup>23</sup>. Die möglichen Lösungsansätze lassen sich dabei in zwei Gruppen einteilen: Eine erste Möglichkeit ist, den zu analysierende Datensatz (zum Beispiel eine Transaktionsdatenbank) in zwei unabhängige Mengen – Trainings- und Testmenge – aufzuteilen. Auf der Trainingsmenge werden im ersten Schritt die zunächst signifikanten Zusammenhänge identifiziert und anschließend auf der Testmenge verifiziert (vgl. auch Evaluation des RCA-Verfahrens in Kapitel 3.3.2). Dabei wird nur im ersten Schritt ein möglicher Zusammenhang mittels eines Signifikanztests überprüft.

Die Aufteilung des vorliegenden Datensatzes in eine Test- und eine Trainingsmenge “löst“<sup>24</sup> das multiple Hypothesenproblem jedoch nicht, sondern umgeht es lediglich. In der Literatur werden stattdessen verschiedene statistische Verfahren verwendet, die das

<sup>23</sup>Ein simpler, und zugleich weit verbreiteter “Ansatz“ im Umgang mit dem multiplen Hypothesenproblem ist das Ignorieren des Problems. Da man sich der Tragweite der Problematik nicht bewusst ist, wird das Multiple Hypothesenproblem in vielen Data Mining Anwendungen nicht beachtet. Dies wird vor allem durch die Tatsache deutlich, dass es nur wenig Arbeiten im Bereich Data Mining gibt, die sich mit dem multiplen Hypothesenproblem beschäftigen. Hingegen ist in der Biostatistik und vor allem in der Genetik das Problem bekannt und dementsprechend etabliert (vgl. [Lallich et al. 2006])

<sup>24</sup>Für das multiple Hypothesenproblem gibt es im eigentlichen Sinn keine Lösung

Problem der Alpha-Kumulierung mit unterschiedlichen Schwerpunkten lösen. Die Grundidee dieser zweiten Gruppe von Lösungsansätzen ist, eine, beim multiplen Testen gemachte Fehlerrate, auf einem gewünschten Niveau zu kontrollieren, in dem das durch den Anwender vorgegebene Signifikanzniveau  $\alpha$  geeignet angepasst wird –  $\alpha \rightsquigarrow \tilde{\alpha}$ . Dabei können unterschiedliche Arten von Fehlerraten betrachtet werden, wobei dementsprechend eine geeignete Testprozedur gewählt wird, welche die Kontrolle der jeweiligen Fehlerrate gewährleistet.

Die bekannteste Fehlerrate im Kontext des multiplen Testens ist die so genannte *Family Wise Error Rate* (FWER). Damit wird die Wahrscheinlichkeit bezeichnet, dass mindestens eine Nullhypothese fälschlicherweise abgelehnt wird. Bei vielen Tests ( $N \gg 1$ ), führt die Kontrolle der FWER zu einer sehr starken Erhöhung des Signifikanzniveaus der einzelnen Tests. Methoden, wie beispielsweise das Bonferroni-Verfahren, welche den Family Wise Error kontrollieren, gelten aus diesem Grund als zu konservativ.

In der Praxis sind Anwender jedoch oftmals bereit, einen, von ihnen vorgegeben, Anteil fälschlicherweise abgelehnter Nullhypothesen zu “akzeptieren“. Mit anderen Worten, ist der erwartete Quotient aus fälschlicherweise abgelehnten Nullhypothesen zu der Gesamtanzahl abgelehnter Hypothesen kleiner, als ein vom Anwender vorgegebener Schwellwert, so werden die abgelehnten Nullhypothesen als “wahre“ Alternativhypothesen eingestuft. Verfahren, die diese Art von Fehlerrate, die so genannten *False Discovery Rate* (FDR), kontrollieren, werden im folgenden Verlauf dieses Kapitels vorgestellt. Da sie weniger konservativ sind im Vergleich zu FWER kontrollierenden Verfahren, werden folglich mehr Hypothesen  $H_0$  abgelehnt, was zu einer Verminderung des Fehlers 2. Art (auch als Powergewinn bezeichnet) führt.

Bei der Assoziationsanalyse werden nach Hypothesentests auch statistische oder subjektive Interessantheitsmaße verwendet, um signifikante Zusammenhänge zu extrahieren. Dadurch wird die Menge der gefundenen Zusammenhänge  $R$  weiter eingeschränkt, d.h. man erhält eine Anzahl  $R^*$  von Assoziationsregeln mit  $R^* < R$ . Denn ein Teil der durch die Hypothesentests (z.B. Binomialtest) gefundenen  $R$  Zusammenhänge ist nicht interessant genug für die Anwendung – sie weisen beispielsweise ein zu geringes Potential auf, oder werden aufgrund von Expertenwissen als nicht interessant eingestuft. Erhält man nach Durchführung einer Assoziationsanalyse, bestehend aus Binomialtest (ohne Anpassung des Signifikanzniveaus zur Vermeidung des multiplen Hypothesenproblems) und Anwendung von Interessantheitsmaßen zur Bewertung von gefundenen Assoziationsregeln  $R^*$  Zusammenhänge, so ist es für den Anwender von Interesse, ob diese vertrauenswürdig sind. Die Frage die es somit zu beantworten gilt: Wurden durch Tests, Interessantheitsmaße und Expertenwissen so viele nicht signifikante Zusammenhänge eliminiert, dass trotz der Kumulierung des Fehlers 1. Art, die verbleibenden Regeln als signifikant eingestuft werden können.

Das Ziel eines solchen alternativen Ansatzes ist es dabei nicht, ein vom Anwender vorgegebenes Fehlerniveau zu gewährleisten, wie bei Verfahren zur Kontrolle von FWER und FDR, sondern ein Maß für die Vertrauenswürdigkeit der gefundenen Zusammenhänge unter einer vermeintlichen Ignoranz des multiplen Hypothesenproblems zu geben. Wie jedoch bei der abschließenden Simulation zum Vergleich der verschiedenen Lösungsansätze empirisch gezeigt wird, ist es unter bestimmten Voraussetzungen möglich, dass mittels des alternativen Ansatzes die FDR implizit kontrolliert wird.

Im weiteren Verlauf dieses Kapitels werden verschiedene Verfahren zur Kontrolle der eben eingeführten Fehlerraten motiviert und analysiert. Darüber hinaus wird ein Ansatz, welcher die Bewertung der Zuverlässigkeit gefundener Zusammenhänge erlaubt, als alternativer Ansatz zur den FWER bzw. FDR Methoden vorgestellt. Den Abschluss bildet ein empirischer Vergleich der vorgestellten Lösungsansätze zur Kontrolle bzw. Vermeidung des multiplen Hypothesenproblem.

## Family Wise Error

Der *Family Wise Error Rate* beschreibt die Wahrscheinlichkeit, die Nullhypothese  $H_0$  mindestens ein Mal abzulehnen, obwohl sie wahr ist (siehe Tabelle 4.5):

$$FWER = P(V > 0) \quad (4.31)$$

Die einfachste und bekannteste Methode den FWER zu kontrollieren ist mittels der *Bonferroni-Korrektur*. Die auf den italienischen Mathematiker Carlo Emilio Bonferroni zurückgehende Methode basiert dabei auf folgender Überlegung: seien  $H_0^1, \dots, H_0^N$  die  $N$  zu überprüfenden Nullhypothesen und  $p_1, \dots, p_N$  die dazugehörigen  $p$ -Werte. Die Wahrscheinlichkeit, keine Nullhypothese abzulehnen, obwohl sie wahr ist (Fehler 1. Art), ist bei unabhängigen Tests zum Signifikanzniveau  $\alpha$  gleich  $(1 - \alpha)^N$ . Folglich ist die Wahrscheinlichkeit, mindestens eine Nullhypothese abzulehnen, obwohl sie wahr ist, gleich

$$\tilde{\alpha} = 1 - (1 - \alpha)^N \quad (4.32)$$

Will man nun für alle Hypothesen die Wahrscheinlichkeit  $\tilde{\alpha}$  steuern, eine Nullhypothese fälschlicherweise abzulehnen, so kann Gleichung (4.32) nach  $\alpha$  aufgelöst werden und man erhält das neue Signifikanzniveau

$$\tilde{\alpha} = 1 - (1 - \alpha)^{\frac{1}{N}}$$

Eine Anpassung des Signifikanzniveaus nach dieser Art findet man in der Literatur unter dem Namen **Dunn-Sidak-Methode** (vgl. [Dudoit et al. 2003]).

Die eigentliche **Bonferroni-Korrektur** erhält man durch die Vereinfachung,  $(1 - \alpha)^N \simeq 1 - N\alpha$ . Damit berechnet sich das neue Signifikanzniveau zu  $\alpha = \tilde{\alpha}/N$ . Dieses Niveau ergibt sich aus einer Abschätzung der FWER im Falle von abhängigen Tests<sup>25</sup>, was hier sinnvoll ist, da im Allgemeinen nicht von unabhängigen Tests ausgegangen werden kann. Speziell im Falle von Assoziationsregeln haben gefundene Zusammenhänge oftmals gemeinsame Itemsets. Ist eine große Anzahl von zu testenden Hypothesen gegeben, so ist sowohl die Dunn-Sidak-Methode als auch die Bonferroni-Korrektur zu konservativ, denn für  $N \rightarrow \infty$  gilt:  $\alpha \rightarrow 0$ .

Die verschiedenen Verfahren zur Kontrolle des FWER lassen sich prinzipiell in drei Kategorien einteilen:

**Single-step** Diese leicht zu implementierenden Verfahren erreichen eine Kontrolle des FWER, indem ein globales Signifikanzniveau gesetzt wird. Der Nachteil ist, dass sie jedoch nicht den Fehler 2. Art kontrollieren und stets als zu konservativ eingestuft werden. Die bereits eingeführte Bonferroni-Korrektur und die Dunn-Sidak-Methode gehören zur Klasse der single-step-Verfahren, da das Signifikanzniveau nur ein Mal angepasst wird. Bei jeder Ablehnung einer Hypothese verringert sich die Gesamtzahl der verbleibenden Hypothesen. Um dies zu berücksichtigen wurden weitere Verfahren entwickelt, die dieses mit einbeziehen.

**Step-down** Um die zu konservative Kontrolle des FWER der single-step-Verfahren zu verbessern, wurden 1979 von S. Holm [Holm 1979] eine zur Bonferroni-Korrektur analoge Methode vorgeschlagen. Seien  $p_1, \dots, p_N$  die aufsteigend sortierten  $p$ -Werte und  $H_0^i, i = 1, \dots, N$  die dazugehörigen Nullhypothesen. Des Weiteren sei

$$i^* = \min\left\{i : p_i > \frac{\alpha}{(N - i + 1)}\right\} \quad (4.33)$$

dann werden alle Hypothesen  $H_0^i$  verworfen mit  $i = 1, \dots, i^* - 1$ . Existiert kein solches  $i^*$ , so wird keine Hypothese verworfen. Setzt man anstelle von (4.33)

$$i^* = \min\left\{i : p_i > 1 - (1 - \alpha)^{\frac{1}{N-i+1}}\right\}$$

so erhält man das Analogon zur Dunn-Sidak-Methode. Ein Nachteil der vorgestellten **Bonferroni-Holm-Methoden** ist, dass sie lediglich für kleine Werte von  $N$  brauchbare Ergebnisse liefern. Für große Werte von  $N$  ist die Adaption inkorrekt (siehe [Lallich et al. 2006], 3.2).

**Step-up** Bei den eben vorgestellten Methoden, werden alle Nullhypothesen bis zum

<sup>25</sup>Falls nicht von Unabhängigkeit ausgegangen werden kann, gilt (siehe [Dudoit et al. 2003]):  $FWER \leq \sum_{i=1}^N P(p_i < \alpha | H_0^1, \dots, H_0^N) = N\alpha$

Stopindex  $i^*$  verworfen, bzw. bis zum ersten Mal gilt:

$$p_i > \frac{\alpha}{(N - i + 1)} \text{ bzw. } p_i > 1 - (1 - \alpha)^{\frac{1}{N-i+1}}$$

Eine Verbesserung dieses Ansatzes mittels Umkehrung der Suchreihenfolge wurde 1988 von Y. Hochberg gemacht (vgl. [Dudoit et al. 2003]): Sei  $H_0^C$  die Hypothese, dass alle Nullhypothesen wahr sind. Dann gilt für unabhängige Hypothesentests, die *Simes-Ungleichung* (siehe [Dudoit et al. 2003]):

$$P(p_i > \alpha \frac{i}{N}, i = 1, \dots, N | H_0^C) \geq 1 - \alpha$$

falls  $p_1 \leq p_2 \leq \dots \leq p_N$ . Diese Ungleichung nutzte Hochberg in seiner Veröffentlichung [Hochberg 1988] um eine weitere, weniger konservative Methode zur Kontrolle des FWER zu konstruieren. Im Gegensatz zu den step-down-Verfahren wird bei der **Simes-Hochberg-Methode** der ‘‘Stopindex‘‘  $i^*$  wie folgt definiert:

$$i^* = \max\{i : p_i > \frac{\alpha}{(N - i + 1)}\}$$

Dann sind alle Nullhypothesen  $H_0^i$  abzulehnen mit  $i = 1, \dots, i^*$ . Im Falle, dass kein solches  $i^*$  gefunden werden kann, sind alle Hypothesen nicht abzulehnen.

Der Vorteil des Simes-Hochberg-Verfahrens ist, dass es eine bessere Eigenschaft im Bezug auf den Fehler 2. Art besitzt als die Step-Down-Methoden (vgl. [Dudoit et al. 2003]). Jedoch kann die Simes-Hochberg-Methode lediglich bei unabhängigen Tests eingesetzt werden, so dass im Zweifelsfalle die Bonferroni-Holm-Methode verwendet werden sollte.

## False Discovery Rate

Ein anderer Ansatz, um das multiple Hypothesenproblem zu lösen, wurde 1995 von Y. Benjamini und Y. Hochberg in [Benjamini und Hochberg 1995] vorgeschlagen. Die Autoren waren der Ansicht, dass in vielen Situationen die eben vorgestellten Methoden zur Kontrolle des FWER zu konservativ sind. Denn in praktischen Anwendungen ist man gewillt, wenige Fehler 1. Art in Kauf zu nehmen, wenn das Verhältnis von fälschlicherweise abgelehnter Hypothesen zu zutreffend abgelehnten Nullhypothesen klein ist (vgl. [Dudoit et al. 2003], 2.5). Diese Überlegungen führten zu dem Ansatz, den erwarteten Anteil fälschlich abgelehnter Nullhypothesen innerhalb der abgelehnten Hypothesen – die so genannte *False Discovery Rate* – zu kontrollieren. Für den Fall, dass alle  $H_0$  wahr sind, ist die FDR gleich dem FWER.

Formal lässt sich die FDR wie folgt definieren: beschreibe  $V$  die Anzahl fälschlich abgelehnter Nullhypothesen  $H_0$  und  $R$  die Gesamtanzahl abgelehnter Hypothesen (vgl. Tabelle 4.5). Der Anteil falscher Ablehnungen ist gegeben durch

$$Q = \begin{cases} \frac{V}{R} & R > 0 \\ 0 & R = 0 \end{cases} \quad (4.34)$$

und somit ist die *False Discovery Rate* definiert als der Erwartungswert  $E(Q)$ . Basierend auf dieser Definition schlugen Benjamini und Hochberg in [Benjamini und Hochberg 1995] eine Methode vor, welche auf der einen Seite die FDR auf einem gewünschten Niveau  $\alpha$  kontrolliert aber auf der anderen Seite die Anzahl abgelehnter Nullhypothesen maximiert (vgl. FWER). Dafür leiteten sie folgendes Step-up-Verfahren für unabhängige Tests her: seien wie im Falle der FWER-Verfahren  $p_1, \dots, p_N$  die geordneten  $p$ -Werte mit  $p_1 \leq p_2 \leq \dots \leq p_N$ . Des Weiteren sei

$$i^* = \max\{i : p_i \leq \frac{i\alpha}{N}\}$$

Dann sind alle Hypothesen  $H_0^i$  mit  $i = 1, \dots, i^*$  abzulehnen und falls kein  $i^*$  gefunden werden kann, sind alle Hypothesen nicht abzulehnen. Dabei kontrolliert das Verfahren die FDR zum Niveau  $\alpha m_0/N$ , unabhängig von der zugrunde liegenden Verteilung der Daten. In [Benjamini und Yekutieli 2001] zeigten Y. Benjamini und Y. Yekutieli, dass die **Benjamini-Hochberg-Methode** auch für spezielle Abhängigkeitsstrukturen die FDR kontrolliert. Für allgemeinere Abhängigkeiten stellten sie in ihrer Arbeit eine konservativere Methode – das **Benjamini-Yekutieli-Verfahren** – vor, wobei  $i^*$  definiert ist durch

$$i^* = \max\{i : p_i \leq \frac{i\alpha}{N \sum_{i=1}^N 1/i}\}$$

Eine Verbesserung der vorgestellten Methoden ist mittels einer Schätzung des unbekanntes Wertes  $m_0$  möglich (vgl. Tabelle 4.5). In [Benjamini und Leshno 2007] und [Benjamini et al. 2006] wird eine zweistufige Anpassung des Signifikanzniveaus für positiv abhängige Teststatistiken und Hypothesentests für Assoziationsregeln vorgestellt. Dabei wird im ersten Schritt der **Two-Stage Linear Step-Up Prozedur** (TST) der unbekanntes Wert für  $m_0$  als die Differenz des Gesamttestumfangs  $N$  und der Anzahl der verworfenen Nullhypothesen zum Niveau  $\tilde{\alpha} = \alpha/(1 + \alpha)$  mittels der Benjamini-Hochberg-Methode bestimmt. Im zweiten Schritt wird wiederum die Benjamini-Hochberg-Methode durchgeführt, jedoch mit

$$\alpha^* = \frac{\tilde{\alpha}N}{m_0}$$

## Alternativer Ansatz

Die im vorherigen Abschnitt vorgestellten Verfahren benötigen zur Ermittlung und Kontrolle des beim multiplen Testen gemachten Fehlers (FWER oder FDR) eine Liste mit den Ergebnissen aller getesteten Zusammenhänge (Hypothesen). Aufbauend auf dieser Liste wird in einem zweiten Schritt die Menge abgelehnter Nullhypothesen unter Berücksichtigung des multiplen Hypothesenproblems ermittelt.

Bei der Assoziationsanalyse werden jedoch neben statistischen Tests zur Identifizierung eines Zusammenhangs zwischen Prämisse und Konklusion auch Interessantheitsmaße und Expertenwissen verwendet (vgl. Testalgorithmus), um den Gehalt (Interessantheit) der gefundenen Zusammenhänge zu bestimmen. Sei  $\mathcal{M}$  die Menge abgelehnter Nullhypothesen, die mittels eines der Verfahren zur Kontrolle von FDR oder FWER bestimmt wird und  $\mathcal{M}^*$  die als interessant eingestuften Zusammenhänge nach Anwendung von Interessantheitsmaßen und Expertenwissen. Die Menge  $\mathcal{M}^*$  entsteht somit durch Überprüfung, ob die in  $\mathcal{M}$  enthaltenen Regeln eine, vom Anwender vorgegebene Schwelle für ein Interessantheitsmaß erfüllen, bzw. durch das Expertenwissen des Anwenders als interessant eingestuft werden. Die so erhaltene Menge  $\mathcal{M}^*$  hat dabei folgende zwei Eigenschaften:

- Die beim multiplen Testen gemachten Fehlerrate wird auf einem vom Anwender gewünschten Niveau kontrolliert
- Aus der subjektiven Sicht des Anwenders enthält  $\mathcal{M}^*$  nur noch interessante Zusammenhänge, da die Regeln eine gewisse Interessantheitsschwelle überschritten haben (beispielsweise ist das Potential größer als ein vorgegebener Schwellwert) und aus technischer Sicht Sinn ergeben<sup>26</sup>

Sei  $\tilde{\mathcal{M}}$  die Menge von Zusammenhänge, die man nach Anwendung eines (nicht angepassten) Signifikanztests zum Niveau  $\alpha$  erhält (das multiple Hypothesenproblem wird ignoriert). Des Weiteren enthalte  $\tilde{\mathcal{M}}^*$  diejenigen Zusammenhänge, die nach Anwendung von Interessantheitsmaßen und Expertenwissen auf  $\tilde{\mathcal{M}}$  als interessant eingestuft werden. Dann ist es von Interesse zu wissen, wie zuverlässig die verbleibenden Assoziationsregeln in  $\tilde{\mathcal{M}}^*$  sind. Das heißt, wurden durch Interessantheitsmaße und Expertenwissen so viele nicht signifikante Zusammenhänge eliminiert, dass die Kumulierung des Fehlers 1. Art nicht weiter beachtet werden muss? Um das abschätzen zu können, benötigt man einen Ansatz, welcher die Menge  $\tilde{\mathcal{M}}^*$  im Hinblick auf das multiple Hypothesenproblem a posteriori bewertet.

---

<sup>26</sup>Zum Beispiel kann ein gefundener Zusammenhang aus technischer Sicht nicht sinnvoll sein, obwohl ein statistischer Test auf einen signifikanten Zusammenhang hinweist

Sei  $N$  die Anzahl möglicher Zusammenhänge (beispielsweise bei einstelligen Assoziationsregeln die Mächtigkeit der Kombinationsmenge von Ziel- und Einflussvariablen). Beschreibe wie in Tabelle 4.5  $m_1$  die *unbekannte* Anzahl gültiger Zusammenhänge (entsprechend der Alternativhypothese) so sind folglich  $N - m_1 = m_0$  Regeln ungültig. Des Weiteren sei das Ergebnis eines Testverfahrens eine Anzahl  $R$  von (signifikanten) Zusammenhänge, dann lässt sich die Anzahl der gefundenen Regeln  $R$  wie folgt aufspalten:

$$\begin{aligned} R &= m_0\alpha + m_1\gamma \\ &= (N - m_1)\alpha + m_1\gamma \\ &= N\alpha + m_1(\gamma - \alpha) \end{aligned} \tag{4.35}$$

Dabei entspricht  $\gamma$  dem Anteil an wahren Alternativhypothesen, die auch richtigerweise identifiziert wurden.

Da  $N\alpha > (N - m_1)\alpha$  lässt sich die Summe der gefundenen Regeln aus Gleichung (4.35) nach oben durch

$$R < N\alpha + m_1\gamma$$

abschätzen, so dass mindestens  $R - N\alpha$  gültige Regeln erwartet werden können. Das Verhältnis  $R$  zu  $N\alpha$  liefert somit eine erste grobe Abschätzung, wie groß der Anteil ungültiger Regeln ist. Eine genauere Schätzung für den Anteil ungültiger Regeln an der Gesamtmenge  $R$  gefundener Zusammenhänge lässt sich wie folgt motivieren:

Bei einem gegebenen Signifikanzniveau  $\alpha$  kann man vom Testalgorithmus im Mittel  $(N - m_1)\alpha$  ungültige Assoziationsregeln erwarten. Diesen Wert sollte mit Hilfe geeigneter Konfidenzintervalle abgesichert werden. Dabei kann die Entscheidung, ob eine Regel ungültig ist oder nicht, als ein Bernoulli-Prozess aufgefasst werden. Folglich ist die Anzahl der als ungültig erkannter Regeln bei Stichprobenumfang  $n$  eine binomialverteilte Zufallsvariable mit Wahrscheinlichkeit  $p = \alpha$ . Sei  $(N - m_1)\alpha$  ein Schätzer für diese unbekannt Anzahl ungültiger Regeln und  $n = (N - m_1)$  der zugrunde liegende Stichprobenumfang. Eine obere Konfidenzschranke lässt sich dann folgendermaßen konstruieren (vgl. auch Standardintervall in Kapitel 4.2.2):

$$C_u^{Bin} = (N - m_1)\alpha + z_{1-\alpha'}\sqrt{(N - m_1)(1 - \alpha)\alpha} \tag{4.36}$$

mit  $z_{1-\alpha'}$  dem  $1 - \alpha'$ -Quantil der Standard-Normalverteilung.

Da  $\alpha \ll 1$ , kann (4.36) zu

$$C_u^{Bin} = (N - m_1)\alpha + z_{1-\alpha'}\sqrt{(N - m_1)\alpha}$$

vereinfacht werden. Allerdings ist die Anzahl der wahren Alternativen  $m_1$  unbekannt, man erhält jedoch eine obere Grenze für die linke Intervallgrenze ungültiger Zusammenhänge mittels

$$L(\alpha) = N\alpha + z_{1-\alpha'}\sqrt{N\alpha} \tag{4.37}$$

$i$	$p_i$	$L(p_i)$
1	0.00001	0.15
2	0.00021	0.74
3	0.00037	1.04
4	0.00059	1.38
5	0.00064	1.45
6	0.00095	1.85
7	0.00201	3.01
8	0.00232	3.31
9	0.00254	3.52
10	0.00297	3.93
11	0.00331	4.23
12	0.00392	4.76
13	0.00454	5.28
14	0.01426	12.47
15	0.04577	32.45

Tabelle 4.6: Nach (4.38) geschätzte Anzahl ungültiger Regeln mit  $\alpha' = 0.05$ 

Basierend auf diesem Ansatz lässt sich nun eine Abschätzung der Anzahl fälschlicherweise abgelehnter Nullhypothesen mit Hilfe des folgenden Verfahrens angeben:

Sei  $N$  die Gesamtzahl der zu testenden Hypothesen und  $p_1, \dots, p_N$  die dazugehörigen, aufsteigend sortierten  $p$ -Werte. Des Weiteren sei  $R$  die Anzahl abgelehnter Hypothesen zum Signifikanzniveau  $\alpha$ . Dann ist der Anteil fälschlicherweise abgelehnter Hypothesen zum Niveau  $p_i, i = 1, \dots, N$  gegeben<sup>27</sup> durch

$$L(p_i) = Np_i + z_{1-\alpha'}\sqrt{Np_i} \quad (4.38)$$

Mit anderen Worten: werden die getesteten Hypothesen zum Niveau  $p_i$  abgelehnt, so kann der Anteil fälschlicherweise abgelehnter Nullhypothesen durch 4.38 abgeschätzt werden.

#### Beispiel 4.1

Sei eine Transaktionsdatenbank gegeben, bei der  $N = 500$  Hypothesen zum Niveau  $\alpha = 0.05$  zu testen sind. Des Weiteren sei  $R = 15$  die Anzahl der Regeln für die gilt:  $p_i \leq \alpha$ . In Tabelle 4.6 ist die geschätzte Anzahl ungültiger Regeln zum Niveau  $p_i, i = 1, \dots, R$  aufgelistet.

<sup>27</sup>Sinnvollerweise werden nur die ersten  $R$  Regeln betrachtet

### 4.3.2 Vergleich der Verfahren

Das im vorherigen Abschnitt neu eingeführte Verfahren zur Umgehung des multiplen Hypothesenproblems bietet im Gegensatz zu den bekannten Verfahren, wie TST-Methode bzw. Bonferroni-Korrektur keine direkte Kontrolle über die in der Literatur verwendeten Fehlerarten FDR oder FWER. Vielmehr wird beim alternativen Ansatz der Anteil fälschlicherweise identifizierter Zusammenhänge abgeschätzt, um somit eine Aussage über die Vertrauenswürdigkeit der gefundenen Zusammenhänge treffen zu können. Wie im Folgenden empirischen Vergleich gezeigt wird, ist es jedoch durch Forderungen an den geschätzten Anteil ungültiger Zusammenhänge  $L(\alpha)$  an der Gesamtzahl gefundener Zusammenhänge  $R$  möglich, mittels des alternativen Ansatzes ähnliche Ergebnisse wie mit Hilfe des TST-Verfahrens zur Kontrolle der FDR zu erhalten. Ein Vergleich mit Verfahren zur Kontrolle der FWER war nicht sinnvoll, da beispielsweise die Bonferroni-Korrektur und ähnliche FWER kontrollierende Verfahren zu restriktiv sind, und deutlich weniger signifikante Regeln identifiziert wurden<sup>28</sup> als mit den Verfahren zur Kontrolle der FDR bzw. dem alternativen Ansatz.

Die Benjamini-Hochberg- und TST-Methode basieren auf folgender, prinzipiellen Vorgehensweise: In einem ersten Schritt wird für alle möglichen Kombinationen der vorliegenden Anwendung ein geeigneter Hypothesentest durchgeführt. Jedoch werden die Nullhypothese weder verworfen noch nicht abgelehnt, sondern lediglich der  $p$ -Wert der Hypothese berechnet. In einem zweiten Schritt wird, in Abhängigkeit der entsprechenden  $p$ -Werte und der Anzahl verbleibender Hypothesen, abgelehnt. Die so abgelehnten Hypothesen bilden die statistisch signifikanten Zusammenhänge (im weiteren Verlauf als *signifikante* Regeln bezeichnet), welche anschließend mittels eines Interessantheitsmaßes wie beispielsweise dem Potential bewertet werden können, um die Regelmenge weiter einzugrenzen bzw. zu sortieren. Die aus der letzten Einschränkung resultierenden Regeln werden im weiteren Verlauf als *interessante* Regeln bezeichnet.

Verwendet man den neuen Ansatz, so wird im ersten Schritt jede Hypothese zum nicht angepassten Niveau  $\alpha$  getestet und, falls die Nullhypothese verworfen werden kann, der Zusammenhang als signifikant eingestuft. In diesem ersten Schritt wird somit die Problematik des multiplen Testens zunächst *ignoriert*. Im zweiten Schritt werden die oben beschriebenen Interessantheitsmaße (z.B. Potential oder Lift) verwendet, um die gefundenen Zusammenhänge weiter einzugrenzen. Zur Überprüfung der Vertrauenswürdigkeit der letzteren Menge, werden die Ergebnisse aufsteigend nach ihren  $p$ -Werte sortiert und die Anzahl ungültiger Regeln mittels Gleichung (4.37) bzw. (4.38) ermittelt. Die hiernach verbleibenden Regeln bilden die *interessanten* Regeln.

<sup>28</sup>Für die in der empirischen Studie betrachteten Datensätze hat man, unter Verwendung der FWER, lediglich 3 signifikante Regeln erhalten

Für den empirischen Vergleich der Verfahren wurden zwei reale Datensätze aus dem Qualitätsmanagement verwendet (siehe Kapitel 4.4) und mögliche Zusammenhänge mit Hilfe von Assoziationsregeln analysiert<sup>29</sup>. Dabei wurde als erster Schritt für jede mögliche Kombination aus Ziel- und Einflussvariablen ( $I \times J$ ) eine Kontingenztabelle (Tabelle 4.2) und der  $p$ -Wert des Binomialtests (vgl. Abschnitt 4.2.1) mit Hilfe der Einträge in die Vierfelder-Tafel ermittelt<sup>30</sup>. Da in vielen praktischen Anwendungen die Anzahl möglicher Kombinationen von Ziel und Einflussvariablen sehr groß ist, wurde die Menge a priori sinnvoll eingegrenzt, indem ein unterer Schwellwert für das gemeinsame Auftreten von Prämisse und Konklusion festgelegt wurde ( $n_{A \wedge B} \geq 3$ ). Die theoretische Anzahl von  $|I \times J|$  wurde somit auf die *Kandidatenanzahl*  $N$  reduziert. Einen Überblick über die verwendeten Auswahlkriterien und Kenngrößen der durchgeführten Assoziationsanalysen ist in Tabelle 4.7 zu finden.

Größe	Datensatz 1	Datensatz 2
Einflussgrößen $I$	744 (208)	108 (68)
Zielgrößen $J$	587 (51)	3789 (884)
$ I \times J $	436728	409212
Kandidaten $N$	444	6998
$\min_{n_{A \wedge B}}$	3	3
$\min_{\text{Potential}}$	4	4

Tabelle 4.7: Verwendete Auswahlkriterien und Kenngrößen der empirischen Studie. Die Werte in Klammern geben die Anzahlen nach Einschränkung durch  $\min_{A \wedge B} \geq 3$

Im zweiten Schritt wurde zur Bestimmung des angepassten Signifikanzniveaus, die Kandidatenmenge mit Hilfe der beiden Verfahren zur Kontrolle der FDR analysiert. Dabei wurde, basierend auf der Kandidatenmenge, die Menge der statistisch signifikanten Regeln mittels BH- und TST-Verfahren ermittelt. Vergleicht man die Ergebnisse in Tabelle 4.9 mit den Ergebnissen ohne eine Korrektur des Signifikanzniveaus (Tabelle 4.8), so ist zunächst wie erwartet zu beobachten, dass die Anzahl der gefundenen interessanten Zusammenhänge geringer ist als ohne Korrektur. Wie weiter in Tabelle 4.9 zu sehen ist, liefert die BH-Methode für den ersten Datensatz ein leicht konservativeres Ergebnis als die TST-Methode (es werden weniger Regeln abgelehnt, da das angepasste Signifikanzniveau  $\tilde{\alpha}$  bei der BH-Methode niedriger ist als bei dem TST-Verfahren). Dabei stufte die BH-Methode 162 Regeln als signifikant ein, das TST-Verfahren hingegen 202 Regeln. Ein ähnliches Verhalten ist für den 2. Datensatz zu beobachten, jedoch weniger stark ausgeprägt. Schränkt man diese Menge weiter auf diejenigen Regeln mit  $\text{Potential} \geq 4$

<sup>29</sup>Grundlage für den Vergleich der Verfahren in diesem Abschnitt bilden die Produktions- und Testdaten für zwei Steuergeräte

<sup>30</sup>Bei der vorliegenden Anwendung wird mittels des Binomialtests überprüft, ob es einen signifikanten Zusammenhang zwischen der Prämisse und der Konklusion einer Regel gibt

ein, so reduziert sich die Anzahl auf 62 (64) (bzw. 100 (102) für den zweiten Datensatz) interessante Regeln. Vergleicht man diese mit den Ergebnissen ohne Korrektur (Tabelle 4.8) so stellt man fest, dass für den 1. Datensatz die Anzahl der interessanten Regeln fast identisch ist. Dies wird auch deutlich durch das verhältnismäßig hohe, korrigierte Signifikanzniveau  $\tilde{\alpha} = 0.017$  für die BH-Methode bzw.  $\tilde{\alpha} = 0.021$  für das TST-Verfahren (im Vergleich zu  $\alpha = 0.05$ ).

	Datensatz 1	Datensatz 2
Signifikante Regeln	225	866
Potential $\geq 4$	68	390
Interessante Regeln	68	312

Tabelle 4.8: Gefundene Zusammenhänge ohne Korrektur zum Niveau  $\alpha = 0.05$

Im Gegensatz zu dem ersten Datensatz ist die Anzahl der zu testenden Hypothesen im zweiten Fall wesentlich höher (insgesamt sind 6998 Hypothesen zu überprüfen). Folglich ist das korrigierte Signifikanzniveau niedriger (für beide Methoden:  $\tilde{\alpha} \approx 0.00087$ ), was sich auch in der Differenz zwischen interessanten Regeln *ohne* Korrektur (Tabelle 4.8) und *mit* Korrektur (Tabelle 4.9) widerspiegelt – im letzteren Fall wurden nur rund ein Viertel der Zusammenhänge als signifikant eingestuft. Betrachtet man nun die gefundenen Zusammenhänge der einzelnen Verfahren, so erhält man wie erwartet folgende Teilmengenbeziehung: in beiden Fällen sind die durch die BH-Methode gefundenen Regeln eine echte Teilmenge der Regelmenge, welche mittels der TST-Methode identifiziert wurde.

	Datensatz 1		Datensatz 2	
	BH	TST	BH	TST
Signifikante Regeln	162	202	110	112
Interessante Regeln	62	64	100	102
Signifikanzniveau $\tilde{\alpha}$	0.017	0.021	0.00087	0.00087
Schnittmenge	62		100	

Tabelle 4.9: Gefundene Zusammenhänge mittels der Bonferroni-Hochberg-Methode und dem TST-Verfahren zum Niveau  $\alpha = 0.05$

Bei dem in Abschnitt 4.3.1 vorgestellten alternativen Ansatz wird keine automatische Korrektur des Signifikanzniveaus vorgenommen, sondern es lässt sich die Vertrauenswürdigkeit der gefundenen Zusammenhänge mittels der Abschätzung aus Gleichung (4.37) bestimmen. Darauf aufbauend kann das Signifikanzniveau ggf. angepasst werden. In dem Fall, dass das multiple Hypothesenproblem vollständig ignoriert wird ( $\alpha = 0.05$ ), sind für den ersten Datensatz von den 444 möglichen Regeln 68 interessant. Jedoch sind davon, wenn man den alternativen Ansatz zu Grunde legt, rund 29 ungültig bzw. zufällig

denn mit

$$L(\alpha) = \max_{\substack{p_i \leq \alpha \\ i=1, \dots, N}} L(p_i) = 29 \quad (4.39)$$

gilt:  $L(\alpha = 0.05) = 29$ . Dies entspricht einem Verhältnis gefundener Regeln zu ungültigen Regeln von 2.303 bzw. einem Anteil von 43 Prozent ungültiger Regeln an der Gesamtzahl gefundener Regeln<sup>31</sup>.

Das Verhalten von  $L(\alpha)$  und  $R(\alpha)$  bei unterschiedlichen Signifikanzniveaus ist für beide Datensätze in Abbildung 4.11 dargestellt. Wie in den Abbildungen deutlich zu sehen ist, ist im Falle des ersten Datensatzes die Anzahl ungültiger Regeln  $L(\alpha)$  in jedem Fall geringer als die gefundenen interessanten Zusammenhänge  $R(\alpha)$ . Da aber keine Korrektur einen zu hohen Anteil ungültiger Regeln ergibt, rund jede zweite Regel wäre zufällig, sollte das Signifikanzniveau angepasst werden. Genauer gesagt muss ein  $\tilde{\alpha}$  gewählt werden, welches den Anteil ungültiger Regeln reduziert. Ein Kompromiss, der auf der einen Seite ein nicht zu kleines  $\tilde{\alpha}$  bedingt und auf der anderen Seite den Anteil ungültiger Regeln gering hält, ist eine Fixierung des Signifikanzniveaus auf  $\tilde{\alpha} = 0.005$ . Dies entspricht einem Verhältnis

$$Q(\alpha) = \frac{L(\alpha)}{R(\alpha)} \approx 0.077$$

bzw. 7.7 Prozent. Fordert man beispielsweise für das angepasste Signifikanzniveau  $\tilde{\alpha}$ , dass

$$\tilde{\alpha} = \arg \max_{\alpha} (Q(\alpha) \leq 0.12) \quad (4.40)$$

so erhält man 62 interessante Zusammenhänge und die Anzahl ungültiger Zusammenhänge  $L(\alpha)$  ergibt sich zu rund 9 Regeln. Dies entspricht dem Ergebnis des TST-Verfahrens (vgl. Tabelle 4.9)

Betrachtet man die Ergebnisse für den zweiten Datensatz, so ist in Abbildung 4.11 zu beobachten, dass im Gegensatz zum 1. Datensatz, die Anzahl ungültiger Regeln  $L(\alpha)$  nicht immer geringer ist als die Anzahl gefundener Zusammenhänge  $R(\alpha)$ . Speziell gilt für den 2. Datensatz:  $L(\alpha) > R(\alpha)$ , falls  $\alpha \geq 0.035$ . Fordert man wie in (4.40), dass der Anteil ungültiger Regeln höchstens 12 Prozent betragen darf, so erhält man, wie im Falle des TST-Verfahrens, die Anzahl von 102 abgelehnten Nullhypothesen. Dies führt zu einem angepassten Signifikanzniveau von  $\tilde{\alpha} = 0.0086$ .

Der alternative Ansatz bietet somit eine einfache und praktische Möglichkeit, die Vertrauenswürdigkeit der gefundenen Zusammenhänge einer Anwendung abzuschätzen bzw. anzugeben. Aufgrund einer solchen Abschätzung ist es dem Anwender möglich, das Signifikanzniveau  $\alpha$  zu erhöhen bzw. zu verringern. Darüber hinaus kann durch Forderungen

<sup>31</sup>Zur besseren Lesbarkeit wird im weiteren Verlauf für die Anzahl ungültiger Regel bzw. für die Anzahl gefundener Regeln bis zu einem Niveau  $\alpha$ , die Notation analog zu Gleichung (4.39) verwendet

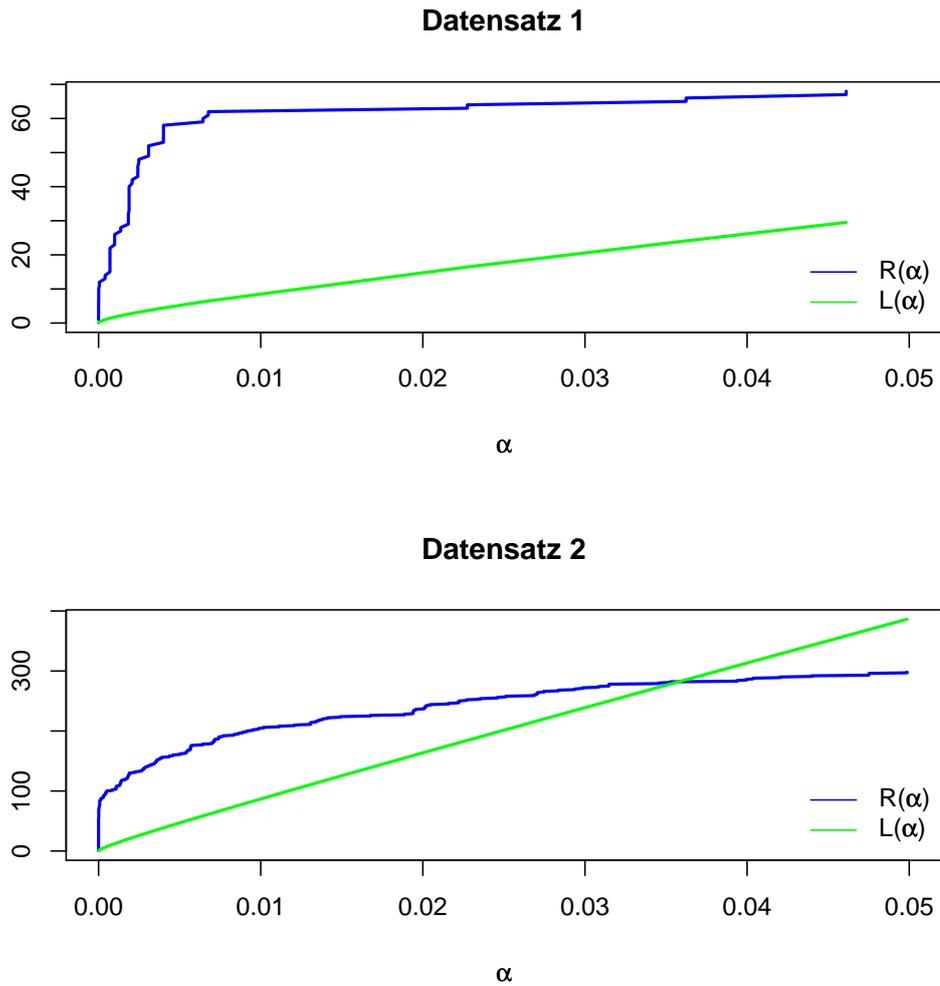


Abbildung 4.11: Verhalten von gefundenen, interessanten  $R(\alpha)$  zu ungültigen Zusammenhängen  $L(\alpha)$

---

an das Verhältnis von fälschlicherweise abgelehnter Nullhypothesen  $L(\alpha)$  zu abgelehnter Nullhypothesen  $R(\alpha) - \tilde{\alpha} = \arg \max_{\alpha} (Q(\alpha) \leq 0.2)$  – ein ähnliches Ergebnis wie beim TST-Verfahren zur Kontrolle der False Discovery Rate erreicht werden. Ist es jedoch primär das Ziel, die FDR zu kontrollieren, so sind die in Abschnitt 4.3.1 vorgestellten Verfahren dem alternativen Ansatz vorzuziehen.

## 4.4 Felddaten Analyse

Die einzige Möglichkeit, die Qualität eines Produktes umfassend zu beurteilen und zu messen, ist sein Verhalten beim Kunden bzw. im Feld zu beobachten und für spätere Analysen geeignet aufzuzeichnen. In der Automobilbranche beispielsweise, werden Probleme und Schäden an einem Fahrzeug während des Besuchs des Kunden in einer Vertragswerkstatt erfasst und zentral gespeichert. Aufgrund dieser Aufzeichnungen steht den Automobilherstellern eine, in der Regel lückenlose Dokumentationen der Reparaturmaßnahmen bzw. des *Feldverhaltens* (im weiteren Verlauf auch Felddaten) der Fahrzeuge für die (gesetzliche) Garantiezeit zur Verfügung<sup>32</sup>.

Ein Ansatz, wie man die Information bezüglich des Verhaltens im Feld in den Qualitätsverbesserungsprozess rückkoppeln kann, ist Zusammenhänge zwischen dem Fertigungsprozess (Produktionsdaten) und den Felddaten zu identifizieren. Ein geeignetes Verfahren, mögliche Zusammenhänge zu extrahieren ist die in Abschnitt 4.1 vorgestellte Assoziationsanalyse. Unter der Annahme, dass mögliche Fehler im Feld ihre Ursache in der Produktion haben, lassen sich Klassifikationsregeln verwenden, um das Feldverhalten mittels Produktionsdaten zu erklären.

Betrachtet man die Fertigungsprozesse in der Automobilbranche, so stellt man fest, dass fahrzeugbezogene Produktionsdaten entlang der gesamten *Wertschöpfungskette*<sup>33</sup> des Fahrzeuges erfasst werden. So werden Informationen wie Produktionsparameter oder Messergebnisse für einzelne Fahrzeugkomponenten während der Produktion aufgezeichnet und analysiert<sup>34</sup>. Diese, in oftmals vorgelagerten Fertigungs- bzw. Messpunkten (in Relation zur Endfertigung) gewonnenen Produktionsdaten bilden eine Art Vorkontrolle, die jede Fahrzeugkomponente passieren muss, bevor sie eingebaut wird. In diesem Zusammenhang spricht man auch von so genannten *Quality Gates*. Doch eine Vorkontrolle, bei der darauf geachtet wird, dass lediglich *die guten* Teile verbaut werden, reicht nicht aus, um die vom Endverbraucher geforderte Qualität zu gewährleisten. Es ist notwendig das Verhalten der verbauten Einzelkomponenten im Zusammenspiel mit anderen Komponenten des Fahrzeugs zu überprüfen.

Eine besondere Herausforderung für Automobilhersteller stellen dabei elektronische Komponenten – speziell *elektronische Steuergeräte* (engl. ECU = *electronic control unit* oder ECM = *electronic control module*) – dar. In modernen Fahrzeugen haben Steuergeräte die Aufgabe, unterschiedliche elektronische Module anzusteuern oder zu regeln, wie bei-

---

<sup>32</sup>Daten, die über die Garantiezeit hinaus gehen, sind in der Regel nur auf wenige Fahrzeuge, wie beispielsweise Flotten, beschränkt

<sup>33</sup>Unter Wertschöpfungskette werden die unterschiedlichen Produktionsstufen und Arbeitsschritte der (verbauten) Komponenten eines Produktes bis hin zum Endverbraucher verstanden

<sup>34</sup>In Kapitel 3 ist diese Komponente die Ölwanne eines Fahrzeuges, die in einer Gießerei gefertigt wurde

spielsweise das Steuergerät für die Motorregelung oder das Anti-Blockier-System (ABS). Dabei sind in einem Fahrzeug des Premiumsegments bis zu 70 verschiedene Steuergeräte verbaut, welche über so genannte Bus-Systeme<sup>35</sup> kommunizieren. Über diese Systeme werden systemweite Informationen über den Zustand des Fahrzeuges und andere relevante Daten ausgetauscht. Um die Funktionsfähigkeit dieser Steuergeräte zu gewährleisten werden komplexe Testsysteme eingesetzt, damit Fehlfunktionen und Interaktionsprobleme in den verbauten Steuergeräten möglichst früh erkannt und behoben werden können. Jedoch ist es aufgrund der hohen Kombinationsvielfalt, der vielfältigen Funktionsmöglichkeiten verbauter Steuergeräte und der beschränkten Zeitkapazität zur Durchführung von Tests nicht möglich, vollumfänglich zu testen. Vielmehr werden lediglich im Vorfeld festgelegte Testroutinen während der Produktion durchgeführt, die einen Teil der möglichen Funktionalität überprüfen. In Zusammenarbeit mit dem Ingenieuren ist die Auswahl der richtigen Prüfroutinen ein wesentlicher Bestandteil der Qualitätssicherung.

Eine Möglichkeit den Testprozess elektrischer Steuergeräte zu verbessern und zu verfeinern ist, Zusammenhänge zwischen dem Verhalten im Feld und möglichen Testergebnissen zu analysieren und daraus gewonnene Informationen in den Qualitätsverbesserungsprozess rückzukoppeln. Die zugrunde liegende Idee dabei ist, Schäden, die im Feld auftreten mittels Testergebnissen, Testabläufen und weitere Produktionsdaten zu erklären. Dabei werden die in Kapitel 4.2 vorgestellten Interessantheitsmaße verwendet, um mögliche Verbesserungen im Sinne der Qualitätssicherung, sowohl nach statistischen als auch nach wirtschaftlichen Aspekten zu identifizieren.

#### 4.4.1 Modellierung

In der automobilen Fertigung ist das Testen verschiedener Komponenten und Funktionen in den Produktionsprozess integriert. Dies ermöglicht es, die Qualitätsmessung bzw. -Sicherung parallel zur Produktion, ohne zusätzlichen Zeitaufwand durchzuführen. Einen integralen Bestandteil der Qualitätssicherung in der Produktion stellt dabei die Überprüfung eingebauter Steuergeräte dar. Bereits in frühen Stadien des Fertigungsprozesses werden die unterschiedliche Steuergeräte auf ihre individuelle und integrierte Funktionalität getestet und die Testergebnisse aufgezeichnet. Diese Test- und Messergebnisse lassen sich nun gemeinsam mit Felddaten wie beispielsweise aufgezeichneten Schäden analysieren, denn letztere sind letztendlich das maßgebliche Kriterium für die Qualität eines gefertigten Produktes. Für dieses Anwendung sind speziell (Feld-)Schäden, die im Zusammenhang mit Elektrik bzw. den Steuergeräten auftreten von Interesse.

---

<sup>35</sup>In modernen Fahrzeugen sind die Steuergeräte über CAN, LIN, MOST oder Flexray Bus-Systeme verbunden

Um die daraus gewonnenen Informationen in den Qualitätsverbesserungsprozess eines Automobilunternehmens rückzukoppeln, müssen somit zwei unterschiedliche Datenquellen, welche jeweils einen Teil des Lebenszyklus (Abschnitt in der Wertschöpfungskette) eines Produktes beschreiben, kombiniert und gemeinsam analysiert werden:

1. **Produktionsdatenbank:** Sie enthält die Ergebnisse vielfältiger Tests und Messungen, die während des Fertigungsprozesses durchgeführt und dokumentiert werden. Darüber hinaus werden fahrzeugspezifische Daten wie beispielsweise das Modell, Motorisierung, verbaute Sonderausstattung etc. dort gespeichert.
2. **Garantie- und Kulanzdatenbank:** Informationen über (Garantie)Schäden der produzierten Fahrzeuge innerhalb der Garantiezeit werden in einem unternehmensweiten Data Warehouse erfasst.

Die Extraktion von signifikanten Zusammenhängen zwischen beiden Datenquellen kann mittels der klassischen Data Mining Methode – der Assoziationsanalyse – realisiert werden. Die so gefundenen Zusammenhänge werden dabei in Form von Klassifikationsregeln dargestellt. Um Assoziationsregeln in großen Datenbeständen zu finden, gibt es in der Literatur eine Vielzahl unterschiedlicher Verfahren und Algorithmen. Dennoch bedingen einige fundamentale Eigenschaften der hier betrachteten Anwendung die Weiterentwicklung und Anpassung vorhandener Verfahren.

Hierbei wird davon ausgegangen, dass in der industriellen Praxis einer solchen Assoziationsanalyse ein geeignetes Projekt vorangeht. In diesem Data Mining Projekt werden die Voraussetzungen geschaffen, dies es später dem Anwender erlauben, die Analyse (beispielsweise) täglich durchzuführen. Im weiteren Verlauf beschreibt die Felddaten Analyse ein solches, einmaliges Data Mining Projekt, welches sich aus unterschiedlichen Projektphasen zusammensetzt. Auf die einzelnen Phasen wird im weiteren Verlauf dieses Kapitels eingegangen, soweit sie für die Interpretation der Ergebnisse und dem Verständnis der Anwendung notwendig sind.

Eine Vorgehensweise zur Umsetzung eines Data Mining Projektes ist der *CRoss Industry Standard Process for Data Mining 1.0* (kurz CRISP-DM Standard). Der Standard wurde Mitte der neunziger Jahre unter der Leitung der Data Mining-Forschungsgruppe der Daimler AG entwickelt und hat sich im Laufe der Jahre als der internationale Industriestandard für Data Mining Projekte und Methoden etabliert<sup>36</sup>. Kernelement von CRISP-DM ist eine Untergliederung des Data Mining Projektes in sechs unterschiedliche

<sup>36</sup>Aktuell erweitert eine Arbeitsgruppe den aktuellen Standard um aktuelle Entwicklungen und Methoden in den CRISP-DM-Prozess zu integrieren (siehe CRISP-2.0). Jedoch ist dieser noch nicht endgültig verabschiedet und findet somit in der praktischen Umsetzung von Data Mining-Projekten zur Zeit keine Verwendung

Phasen (Schritte), welche iterativ durchlaufen werden (vgl. Abbildung 4.12).

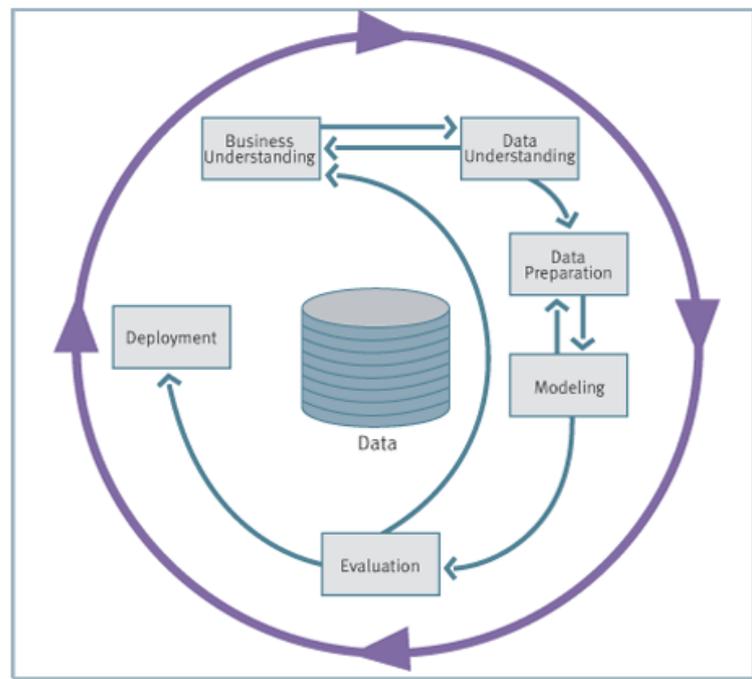


Abbildung 4.12: Phasen des CRISP-DM Prozessmodells nach [Chapman und Clinton 2000]

Den Ausgangspunkt des CRISP-DM-Prozesses bildet das so genannte *Business Understanding*, in dem der Data Mining Experte das zugrunde liegende Problem aus Sicht des Anwenders verstehen und interpretieren soll. Aufbauend darauf sind in der *Data Understanding*-Phase die vorhandenen Datenquellen, mit dessen Hilfe das im vorherigen Schritt erarbeitete Problem gelöst werden soll, zu analysieren. Ist dies geschehen, so müssen die nun relevanten Informationen im *Data Preparation*-Schritt ermittelt und bereinigt werden<sup>37</sup> um anschließend einen Lösungsansatz (z.B. die Extraktion von Wissen bzw. Identifizierung von Zusammenhängen mittels Assoziationsanalyse) zu *modellieren*. Abschließend müssen die gefundene Ergebnisse verifiziert bzw. *evaluiert* werden und die Data Mining-Anwendung an den jeweiligen Anwender mittels geeigneter Schulungen übergeben werden. Offensichtlich handelt es sich bei den ersten fünf Phasen um Schritte, die mehrfach durchlaufen werden, da Wissen und Probleme aus nachgelagerten Phasen ein weiteres Durchlaufen vorgelagerter Phasen notwendig machen können.

Im Folgenden werden die, für die Identifikation von signifikanten Zusammenhängen zwischen Feld- und Produktionsdaten notwendigen Schritte eines Data Mining Projektes

<sup>37</sup>Unter Datenbereinigung wird eine Vorverarbeitung der Daten zur besseren Analyse verstanden – es werden so zum Beispiel fehlende Werte ersetzt

nach dem CRISP-DM Modell näher erläutert. Dabei wird jedoch nur auf diejenigen Phasen, die für die vorliegende Arbeit relevant sind, eingegangen und geeignet zusammengefasst. Die Ergebnisse der Felddaten Analyse werden im Zusammenhang mit der Evaluationsphase vorgestellt.

## Business- und Data-Understanding

Kernelement jedes Data Mining Projektes ist es, die zu analysierenden Datenquellen und die *Umgebungen*, welche die Daten erzeugen, zu verstehen. In der vorliegenden Anwendung, sind die Produktionsdaten Ergebnisse einer komplexen Testprozedur und die Felddaten sind Schadensmeldungen in Form von *Schadensschlüsseln*. Dabei ist der Schadensschlüssel als ein Nummernsystem aufgebaut, mit dessen Hilfe einzelne Fahrzeugteile im Garantie- und Kulanzfall verschlüsselt werden. Genauer gesagt, setzt sich der Schadensschlüssel aus einem fünfstelligen Fehlerort (Schadensteil)<sup>38</sup> und einer zweistelligen Fehlerart (Schadensbild) zusammen. Bei jedem Garantiefall werden die defekten bzw. ausgetauschten Fahrzeugteile mit ihrem dazugehörigen Schadensteil und Schadensbild erfasst und abgespeichert.

Im Gegensatz zu den Felddaten, setzten sich die Produktionsdaten aus verschiedenen Informationsquellen zusammen. Die verwendete Logik zu Erstellung der Daten ist dabei komplexer als die der Felddaten bzw. Schadensschlüsseln. Ein erster Teil der Produktionsdatenbank besteht aus der Fahrzeug Dokumentation, welche sich zum einen aus fahrzeugspezifischen Werten wie dem Modell, der Motorisierung und zum anderen aus den verbauten Sonderausstattungen zusammensetzt. Unter Sonderausstattungen versteht man in der Automobilbranche die Ausstattungsmerkmale eines Fahrzeuges, wie beispielsweise Klimaanlage oder Navigationsgerät, die nicht in der Grundausstattung (Basismodell) enthalten sind. Umgangssprachlich werden diese auch *Extras* genannt. In Kombination mit den fahrzeugspezifischen Merkmalen spezifizieren sie, welche elektrischen Steuergeräte in dem jeweiligen Fahrzeug eingebaut sind.

Der zweite Teil der Produktionsdatenbank beschreibt die Ergebnisse der durchgeführten Qualitätstests für die verbauten elektrischen Steuergeräte. Während der Produktion wird jedes Steuergeräte des Fahrzeuges mehrfach getestet. Dazu wird jedes Fahrzeug bereits in einem frühen Stadium der Produktion mit einer Schnittstelle zur Kommunikation mit den verschiedenen Steuergeräten über das integrierte BUS-System, ausgestattet. Diese Schnittstelle ermöglicht es, Steuergeräte explizit anzusprechen, Informationen auszulesen und somit den Fahrbetrieb eines Autos zu simulieren.

---

<sup>38</sup>Der Schadensteil lässt sich wiederum in eine dreistellige Konstruktionsgruppe und in das tatsächliche Schadensteil unterteilen

Um nun Fehler in den Steuergeräten zu erkennen, werden an verschiedenen Stationen während des Innenausbaus, in Abhängigkeit des Fertigungsgrades des Fahrzeuges und verbauter Steuergeräte, verschiedene Tests durchgeführt. Dabei wird jeder Test entweder vollautomatisch oder mit Hilfe eines Arbeiters, der zum Beispiel den Fensterheber oder andere Knöpfe im Auto betätigen muss, durchgeführt. Fehler während des Tests können somit entweder durch mechanisches Versagen (z.B. das Fenster senkt sich nicht bei der Betätigung des entsprechenden Knopfes) oder durch Fehlercodes in den Steuergeräten charakterisiert werden. Fehlerzustände oder Kommunikationsprobleme eines Steuergeräts werden in den Speicher des Steuergerätes geschrieben und können als codierte Fehler oder Warnungen<sup>39</sup> mit Hilfe der Schnittstelle ausgelesen werden. Prinzipiell hat jeder Test zwei mögliche Ausgänge:

1. Der Test war erfolgreich, sprich es wurden keine Fehler sondern höchstens Warnungen gefunden. Das Ergebnis ist somit eine leere Fehlerliste, und eine eventuell nicht leere Liste mit Warnungen.
2. Durch den Test wurden Fehler entdeckt und in Form einer Fehlerliste ausgegeben. Sind zusätzlich Warnungen aufgetreten, werden diese Informationen ebenfalls in eine Liste geschrieben.

Die einzelnen Ergebnisse der Tests werden erfasst, in eine Datenbank gespeichert und dem Arbeiter am Band gemeldet (siehe Abbildung 4.13). Basierend auf den Testergebnissen muss der Bandarbeiter nun weitere Maßnahmen ableiten:

- Wurde der Test mit einem Fehler abgebrochen, so muss untersucht werden, ob die Ursache noch am Band behoben werden kann, wie beispielsweise ein loses Kabel oder eine nicht richtig gesteckte Verbindung.
- War das Ergebnis des Tests eine Liste von so genannten "hidden errors", kann der Arbeiter das fehlerhafte Ergebnis ignorieren.
- Ist der Test abgebrochen worden, so kann eine Wiederholprüfung durchgeführt werden.
- Muss das betroffene Steuergerät ausgetauscht werden, so wird anschließend der Test wiederholt.
- Ist der gefundene Fehler so gravierend, dass er nicht am Band behoben werden kann, so muss das Fahrzeug aus der Linie genommen werden und zur Nacharbeit

---

<sup>39</sup>Warnungen sind Fehler die vom Gerät als nicht kritisch eingestuft werden. Somit ist es möglich das ein und der selbe Sachverhalt als Fehler oder als Warnung auftreten kann

gebracht werden. Dort kann dann ohne die Zeittaktvorgabe der Linie, der Fehler behoben werden.

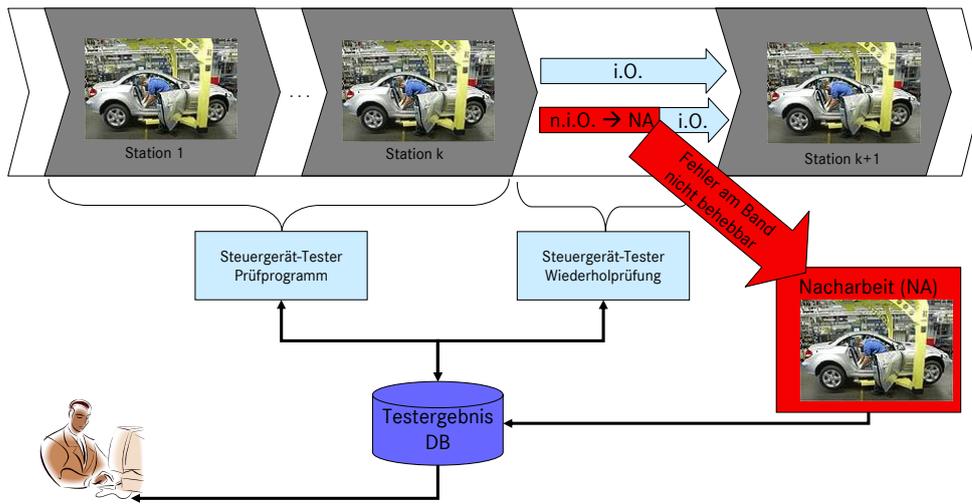


Abbildung 4.13: Testprozedur Felddaten Analyse: Fehlerhafte Ergebnisse (n.i.o) führen entweder zu einer Wiederholprüfung, einer Behebung am Band oder zur Herausnahme des Fahrzeuges in die Nacharbeit (NA). Erst wenn das Fahrzeug fehlerfrei getestet (i.o.) wurde, wird es für die nächste Montagestufe freigegeben

Diese Vielzahl an möglichen Ausgängen und Maßnahmen bedingt, dass ein und der selbe Test mehrfach ausgeführt werden kann, wobei er unter Umständen jedes Mal ein anderes Ergebnis liefert, welches individuell bewertet werden muss. Zum Beispiel verliert eine Fehlermeldung eines Steuergerätes, welches später ausgetauscht wurde, an Relevanz gegenüber einer Fehlermeldung, die nach dem Neueinbau des Steuergerätes ausgelesen wurde. Denn speziell für die Verbundanalyse mit Felddaten ist es notwendig nur Testergebnisse (Produktionsdaten) von Komponenten zu verwenden, die auch letztendlich im Auto verbaut wurden.

## Data Preparation und Modellierung

Wie eingangs erläutert ist es das Ziel Anwendung, mögliche interessante Zusammenhänge zwischen dem Verhalten von produzierten Fahrzeugen beim Kunden (Felddaten) und Produktionsereignissen (Testergebnissen / Produktionsdaten) zu identifizieren. Dies kann mit Hilfe von Klassifikationsregeln geschehen, wobei jedoch charakteristische Eigenschaften der zu Grunde liegenden Anwendung beachtet werden müssen. Dabei gilt, dass die hier aufgeführten Charakteristika spezifisch für die Anwendung, jedoch nicht restringiert auf das hier vorliegende Problem sind. Vielmehr gibt eine Vielzahl von industriellen Anwendungen – nicht nur im automobilen Bereich – welche ähnliche Eigenschaften besitzen.

1. Aufgrund der Vielfalt der möglichen Sonderausstattungen bzw. fahrzeugspezifischen Merkmalen<sup>40</sup> und einer ebenfalls hohen Anzahl an möglichen Schadensschlüsseln, ist eine Vielzahl von möglichen Kombinationen zu analysieren. Trotz dieser Kombinationsvielfalt sind gemeinsam auftretende Ereignisse von Schadensschlüsseln und Testergebnissen selten. Dies erfordert die Verwendung von statistische Methoden, welche auch bei geringen Kombinationshäufigkeiten interessante von uninteressanten Zusammenhängen trennen können.
2. Da die Ergebnisse der Zusammenhangsanalyse in den Qualitätsverbesserungsprozess rückgekoppelt werden sollen, ist es notwendig, Wirtschaftlichkeitsaspekte bei der Generierung von Assoziationsregeln zu beachten: in dem vorliegenden Kontext definieren sich interessante Regeln nicht ausschließlich über ihre statistische Signifikanz sondern ebenfalls über ihr wirtschaftliches *Potential*.
3. Die hohe Anzahl von Einflussgrößen und Zielvariablen führt im Allgemeinen zu einem hohen Umfang von signifikanten Zusammenhängen. Durch die Verwendung von Hypothesentests zur Entscheidung, ob Prämisse und Konklusion nicht unabhängig sind, ist ein gewisser Anteil dieser Zusammenhänge zufällig (multiples Hypothesenproblem). Eine Abschätzung dieses Anteils ist wünschenswert, um Aussagen über die Verlässlichkeit der gefundenen Regeln treffen zu können.

Um eine Verbundanalyse von Feld- und Produktionsdaten durchzuführen, müssen zum einen die jeweiligen Datenquellen in eine geeignete Form transformiert und zum anderen alle relevanten Informationen der produzierten Fahrzeuge geeignet abgebildet werden. Eine Skizzierung der benötigten Datenquellen und der Zusammenführung dieser Daten

---

<sup>40</sup>Durch die hohe Flexibilität der Konfigurationsmöglichkeit von Fahrzeugen der Premiumklasse ist die Produktion “identischer“ Fahrzeuge äußerst selten. So werden beispielsweise bei Daimler im Jahr lediglich zwei identische PKWs produziert

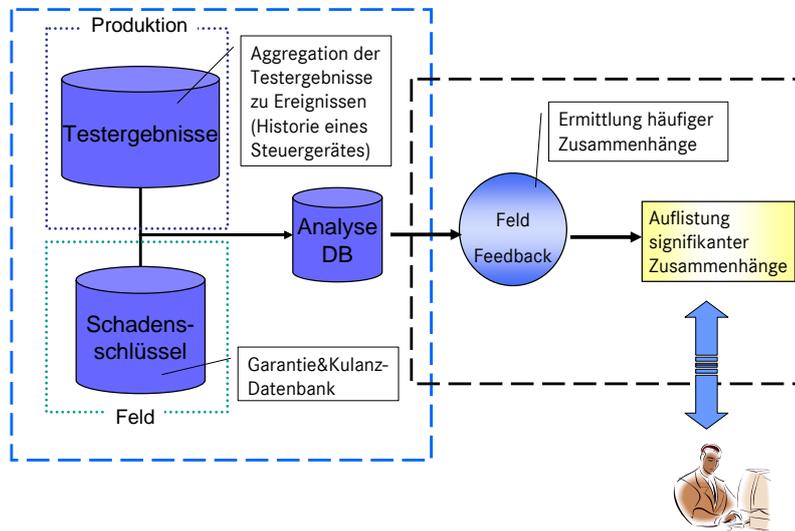


Abbildung 4.14: Schematische Darstellung der benötigten Datenquellen für die Felddaten Analyse

ist in Abbildung 4.14 wiedergegeben. Zur Abbildung der benötigten Informationen wurde folgende Darstellungsart gewählt: Die Darstellung der Felddaten ist mittels binärer Vektoren möglich. Sei  $k$  die Anzahl der in der *Analysedatenbank*<sup>41</sup> vorhandenen Fahrzeuge und  $\mathcal{S}$  die Menge der möglichen Schadensschlüssel. Dann kann für jedes Fahrzeug ein Vektor dargestellt werden, dessen  $i$ -ter Eintrag den Wert eins annimmt, falls bei dem Fahrzeug der  $i$ -te Schadensschlüssel zutrifft. Analog zu den Felddaten, lässt sich für jedes Fahrzeug ein Prämissenvektor mit fahrzeugspezifischen Merkmalen, Kombinationen aus Fehler, Warnungen und abgeleiteten Maßnahmen (Ereignissen) aufbauen.

Beispielsweise könnte der  $i$ -te Eintrag den Wert eins annehmen, falls das Ereignis *Error=X604 im Dieselmotor-Steuergerät* im Fahrzeug aufgetreten ist. Ein solcher Prämissenvektor, der alle möglichen Ereignisse für alle Steuergeräte abbilden würde, wäre allerdings sehr groß, da in einem solchen Fall stets  $n = |\mathcal{E}|$  gilt, wobei  $\mathcal{E}$  die Menge der möglichen Einflussereignisse beschreibt. Da  $n$  in der Felddaten Analyse sehr groß ist, ist es sinnvoll, die Informationen nach unterschiedlichen Steuergeräten aufzugliedern. Seien

<sup>41</sup>Der Begriff Analysedatenbank wird im weiteren Verlauf für eine gemeinsame Datenbank, bestehend aus Produktions- und Felddaten, die über einen Schlüssel, in der Regel der Fahrzeugidentifikationsnummer (FIN), verbunden sind, verwendet

$m$  verschiedenen Steuergeräte gegeben und existiere für jedes Steuergerät  $n_i$ ,  $i = 1, \dots, m$  Einflussereignisse, dann setzt sich die Menge aller Prämissen aus

$$\mathcal{E} = \bigcup_{i \in \{1, \dots, m\}} \mathcal{E}_i$$

zusammen und es können zur Vereinfachung anstatt einem “großen“ Datensatz,  $m$  kleine Datensätze auf Steuergeräteebene analysiert werden. Neben der reduzierten Größe der Einflussvektoren hat dies den Vorteil, dass eine Parallelisierung der Assoziationsanalyse durch *verteiltes Rechnen* möglich ist.

Welche Information nun konkret in den Analyseprozess einfließt, ist das Kernproblem jeder Data Mining Anwendung, da es im Allgemeinen nicht möglich ist, alle vorhandenen Daten zu verwenden. Vielmehr müssen Informationen aggregiert werden, um daraus das für die Anwendung nützliche Wissen zu extrahieren. Entscheidend, neben einem möglichst vollständigem Verständnis der datengenerierenden Umgebung (Domäne), ist die enge Zusammenarbeit mit dem Experten bzw. Anwender. Um mittels der Verbundanalyse den Testprozess der Steuergeräte zu verbessern wurden folgende Ereignisse identifiziert, die es einem Ingenieur erlauben, auf Mängel oder nicht ausreichend getestete Bereiche (Steuergeräte) im Testprozess Rückschlüsse zu schließen und den Testprozess entsprechend anzupassen:

- Welche Fehler und Warnungen auf Steuergeräteebene führten zu einer signifikanten Erhöhung von Schäden im Feld
- Trat ein gleichartiger Fehler in einem anderen Test bereits auf
- Führte ein Fehler dazu, dass das Fahrzeug in die Nacharbeit musste
- Trat ein Fehler als eine Warnung in einem anderen Test auf

Die vier Eigenschaften lassen sich als die Historie eines Steuergerätes auffassen und werden mit einem binären Vektor abgebildet. Zusätzlich dazu wurde der Wochentag und der Monat als Merkmale mit aufgenommen, um zeitbedingte Fluktuationen zu berücksichtigen.

Eng verbunden mit der in den vorherigen Abschnitten eingeführten Phase der Datenvorbereitung (Data Preparation), ist die eigentliche Modellierung der Wissensextraktion. Um interessante Zusammenhänge zwischen der Historie eines Steuergerätes und Schadensfällen im Feld zu identifizieren, werden Klassifikationsregeln benutzt. Dabei setzen sich die Produktionsdaten (Steuergeräthistorie, Wochentag und Monat) aus den unabhängigen Prämissen (Produktionsdaten) und die Schadensschlüssel aus den zu erklärenden Zielvariablen (Konklusionen) zusammen.

In der Literatur wie beispielsweise [Hipp 2004], [Agrawal et al. 1993] findet man unterschiedliche Algorithmen und Verfahren, mit deren Hilfe, große Datenmenge effizient nach Assoziationsregel durchsucht werden können. Die Grundidee der meisten Verfahren basiert auf dem in [Agrawal et al. 1993] eingeführten *Support-Konfidenz-Konzept*, bei dem für Assoziationsregeln ein Mindestwert für die Konfidenz (minconf) und den Support (minsupp) vorgegeben wird.

Die auf dem Support-Konfidenz-Ansatz basierenden Verfahren unterteilen dabei die Regelgenerierung in zwei aufeinander bauenden Schritte: In einem ersten Schritt werden häufige *Ereigniskombinationen* bestimmt und in einem zweiten Schritt werden Assoziationsregeln aus den häufigen Ereigniskombinationen (Itemsets) abgeleitet (vgl. [Hipp 2004], Kapitel 3.1). In diesem Zusammenhang ist unter einer Ereigniskombination eine mehrstellige Assoziationsregel zu verstehen, die nur mit einem logischen UND verknüpft sind (vgl. Seite 99).

Sei  $T = (T_1, \dots, T_n)$  eine Transaktionsdatenbank über den Ereignismengen  $I \times J$ . Dann ist eine Ereigniskombination  $(A, B)$ ,  $A \subset I, B \subset J$  genau dann häufig, wenn gilt:

$$(A, B) \text{ häufig} \Leftrightarrow \text{supp}_T(A, B) = \text{supp}_T(A \rightarrow B) \geq \text{minsupp}$$

Dabei gilt, dass jede nicht leere Teilmenge häufiger Ereigniskombinationen wiederum häufig ist. Denn sei  $(A, B)$  eine häufige Ereigniskombination und  $\tilde{A} \subset A, \tilde{B} \subset B$ , dann folgt unmittelbar aus der Supportdefinition 4.1 (vgl. auch [Hipp 2004], Kapitel 3):

$$\text{supp}_T(A, B) \leq \text{supp}_T(\tilde{A}, \tilde{B})$$

Mit anderen

Im zweiten Schritt werden nun aus der Menge der häufigen Ereigniskombinationen diejenigen Regeln abgeleitet, welche die vorgegebenen Schwellenwerte für Konfidenz und andere Interessantheitsmaße erfüllen. Das Laufzeitverhalten von Algorithmen zur Identifizierung von signifikanten Assoziationsregeln wird dabei typischerweise dominiert durch die Generierung von häufigen Ereigniskombinationen. Somit lässt sich im Allgemeinen die Erzeugung von Assoziationsregeln auf das Finden von häufigen Ereigniskombinationen reduzieren. Eine Übersicht verschiedener Verfahren zur Bestimmung von häufigen Ereigniskombinationen ist in [Hipp 2004], Kapitel 3.2.1 zu finden. Bekanntester Algorithmus zur Erzeugung von häufigen Ereigniskombinationen ist der 1993 von Agrawal in [Agrawal et al. 1993] vorgestellte Apriori-Algorithmus.

Zur Analyse von Produktions- und Felddaten wird eine Implementierung des Apriori-Algorithmus von [Borgelt und Kruse 2002] verwendet. Dabei musste das Verfahren wie folgt angepasst werden:

- Durch die Eigenschaften der Anwendung ist vorgegeben, dass die Einflussgrößen lediglich einen Zielwert beschreiben. Das heißt, dass Konklusionen stets aus genau *einem* Ereignis bestehen.
- Die Ursache für einen Schaden kann jedoch eine komplexere Ursache haben, wie beispielsweise ein Schaden tritt dann auf, wenn ein Ereignis  $A_1$  *stattgefunden* und ein Ereignis  $A_2$  *nicht stattgefunden* hat. Seien  $A_1 \rightarrow B$ ,  $A_2 \rightarrow B$  zwei Assoziationsregeln deren Support und Konfidenz die vorgegebenen Schwellwerte erfüllen. Dann werden aus den Prämissen  $A_1$ ,  $A_2$  mittels logischer Operatoren Assoziationsregeln der Form  $\varphi(A_1, A_2) \rightarrow B$  gebildet (vgl. komplexe Assoziationsregeln, Seite 101)<sup>42</sup>.

Liegen die Daten für das Feldverhalten und die Historie der Steuergeräte in der oben beschriebenen Form vor, kann die Extraktion von Wissen mittels eines dreistufigen Testalgorithmus durchgeführt werden.

Sei  $m$  die Anzahl unterschiedlicher Steuergeräte in der Analysedatenbank,  $\mathcal{S}_i$  die Menge von Schadensschlüsseln, die bei den Fahrzeugen mit Steuergerät  $i$  im Feld aufgetreten sind,  $\mathcal{E}_i$  die Menge an Ereignissen welche die Historie des  $i$ -ten Steuergerätes beschreibt und  $n_i$  die Anzahl der Fahrzeuge mit Steuergerät  $i$ . Des Weiteren sei  $T_i = (T_1, \dots, T_{n_i})$  die binäre Transaktionsdatenbank für das  $i$ -te Steuergerät über  $I \times J$  mit  $I = \{0, 1\}^{|\mathcal{E}_i|}$  und  $J = \{0, 1\}^{|\mathcal{S}_i|}$ . Die Menge  $I$  beschreibt dabei die binären Einflussgrößen – Historie eines Steuergerätes – und  $J$  die Schadensschlüssel. Dann kann für alle Steuergeräte folgender Testalgorithmus (parallel) ausgeführt werden:

- 1. Generierung häufiger Ereigniskombinationen** Wie oben beschrieben, werden mittels des Apriori-Algorithmus häufige Ereigniskombinationen generiert. Dabei kann minsupp wie in Kapitel 4.3.2 gewählt werden, so dass gilt:

$$\text{minsupp} = \frac{3}{n_i}$$

- 2. Bestimmung interessanter Assoziationsregeln** Basierend auf der Menge häufiger, einstelliger Ereigniskombinationen wird nun mittels des Binomialtests zum Niveau  $\alpha$  bestimmt, ob ein signifikanter Zusammenhang zwischen der Prämisse besteht (vgl. Kapitel 4.2.1, Seite 110). Ist eine Regel (statistisch) signifikant, wird ihre Interessantheit für die Anwendung mit Hilfe der Gütemaße SafeLift und Potential bestimmt. Erfüllt eine Regel die vorgegebenen Schwellenwerte für die Interessantheitsmaße SafeLift und Potential, so spricht man von interessanten Regeln. Dafür wird ein  $\tilde{\alpha}$  für den SafeLift vorgegeben und die Schwellenwerte entsprechend der

<sup>42</sup>Zur Konstruktion solcher komplexen genügt es, die Kontingenztafel der Regel  $A_1 \cup A_2 \rightarrow B$  zu berechnen

Anwendung festgelegt. Zum Beispiel kann der Mindestwert für den SafeLift wie in Abschnitt 4.4.2 auf den Wert 1.3 und für das Potential auf 3 gesetzt werden.

Ähnlich wird mit zwei- und mehrstelligen Ereigniskombinationen verfahren: Seien  $A_1 \rightarrow B$ ,  $A_2 \rightarrow B$  zwei Assoziationsregeln, dann werden mittels der zweistelligen Ereigniskombination  $(A_1 \wedge A_2, B)$  und den einstelligen Ereigniskombinationen  $(A_1, B)$ ,  $(A_2, B)$  Kontingenztabelle für folgende logische Verknüpfungen der Prämisse gebildet:  $\neg A_1 \wedge A_2$ ,  $A_1 \wedge \neg A_2$ ,  $\neg A_1 \wedge \neg A_2$ ,  $A_1 \vee A_2$ ,  $\neg A_1 \vee A_2$ ,  $A_1 \vee \neg A_2$  und  $\neg A_1 \vee \neg A_2$ . Zum Beispiel lassen sich die Einträge der Kontingenztabelle der Assoziationsregel  $(A_1 \wedge \neg A_2) \rightarrow B$  wie folgt berechnen:

$$\begin{aligned} n_{A_1 \wedge \neg A_2} &= n_{A_1} - n_{A_1 \wedge A_2} \\ n_{A_1 \wedge \neg A_2 \wedge B} &= n_{A_1 \wedge B} - n_{A_1 \wedge A_2 \wedge B} \\ n_B &= n_B \\ n &= n \end{aligned}$$

	$B$	$\neg B$	
$A_1 \wedge A_2$	$n_{A_1 \wedge A_2 \wedge B}$	$n_{A_1 \wedge A_2 \wedge \neg B}$	$n_{A_1 \wedge A_2}$
$\neg A_1 \wedge \neg A_2$	$n_{\neg A_1 \wedge \neg A_2 \wedge B}$	$n_{\neg A_1 \wedge \neg A_2 \wedge \neg B}$	$n_{\neg A_1 \wedge \neg A_2}$
	$n_B$	$n_{\neg B}$	$n$

Tabelle 4.10: Kontingenztabelle für die Regel  $(A_1 \wedge A_2) \rightarrow B$

	$B$	$\neg B$	
$A_1 \wedge \neg A_2$	$n_{A_1 \wedge A_2} - n_{A_1 \wedge A_2 \wedge B}$	$n_{A_1 \wedge \neg A_2 \wedge \neg B}$ *	$n_{A_1} - n_{A_1 \wedge A_2}$
$\neg(A_1 \wedge \neg A_2)$	$n_{\neg(A_1 \wedge \neg A_2) \wedge B}$ *	$n_{\neg(A_1 \wedge \neg A_2) \wedge \neg B}$ *	$n_{\neg(A_1 \wedge \neg A_2)}$ *
	$n_B$	$n_{\neg B}$	$n$

Tabelle 4.11: Kontingenztabelle für die Regel  $(A_1 \wedge \neg A_2) \rightarrow B$  mit Berechnungsvorschrift, wie sie aus der Kontingenztabelle für  $(A_1 \wedge A_2) \rightarrow B$  abgeleitet werden kann. Die mit \* gekennzeichneten Felder sind für die Berechnung von Interessantheitsmaßen irrelevant, ergeben sich jedoch implizit aus den übrigen Feldern

Mit Hilfe dieser Werte können Interessantheitsmaße wie Potential und SafeLift analog zum “einfachen“ Fall berechnet werden.

Bei der Wahl des Support-Konfidenz-Ansatzes zur Bestimmung von Assoziationsregeln, kann es passieren, dass zwei Kombinationen  $(A_1, B)$ ,  $(A_2, B)$  häufig sind jedoch  $(A_1 \wedge A_2, B)$  nicht häufig ist. Ist dies der Fall, so lässt sich die benötigte Kontingenztabelle nicht ableiten, da die Kontingenztabelle für  $(A_1 \wedge A_2, B)$  im Support-Konfidenz-Ansatz nicht erstellt wurde. Folglich lassen sich in einem solchen Fall

auch nicht die Interessanztheit der komplexen Regeln, die durch die Verknüpfung mit anderen logischen Operatoren entstehen bestimmen. Einzige Möglichkeit dieses zu umgehen, ist die Bestimmung aller Assoziationsregeln mit ein und mehr Einflussfaktoren ohne die Verwendung des oben beschriebenen Minimum-Support-Prinzipes.

**3. Abschätzung der Vertrauenswürdigkeit gefundener Zusammenhänge** Ein gewisser Anteil der gefundenen Zusammenhänge ist aufgrund der vielen, im ersten durchgeführten Hypothesentests, zufällig. Um nun eine Aussage über den Anteil zufälliger, fälschlicherweise abgelehnter Nullhypothesen treffen zu können, wird das in Kapitel 4.3 vorgestellte, alternative Verfahren verwendet. Stellt sich dabei heraus, dass aufgrund von Gleichung 4.37 dieser (unbekannte) falsche Anteil einen zu großen prozentualen Anteil an der Gesamtzahl abgelehnter Hypothesen bildet, so kann das Konfidenzniveau im Nachhinein angepasst werden.

#### 4.4.2 Evaluierung und Ergebnisse

Der im vorherigen Abschnitt beschriebene Testalgorithmus zur Identifizierung von interessanten Zusammenhängen für die Verbesserung des Steuergerätestprozesses (Felddaten Analyse), wird in der Praxis in regelmäßigen Abständen von den verschiedenen Anwendern für alle PKW-Baureihen durchgeführt. Da die Darstellung der Ergebnisse einer Analyse der gesamten Daimler PKW-Flotte zu umfangreich wäre, wird anhand einer ausgewählten Baureihe (Mercedes-Benz C-Klasse) zum einen eine Evaluation der oben beschriebenen Anwendung und zum anderen die prinzipielle Vorgehensweise bei der Felddaten Analyse vorgestellt. Dazu wird zunächst das verwendete Mengengerüst eingeführt und anschließend die zur Identifizierung interessanter Zusammenhänge notwendigen Schritte erläutert (Schritt 1 und 2 des Testalgorithmus). Anhand dieser Ergebnisse wird im darauf folgenden Abschnitt auf das multiple Hypothesenproblem, bzw. auf die Vertrauenswürdigkeit der gefundenen Zusammenhänge eingegangen.

Grundlage der Analyse sind die, im Zuge der Felddaten Analyse vom Anwender zur Verfügung gestellten, Produktions- und Felddaten. Diese, zur Evaluation des Verfahrens genutzten Daten, wurden dabei so gewählt, dass zum einen eine Verifizierung des implementierten Verfahrens möglich ist und zum anderen alle betrachteten Fahrzeuge mindestens zwei Jahre alt sind. Die Verifizierung des Verfahrens erfolgte dadurch, dass die Daten Zusammenhänge enthielten, die dem Anwender bereits bekannt waren und deren Implikationen größtenteils schon umgesetzt wurden. Durch die Verwendung eines "alten" Datensatzes konnte darüber hinaus sichergestellt werden, dass alle Fahrzeuge die volle Garantiezeit ausschöpfen konnten. Somit hatten alle Fahrzeuge den gleichen zeitlichen *Feldhorizont* (alle sind mindestens zwei Jahre gelaufen), was die Datenvorbereitung

an dieser Stelle vereinfacht. Würde man auch Fahrzeuge einbeziehen, deren Garantiezeit noch nicht abgelaufen ist, so müssten anderen Kriterien zur Schaffung einer gemeinsamen Analysegrundlage verwendet werden. In einem solchen Fall kann die Laufleistung der Fahrzeuge genutzt werden, um homogene Gruppen bilden zu können.

## Häufige Ereigniskombinationen

Die Kernproblematik der Anwendung ist die hohe Vielfalt der Ereignisse (Prämissen und Konklusionen), die durch die umfangreichen Konfigurationsmöglichkeiten der Fahrzeuge gegeben ist. Dies bedingt beispielsweise, dass die gleiche Fahrzeugkonfiguration im Schnitt nur ein Mal im Jahr auftritt. Folglich sind die beobachteten Schadensarten und -bilder, wie auch die Testergebnisse, dementsprechend unterschiedlich. Aufgrund dessen ergibt sich im vorliegenden Fall folgendes Mengengerüst (vgl. auch Tabelle 4.12):

- Im betrachteten Zeitraum wurden insgesamt 38529 C-Klassen produziert und getestet
- In der Garantie- und Kulanzdatenbank sind 3789 unterschiedliche Schadensschlüssel für die betrachtete Baureihe erfasst.
- In der Baureihe sind pro Fahrzeug bis zu 61 verschiedene Steuergeräte verbaut.
- Der im vorherigen Abschnitt beschriebene Testprozess zur Überprüfung der elektrischen Steuergeräte, bildet die Historie der Steuergeräte auf insgesamt 6290 unterschiedliche Ereignisse ab.

Größe	Wert
Schadensschlüssel	3789
Einflussereignisse	6260
Steuergeräte	61
Fahrzeuge	38529
Zeitraum	1. Halbjahr 2005

Tabelle 4.12: Mengengerüst Felddaten Analyse

Die oben angeführten Anzahlen der zu analysierenden Schadensschlüsseln und Produktionsereignissen würde auf rund  $24 \cdot 10^6$  einfachen Regeln (Kombinationen von Testergebnissen und Schadensschlüsseln) führen. Um diese Menge sinnvoll einzuschränken, werden in der Anwendung nur solche Kombinationen untersucht, bei denen das gemein-

same Auftreten von Prämisse und Konklusion mindestens 4 ist ( $n_{A \wedge B} \geq 4$ )<sup>43</sup>. Dies führte zu 261198 Zusammenhängen, die auf ihre statistische Signifikanz mittels des Binomialtests (vgl. 110) analysiert werden. Wird dieser zum Niveau  $1 - \alpha = 0.95$  ausgeführt, so erhält man für den Beispieldatensatz 25986 Regeln.

## Interessante Zusammenhänge

Dem nächsten Schritt des Testalgorithmus auf Seite 161 folgend, werden auf der Menge der signifikanten Regeln aufbauend, die für die Anwendung interessante Zusammenhänge identifiziert. Dabei kann die Ursache auf einen Schaden im Feld einen komplexeren Grund als ein einzelnes Ereignis haben. Denn es können zwei oder mehrere Ereignisse im Testprozess einen signifikanten Einfluss auf das Verhalten im Feld aufweisen. Für diese können mit Hilfe der logischen Verkettungsoperatoren *UND*, *ODER* und deren Negation komplexe Assoziationsregeln gebildet werden (siehe auch Seite 101). In der vorliegenden Anwendung ist die Generierung komplexer Regeln aufgrund der abnehmenden Häufigkeit gemeinsam auftretender Ereignisse und damit der geringen technischen Relevanz (aus Sicht des Kunden), auf maximal zwei Einflussgrößen beschränkt.

Um die Interessanztheit eines Zusammenhangs für die Anwendung zu bewerten, werden im letzten Schritt zwei unterschiedliche Interessanztheitsmaße verwendet. Ausgangspunkt ist der Lift, welcher in der Literatur als das bekannteste Maß zur statistischen Bewertung von Regeln gilt. Der Lift gibt an, ob ein Zusammenhang zwischen der Prämisse und der Konklusion besteht, jedoch nicht wie stark dieser Zusammenhang ist. Eine Möglichkeit, die Stärke des Zusammenhangs auszudrücken ist mit Hilfe des in Kapitel 4.2.2 eingeführten SafeLifts. Betrachtet man die auf dem Wilson-Intervall basierende Definition des SafeLifts 4.22, so setzt sich der SafeLift aus dem Lift abzüglich der Varianz des Liftes, multipliziert mit einem Quantil der Normalverteilung, zusammen.

Größe	Wert
$\min_{n_{A \wedge B}}$	4
Minimum SafeLift	1.3
Minimum Potential	3
$\alpha$ für SafeLift	0.05
$\alpha$ für Binomialtest	0.05

Tabelle 4.13: Vorauswahlkriterien Felddaten Analyse

Interpretiert man nun die Varianz als ein Maß für die Unsicherheit des Lifts, so lässt

<sup>43</sup>Im weiteren Verlauf wird  $n_{A \wedge B}$  auch als absoluter Support bezeichnet

sich der SafeLift als ein um die Unsicherheit bereinigter Lift auffassen<sup>44</sup>. Wendet man die Kriterien aus Tabelle 4.13 auf die gefundenen Zusammenhänge an, so werden insgesamt 3043 (interessante) Assoziationsregeln identifiziert, wobei 2882 einfachen Regeln (mit nur einem Ereignis als Einflussvariable) und 221 komplexen Regeln (Regeln mit zwei Einflussgrößen) entsprechen. Zu den komplexen Regeln ist zu bemerken, dass lediglich fünf Assoziationen durch eine andere logische Verknüpfung als die logische Konjunktion ( $A_1 \wedge A_2 \rightarrow B$ ) entstanden sind. Auszüge aus den gefundenen Zusammenhänge der Felddaten Analyse sind in Tabelle 4.14 dargestellt. Die Schadensschlüssel wurden in der Tabelle nicht in Form der siebenstelligen Nummer dargestellt, sondern es wurde lediglich die jeweilige Konstruktionsgruppe des Schlüssels wiedergegeben.

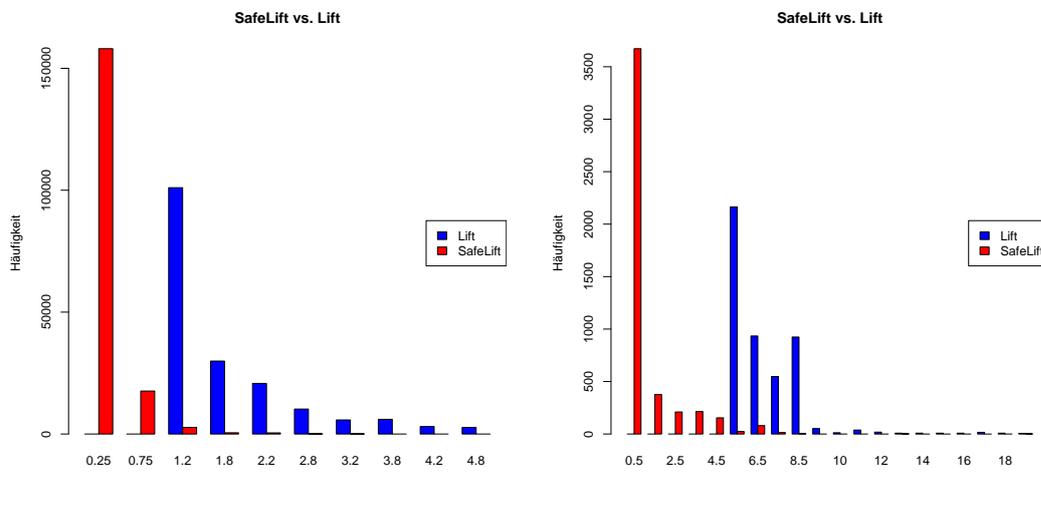


Abbildung 4.15: Histogramm für Lift und SafeLift für Assoziationsregeln mit Lift  $\leq 5$  (links) und Lift  $> 5$  (rechts)

Um den Vorteil des SafeLifts gegenüber dem herkömmlichen Lift zu zeigen, wurden in Abbildung 4.15 die Histogramme von Lift und SafeLift aller häufiger Ereignisse miteinander verglichen. Wie aus den Abbildungen hervorgeht, gibt es eine Vielzahl von Regeln, die einen Lift größer als 1.0 haben (Regeln mit einem Lift kleiner als 1.0 wurden nicht dargestellt). Folglich existiert für jede im Histogramm abgebildete Regel ein Zusammenhang zwischen der Regelprämisse und -konklusion (vgl. Kapitel 4.2.2). Betrachtet man jedoch die Häufigkeitsverteilung des SafeLifts so wird deutlich, dass die wenigsten dieser Regeln statistisch signifikant sind. Ein großer Anteil der in 4.15 abgebildeten Zusammenhänge besitzt einen SafeLift der kleiner als 1.0 ist.

<sup>44</sup>Diese Interpretationsweise lässt sich auch auf die Definition des SafeLifts mittels des Agresti-Cull-Intervalls übertragen

Eine weitere Facette bei der Identifizierung von interessanten Zusammenhängen ist die Berücksichtigung von wirtschaftlichen Aspekten. Die Suche nach wirtschaftlich “sinnvollen“ Regeln bedingt dabei einen Zielkonflikt zwischen einer quantitativen und einer qualitativen Betrachtungsweise gefundener Zusammenhänge. Denn zum einen können viele statistisch signifikante Regeln umgesetzt werden, auf der anderen Seite gilt es diejenigen Zusammenhänge zu identifizieren, die für die Anwendung einen möglichst großen (subjektive) Benefit haben (siehe Kapitel 4.2.3). Das in (4.30) definierte Potential bietet eine Möglichkeit, wirtschaftlich sinnvolle Regel zu identifizieren. Ordnet man die gefundenen, signifikanten Regeln anhand ihres absteigenden Potentials, so erlaubt dies, eine aus wirtschaftlichem Gesichtspunkt sinnvolle Schwelle (Cut-Off) zu bestimmen, die die gefundenen Zusammenhänge in zwei Mengen einteilt:

1. Statistisch signifikante Regeln, deren Umsetzung wirtschaftlich sinnvoll ist.
2. Statistisch signifikante Regeln, deren Umsetzung nicht wirtschaftlich sinnvoll ist.

Im Kontext der Felddaten Analyse versteht man unter der Umsetzen einer Regel die Veränderung bzw. Verbesserung des Testprozesses aufgrund eines identifizierten Zusammenhangs. Betrifft dieser viele Fahrzeuge, bzw. sind die Kosten einer Umsetzung geringer als die, den Fehler nicht zu beheben, so kann man in diesem Zusammenhang von einer wirtschaftlich sinnvollen Regel sprechen.

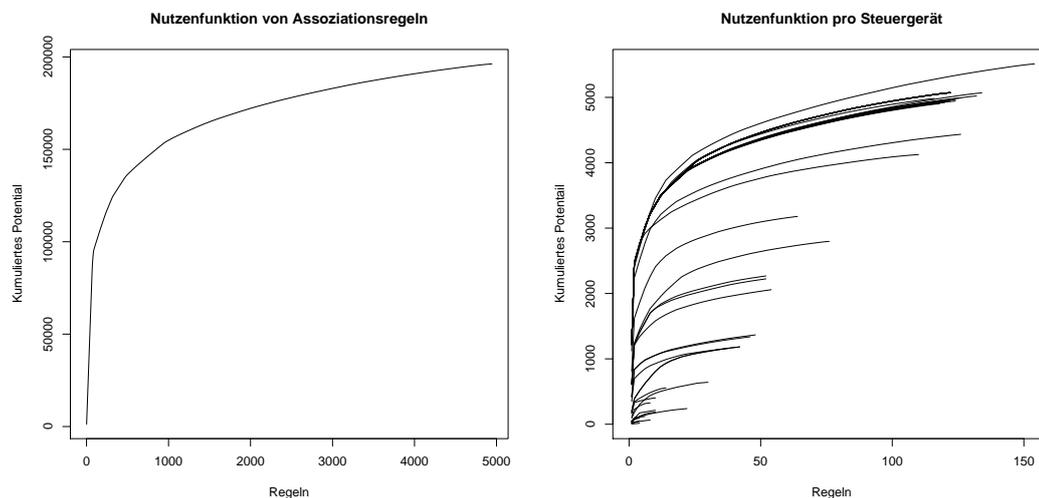


Abbildung 4.16: Nutzenfunktion der gefundenen Zusammenhänge (kumulierte Anzahl von Assoziationsregeln vs. kumuliertes Potential) aggregiert über alle Steuergeräte (links) und für jedes Steuergerät einzeln (rechts)

In der rechten Grafik der 4.16 ist für eine Wirtschaftlichkeitsbetrachtung<sup>45</sup> das kumulierte Potential aller gefundenen, interessanten Regeln dargestellt. Die abgebildete Funktion lässt sich dabei als die Nutzenfunktion einer weiteren Assoziationsregel interpretieren und die Steigung als der abnehmende Grenznutzen. Die Steigung der Nutzenfunktion kann beispielsweise mit den durchschnittlichen Kosten der Umsetzung einer Regel verglichen werden. Ein optimaler Cut-Off Punkt wäre diejenige Regel, bei der die Steigung der Grenznutzenfunktion gleich der der kumulierten Kostenfunktion ist. Regeln die links von dem Cut-Off Punkt liegen, wären wirtschaftlich sinnvoll, rechts davon liegenden Regeln entsprechend nicht. Dabei ist zu beachten, dass in der industriellen Praxis eine derartige, strikte Trennung nicht immer möglich sein wird, da Faktoren wie beispielsweise Imageschaden diese Art von wirtschaftlichen Überlegungen überwiegen.

Da die Kosten für Umsetzung einer Regel sich für die unterschiedlichen Steuergeräte stark voneinander unterscheiden, ist es sinnvoll, die Steigung der Grenznutzenfunktion der einzelnen Steuergeräte mit den jeweiligen unbekanntem, kumulierten Umsetzungs- bzw. Behebungskosten zu vergleichen. Da dies in der vorliegenden Anwendung angesichts von fehlenden Informationen über anfallende (Umsetzungs-)Kosten nicht möglich ist, wurde im linken Graph der Abbildung 4.16 das kumulierte Potential für die 61 verschiedenen Steuergeräte einzeln dargestellt. Dabei wird ersichtlich, dass die Nutzenfunktionen für die verschiedenen Steuergeräte unterschiedlich sind. Somit würde sich unter der ausschließlichen Berücksichtigung von Wirtschaftlichkeitsaspekten, die Umsetzung gefundener Zusammenhänge für manche Steuergeräte nicht lohnen.

In Tabelle 4.14 sind beispielhaft 11 interessante Assoziationsregeln, die vom beschriebenen Testalgorithmus identifiziert wurden, aufgelistet. Zur besseren Lesbarkeit sind die Ergebnisse für ein Steuergerät (Dieselsteuergerät) und als Zielgröße, die im Feld aufgetretenen Probleme mit der Ölpumpe dargestellt. Die Einflussgrößen spiegeln dabei die verschiedenen Ereignisse der Historie eines Steuergerätes wider (Fehler sowie Ergebnisse, die im Laufe der Tests parallel zur Produktion erfasst wurden), wobei die folgende Notation zugrunde liegt:

- *Fehler* oder *Warnungen*, die während des Tests aufgetreten sind, sind durch Nummern codiert. Warnungen unterscheiden sich von Fehlern durch ein vorangestelltes "X". Bis auf Zeile 9 in Tabelle 4.14 stellen alle Einflussgrößen gefundene Fehler bzw. Warnungen dar
- Logische Operatoren zur Bildung komplexer Assoziationsregeln sind *UND* bzw. *ODER* und deren Negation (Zeile 5)

<sup>45</sup>Konkrete Kosten für die Behebung bzw. Umsetzung eines gefundenen Zusammenhangs sind nicht bekannt. Gleiches gilt für die Folgekosten. Aus diesem Grund wird im letzteren Fall mit dem Wert ein für die Folgekosten gerechnet

- Durch das Schlüsselwort *Multi* sind Fehler (Warnungen) gekennzeichnet, die auch in anderen Tests für das betrachtete Steuergerät aufgetreten sind
- Führte ein Fehler in einem Steuergerät unmittelbar dazu, dass das Fahrzeug aus der Produktionslinie genommen und zur Nacharbeit gebracht wurde, so wird dies durch das vorangestellte Schlüsselwort *NA* dargestellt

ID	Einfluss	$n_{AB}$	$n_B$	$n_A$	$n$	SafeLift	Potential
1	0115.1 UND X341	5	1671	9	22170	3.53	4.32
2	X341 ODER X332	12	1671	56	22170	3.14	4.25
3	115.1	5	1671	13	22170	2.33	4.02
4	X341	7	1671	22	22170	2.15	5.34
5	$\neg$ 1436.1 & $\neg$ NA 559	5	1671	15	22170	1.99	3.87
6	Multi X902	13	1671	56	22170	1.85	8.78
7	X902	18	1671	90	22170	1.72	11.22
8	X332	9	1671	38	22170	1.70	6.14
9	Montag	12	4644	27	22170	1.32	6.34
10	1436.1	6	1671	28	22170	1.31	3.89
11	NA 559	5	1671	20	22170	1.43	3.49

Tabelle 4.14: Beispielergebnisse Felddaten Analyse

## Abschätzung der Vertrauenswürdigkeit

Bei dem im zweiten Schritt des Testalgorithmus durchgeführten Test auf Signifikanz, wird für jede häufige Ereigniskombination ein Hypothesentest durchgeführt. Da eine große Anzahl von Kombinationen und damit möglichen Zusammenhängen auf diese Weise bewertet wird, kommt es zum Problem der *Alpha-Kumulierung*, bzw. dem so genannte *multiplen Hypothesenproblem*. In der Literatur werden unterschiedliche Verfahren vorgestellt, mit dessen Hilfe das multiple Hypothesenproblem bzw. die dabei gemachte Fehlerrate kontrolliert wird. In Kapitel 4.3 wurden dazu die in der Literatur gängigen Methoden und Verfahren erläutert. Für den Anwender ist jedoch letztendlich entscheidend zu wissen, wie zuverlässig bzw. unzuverlässig die gefundenen Zusammenhänge sind. Der in Abschnitt 4.3.1 vorgestellte *alternative Ansatz* kontrolliert nicht wie die bekannten Verfahren eine Fehlerrate (FWER oder FDR), sondern bewertet die Zuverlässigkeit der gefundenen Zusammenhänge. Mit anderen Worten: das alternative Verfahren liefert eine Abschätzung für den Anteil fälschlicherweise abgelehnter Nullhypothesen.

Im weiteren Verlauf wird die Abschätzung der Vertrauenswürdigkeit der Felddaten Ana-

lyse Ergebnisse für einfache Assoziationsregeln vorgestellt. Dabei werden verschiedene Szenarien betrachtet, um das prinzipielle Zusammenspiel zwischen interessanten und zuverlässigen Regeln, in Abhängigkeit der jeweils getroffenen Voraussetzungen zu erläutern. Zum besseren Verständnis, seien an dieser Stelle die Bezeichnungen für die betrachteten Regelmengen wiederholt:

Unter interessanten Regeln werden all diejenigen Regeln zusammengefasst, für die gilt:

- Aufgrund des Hypothesentests lässt sich auf einen signifikanten Zusammenhang zwischen Prämisse und Konklusion schließen. Im vorliegenden Fall sind es 25986 Regeln, welche dieses Kriterium erfüllen.
- Die Assoziationsregeln erfüllen die Mindestanforderungen an SafeLift oder Potential (z.B. 4.13).

Eine Teilmenge der interessanten Regeln sind die zuverlässigen Regeln. Sie erfüllen zusätzlich die Anforderungen, die sich aus dem alternativen Ansatz ergeben.

Implementiert man das auf Seite 141 vorgestellte Verfahren zur Abschätzung der Zuverlässigkeit (alternativer Ansatz) für die Felddaten Analyse, so stellt man fest, dass das Verhältnis von zufälligen Zusammenhängen  $L(\alpha)$  zu gefundenen Regeln  $R(\alpha)$  für die einzelnen Steuergeräte variiert. Zum einen gibt es Steuergeräte, bei denen der Testalgorithmus nur wenige zufällige Regeln identifiziert hat. Folglich muss in einem solchen Fall das Signifikanzniveau nicht weiter angepasst werden. Es können somit alle gefundenen interessanten Zusammenhänge als vertrauenswürdig eingestuft werden. Zum anderen gibt es Steuergeräte, bei denen eine Anpassung des Signifikanzniveaus notwendig ist, da der Anteil zufälliger Regeln über dem vom Anwender akzeptierten Niveau liegt. Da für die vorliegende Anwendung keine konkreten Informationen zu diesem "Anwendungsniveau" vorliegen, wird die in Gleichung (4.40) gegebene Schranke für das Verhältnis von zufälligen Regeln an der Gesamtanzahl gefundener interessanter Regeln verwendet<sup>46</sup>.

Zur Verdeutlichung ist in Abbildung 4.17 die Anzahl sowohl für interessante als auch zuverlässige Regeln der betrachteten Steuergeräte abgebildet. Wie aus dem Balkendiagramm ersichtlich ist, ist die Anzahl interessanter Zusammenhänge für die einzelnen Steuergeräte stark unterschiedlich und somit auch die geschätzte Anzahl von Regeln, die als zufällig geschätzt werden. Letztere wurden in der Grafik nicht absolut, sondern auf der Sekundärachse als das Verhältnis  $Q(\alpha)$  von unzuverlässigen Regeln  $L(\alpha)$  zu gefundenen Regeln  $R(\alpha)$  dargestellt. Fasst man die horizontale, rote Linie als die obere

<sup>46</sup>In Abschnitt 4.3.2 wurde gezeigt, dass dieses Verhältnis ein ähnliches Ergebnis wie die Anpassung des Signifikanzniveaus mit Hilfe des TST-Verfahrens liefert. Dabei wurde ein ebenfalls Signifikanzniveau  $\alpha = 0.05$  zugrunde gelegt

Schranke für  $Q(\alpha)$  auf (analog zu (4.40)), so erfüllen lediglich 11 Steuergeräte dieses Kriterium nicht. Mit anderen Worten, in 50 Fällen<sup>47</sup> ist der Anteil von zufälligen Regeln kleiner oder gleich 12 Prozent, so dass das Signifikanzniveau nicht angepasst werden muss ( $\tilde{\alpha} = 0.05$ ). In den übrigen zwölf Fällen muss das Signifikanzniveau geeignet angepasst werden. Führt man die auf Seite 4.3.2 beschriebene Anpassung des Niveaus durch, so erhält man letztendlich 2777 zuverlässige Regeln (vgl. grau schraffierte Fläche im Balkendiagramm) zu einem durchschnittlichen Signifikanzniveau  $\tilde{\alpha} = 0.00037$ . Ohne die Anpassung würden 2822 Zusammenhänge als interessant eingestuft werden.

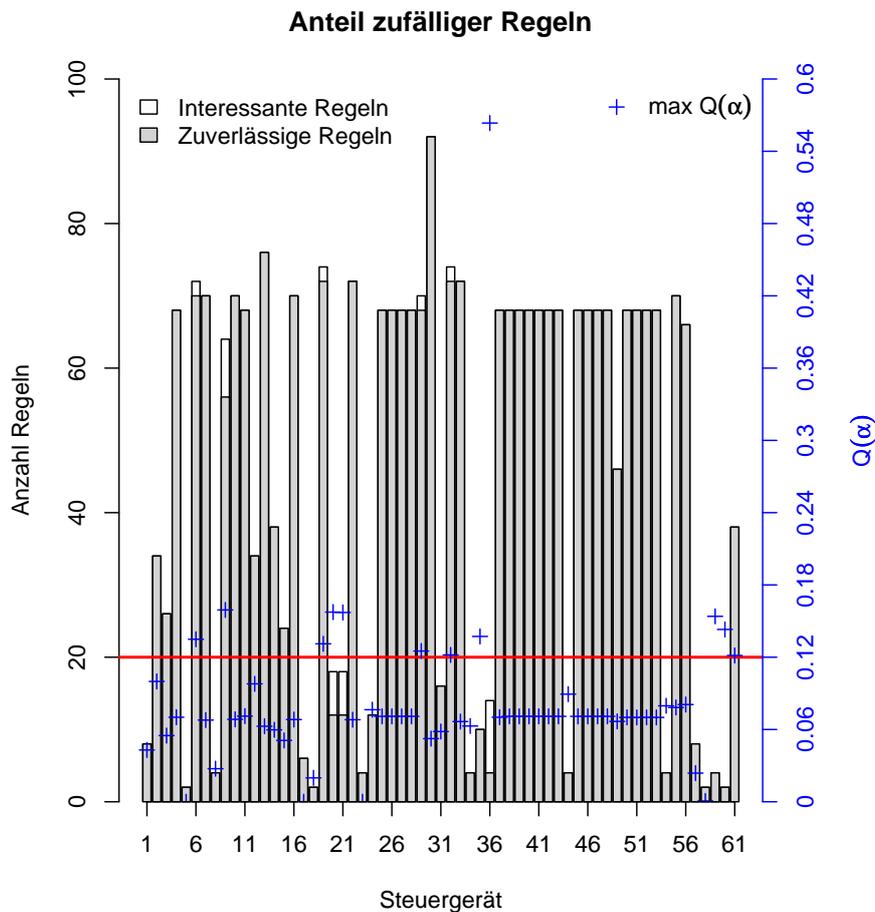


Abbildung 4.17: Gefundene, interessante Regeln vs. zuverlässige Regeln für SafeLift > 1.3 und Potential > 3

Der Anteil zuverlässiger Regeln hängt dabei auch von den Interessantheitsmaßen, die zur

<sup>47</sup>Insgesamt wurden 61 Steuergeräte untersucht

Einschränkung der Zusammenhänge verwendet werden, ab. Bis jetzt wurden nur diejenigen Regeln auf ihre Zuverlässigkeit im Kontext des multiplen Hypothesenproblems untersucht, die den Forderungen aus Tabelle 4.13 genügen (Szenario 1). Fordert man hingegen, dass eine Regel genau dann interessant ist, wenn ihr SafeLift größer als 2 ist (unabhängig von einem subjektiven Interessantheitsmaß wie dem Potential), so musste für den betrachteten Datensatz das Signifikanzniveau  $\alpha$  nicht weiter angepasst werden (Szenario 2). Mit anderen Worten: alle Regeln, welche diese Bedingung erfüllen, besitzen einen so geringen  $p$ -Wert, dass der Anteil an fälschlicherweise abgelehnten Zusammenhängen vernachlässigbar bzw. vom Anwender im Sinne der Forderung (4.40) tolerierbar ist. Wendet man dieses Kriterium auf den vorliegenden Datensatz an, so erhält man 1398 interessante und eben so viele zuverlässige Regeln.

Anders verhält es sich, wenn man die Forderung nach einem Mindestwert für den SafeLift einer Regel durch einen Mindestwert für das Potential ersetzt (Szenario 3). In diesem Fall werden zwar im ersten Schritt des Testalgorithmus diejenigen Assoziationsregeln identifiziert, die aufgrund des statistischen Unabhängigkeitstests als signifikant bewertet werden, jedoch wird die Stärke des Zusammenhangs nicht wie beim SafeLift statistisch gemessen. Folglich werden mehr Regeln als interessant identifiziert, die jedoch einen geringeren Anteil vertrauenswürdiger Regeln im Hinblick auf die Kumulierung des Fehlers 1. Art bedingen. Dies wird in Abbildung 4.18 deutlich. Dort wird ersichtlich, dass bis auf drei Ausnahmen, für alle Steuergeräte Regeln identifiziert wurden, die nach dem vorliegenden Kriterium nicht zuverlässig sind. Zur Übersicht sind in Tabelle 4.15 die Mengengerüste, durchschnittliche Interessantheitsmaße sowie weitere charakteristische Eigenschaften für die betrachteten Szenarien aufgelistet.

Abschließend lässt sich zusammenfassen, dass Szenario 1 ein guter Kompromiss zu sein scheint, da mittels des SafeLifts genügend nicht signifikante Regeln eliminiert werden, jedoch die Möglichkeit besteht, mittels subjektiver Interessantheitsmaße wie beispielsweise dem Potential, die Identifizierung von Zusammenhängen zu steuern. Eine Anpassung des Signifikanzniveaus ist in wenigen Fällen trotzdem notwendig.

	Szenario 1		Szenario 2		Szenario 3	
	Interes.	Zuverl.	Interes.	Zuverl.	Interes.	Zuverl.
Anzahl Regeln	2822	2777	1398	1398	15166	3925
Potential	55.55	56.34	87.60	87.60	16.14	46.25
SafeLift	2.88	2.90	4.38	4.38	1.04	2.35
p-Wert	0.00050	0.00030	0.00003	0.00003	0.04000	0.00100
max Potential	1245.36	1245.36	1245.36	1245.36	1245.36	1245.36
min Potential	3.24	3.70	2.83	2.83	3.00	3.30
max SafeLift	23.72	23.72	28.26	28.26	23.72	23.72
min SafeLift	1.30	1.30	2.00	2.00	0.00	0.00

Tabelle 4.15: Charakteristische Eigenschaften für interessante und zuverlässige Assoziationsregeln der Szenarien 1 bis 3

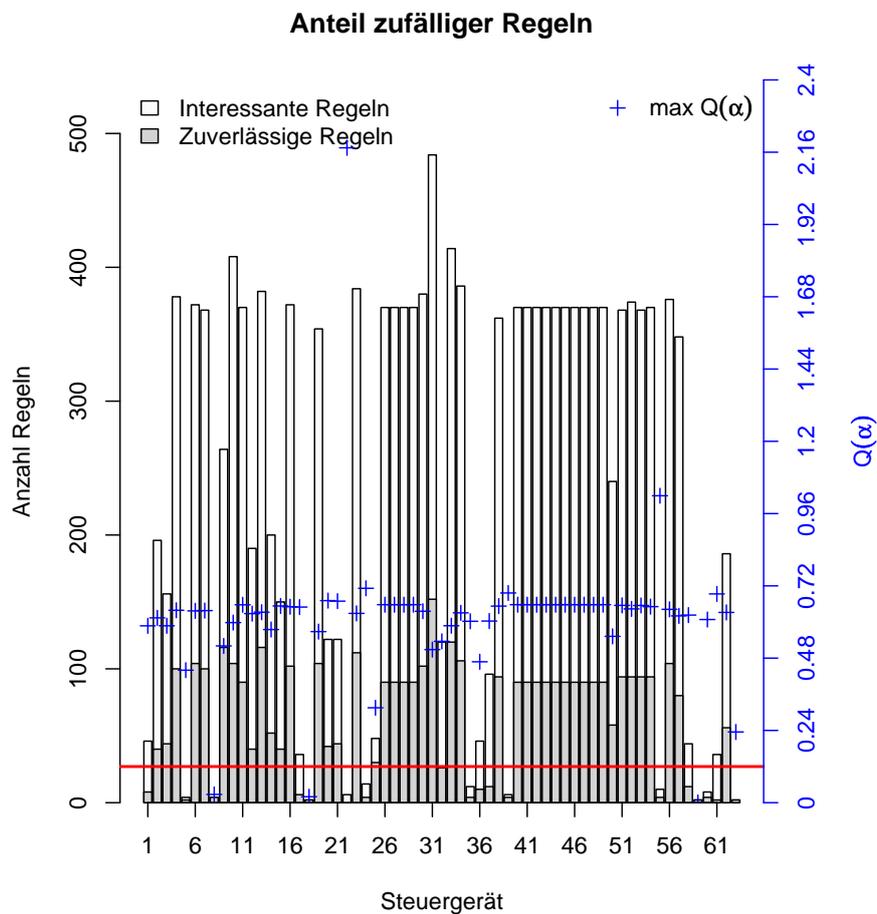


Abbildung 4.18: Gefundene, interessante Regeln vs. zuverlässige Regeln.

## 4.5 Zusammenfassung und Bewertung

In vielen Anwendungen des Qualitätsmanagements sind die Daten bzw. Variablen, welche die Qualität der gefertigten Produkte messen nicht wie in den ersten zwei Kapiteln dieser Arbeit numerischer Natur, sondern nominal. Will man nun diese nominalen, qualitätsmessenden Daten durch nominale (Produktions-)Faktoren beschreiben, so kann dies mittels der Assoziationsanalyse geschehen. Grundlage dafür sind die in Abschnitt 4.1 eingeführten Assoziationsregeln, mit deren Hilfe Zusammenhänge zwischen Einflussfaktoren (Prämisse oder Einflussvariable) und qualitätsmessenden Zielgrößen (Konklusion) in Form von Implikationen beschrieben werden können.

Die Kernproblematik bei der Extraktion von Wissen mittels Assoziationsregeln ist, dass es eine Vielzahl unterschiedlicher Zusammenhänge zwischen den zu analysierenden Ziel- und Einflussgrößen gibt. Jedoch sind nicht alle Regeln, die einen Zusammenhang zwischen Prämisse und Konklusion beschreiben, für die Anwendung relevant. Vielmehr sind nur diejenigen Regeln von Belang, die sowohl *statistisch signifikant* als auch *interessant* für die Anwendung sind. Um solche Informationen aus den Daten zu extrahieren wurden in diesem Kapitel verschiedene bekannte und auch neue statistische Verfahren zur Generierung interessanter Assoziationsregeln motiviert und kombiniert. Mit Hilfe von Simulationsstudien wurden anschließend diejenigen Methoden identifiziert, welche am geeignetsten für die Analyse nominaler Daten sind. Die so ermittelten Verfahren wurden dann zu einem Testalgorithmus kombiniert, mit dessen Hilfe es möglich ist, interessante Informationen aus nominalen Daten zu extrahieren. Der Testalgorithmus besteht im Wesentlichen aus den folgenden drei Schritten:

Um die Menge an möglichen Assoziationsregeln sinnvoll einzuschränken, werden im ersten Schritt des Testalgorithmus diejenigen Regeln identifiziert, für die ein statistischer Zusammenhang zwischen der Prämisse und der Konklusion existiert. Wie gezeigt wurde, ist es dazu notwendig, zunächst all diejenigen Kombinationen von Ziel- und Einflussgrößen zu ermitteln, die häufig sind und anschließend mittels eines Hypothesentests auf ihre statistische Signifikanz hin zu untersuchen. Eine Möglichkeit, einen statistischen Zusammenhang zu identifizieren, ist die Verwendung von Unabhängigkeitstests, wie beispielsweise der  $\chi^2$ - oder Binomialtests (vgl. Seite 109). Der Nachteil dieser Tests ist jedoch, dass sie lediglich eine Aussage über die Existenz eines Zusammenhangs zwischen Prämisse und der Konklusion einer Regel treffen können, jedoch nichts darüber aussagen, wie stark dieser Zusammenhang ist. Um diese Stärke bewerten zu können, müssen im zweiten Schritt des Testalgorithmus die in 4.2 eingeführten Interessantheitsmaße verwendet werden. Der dritte und letzte Schritt berücksichtigt die statistischen Implikationen bei der Bewertung von Zusammenhängen in großen Datenmengen, mittels des in Abschnitt 4.3 neu eingeführten Modells.

Interessantheitsmaße lassen sich prinzipiell in zwei Kategorien einteilen. Zum einen in statistische Maße, deren Aussagekraft in objektiven, statistischen Eigenschaften der Assoziationsregel begründet liegt und zum anderen in subjektive Maße. Das in der Literatur bekannteste statistische Interessantheitsmaß ist der Lift. Er gibt an, um wie viel häufiger die Konklusion und Prämisse gemeinsam in einem Datensatz enthalten sind gegenüber der Häufigkeit, dass ein Datensatz lediglich die Konklusion enthält. Folglich kann der Lift als ein Maß für die Abweichung von der Unabhängigkeit zweier Variablen interpretiert werden. Wie in Abschnitt 4.2 dargestellt wurde, besitzen die betrachteten Maße den Nachteil, dass sie statistische Schätzer sind und folglich mit Unsicherheit behaftet sind. Somit ist es notwendig, den Anteil an Unsicherheit zu reduzieren wie beispielsweise durch die Konstruktion eines Konfidenzintervalls für das Maß. Um eine statistisch abgesicherte Aussage sowohl über die Stärke als auch die Signifikanz einer Assoziationsregel treffen zu können, kann dann die untere Grenze des Intervalls als konservative Schätzung verwendet werden. Im Kontext der Assoziationsanalyse liegt den betrachteten Schätzer für die Interessantheit (z.B. Lift oder Support) stets eine binomial-verteilte Grundgesamtheit zugrunde. Für binomial-verteilte Zufallsvariablen existieren in der Literatur eine Vielzahl unterschiedlicher Konfidenzintervalle. Trotz dieser Vielfalt gibt es kein Intervall, welches allgemein als das Beste anerkannt ist, da jedes Stärken und Schwächen besitzt. Um für die Bedürfnisse der Assoziationsanalyse ein bestes Intervall zu identifizieren wurde eine Evaluation der vorgestellten Intervalle vorgenommen, wobei als Gütekriterien die Überdeckungswahrscheinlichkeit  $CP(n, p)$ , das Fehlerverhalten  $Err_{\alpha, n}(p)$  und die leichte Implementierbarkeit verwendet wurden. Als Ergebnis dieser Evaluation wurde das Agresti-Coull-Konfidenzintervall als am geeignetsten identifiziert, da es zum einen eine hohe Überdeckungswahrscheinlichkeit besitzt, ein besseres Fehlerverhalten aufweist und vor allem leicht zu implementieren ist. Mit Hilfe der linken Intervallgrenze dieses Intervalls für den Lift wurde in (4.29) ein neues Maß definiert, welches die Eigenschaften eines Unabhängigkeitstests (Binomialtest) und des Lifts zu einem neuen Interessantheitsmaß, dem so genannten SafeLift kombiniert.

Aus Sicht des Anwenders spielen subjektive Interessantheitsmaße oftmals eine ebenso wichtige Rolle wie statistische Maße. Wie der Begriff Subjektivität suggeriert, können diese Maße in unterschiedliche Facetten, welche jeweils eine (subjektive) Sicht des Anwenders auf gefundene Assoziationsregeln widerspiegeln, gruppiert werden. So sind beispielsweise für das Qualitätsmanagement Maße, welche den monetären Wert einer Regel bewerten, ähnlich bedeutend wie Lift, Support oder SafeLift. Denn bei jedem gefundenen, statistischen signifikanten Zusammenhang muss sich der Anwender mit der Frage nach den Kosten für die Umsetzung (Behebung) des Problems, welches mittels der Assoziationsregel identifiziert wurde, auseinandersetzen. Im Allgemeinen sind die Kosten für die Behebung eines Fehlers (im Sinne eines identifizierten Zusammenhangs) nicht bekannt, oder nur sehr schwer abzuschätzen. Um trotzdem die Möglichkeit zu haben Wirtschaftlichkeitsaspekte bei der Extraktion von interessanten Zusammenhängen zu berücksichtigen, wurde in Abschnitt 4.2.3 ein neues Maß eingeführt, welches es er-

möglichst, auch ohne konkrete Informationen über Behebungskosten, das *Potential* einer Regel zu bestimmen.

Im ersten Schritt des Testalgorithmus wird mittels statistischer Unabhängigkeitstests überprüft, ob zwischen Prämisse und Konklusion einer Assoziationsregel ein signifikanter Zusammenhang existiert. Folglich muss für jeden gefundenen Zusammenhang ein statistischer Test, wie beispielsweise der Binomialtest, durchgeführt werden. In den meisten Anwendungen, bei denen Wissen mittels Assoziationsregeln extrahiert werden soll, werden die betrachteten Datensätze durch eine Anzahl von unterschiedlichen Ziel- und Einflussgrößen beschrieben. Aufgrund der hohen Anzahl durchgeführter Tests kommt es zu dem in Abschnitt 4.3 beschriebenen Problem der Kumulierung des Fehlers 1. Art bzw. dem multiplen Hypothesenproblem. In der Literatur gibt es eine Vielzahl von Verfahren, mit deren Hilfe die Auswirkungen des multiplen Hypothesenproblems eingeschränkt werden können. Allen hier vorgestellten Verfahren liegt dabei die Idee zugrunde, eine, beim multiplen Testen gemachte Fehlerrate (FDR oder FWER), auf einem, vom Anwender vorgegebenen Niveau zu kontrollieren. Mit Hilfe dieser Methoden zur Fehlerratenkontrolle wird das ursprüngliche Signifikanzniveau  $\alpha$  a posteriori modifiziert und es werden nur diejenigen Zusammenhänge als signifikant eingestuft, die dem angepassten Niveau genügen.

Im Gegensatz zu anderen Anwendungen, wie beispielsweise Microarray Studien<sup>48</sup>, bei denen ebenfalls das multiple Hypothesenproblem auftritt, werden bei der Assoziationsanalyse, neben Unabhängigkeitstests, auch Interessantheitsmaße und ggf. Expertenwissen zur Identifizierung von interessanten Zusammenhängen verwendet. Unter Umständen reduzieren diese die gefundenen Regeln in einer solchen Art und Weise, dass das multiple Hypothesenproblem in solchen Fällen nicht (mehr) relevant ist. Für den Anwender stellt sich somit die Frage, wie vertrauenswürdig sind die gefundenen Zusammenhänge, bzw. wie groß ist der Anteil zufälliger Regeln an denjenigen Regeln, die der Testalgorithmus nach Verwendung von Binomialtest, statistischer und subjektiver Interessantheitsmaße, identifiziert hat. Um dies einschätzen zu können, wurde auf Seite 141 und ff. ein Modell zur Abschätzung der Vertrauenswürdigkeit, der durch den Testalgorithmus gefundenen Zusammenhänge, motiviert und entwickelt. Die Grundidee dieses *alternativen Ansatzes* ist, den Anteil fälschlicherweise identifizierter Zusammenhänge mittels eines Konfidenzintervalls geeignet abzuschätzen und somit eine Aussage über die Vertrauenswürdigkeit der gefundenen Regeln abzuleiten. Der Nachteil des alternativen Ansatzes ist, dass es mit seiner Hilfe nicht möglich ist, Fehlerraten wie beispielsweise die FDR explizit zu kontrollieren. Jedoch wurde in dem anschließend durchgeführten Vergleich der verschiedenen Verfahren empirisch gezeigt, dass wenn der Anteil an zufälligen Zusammenhängen zu tatsächlich gefundenen Zusammenhängen einen vorgegeben Pro-

<sup>48</sup>Ein Sammelbegriff für bio-statistische Verfahren, mit deren Hilfe Informationen, die in Chromosomensätzen enthalten sind, identifiziert werden können

zentsatz nicht überschreitet, man ähnliche Ergebnisse wie bei der Kontrolle der FDR durch das TST-Verfahren erhält. Für den Anwender ist der alternative Ansatz eine einfache Möglichkeit, die Vertrauenswürdigkeit der gefundenen Zusammenhänge abzuschätzen und ggf. das Signifikanzniveau anzupassen und bildet somit den dritten Schritt des Testalgorithmus.

Im letzten Teil dieses Kapitels wurde der Testalgorithmus anhand einer Qualitätsanwendung aus dem Automobilssektor evaluiert. Ziel dieser Anwendung ist es, Fehler, die während des Gebrauches des Fahrzeuges beim Kunden aufgetreten sind, anhand von Produktionsdaten zu beschreiben, qualitätsrelevante Zusammenhänge zu identifizieren und anschließend in den Qualitätsverbesserungsprozess rückzukoppeln. Der dazu verwendete Datensatz, bestehend aus qualitätsmessenden Felddaten und qualitätsbeschreibenden Produktionsdaten, wurde dabei so gewählt, dass zum einen der Testalgorithmus verifiziert, und zum anderen die prinzipielle Vorgehensweise exemplarisch aufgezeigt werden konnte. Ersteres war möglich, indem ein Datensatz gewählt wurde, der für den Anwender schon bekannte Zusammenhänge enthielt. Da die Menge, die durch den Algorithmus identifizierten Regeln, eine Teilmenge der schon bekannten Zusammenhänge beschreibt, konnte der Anwender die Richtigkeit des Testalgorithmus verifizieren. Zur Demonstration der Vorgehensweise wurden die Ergebnisse der sukzessiven Reduktion gefundener Zusammenhänge, die der Testalgorithmus in den einzelnen Schritten produziert, verwendet. Wie in Abschnitt 4.4.2 dargestellt, wurden mittels des Algorithmus die Anzahl der Regeln derart reduziert, dass letztendlich noch 1.1 Prozent der ursprünglichen Zusammenhänge übrig blieben und als interessant für die Anwendung eingestuft werden konnten.

Aufgrund der hohen Anzahl von Ziel- und Einflussgrößen, die analysiert wurden, musste das Signifikanzniveau für die im ersten Schritt verwendeten Unabhängigkeitstests a posteriori angepasst werden, um die beim multiplen Testen gemachte Fehlerrate zu kontrollieren. Dazu wurde die Zuverlässigkeit der Regeln, die man nach dem 2. Schritt des Algorithmus erhält, mittels des alternativen Ansatzes bewertet. Die Ergebnisse der Evaluation bestätigten die eingangs getroffene Vermutung, dass in den Fällen, in denen eine Regel nur dann interessant ist, wenn der Wert des SafeLifts größer oder gleich einer vorgegebenen Schranke ist, eine Anpassung des Niveaus nicht zwingend notwendig ist. Aufgrund der Eigenschaften des SafeLifts werden nur diejenigen Regeln als interessant bewertet, deren Zusammenhang sowohl statistisch signifikant als auch stark ist. Je höher dieser Mindestwert für den SafeLift gewählt wird, desto unwahrscheinlicher ist es, dass Zusammenhänge identifiziert werden, die fälschlicherweise abgelehnt wurden.

Anders verhält es sich bei der Verwendung von subjektiven Interessanztheitsmaßen, wie dem in Abschnitt 4.2.3 neu eingeführten Potential. Basiert die Entscheidung, ob eine Assoziationsregel für die Anwendung interessant ist oder nicht lediglich auf dem Potential, so muss das Signifikanzniveau wie in Abschnitt 4.3.1 gezeigt, angepasst werden. In

der Praxis wird man jedoch selten nur eine Art von Interessantheitsmaßen verwenden, so dass eine Kombination von subjektiven wie auch statistischen Maßen sinnvoll ist. Ideal für die hier vorgestellte Anwendung ist der Einsatz von SafeLift und dem Potential (als Repräsentant für subjektive Interessantheitsmaße) in Verbindung mit einem jeweiligen Mindestwert. Durch die Forderung, dass Regeln für beide Maße einen Mindestwert überschreiten müssen (z.B. SafeLift größer 1.3 und Potential größer 3), enthält die verbleibende Menge an Assoziationsregeln nur diejenigen Zusammenhänge, die statistisch Signifikant, interessant für die Anwendung aber auch ausreichend zuverlässig im Sinne des multiplen Hypothesenproblems sind.

# 5 Zusammenfassung und Ausblick

In der vorliegenden Ausarbeitung wurden zwei neue Konzepte und Modelle zur Bestimmung von qualitätsrelevanten, signifikanten Zusammenhängen dargestellt. Dazu wurden verschiedene Ansätze und Modelle aus den Bereichen Statistik, Mathematik, Qualitätsmanagement, Operations Research und Data Mining kombiniert, um unter den, im Eingang dieser Arbeit beschriebenen Rahmenbedingungen, effiziente Verfahren zur Qualitätsoptimierung zu entwickeln. Im Folgenden wird eine Zusammenfassung der Ergebnisse gegeben, wobei unter anderem auf die Anforderungen, die den wissenschaftlichen Mehrwert dieser Arbeit widerspiegeln, eingegangen wird. Den Abschluss bildet ein Ausblick auf mögliche weitere Themen und Stoßrichtungen, die sich aufgrund dieser Ausarbeitung ergeben.

Die Grundidee dieser Arbeit ist, Daten, die entlang des Wertschöpfungsprozesses eines Produkt erfasst werden, geeignet zu kombinieren und entsprechend zu analysieren, so dass die aus der Analyse gewonnenen Informationen in den unternehmensinternen Qualitätsverbesserungsprozess rückgekoppelt werden können. Dabei liegt die Betrachtungsweise zugrunde, dass sich die zu analysierenden Daten in zwei unterschiedliche Klassen einteilen lassen. Die erste Klasse bilden die qualitätsbeschreibenden Größen (Produktionsfaktoren bzw. -parameter). Sie werden im Laufe der Produktion erfasst und beschreiben den Zustand des Produktes zu einem gewissen Zeitpunkt des Entstehungsprozesses. Die zweite Klasse wird durch die qualitätsmessenden Größen beschrieben. Diese Werte werden zu einem beliebigen Zeitpunkt in der Wertschöpfungskette erfasst und messen die Qualität des zugrunde liegenden Produktes. Die Kernproblematik bei der Entwicklung von Modellen zur Analyse der oben beschriebenen Daten ist, dass die verwendeten Verfahren abhängig sind von der Art der zugrunde liegenden Daten. Diese Abhängigkeit bedingt die Gliederung der Arbeit in zwei Hauptteile.

Die Thematik des ersten Hauptteils ist die Entwicklung eines Verfahrens, mit dessen Hilfe numerische Messwerte, deren qualitätsbeschreibende Größen nominaler Natur sind, analysiert werden können. Dazu wurden zunächst statistische Schätzer zur Messung der Qualität, – *Qualitätsfähigkeitsindizes* – eingeführt. Als Ausgangspunkt wurden die in der Literatur bekannten Qualitätsfähigkeitsindizes vorgestellt. Sie gehen auf die Six-Sigma-Betrachtung der Statistischen Prozesslenkung (SPC) zurück und fordern, dass die zugrunde liegenden Daten einer bestimmten Verteilungsart folgen. Da in vielen prakti-

schen Anwendungen, die Verteilung der Messdaten unbekannt ist, wurde in Abschnitt 2.2 ein allgemeiner, nicht-parametrischer Ansatz modelliert, der die Einteilung von Produktionsparametern und den dazugehörigen Messwerten mit Hilfe eines Wahrscheinlichkeitsraumes und eines Zufallsprozesses formalisiert. Diese verteilungsfreie Modellierung bietet die theoretische Grundlage, die Ursachen-Wirkungs-Beziehungen zwischen den Messwerten und ihren Einflussgrößen in Form eines *Produktionsprozesses* bzw. *Sub-Prozesses* darzustellen. Aufbauend auf diesen theoretischen Grundlagen wird mit dem  $E_{ci}$  ein nicht-parametrischer Qualitätsfähigkeitsindex definiert, der eine Abschätzung der Qualität ohne Verteilungsvoraussetzung der zugrunde liegenden Daten ermöglicht.

Aufgrund der Tatsache, dass die in Kapitel 2 vorgestellten Qualitätsfähigkeitsindizes statistische Schätzer sind, ist es notwendig, ihre Aussagen geeignet abzusichern. Eine Möglichkeit dies zu tun, ist die Verwendung von Konfidenzintervallen. Wie gezeigt wurde, ist die Konstruktion von analytischen Konfidenzintervallen für parametrische Indizes oftmals schwierig, im Allgemeinen nur approximativ und für nicht-parametrische Indizes lediglich mit Hilfe der Bootstrap-Verfahrens möglich. Zur Bestimmung eines Bootstrap-Konfidenzintervalles können verschiedene Methoden verwendet werden, die, je nach Art des verwendeten Qualitätsschätzers, unterschiedlich geeignet sind. Eine abschließend durchgeführte Simulation (vgl. Abschnitt 2.3.3) fasst diese Thematik auf, indem die beste Methode zur Konstruktion von Konfidenzintervallen in Abhängigkeit von dem verwendeten Qualitätsfähigkeitsindex anhand verschiedener, künstlich generierter Produktionsprozesse bestimmt wird. Die Simulation ergab, dass es für den neu eingeführten, nicht-parametrischen Fähigkeitsindex  $E_{ci}$  nicht nötig ist, Bootstrap-Methoden zur Konstruktion eines Konfidenzintervalles zu verwenden. Für die in der Literatur verwendeten Qualitätsfähigkeitsindizes hingegen liefert die Verwendung von Bootstrap-Verfahren bessere Konfidenzintervalle.

Wie in den eingangs dieser Ausarbeitung beschriebenen Anforderungen, müssen die verwendeten Modelle zur Messung von Qualität den industriellen Gegebenheiten genügen. Durch die Einführung des nicht-parametrischen Qualitätsindex  $E_{ci}$  ist es nun möglich, unabhängig von der zugrunde liegenden Verteilung der (Mess-)Daten, die Qualität gefertigter Produkte zu messen. Doch erlauben Indizes alleine es nicht, die für die nachhaltige Qualitätssteigerung notwendigen Zusammenhänge zwischen qualitätsmessenden und -beschreibenden Größen zu bestimmen. Dieses übergeordnete Ziel der Arbeit ist Gegenstand von Kapitel 3. Dazu wurden zunächst klassische, bekannte Ansätze vorgestellt, die jedoch aufgrund ihrer restriktiven Vorgaben und ihrer statistischen Implikationen für die vorliegende Problemstellung ungeeignet sind. Aus diesem Grund wurde, aufbauend auf der im vorherigen Kapitel eingeführten Theorie der *Produktionsprozesse*, das *Root-Cause-Analysis*-Verfahren entwickelt. Legt man diese Theorie zugrunde, so kann ein zu analysierender Fertigungsprozess als eine zufällige Stichprobe interpretiert werden. Hierbei werden qualitätsbeschreibende und die dazugehörigen, qualitätsmessenden Größen als ein zufälliges Tupel auf einem Wahrscheinlichkeitsraum aufgefasst.

Damit lässt sich die Ursachen-Wirkungs-Beziehung zwischen numerischen Messwerten und nominalen Produktionsparametern als *Sub-Prozess* oder als Einschränkung der Variation der möglichen Produktionsabläufe darstellen. Sub-Prozesse lassen sich dabei als Wenn-Dann-Beziehungen, ähnlich zu den in Kapitel 4 eingeführten Assoziationsregeln, interpretieren. Unter Zuhilfenahme von Qualitätsfähigkeitsindizes ist es dann möglich, Qualität in Abhängigkeit von Teilmengen der Produktionsparameter zu messen. Mit Hilfe dieses Ansatzes konnte in einem nächsten Schritt gezeigt werden, dass das Ziel, die Extraktion von signifikanten, qualitätsrelevanten Zusammenhängen äquivalent zu der Identifikation geeigneter Sub-Prozesse ist.

Zur Identifikation eines geeigneten Sub-Prozesses wurde in Abschnitt 3.2 der Support als weiteres Kriterium für die Signifikanz eines Zusammenhangs eingeführt. Denn wie erläutert wurde, beschreiben Sub-Prozesse, die sowohl den höchsten Support als auch nach qualitativen Gesichtspunkten besser sind als andere Sub-Prozesse, einen solchen signifikanten Zusammenhang. Integraler Bestandteil dieser Betrachtungsweise ist, dass der Support als Kriterium für Wirtschaftlichkeit interpretiert werden kann und in Kombination mit der gemessenen Qualität in Form von Qualitätsfähigkeitsindizes eine Vergleichbarkeit der Sub-Prozesse ermöglicht. Diese Tatsache erlaubt es, die Suche nach einem signifikanten Sub-Prozess in ein Optimierungsproblem zu überführen, welches mit Hilfe von Branch and Bound Techniken effizient gelöst werden kann. Den Abschluss des Kapitels 3 bildet eine Validierung des Verfahrens. Dazu wurde ein realer Datensatz aus einer Gießerei verwendet und mit Hilfe des RCA-Verfahrens interessante Zusammenhänge zwischen den Produktionsparametern auf der einen und den gemessenen Qualitätswerten auf der anderen Seite bestimmt. Dabei konnten zum einen die Ergebnisse des RCA-Verfahrens durch einen Experten validiert und zum anderen die Verwendbarkeit in der industriellen Praxis erfolgreich getestet werden.

Sind im ersten Teil der Ausarbeitung die zu beschreibenden, qualitätsmessenden Größen numerischer Natur, so widmet sich der zweite Teil der Analyse rein nominaler Daten. Mit Hilfe eines mehrstufigen Testalgorithmus wurde eine Abfolge von interdisziplinären Verfahren und Methoden präsentiert, die es erlaubt signifikante, qualitätsrelevante Informationen aus dieser Art von Daten zu extrahieren. Dabei wurden die verwendeten Verfahren durch eine industrielle Anwendung zur Bewertung von Zusammenhängen zwischen Garantie- und Kulanzdaten auf der einen, und Produktionsdaten auf der anderen Seite motiviert. Diese Felddaten Analyse bildet auch die Grundlage für die im Anschluss durchgeführte Evaluation des Testalgorithmus.

Ausgangspunkt für die Entwicklung des Testalgorithmus ist die Darstellung von Ursache-Wirkungs-Beziehungen in Form von Assoziationsregeln. Speziell Klassifikationsregeln und deren Erweiterung – komplexe Assoziationsregeln – bieten eine geeignete Methode um die gesuchten Zusammenhänge zu beschreiben. Wie auch im Falle des RCA-Verfahrens ist die zentrale Herausforderung, die Vielzahl möglicher Regeln derart einzu-

schränken, dass letztendlich nur noch wenige signifikante Zusammenhänge übrig bleiben. Um dies zu gewährleisten, werden vom Testalgorithmus im Wesentlichen zwei Techniken verwendet: Zum einen statistische Unabhängigkeitstests und zum anderen *Interessantheitsmaße*. Letztere bieten die Möglichkeit, die Stärke eines Zusammenhangs entweder subjektiv oder statistisch zu bewerten.

Ähnlich zu den Qualitätsfähigkeitsindizes sind statistische Interessantheitsmaße Schätzer, deren Aussage mit Hilfe von Konfidenzintervallen abgesichert werden sollte. Dazu wurden in Abschnitt 4.2 unterschiedliche Methoden zur Konstruktion geeigneter Konfidenzintervalle für die eingeführten statistischen Interessantheitsmaße gegenübergestellt. Als Ergebnis konnte ein neues Maß, der so genannte *SafeLift*, motiviert werden, welcher die Eigenschaften eines statistischen Unabhängigkeitstests und des weit verbreiteten Lifts vereinigt. Des Weiteren wurde mit dem *Potential*, einem neu eingeführten, subjektiven Interessantheitsmaß, die Möglichkeit geschaffen, Assoziationsregeln nicht nur nach statistischen Gesichtspunkten, sondern auch unter Berücksichtigung von anfallenden Kosten zu bewerten. Dieses erlaubt die in der Einleitung formulierte Anforderung, qualitätsrelevante Zusammenhänge nach wirtschaftlichen Aspekten zu gewichten.

Basiert die Entscheidung, ob ein signifikanter Zusammenhang zwischen qualitätsmessenden und -beschreibenden Größen existiert auf dem Ergebnis eines Unabhängigkeitstests, so muss im Falle von großen Datenmengen die Folgen des multiplen Hypothesenproblems in Betracht gezogen werden. Ein neuer, *alternativer Ansatz*, der in der Lage ist, den Anteil fälschlicherweise abgelehnter Nullhypothesen geeignet abzuschätzen und sich leicht in den hier vorgestellten Testalgorithmus integrieren lässt, wurde in Abschnitt 4.3 modelliert.

Für beide in dieser Ausarbeitung vorgestellten Ansätze ist es gelungen, die in der Einleitung formulierten Anforderungen zu erfüllen. Sowohl das RCA-Verfahren als auch der Testalgorithmus kombinieren Modelle und Lösungsansätze die es erlauben, unabhängig von der zugrunde liegenden Verteilungsart der Daten und unter Berücksichtigung von statistischen wie auch wirtschaftlichen Gesichtspunkten, signifikantes und qualitätsrelevantes Wissen zu extrahieren.

Trotz der engen Verknüpfung mit Praxisanwendungen und der Forderung möglichst allgemeingültige bzw. nicht-parametrische Verfahren zu implementieren, lassen sich Möglichkeiten aufzeigen, welche die Genauigkeit und Effizienz der hier vorgestellten Wissensextraktionsmodelle verbessern. Dabei können prinzipiell drei Stoßrichtungen verfolgt werden. Auf der einen Seite können die verwendeten Maße zur Messung von Qualität bzw. Interessantheit eines Zusammenhangs erweitert werden. Vor allem bei der Verwendung von subjektiven Interessantheitsmaßen und den eingeführten Interessantheitsfacetten besteht die Möglichkeit, für die jeweilige Anwendung passende (subjektive) Maße zu entwickeln. Auch können die Methoden zur Konstruktion von Konfidenzintervallen ver-

feinert werden, da speziell Intervalle für binomialverteilte Größen ständig weiterentwickelt werden.

Auf der anderen Seite können Verallgemeinerungen der Ursachen-Wirkungs-Beziehungen, die in dieser Arbeit in Form von Assoziationsregeln bzw. Sub-Prozessen dargestellt wurden, die Möglichkeiten zur Identifizierung von Optimierungspotential steigern. So können für Assoziationsregeln beispielsweise Konzepte wie Data-Cubes [Blumenstock et al. 2007] verwendet werden, um Qualitätszusammenhänge in Kombination mit den hier vorgestellten Interessantheitsmaßen abzubilden. Das RCA-Verfahren bzw. die zugrunde liegende Beschreibung von Ziel- und Einflussgrößen als Sub-Prozesse, kann mit Hilfe des Konzeptes komplexer Assoziationsregeln erweitert werden. Einzig die Möglichkeit einer effizienten Implementierung verhindert aktuell die Berücksichtigung dieser Art von Zusammenhängen in dem Ursachenanalyse-Verfahren.

Die dritte Stoßrichtung ergibt sich durch die Entwicklungen auf dem Forschungsgebiet des multiplen Hypothesenproblems. Durch die zunehmende Leistungsfähigkeit moderner Computersysteme und die stetige Zunahme der Datenmenge, die mit unterschiedlichsten, statistischen Verfahren analysiert wird, nimmt dieses, oft stiefmütterlich behandelte Thema, in der Zukunft eine bedeutendere Rolle ein. Hier werden sicherlich Arbeiten aus den Bereichen der Biostatistik, speziell der Micro Array-Studien wegweisend sein.



# A Anhang

## A.1 Sätze und Definitionen

Für die in Kapitel 2.1 gezeigten Eigenschaften werden die wichtigsten Theoreme und Definitionen im Folgenden aufgeführt. Für einen jeweiligen Beweis sei auf die entsprechende Literaturquelle verwiesen.

**Satz A.1** (Mittelwertsatz für reellwertige Funktionen mehrerer Variablen)

Sei  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  und  $f$  differenzierbar auf einer offenen Menge  $G \subseteq D(f)$ . Des Weiteren seien  $x_1, x_2 \subseteq G$  und  $\overline{x_1, x_2} \subseteq G$ . Dann  $\exists x_0 \in \overline{x_1, x_2}$  mit  $x_0 \neq x_1 \wedge x_0 \neq x_2$  und es gilt:

$$f(x_1) - f(x_2) = \nabla f(x_0)(x_2 - x_1)$$

### Definition A.1

Sei  $X_1, X_2, \dots$  eine beliebige Folge von Zufallszahlen. Dann gilt:

- $\{X_n\}$  konvergiert fast sicher gegen eine Zufallsvariable  $X$ , falls gilt:  
 $\forall \epsilon > 0, P(\lim_{n \rightarrow \infty} |X_n - X| < \epsilon) = 1$ .
- $\{X_n\}$  konvergiert in Wahrscheinlichkeit gegen eine Zufallsvariable  $X$  falls gilt:  
 $\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$ . Notation:  $X_n \xrightarrow{p} X$ .
- $\{X_n\}$  konvergiert in Verteilung gegen eine Zufallsvariable  $X$ , falls für alle  $\epsilon > 0$ ,  
 $\lim_{n \rightarrow \infty} P(|X_n - X| \geq \epsilon) = 0$ . Notation:  $X_n \xrightarrow{d} X$ .

**Satz A.2** (Asymptotische Eigenschaften empirischer Quantile)

Für die in Def. 2.3 definierten empirischen Quantile gelten folgende Eigenschaften:

1. Falls  $F$  eine Dichte  $f$  in einer Umgebung  $\mu_\alpha$  mit  $0 < \alpha < 1$  besitzt und  $f$  positiv und stetig an der Stelle  $\mu_\alpha$  ist, dann gilt:

$$\sqrt{n}(\hat{\mu}_\alpha - \mu_\alpha) \xrightarrow{d} N\left(0, \frac{\alpha(1-\alpha)}{f^2(\mu_\alpha)}\right) \quad (\text{A.1})$$

2. Sei  $0 < \alpha_1 < \dots < \alpha_k < 1$ . Des Weiteren habe  $F$  eine Dichte  $f$  in einer Umgebung von  $\mu_{\alpha_1}, \dots, \mu_{\alpha_k}$  und  $f$  sei positiv und stetig an der Stelle  $\mu_{\alpha_1}, \dots, \mu_{\alpha_k}$ . Dann ist  $\hat{\mu}_{\alpha_1}, \dots, \hat{\mu}_{\alpha_k}$  asymptotisch normal mit Mittelwertvektor  $(\mu_{\alpha_1}, \dots, \mu_{\alpha_k})$  und Kovarianzen  $\sigma/n$ , mit

$$\sigma_{ij} = \begin{cases} \frac{\alpha_i(1-\alpha_j)}{f(\mu_{\alpha_i})f(\mu_{\alpha_j})}, & \text{für } i \leq j \\ \sigma_{ij} = \sigma_{ji} & \text{für } i > j \end{cases} \quad (\text{A.2})$$

3. Es gilt für den Erwartungswert und die Varianz von  $\mu_\alpha$ :

$$\mathbb{E}(\hat{\mu}_\alpha) = \mu_\alpha \quad (\text{A.3})$$

$$\text{Var}(\hat{\mu}_\alpha) \xrightarrow{n \rightarrow \infty} 0 \quad (\text{A.4})$$

Beweise siehe [Serfling 1980], Kapitel 2.3.

**Satz A.3** (Satz von Slutsky)

Seien  $\{X, X_n, Y_n, n \geq 1\}$  Zufallszahlen. Falls  $X_n \xrightarrow{d} X$  und  $Y_n \xrightarrow{p} c$ , dann gilt:

$$Y_n X_n \xrightarrow{d} cX$$

Beweis vergleiche [Resnick 2001], Kapitel 2.8.

**Satz A.4**

Sei  $X = (X_1, \dots, X_n)$  ein multivariat normalverteilter Zufallsvektor mit  $\mathbb{E}(X) = \mu = (\mu_1, \dots, \mu_n)$  und positiv definiter Kovarianzmatrix  $\Sigma$ . Des Weiteren sei  $A \in \mathbb{R}^{n \times m}$ . Dann ist  $AX$  wiederum multivariat normalverteilt mit  $E(AX) = A\mu$  und Kovarianzmatrix  $A\Sigma A^T$ .

Beweise siehe [Kendall et al. 1987], Kapitel 15.4.

## A.2 Tabellen und Ergebnisse

Auf den folgenden Seiten sind zunächst die Ergebnisse der Simulationsstudie aus Kapitel 2.3.3 und anschließend die zur Beispielergebnisse des Root-Cause-Analysis-Verfahrens (Kapitel 3.3.1) gehörigen Ergebnisse tabellarisch dargestellt.

Qualität	$N$	$\hat{E}_{ci}$	$E_{ci}$	BCa	Perzentil	SBBias	BCPB	SB
Schlecht	50	0.80	0.80	0.65	0.70	0.70	0.73	0.70
Schlecht	120	0.79	0.79	0.71	0.73	0.73	0.75	0.73
Schlecht	400	0.79	0.79	0.75	0.76	0.76	0.76	0.76
Schlecht	900	0.79	0.79	0.77	0.77	0.77	0.77	0.77
Gut	50	0.87	0.87	0.73	0.78	0.78	0.82	0.79
Gut	120	0.87	0.87	0.79	0.81	0.81	0.83	0.81
Gut	400	0.87	0.87	0.83	0.84	0.84	0.84	0.84
Gut	900	0.87	0.87	0.84	0.85	0.85	0.85	0.85
Sehr gut	50	0.97	0.97	0.88	0.93	0.94	0.96	0.94
Sehr gut	120	0.98	0.98	0.93	0.96	0.96	0.97	0.96
Sehr gut	400	0.98	0.98	0.96	0.97	0.97	0.97	0.97
Sehr gut	900	0.98	0.98	0.97	0.97	0.97	0.98	0.97

Tabelle A.1: Simulationsergebnisse  $E_{ci}$

QFI	N	Bca	Perzentil	SBBias	BCPB	SB	Bca	Perzentil	SBBias	BCPB	SB
$C'_p$	50	1.54	1.54	1.23	1.54	1.33	61.80%	61.80%	73.70%	61.80%	79.10%
$C_p$	120	1.37	1.37	1.17	1.37	1.24	95.70%	95.70%	81.30%	95.70%	89.60%
$C'_p$	400	1.21	1.21	1.13	1.21	1.16	97.70%	97.70%	86.60%	97.70%	93.00%
$C_p$	900	1.13	1.14	1.12	1.15	1.13	93.20%	95.00%	91.30%	96.60%	94.40%
$C'_{pk}$	50	1.21	1.32	1.15	1.29	1.22	81.70%	92.80%	70.00%	87.70%	79.50%
$C_{pk}$	120	1.16	1.24	1.11	1.21	1.16	87.50%	97.50%	74.10%	94.90%	83.30%
$C'_{pk}$	400	1.10	1.14	1.07	1.13	1.10	86.60%	94.10%	76.00%	92.40%	85.20%
$C_{pk}$	900	1.07	1.08	1.07	1.09	1.08	82.30%	86.30%	79.30%	89.60%	83.90%
$C'_{pm}$	50	1.19	1.28	1.19	1.30	1.23	82.30%	96.30%	76.60%	90.90%	85.30%
$C_{pm}$	120	1.21	1.29	1.17	1.28	1.21	95.20%	99.20%	82.20%	98.50%	89.90%
$C'_{pm}$	400	1.18	1.20	1.13	1.20	1.16	97.10%	97.70%	87.10%	97.50%	93.10%
$C_{pm}$	900	1.13	1.14	1.12	1.15	1.13	93.60%	94.90%	91.70%	96.70%	94.20%
$C'_{pmk}$	50	1.03	1.09	1.08	1.14	1.08	51.30%	68.70%	61.40%	68.10%	66.70%
$C_{pmk}$	120	1.07	1.16	1.09	1.15	1.12	68.10%	90.60%	70.50%	84.10%	79.20%
$C'_{pmk}$	400	1.09	1.13	1.07	1.12	1.10	81.40%	93.50%	76.40%	90.80%	84.10%
$C_{pmk}$	900	1.07	1.08	1.07	1.09	1.07	81.70%	85.50%	80.00%	89.30%	83.70%

Tabelle A.2: Simulationsergebnisse für "guten" Prozess

QFI	$N$	BCa	Perzentil	SBBias	BCPB	SB	BCa	Perzentil	SBBias	BCPB	SB
$C_p$	50	2.76	2.76	1.67	2.76	2.00	62.30%	62.30%	48.10%	62.30%	56.70%
$C_p$	120	2.24	2.24	1.56	2.24	1.77	89.70%	89.70%	50.00%	89.70%	63.50%
$C_p$	400	1.81	1.81	1.52	1.81	1.62	75.80%	75.80%	45.20%	75.80%	54.30%
$C_p$	900	1.59	1.60	1.51	1.61	1.54	50.20%	51.70%	39.70%	52.40%	42.10%
$C_{pk}$	50	2.47	2.60	1.62	2.55	1.92	76.20%	69.80%	46.60%	72.60%	56.10%
$C_{pk}$	120	2.08	2.15	1.52	2.12	1.71	86.60%	88.20%	48.30%	88.00%	59.40%
$C_{pk}$	400	1.73	1.76	1.48	1.75	1.57	66.90%	69.60%	40.60%	68.70%	49.10%
$C_{pk}$	900	1.55	1.56	1.47	1.57	1.50	43.60%	45.40%	35.10%	46.60%	36.80%
$C_{pm}$	50	2.16	2.37	1.70	2.31	1.92	89.60%	87.70%	56.80%	88.60%	66.80%
$C_{pm}$	120	1.99	2.11	1.58	2.07	1.75	88.70%	90.90%	51.60%	90.50%	64.30%
$C_{pm}$	400	1.74	1.77	1.51	1.76	1.60	68.70%	73.20%	44.50%	72.10%	52.60%
$C_{pm}$	900	1.57	1.58	1.50	1.59	1.52	46.80%	49.00%	38.10%	49.50%	39.30%
$C_{pmk}$	50	2.04	2.28	1.67	2.23	1.87	83.10%	88.60%	54.70%	87.60%	65.30%
$C_{pmk}$	120	1.93	2.06	1.54	2.03	1.71	85.60%	88.90%	50.50%	88.50%	61.20%
$C_{pmk}$	400	1.71	1.73	1.47	1.73	1.56	65.40%	67.30%	40.20%	67.00%	48.50%
$C_{pmk}$	900	1.53	1.54	1.46	1.55	1.48	40.50%	42.30%	32.80%	43.20%	34.60%

Tabelle A.3: Simulationsergebnisse für "sehr guten" Prozess

QFI	N	Bca	Perzentil	SBBias	BCPB	SB	Bca	Perzentil	SBBias	BCPB	SB
$C'_p$	50	1.10	1.10	0.74	1.10	0.85	32.60%	32.60%	78.10%	32.60%	70.30%
$C'_p$	120	0.94	0.94	0.73	0.94	0.79	67.70%	67.70%	91.90%	67.70%	88.00%
$C'_p$	400	0.79	0.79	0.70	0.79	0.73	98.80%	98.80%	99.70%	98.80%	99.50%
$C'_p$	900	0.73	0.73	0.70	0.73	0.71	100.00%	100.00%	100.00%	100.00%	100.00%
$C'_{pk}$	50	0.79	0.88	0.64	0.84	0.73	88.60%	77.60%	90.70%	82.50%	88.80%
$C'_{pk}$	120	0.72	0.78	0.63	0.76	0.68	99.20%	97.20%	99.30%	98.50%	98.90%
$C'_{pk}$	400	0.66	0.68	0.61	0.67	0.64	100.00%	100.00%	100.00%	100.00%	100.00%
$C'_{pk}$	900	0.63	0.64	0.62	0.64	0.63	100.00%	100.00%	100.00%	100.00%	100.00%
$C'_{pm}$	50	0.88	0.97	0.76	0.95	0.83	77.90%	58.70%	80.80%	63.30%	76.90%
$C'_{pm}$	120	0.86	0.90	0.73	0.89	0.79	89.70%	79.70%	92.40%	83.00%	90.30%
$C'_{pm}$	400	0.77	0.79	0.70	0.78	0.73	99.40%	99.10%	99.70%	99.20%	99.50%
$C'_{pm}$	900	0.73	0.73	0.70	0.73	0.71	100.00%	100.00%	100.00%	100.00%	100.00%
$C'_{pmk}$	50	0.68	0.78	0.64	0.76	0.69	96.40%	91.50%	93.10%	91.40%	93.60%
$C'_{pmk}$	120	0.68	0.75	0.63	0.73	0.67	99.70%	98.90%	99.00%	98.90%	99.00%
$C'_{pmk}$	400	0.65	0.68	0.61	0.67	0.64	100.00%	100.00%	100.00%	100.00%	100.00%
$C'_{pmk}$	900	0.63	0.64	0.62	0.64	0.63	100.00%	100.00%	100.00%	100.00%	100.00%

Tabelle A.4: Simulationsergebnisse für "schlechter" Prozess

Sub-Prozess	$N_{Test}$	$Q_{Test}$	$N_{Train}$	$C_l^{SB}$
BAZ in (1,2)	244	0.81	210	0.81
BAZ in (2,3) and Spindel in (2)	124	0.85	127	0.72
BAZ in (2,4)	225	0.88	236	0.81
BAZ not in (1)	356	0.81	377	0.77
BAZ not in (1) and Center in (rechts)	190	0.85	183	0.75
BAZ not in (1) and Spindel in (2)	178	0.85	194	0.78
BAZ not in (3)	353	0.81	343	0.79
BAZ not in (3) and Center in (rechts)	174	0.87	174	0.80
Center in (rechts)	243	0.83	244	0.77

Tabelle A.5: Beispielergebnisse: Gemeinsame Sub-Prozesse für  $E_{ci}$ 

Sub-Prozess	$N_{Test}$	$Q_{Test}$	$N_{Train}$	$C_l^{BCPB}$
BAZ in (1,4) and Spindel in (1)	127	1.10	118	1.01
BAZ in (2,3) and Spindel in (2)	124	1.03	127	1.02
BAZ in (3,4) and Center in (rechts)	131	1.05	139	1.19

Tabelle A.6: Beispielergebnisse: Gemeinsame Sub-Prozesse für  $C_{pk}$

Menge	$N_{\mathcal{X}}$	$Q_{\mathcal{X}}$	$C_l^{SB}$	Sub-Prozess $\mathcal{X}$
Training	126	0.86	0.80	Spindel in (2) and Center in (rechts)
Training	249	0.83	0.79	Spindel in (2)
Training	244	0.81	0.77	Center in (rechts)
Training	182	0.81	0.76	BAZ not in (4) and Spindel in (2)
Training	175	0.87	0.83	BAZ not in (3) and Spindel in (2)
Training	174	0.83	0.80	BAZ not in (3) and Center in (rechts)
Training	169	0.83	0.78	BAZ not in (3) and Center in (links)
Training	343	0.83	0.79	BAZ not in (3)
Training	194	0.85	0.78	BAZ not in (1) and Spindel in (2)
Training	183	0.81	0.75	BAZ not in (1) and Center in (rechts)
Training	377	0.81	0.77	BAZ not in (1)
Training	133	0.83	0.83	BAZ in (4)
Training	123	0.85	0.81	BAZ in (2,4) and Center in (links)
Training	236	0.85	0.81	BAZ in (2,4)
Training	127	0.83	0.72	BAZ in (2,3) and Spindel in (2)
Training	130	0.81	0.75	BAZ in (2,3) and Center in (links)
Training	122	0.83	0.78	BAZ in (1,4) and Spindel in (2)
Training	130	0.83	0.77	BAZ in (1,4) and Center in (rechts)
Training	240	0.81	0.76	BAZ in (1,4)
Training	210	0.84	0.81	BAZ in (1,2)
Test	132	0.82	-	Spindel in (1) and Center in (rechts)
Test	250	0.79	-	Spindel in (1)
Test	243	0.83	-	Center in (rechts)
Test	195	0.79	-	BAZ not in (4) and Spindel in (1)
Test	181	0.81	-	BAZ not in (4) and Center in (rechts)
Test	184	0.84	-	BAZ not in (3) and Spindel in (1)
Test	174	0.87	-	BAZ not in (3) and Center in (rechts)
Test	353	0.81	-	BAZ not in (3)
Test	184	0.78	-	BAZ not in (2) and Center in (rechts)
Test	178	0.85	-	BAZ not in (1) and Spindel in (2)
Test	190	0.85	-	BAZ not in (1) and Center in (rechts)
Test	356	0.81	-	BAZ not in (1)
Test	131	0.79	-	BAZ in (3,4) and Center in (rechts)
Test	225	0.88	-	BAZ in (2,4)
Test	124	0.85	-	BAZ in (2,3) and Spindel in (2)
Test	128	0.83	-	BAZ in (2,3) and Center in (rechts)
Test	247	0.81	-	BAZ in (2,3)
Test	127	0.81	-	BAZ in (1,4) and Spindel in (1)
Test	129	0.87	-	BAZ in (1,2) and Spindel in (1)
Test	244	0.81	-	BAZ in (1,2)

Tabelle A.7: Beispielergebnisse: Trainings- und Test-Menge für  $E_{ci}$

Menge	$N_{\mathcal{X}}$	$Q_{\mathcal{X}}$	$C_l^{BCPB}$	Sub-Prozess $\mathcal{A}$
Training	87	1.21	0.86	BAZ not in (3) and Spindel in (2) and Center in (links)
Training	93	1.26	1.24	BAZ not in (2) and Spindel in (1) and Center in (rechts)
Training	91	1.20	1.05	BAZ not in (1) and Spindel in (1) and Center in (rechts)
Training	183	1.09	1.02	BAZ not in (1) and Spindel in (1)
Training	66	1.22	1.02	BAZ in (4) and Spindel in (1)
Training	69	1.36	1.20	BAZ in (4) and Center in (rechts)
Training	133	1.16	1.11	BAZ in (4)
Training	73	1.27	1.08	BAZ in (3,4) and Spindel in (2) and Center in (rechts)
Training	139	1.23	1.19	BAZ in (3,4) and Center in (rechts)
Training	274	1.12	1.10	BAZ in (3,4)
Training	74	1.29	1.20	BAZ in (3) and Spindel in (2)
Training	67	1.13	1.10	BAZ in (3) and Spindel in (1)
Training	70	1.26	1.18	BAZ in (3) and Center in (rechts)
Training	71	1.12	1.10	BAZ in (3) and Center in (links)
Training	141	1.12	1.12	BAZ in (3)
Training	116	1.20	1.02	BAZ in (2,4) and Spindel in (1)
Training	70	1.14	1.02	BAZ in (2,3) and Spindel in (2) and Center in (links)
Training	127	1.09	1.02	BAZ in (2,3) and Spindel in (2)
Training	118	1.14	1.01	BAZ in (1,4) and Spindel in (1)
Training	105	1.19	0.72	BAZ in (1,2) and Center in (links)
Test	132	0.99	-	Spindel in (1) and Center in (rechts)
Test	243	0.86	-	Center in (rechts)
Test	169	0.91	-	BAZ not in (3) and Spindel in (2)
Test	184	0.99	-	BAZ not in (3) and Spindel in (1)
Test	174	1.01	-	BAZ not in (3) and Center in (rechts)
Test	179	0.87	-	BAZ not in (3) and Center in (links)
Test	353	0.90	-	BAZ not in (3)
Test	184	0.93	-	BAZ not in (2) and Center in (rechts)
Test	178	0.91	-	BAZ not in (1) and Spindel in (2)
Test	190	1.04	-	BAZ not in (1) and Center in (rechts)
Test	131	1.05	-	BAZ in (3,4) and Center in (rechts)
Test	225	1.04	-	BAZ in (2,4)
Test	124	1.03	-	BAZ in (2,3) and Spindel in (2)
Test	128	0.96	-	BAZ in (2,3) and Center in (rechts)
Test	127	1.10	-	BAZ in (1,4) and Spindel in (1)
Test	122	0.90	-	BAZ in (1,4) and Center in (links)
Test	237	0.91	-	BAZ in (1,4)
Test	122	0.83	-	BAZ in (1,3) and Center in (rechts)
Test	129	0.87	-	BAZ in (1,2) and Spindel in (1)
Test	244	0.82	-	BAZ in (1,2)

Tabelle A.8: Beispielergebnisse: Trainings- und Test-Menge für  $C_{pk}$



# Literaturverzeichnis

- [Agrawal et al. 1993] Agrawal, R. Imielinski, T. und Swami, A. N. (1993). Mining association rules between sets of items in large databases. In Buneman, P. und Jajodia, S., editors *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data* pages 207–216 Washington, D.C.
- [Agresti 1990] Agresti, A. (1990). *Categorical Data Analysis*. Wiley Series in Probability and Mathematical Statistics 1. ed. edition.
- [Albrecht 2005] Albrecht, C. (2005). Automatisierte Daten-Transformation im Gala-Project. *Beiträge zum Wissenschaftlichen Rechnen Ergebnisse des Gaststudentenprogramms 2005 des John von Neumann-Instituts für Computing*.
- [Balamurali und Kalyanasundaram 2002] Balamurali, S. und Kalyanasundaram, M. (2002). Bootstrap lower confidence limits for the process capability indices cp, cpk and cpm. *International Journal of Quality & Reliability Management* 19:1088 – 1097.
- [Benjamini und Hochberg 1995] Benjamini, Y. und Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society* (57):289–300.
- [Benjamini et al. 2006] Benjamini, Y. Krieger, A. und Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* 93(3):491–507.
- [Benjamini und Leshno 2007] Benjamini, Y. und Leshno, M. (2007). Statistical methods for data mining. <http://www.math.tau.ac.il/~ybenja/MyPapers.html>.
- [Benjamini und Yekutieli 2001] Benjamini, Y. und Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics* 29(4):1165–1188.
- [Berry und Linoff 1997] Berry, M. und Linoff, G. (1997). Data mining techniques - for

- marketing, sales and customer support. In *Data Mining - Theoretische Aspekte und Anwendungen*. Wiley Computer Publishing, New York.
- [Bissel 1990] Bissel, A. (1990). How reliable is your capability index. *Applied Statistics* 39(3):331–340.
- [Blumenstock et al. 2007] Blumenstock, A. Schweiggert, F. und Muller, M. (2007). Rule cubes for causal investigations. In *ICDM '07: Proceedings of the 2007 Seventh IEEE International Conference on Data Mining* pages 53–62 Washington, DC, USA. IEEE Computer Society.
- [Bollinger 1996] Bollinger, T. (1996). Assoziationsregeln - Analyse eines Data Mining Verfahrens. *Informatik Spektrum* 5/19:257–261.
- [Borgelt und Kruse 2002] Borgelt, C. und Kruse, R. (2002). Induction of association rules: Apriori implemenation. In *15th Conference on Computational Statistics*. Physica Verlag, Heidelberg, Germany.
- [Boyles 1991] Boyles, R. (1991). The taguchi capability index. *Journal of Quality Technology* 23(1):17–26.
- [Brown et al. 1999] Brown, L. D. Cai, T. T. und Dasgupta, A. (1999). Interval estimation for a binomial proportion. *Statistical Science* 16:101–133.
- [Casella und Berger 2002] Casella, G. und Berger, R. (2002). *Statistical inference*. Duxbury 2. ed. edition.
- [Chapman und Clinton 2000] Chapman, P. und Clinton, J. (2000). Crisp-dm 1.0. <http://www.crisp-dm.org/>.
- [Clemens 1989] Clemens, J. (1989). Process capability calculations for non-normal distributions. *Quality Progress* 22:523–544.
- [Dietrich und Schulze 2005] Dietrich, E. und Schulze, A. (2005). *Statistische Verfahren zur Maschinen- und Prozessqualifikation*. Hanser Verlag 5., aktualis. Aufl. edition.
- [DIN Deutsches Institut für Normung e.V. 2002] DIN Deutsches Institut für Normung e.V. (2002). Deutsche-Norm: Qualitätsfähigkeitskenngrößen DIN 55319.
- [DIN Deutsches Institut für Normung e.V. 2007] DIN Deutsches Institut für Normung e.V. (2007). *Statistische Verfahren - Prozessleistungs- und Prozessfähigkeitskenngrößen für kontinuierliche Qualitätsmerkmale*. Beuth Verlag GmbH.

- [Dong und Li 1998] Dong, G. und Li, J. (1998). Interestingness of discovered association rules in terms of neighborhood-based unexpectedness. In Wu, X. Kotagiri Ramamohanarao und Korb, K. B., editors *Research and Development in Knowledge Discovery and Data Mining, Proc. 2nd Pacific-Asia Conf. Knowledge Discovery and Data Mining, PAKDD* volume 1394 pages 72–86. Springer.
- [Dudoit et al. 2003] Dudoit, S. Popper-Shaffer, J. und Boldrick, J. (2003). Multiple hypothesis testing in microarray experiments. *Statistical Science* 18(1):71–103.
- [Efron und Tibshirani 1993] Efron, B. und Tibshirani, R. (1993). *An introduction to the bootstrap*. Chapman & Hall.
- [Franklin und Wasserman 1992] Franklin, L. und Wasserman, G. (1992). Bootstrap lower confidence limits for capability indices. *Journal of quality technology* 24(4):196–210 (16 ref.).
- [Hartung 1999] Hartung, J. (1999). *Statistik: Lehr- und Handbuch der angewandten Statistik*. Oldenbourg.
- [Held 2007] Held, L. (2007). *Moderne Statistische Inferenz*. Master’s thesis Universität Zürich, Schweiz.
- [Hettich und Hippner 2001] Hettich, S. und Hippner, H. (2001). Assoziationsanalyse, Kapitel 11. [http://www.ku-eichstaett.de/Fakultaeten/WWF/Lehrstuehle/WI/Lehre/dm\\_v/HF\\_sections/content/DM%205.pdf](http://www.ku-eichstaett.de/Fakultaeten/WWF/Lehrstuehle/WI/Lehre/dm_v/HF_sections/content/DM%205.pdf).
- [Hipp 2004] Hipp, J. (2004). *Wissensentdeckung in Datenbanken mit Assoziationsregeln*. Akademische Verlagsgesellschaft Aka GmbH, Berlin 1. ed. edition.
- [Hjorth 1994] Hjorth, J. (1994). *Computer intensive statistical methods*. Chapman & Hall 1. ed. edition.
- [Hochberg 1988] Hochberg, Y. (1988). A sharper bonferroni procedure for multiple tests of significance. *Biometrika* 75(4):800–802.
- [Holm 1979] Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 6:65–70.
- [Kendall et al. 1987] Kendall, M. G. Stuart, A. und Ord, J. K., editors (1987). *Kendall’s advanced theory of statistics*. Oxford University Press, Inc. New York, NY, USA.

- [Kotz und Johanson 2002] Kotz, S. und Johanson, N. (2002). Process capability indices - a review, 1992 - 2000. *Journal of Quality Technology* Vol. 34.
- [Lallich et al. 2006] Lallich, S. Teytaud, O. und Prudhomme, E. (2006). Association rule interestingness: measure and statistical validation. In Hamilton, G., editor *Quality measures in data mining* page 25. Springer.
- [Müller et al. 1998] Müller, M. Hausdorf, C. und Schneeberger, J. (1998). Zur Interessantheit bei der Entdeckung von Wissen in Datenbanken. In *Data Mining - Theoretische Aspekte und Anwendungen* pages 248–264. Physica, Heidelberg.
- [Montgomery 1991] Montgomery, D. (1991). *Introduction to statistical quality control*. Wiley 2. ed. edition.
- [Nonnemacher 2006] Nonnemacher, S. (2006). Verfahren zur Verbesserung der Prozessfähigkeit mit Kovariablen. Master's thesis Universität Ulm.
- [Perakis und Xakalaki 2004] Perakis, M. und Xakalaki, E. (2004). A new method for constructing confidence intervals for the index cpm. *Quality and Reliability Engineering International* 20:651–665.
- [Resncik 2001] Resncik, S. (2001). *A Probability Path*. Birkenhäuser.
- [Rinne und Mittag 2002] Rinne, H. und Mittag, H. (2002). *Statistische Methoden der Qualitätssicherung*. Hanser Fachbuch.
- [Schmidt 2004] Schmidt, V. (2004). Vorlesungsskript: Statistik 1. Universität Ulm.
- [Schmidt 2006] Schmidt, V. (2006). Vorlesungsskript: Statistik 2. Universität Ulm.
- [Serfling 1980] Serfling, R. (1980). *Approximation theorems of mathematical statistics*. Wiley series in probability and mathematical statistics.
- [Shewhart 1931] Shewhart, W. (1931). *Economic control of quality of manufactured product*. Van Nostrand.
- [Tan et al. 2002] Tan, P. Kumar, V. und Srivastava, J. (2002). Selecting the right interestingness measure for association patterns. In Proceedings of the Eight ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
- [Vännman 1995] Vännman, K. (1995). A unified approach to capability indices. *Statistica Sina* 5:805–820.

- 
- [Vogel 2005] Vogel, F. (2005). *Beschreibende und schließende Statistik*. Oldenbourg Verlag.
- [Wang 2006] Wang, W. (2006). Weizhen wang. *Journal of Statistical Planning and Inference* 136(12):4293–4306.
- [Zhang et al. 1990] Zhang, N. F. Stenback und Wardrop, G. A. (1990). Interval estimation of process capability index  $c_{pk}$ . *Communications in statistics. Theory and methods (Commun. stat., Theory methods)* 19(12): 4455–4470.