

Karlsruher Schriften  
zur Anthropomatik

Band 47



Jennifer Sander

**Ansätze zur lokalen Bayes'schen  
Fusion von Informationsbeiträgen  
heterogener Quellen**



Jennifer Sander

**Ansätze zur lokalen Bayes'schen  
Fusion von Informationsbeiträgen  
heterogener Quellen**

Karlsruher Schriften zur Anthropomatik

Band 47

Herausgeber: Prof. Dr.-Ing. habil. Jürgen Beyerer

Eine Übersicht aller bisher in dieser Schriftenreihe  
erschienenen Bände finden Sie am Ende des Buchs.

# **Ansätze zur lokalen Bayes'schen Fusion von Informationsbeiträgen heterogener Quellen**

von  
Jennifer Sander

Karlsruher Institut für Technologie  
Institut für Anthropomatik und Robotik

Ansätze zur lokalen Bayes'schen Fusion von  
Informationsbeiträgen heterogener Quellen

Zur Erlangung des akademischen Grades eines Doktors der  
Naturwissenschaften von der KIT-Fakultät für Informatik des  
Karlsruher Instituts für Technologie (KIT) genehmigte Dissertation

von Jennifer Sander

Tag der mündlichen Prüfung: 24. Juni 2020  
Erster Gutachter: Prof. Dr.-Ing. habil. Jürgen Beyerer  
Zweiter Gutachter: Prof. Dr.-Ing. Michael Heizmann

#### Impressum



Karlsruher Institut für Technologie (KIT)  
KIT Scientific Publishing  
Straße am Forum 2  
D-76131 Karlsruhe

KIT Scientific Publishing is a registered trademark  
of Karlsruhe Institute of Technology.  
Reprint using the book cover is not allowed.

[www.ksp.kit.edu](http://www.ksp.kit.edu)



*This document – excluding the cover, pictures and graphs – is licensed  
under a Creative Commons Attribution-Share Alike 4.0 International License  
(CC BY-SA 4.0): <https://creativecommons.org/licenses/by-sa/4.0/deed.en>*



*The cover page is licensed under a Creative Commons  
Attribution-No Derivatives 4.0 International License (CC BY-ND 4.0):  
<https://creativecommons.org/licenses/by-nd/4.0/deed.en>*

Print on Demand 2021 – Gedruckt auf FSC-zertifiziertem Papier

ISSN 1863-6489  
ISBN 978-3-7315-1062-8  
DOI 10.5445/KSP/1000125447







# Kurzfassung

Die Lösung diverser Aufgabenstellungen aus unterschiedlichsten Anwendungsbereichen profitiert von der Informationsfusion oder setzt diese sogar voraus, da sich nur so eine ausreichend hochwertige Informationsgrundlage schaffen lässt. Die Fusion der Beiträge heterogener Informationsquellen bietet dabei oftmals besonderes Potential, da diese sich grob gesagt aufgrund unterschiedlicher Stärken und Schwächen besonders gut gegenseitig ergänzen können.

Zur adäquaten Umsetzung der IT-gestützten Informationsfusion bedarf es mathematischer Modelle und Verfahren, um unterschiedliche Informationsbeiträge in geeigneter Weise in eine einheitliche formale Beschreibung zu überführen und sie anschließend zu kombinieren. Dabei muss es methodisch gesehen auch möglich sein, das Fusionsergebnis auf spezielle Einzelfragestellungen zu konzentrieren und es müssen in fundierter und für den Nutzer auch verständlicher Weise Entscheidungen von ihm ableitbar sein.

Die Bayes'sche Theorie liefert oft ausreichend mächtige Werkzeuge zur Umsetzung der Informationsfusion auf Basis Bayes'scher Inferenz. Der Informationsstand in Bezug auf alle in eine Fusionsaufgabe involvierten Größen wird dabei probabilistisch mittels Wahrscheinlichkeitsverteilungen in der sog. Degree-of-Belief-Interpretation von Wahrscheinlichkeit repräsentiert. Die Bayes'sche Fusionsmethodik ist äußerst fundiert und erfüllt alle essentiellen Anforderungen an eine sinnvolle Methodik zur Fusion der Beiträge heterogener Informationsquellen. Die weitreichende Interpretation von Wahrscheinlichkeit als Degree-of-Belief, welche auch oftmals dem intuitiven Verständnis von Unsicherheit entspricht, das gerade auch nicht speziell geschulte Personen mitbringen, eröffnet ihr ein breites Anwendungsspektrum.

In praktisch relevanten Aufgabenstellungen der Informationsfusion ist jedoch der Wertebereich der Größen des Interesses, welche letztlich den im Rahmen der Fusion interessierenden Sachverhalt mathematisch fassen, oft sehr umfangreich und von recht hoher Dimension. In solchen Situationen verursachen Bayes'sche Verfahren in der Regel hohe, in der Praxis oftmals nicht tragbare Rechenzeiten und immensen Speicherbedarf. Dies begründet sich vor allem aus der Tatsache, dass gemäß dem Bayes'schen Formalismus immer die komplette A-Posteriori-Verteilung über dem gesamten Wertebereich der Größen des Interesses berechnet werden muss.

Die vorliegende Arbeit führt neue Ansätze zur Bewältigung Bayes'scher Fusion beim Vorliegen eines umfangreichen, ggf. hochdimensionalen Zielgrößenraums ein. Der Grundgedanke lokaler Bayes'scher Fusionsansätze leitet sich daraus ab, dass in vielen Aufgabenstellungen der Informationsfusion der durch die A-Posteriori-Verteilung in probabilistischer Weise verkörperte vollständige, d. h. maximal umfassende Informationsstand im Hinblick auf die Größen des Interesses eigentlich gar nicht benötigt wird. Das Ziel lokaler Bayes'scher Fusion ist es dementsprechend, die Bayes'sche Fusion möglichst konzentriert auf die Anteile des Wertebereichs der Größen des Interesses durchzuführen, in denen sich tatsächlich Aufgabenrelevantes abspielt – in dem Sinne, dass eine detaillierte Betrachtung dieser Anteile tatsächlich notwendig ist zur Beantwortung der eigentlichen Fragen.

Das primäre Ziel der Arbeit besteht darin, diese Ansätze formal mathematisch zu formulieren und sie grundlegend zu untersuchen, insbesondere im Hinblick auf ihre Umsetzbarkeit gerade auch für die Fusion der Beiträge heterogener Informationsquellen, ihre Ausdrucksstärke und letztlich ihre Fähigkeit, die Beantwortung von im Rahmen der Informationsfusion vorliegenden Fragestellungen in ausreichender Form zu ermöglichen. Die Arbeit zeigt insbesondere auf, dass die anvisierte Einnahme eines lokalen Standpunkts bei der Bayes'schen Fusion konsistent in den Bayes'schen Kalkül integrierbar ist und stellt das hierfür notwendige Werkzeug in konzeptioneller und methodischer Hinsicht zur Verfügung.

Für die Umsetzung lokaler Bayes'scher Fusion kann auf das Konzept für eine agentenbasierte Fusionsarchitektur, in welcher unterschiedliche Agenten in

kooperativer Weise Spuren im Hinblick auf den tatsächlich vorliegenden („wahren“) Wert der Größen des Interesses in den Informationsbeiträgen identifizieren, auswerten sowie die damit korrespondierenden Hypothesen durch Fusion lokal weiterentwickeln und letztlich zur Zusammenschau bereitstellen, zurückgegriffen werden.



# Inhaltsverzeichnis

<b>Kurzfassung</b> . . . . .	<b>i</b>
<b>Notation</b> . . . . .	<b>ix</b>
<b>1 Einleitung</b> . . . . .	<b>1</b>
1.1 Motivation . . . . .	1
1.2 Zielsetzung . . . . .	6
1.3 Aufbau der Arbeit . . . . .	8
<b>2 Informationsfusion</b> . . . . .	<b>13</b>
2.1 Grundprinzip . . . . .	13
2.2 Zentrale Begrifflichkeiten . . . . .	18
2.3 Anforderungen an eine Fusionsmethodik . . . . .	23
2.4 Möglichkeiten zur Einordnung . . . . .	26
2.4.1 Zusammenhang der Informationsbeiträge . . . . .	26
2.4.2 Abstraktionsebenen . . . . .	27
2.4.3 Fusionsmodelle . . . . .	30
2.4.4 Topologie und Aufgabenverteilung . . . . .	42
<b>3 Bayes'sche Fusion</b> . . . . .	<b>53</b>
3.1 Festlegung der Basisgrößen . . . . .	53
3.2 Fusion als statistische Inferenz . . . . .	54
3.3 Wahrscheinlichkeit . . . . .	57
3.3.1 Mathematische Festlegung . . . . .	57
3.3.2 Interpretationsmöglichkeiten . . . . .	60
3.4 Ansätze zur statistischen Inferenz . . . . .	65

3.4.1	Klassische Inferenz . . . . .	65
3.4.2	Bayes'sche Inferenz . . . . .	67
3.4.3	Likelihood-Inferenz . . . . .	70
3.5	Erfüllung der Basisanforderungen . . . . .	75
3.5.1	Fähigkeit zur Transformation . . . . .	75
3.5.2	Fähigkeit zur Fusion . . . . .	83
3.5.3	Fähigkeit zur Fokussierung . . . . .	86
<b>4</b>	<b>Durchführung der Fusion . . . . .</b>	<b>89</b>
4.1	Herausforderung im Hinblick auf die Berechnungen . . . . .	89
4.2	Ansätze in der Fachliteratur . . . . .	91
4.2.1	Konjugierte Verteilungsfamilien . . . . .	91
4.2.2	Deterministische Approximation . . . . .	95
4.2.3	Nichtdeterministische Approximation . . . . .	96
4.3	Ansatzpunkt der Arbeit . . . . .	104
4.3.1	Bezugspunkte . . . . .	104
4.3.2	Agentenbasierte Fusionsarchitektur . . . . .	110
<b>5</b>	<b>Mathematische Basis lokaler Bayes'scher Fusion . . . . .</b>	<b>119</b>
5.1	Einführung . . . . .	119
5.2	Lokale Bayes'sche Modelle . . . . .	121
5.2.1	Bayes'sche Experimente . . . . .	121
5.2.2	Beschreibbare Informationsstände . . . . .	123
5.2.3	Lokalität durch Einschränkung . . . . .	125
5.2.4	Lokalität durch Vergrößerung . . . . .	136
5.2.5	Zusammenfassende Betrachtung . . . . .	147
5.2.6	Verteilte Umsetzung . . . . .	151
5.3	Beispiel für eine qualitative Analyse . . . . .	156
5.3.1	Einführung . . . . .	156
5.3.2	(Globale) Bayes'sche Modellierung . . . . .	159
5.3.3	Lokale Bayes'sche Fusion . . . . .	165
5.3.4	Evaluation und Diskussion . . . . .	170

---

<b>6</b>	<b>Mathematische Kriterien zur Wahl des lokalen Kontexts</b>	<b>175</b>
6.1	Einführung	175
6.2	Prinzipien statistischer Evidenz	176
6.2.1	Irreführende statistische Evidenz	176
6.2.2	Anwendung für die lokale Bayes'sche Fusion	179
6.2.3	Anwendungsbeispiel	185
6.3	Informationstheoretische Betrachtung	194
6.3.1	Informationsverarbeitungsschema nach Zellner	195
6.3.2	Minimierungsregel nach Walker	199
6.3.3	Von den Informationsbeiträgen gelieferte Information	200
6.3.4	Anwendung für die lokale Bayes'schen Fusion	203
<b>7</b>	<b>Einbettung in den globalen Kontext</b>	<b>215</b>
7.1	Einführung	215
7.2	Wahrscheinlichkeitsintervall-Schema	217
7.3	Anwendung des Prinzips der Maximalen Entropie	224
7.3.1	Analytische Bestimmung der (Maximum-Entropie) A-Posteriori-Verteilung	224
7.3.2	Berechnungsbeispiele	230
7.4	Anwendung der Entscheidungstheorie	237
7.4.1	Entscheidungen unter Risiko, partieller Information und linearer partieller Information	237
7.4.2	Entscheidungskriterien unter linearer partieller Information	243
7.4.3	Verbesserung der Entscheidungsgrundlage	249
7.4.4	Berechnungsbeispiel	251
<b>8</b>	<b>Zusammenfassung und Ausblick</b>	<b>255</b>
8.1	Zusammenfassung	255
8.2	Ausblick	258

<b>Literatur</b> . . . . .	<b>261</b>
<b>Publikationen</b> . . . . .	<b>291</b>
<b>Betreute studentische Arbeiten</b> . . . . .	<b>299</b>
<b>Abkürzungsverzeichnis</b> . . . . .	<b>301</b>



# Notation

Nachfolgend ist die in der vorliegenden Arbeit verwendete Notation zusammengestellt. Spezielle Notationen, welche nur für ein bestimmtes Anwendungsbeispiel oder zur Nutzung innerhalb eines mathematischen Beweises verwendet werden, sind nur mit angeführt, wenn dies nötig ist, um z. B. Missverständnisse gegenüber der sonstigen Notation zu vermeiden.

## Allgemeine Notationen

Mengen	Großbuchstaben, kursiv	$A, B, \dots$
Mengensysteme	Großbuchstaben, kalligraphisch	$\mathcal{A}, \mathcal{B}, \dots$
Funktionen	Kleinbuchstaben, kursiv	$f(x), g(x), \dots$
Funktionale	Großbuchstaben, nicht-kursiv	$F[f(x)], \dots$
Matrizen	Großbuchstaben, kursiv	$B, C, \dots$
Vektoren	Kleinbuchstaben, kursiv	$b, c, \dots$
Parameter (Schranken, Gewichtungsfaktoren etc.)	Griechische Kleinbuchstaben	$\alpha, \beta, \dots$
Ganzzahlige Indexvariablen	Kleinbuchstaben, kursiv	$i, j, \dots$
Zufallsvariablen	Großbuchstaben, kursiv	$X, Y, \dots$
Realisierungen von Zufallsvariablen	Kleinbuchstaben, kursiv	$x, y, \dots$
Wahrscheinlichkeitsverteilungen	Kleinbuchstaben, kursiv	$p(x), q(x), \dots$

## Konventionen

Die vorliegende Arbeit folgt der in der Bayes'schen Theorie üblichen Notation, gemäß welcher i. A. nicht zwischen Zufallsvariablen und deren Realisierung unterschieden wird. Zufallsvariablen werden ausschließlich dann mit Großbuchstaben bezeichnet, wenn dies unabdingbar für die Darstellung ist, ansonsten werden sie wie ihre Realisierungen mit Kleinbuchstaben notiert.

Es wird grundsätzlich angenommen, dass die betrachteten Zufallsvariablen diskret, absolut stetig oder gemischt diskret absolut stetig sind und durch geeignete Wahrscheinlichkeitsverteilungen (d. h. Wahrscheinlichkeitsfunktionen, Wahrscheinlichkeitsdichtefunktionen, Mischungen derselben) beschrieben werden können (vgl. hierzu auch Abschnitt 3.2). Um unnötige Fallunterscheidungen zu vermeiden, wird meist nur die kontinuierliche Schreibweise verwendet.

## Allgemeine Symbole

$\infty$	Unendlich
$\propto$	Proportional zu
$\succsim$	Präferenzrelation
$\Leftrightarrow$	Äquivalenzrelation
$**$	Faltung (zweidimensional)
$ \cdot $	Kardinalität
$\times$	Kartesisches Produkt
$\vee$	Logisches oder
$\in$	$a \in A$ bedeutet, dass $a$ Element der Menge $A$ ist.
$\emptyset$	Leere Menge
$\cup$	Vereinigung von Mengen
$\cap$	Schnitt von Mengen

$\setminus$	Differenz zweier Mengen
$\bar{A}$	Komplement einer Menge $A$
$\mathbf{1}_A(x)$	Indikatorfunktion, d. h. es gilt $\mathbf{1}_A(x) = 1$ , wenn $x \in A$ , und $\mathbf{1}_A(x) = 0$ , wenn $x \notin A$ .
$b^\top$	Transponierte, hier eines Vektors $b$
arg max	Argument des Maximums
arg min	Argument des Minimums
$c, c_d$	Konstanten
$I_2$	Zweidimensionale Einheitsmatrix
max	Maximum einer Menge oder Funktion
min	Minimum einer Menge oder Funktion
$\mathbb{N}$	Menge der natürlichen Zahlen exklusive Null
$O(f(N))$	Funktionenklasse asymptotischer oberer Schranken (O-Kalkül): $O(f(N)) := \{g(N) \mid \exists c > 0, N_0 > 0, \forall N \geq N_0, g(N) \leq c f(N)\}$
$\mathbb{R}$	Menge der reellen Zahlen inklusive Null
$\mathbb{R}_{>0}$	Menge der reellen Zahlen größer Null

## Wahrscheinlichkeitstheorie

$\mathcal{A}$	$\sigma$ -Algebra
$\mathcal{A} _B$	Spur- $\sigma$ -Algebra $\{A \cap B \mid A \in \mathcal{A}\}$ von $\mathcal{A}$ in $B$ für $\emptyset \neq B \subset \Omega$
$\mathcal{B}(\mathbb{R}^k)$	Borel'sche $\sigma$ -Algebra über $\mathbb{R}^k$
$E_{p(x)}[x]$	Erwartungswert einer gemäß $p(x)$ verteilten Zufallsvariablen $x$

$E_{p(x)}[f(x)]$	Erwartungswert der Funktion $f(x)$ , wenn die Zufallsvariable $x$ gemäß $p(x)$ verteilt ist.
$h_k(A)$	Anzahl des Eintretens des Ereignisses $A$ bei $k$ Wiederholungen eines Zufallsexperiments
$\Omega$	Grundraum
$\omega_j$	Elementarereignis aus $\Omega$
$(\Omega, \mathcal{A})$	Messbarer Raum
$(\Omega, \mathcal{A}, P)$	Wahrscheinlichkeitsraum
$\sigma(A, B)$	Kleinste $\sigma$ -Algebra (über einem bestimmten Grundraum), welche die im Argument angegebenen Mengen (hier: $A$ und $B$ ) enthält.
$\sigma(\mathcal{A} \times \mathcal{B})$	Kleinste $\sigma$ -Algebra (über einem bestimmten Produktraum), die alle Mengen der Form $A \times B$ mit $A \in \mathcal{A}$ und $B \in \mathcal{B}$ enthält.
$P$	Wahrscheinlichkeitsmaß
$P(A)$	Wahrscheinlichkeit des Ereignisses $A$
$P(A B)$	Bedingte Wahrscheinlichkeit des Ereignisses $A$ gegeben das Ereignis $B$
$\mathcal{P}(\Omega)$	Potenzmenge von $\Omega$
$p(x)$	Wahrscheinlichkeitsverteilung der Zufallsvariablen $x$
$p(x y)$	Bedingte Wahrscheinlichkeitsverteilung der Zufallsvariablen $x$ gegeben die Zufallsvariable $y$
$r_k(A)$	Relative Häufigkeit des Ereignisses $A$ bei $k$ Wiederholungen eines Zufallsexperiments
$(X_t)_{t \in \mathbb{N}}$	Zeitdiskreter stochastischer Prozess
$\zeta$	Abstrakte Notation für einen interessierenden Sachverhalt

## Parametrische Verteilungsfamilien

$\mathcal{C}$	Cauchy-Verteilung
$\mathcal{N}$	Normalverteilung
$\mathcal{U}$	Gleichverteilung

## Statistische Inferenz

$d = (d_1, \dots, d_S) \in D$	Informationsbeiträge
$D = D_1 \times \dots \times D_S$	Wertebereich der Informationsbeiträge
$\mathcal{D}$	$\sigma$ -Algebra über $D$
$(D, \mathcal{D}, \{P_z   z \in Z\})$	Statistisches Experiment
$d_0$	Im Hinblick auf $z$ verfügbares Vorwissen (wird nur in Ausnahmefällen explizit notiert)
$d_s \in D_s$	Informationsbeitrag von Quelle Nummer $s$
$D_s$	Wertebereich des von Quelle Nummer $s$ gelieferten Informationsbeitrags
$I$	Information, welche im Hinblick auf ein bestimmtes Ereignis $A$ vorliegt. Bei Anwendung des Prinzips der Maximalen Entropie bzw. des Prinzips der Minimalen Information zu berücksichtigende Information
$\tilde{I}$	Analog zu $I$
$l(d z)$	Likelihood-Funktion bezüglich $d$ (fest) für $z$ : $l(d z) = c_d \cdot p(d z)$ , wobei $c_d$ eine beliebige, jedoch für alle Werte von $z$ gleich zu wählende Konstante ist.

$l(d_0 z)$	Likelihood-Funktion, welche das verfügbare Vorwissen repräsentiert (bei entsprechender technischer Ummodellierung des Bayes'schen Modells).
$l(d_s z)$	Analogon zu $l(d z)$ bei alleiniger Berücksichtigung des Informationsbeitrags $d_s$
$l_{st}(d z)$	Standardisierte Likelihood-Funktion bezüglich $d$ (fest) für $z$ , resultiert durch Skalierung von $l(d z)$ auf einen Maximalwert von Eins
$l_{st}(d_s z)$	Analogon zu $l_{st}(d z)$ bei alleiniger Berücksichtigung des Informationsbeitrags $d_s$
$\nu_i$	$i$ -tes Elements von $Z$ , wenn $Z$ unabhängig von der exakten Struktur als diskrete Menge aufgefasst wird
$\Pi$	Produktwahrscheinlichkeitsmaß bei einem Bayes'schen Experiment
$p(d)$	Für variables $d$ : Wahrscheinlichkeitsverteilung von $d$ Für festes $d$ : Wahrscheinlichkeitsverteilung von $d$ ausgewertet für den konkreten Wert
$p(d z)$	Für festes $z$ : Wahrscheinlichkeitsverteilung von $d$ unter der Annahme, dass $z$ der „wahre“ Wert der Größen des Interesses ist. Für festes $d$ : bezüglich $d$ normierte Likelihood-Funktion für $z$
$\{p(d z) z \in Z\}$	Mit der Festlegung der Likelihood-Funktion $l(d z)$ korrespondierende Verteilungsannahme
$p(d_s z)$	Analogon zu $p(d z)$ bei alleiniger Berücksichtigung des Informationsbeitrags $d_s$
$p(z)$	A-Priori-Verteilung von $z$
$p(z d_0)$	A-Priori-Verteilung von $z$ in Fällen, in denen die Abhängigkeit vom Vorwissen $d_0$ explizit betont werden soll.

---

$p(z d)$	A-Posteriori-Verteilung von $z$ nach Bekanntwerden von $d$
$p(z d,d_0)$	A-Posteriori-Verteilung von $z$ in Fällen, in denen die Abhängigkeit vom Vorwissen $d_0$ explizit betont werden soll.
$p(z,d)$	Gemeinsame Wahrscheinlichkeitsverteilung von $z$ und $d$
$p_{\text{ME}}(z)$	Wahrscheinlichkeitsverteilung auf $Z$ , welche aus der Anwendung des Prinzips der Maximalen Entropie resultiert.
$p_{\text{MI}}(z)$	Wahrscheinlichkeitsverteilung auf $Z$ , welche aus der Anwendung des Prinzips der Minimalen Information resultiert.
$P_z$	Wahrscheinlichkeitsmaß über $(D,\mathcal{D})$ , wenn hinsichtlich $d$ die Verteilungsklasse $\{p(d z) z \in Z\}$ angenommen wurde und $z$ der „wahre“ Wert der Größen des Interesses ist.
$\{P_z z \in Z\}$	Menge von Wahrscheinlichkeitsmaßen über $(D,\mathcal{D})$ , welche mit der Likelihood-Funktion korrespondiert.
$\mathfrak{P}_z$	Menge von Wahrscheinlichkeitsverteilungen über $Z$
$P_{z^{**}}(A)$	Wahrscheinlichkeit für das Ereignis $A \subset D$ , wenn hinsichtlich $d$ die Verteilungsklasse $\{p(d z) z \in Z\}$ angenommen wurde und $z^{**}$ der „wahre“ Wert der Größen des Interesses ist.
$v(e,z)$	Verlustfunktion: quantifiziert den mit der Auswahl eines Wertes $e \in Z$ verbundenen Verlust, wenn tatsächlich $z$ der „wahre“ Wert der Größen des Interesses ist.
$z = (z_1, \dots, z_N) \in Z$	Größen des Interesses
$Z = Z_1 \times \dots \times Z_N$	Wertebereich der Größen des Interesses

$\mathcal{Z}$	$\sigma$ -Algebra über $Z$
$z^*, z^{**}$	Zwei konkrete mögliche Werte für $z$
$\hat{z}$	Schätzwert für den „wahren“ Wert von $z$
$z^{(i)}$	$i$ -tes erzeugtes Sample bei stochastischer Simulation im Hinblick auf die Zufallsvariable $z$
$z_n \in Z_n$	$n$ -te Größe des Interesses
$Z_n$	Wertebereich der $n$ -ten Größe des Interesses
$(Z \times D, \sigma(\mathcal{Z} \times \mathcal{D}), \Pi)$	Bayes'sches Experiment
$(Z, \mathcal{Z}, P)$	Mit der A-Priori-Verteilung korrespondierender Wahrscheinlichkeitsraum

## Lokale Bayes'sche Fusion

$\beta, \tilde{\beta}$	In Abschnitt 7.2 hergeleitete untere Schranken für $P(U d)$
$c_{s,t}$	Argument eines lokalen Maximums von $o_s(\cdot)$
$d(A)$	Länge des Wahrscheinlichkeitsintervalls für das Ereignis $A$
$[l(A), r(A)]$	Wahrscheinlichkeitsintervall für das Ereignis $A$
$l(d \bar{U})$	Siehe Gleichung (5.30)
$o(A,B)$	(Lokale und globale) Posterior-Odds zweier Ereignisse $A, B \subset U$
$o_s(\cdot)$	Quellenspezifischer Operator zur Vorauswertung des Informationsbeitrags $d_s$
$\pi$	Parameter, welcher zur Angabe der globalen (Maximum-Entropie) A-Posteriori-Verteilung gemäß Theorem 7.4 verwendet wird.
$p(d \bar{U})$	Siehe Gleichung (5.29)



$p(z U)$	Lokale A-Priori-Verteilung von $z$ , wenn die lokale Bayes'sche Fusion eingeschränkt auf $U$ erfolgt.
$p_{U(c_{s,t})}(z d)$	Lokale A-Posteriori-Verteilung bezüglich des Anteils $U(c_{s,t})$ von $U$
$p_{U(c_{s,t})}(z d_s)$	Analogon zu $p_{U(c_{s,t})}(z d)$ , wenn nur der Informationsbeitrag $d_s$ berücksichtigt wird.
$p(z d,U)$	Lokale A-Posteriori-Verteilung von $z$ , wenn die lokale Bayes'sche Fusion eingeschränkt auf $U$ erfolgt.
$p_C(x)$	Wahrscheinlichkeitsverteilung über $X := \{U, \bar{U}\}$ , welche $U$ gerade die Wahrscheinlichkeit $P(U)$ zuweist.
$p_C(x d)$	Wahrscheinlichkeitsverteilung über $X := \{U, \bar{U}\}$ , welche $U$ gerade die Wahrscheinlichkeit $P(U d)$ zuweist.
$P_{\bar{U}}$	Wahrscheinlichkeitsmaß, welches so beschaffen ist, dass es die Wahrscheinlichkeit der Ereignisse $A \in \mathcal{D}$ unter der Hypothese, dass $z \in \bar{U}$ gilt, widerspiegelt.
$\{P_z   z \in U\}$	Menge von Wahrscheinlichkeitsmaßen über $(D, \mathcal{D})$ , welche bei Lokalität durch Einschränkung mit der Likelihood-Funktion korrespondiert.
$\{P_z   z \in U\} \cup \{P_{\bar{U}}\}$	Menge von Wahrscheinlichkeitsmaßen über $(D, \mathcal{D})$ , welche bei Lokalität durch Vergrößerung mit der Likelihood-Funktion korrespondiert.
$\rho_{(j_1, j_2)}$	Faktor, anhand dessen Agent Nummer $j_1$ die Fusionsergebnisse von Agent Nummer $j_2$ zu seinen in Bezug setzen kann.
$U = U_1 \times \dots \times U_N$	Lokaler Kontext $U \subset Z$
$U(c_{s,t})$	Mit der Spur $c_{s,t}$ korrespondierender Anteil des lokalen Kontexts $U$
$U^{(j)}$	$j$ -ter Bestandteil des lokalen Kontexts $U$

$V$	Alternativer lokaler Kontext $V \subset Z$ , z. B. bei vergleichender Betrachtung
$\mathcal{Z} _U$	Spur- $\sigma$ -Algebra von $\mathcal{Z}$ in $U$

## Informationstheorie

$F_1[q(z d)]$	Von Zellner formuliertes Funktional (vgl. Abschnitt 6.3.1)
$F_2[q(z d)]$	Von Walker formuliertes Funktional (vgl. Abschnitt 6.3.2)
$H[p(x)]$	(Shannon'sche Entropie oder Boltzmann) Entropie von $p(x)$
$I_{p(x)}[A]$	Maß für den Informationszuwachs im Hinblick auf die Zufallsvariable $x$ , wenn der bezüglich dieser vorliegende Informationsstand durch die Wahrscheinlichkeitsverteilung $p(x)$ beschrieben ist und die hier durch $A$ notierte Information zusätzlich verfügbar wird.
$\tilde{I}_{p(x)}[A]$	Alternative Definition von für $I_{p(x)}[A]$ , welche der Vollständigkeit halber erwähnt wird.
$J[x,y]$	Transinformation der Zufallsvariablen $x$ und $y$
$KD[p(x),q(x)]$	Kullback-Leibler-Divergenz von $p(x)$ relativ zu $q(x)$

## Entscheidungen unter Risiko und unter (linearer) partieller Information

$a \in A$	Mögliche Aktion
$A$	Menge der möglichen Aktionen

---

$a^*, a^{**}$	Zwei konkrete mögliche Aktionen
$a_i$	Aktion Nummer $i$
$a_{\text{opt}}$	Optimale Aktion, welche den a posteriori erwarteten Nutzen maximiert.
$\bar{E}_{W_{\text{LF}}}[u_{\text{LF}}(a, \nu_i)]$	Maximaler Wert für den erwarteten Nutzen $\bar{E}_p[u_{\text{LF}}(a, \nu_i)]$ der Aktion $a$ , wenn $\nu_i$ eine Verteilung $p \in W_{\text{LF}}$ besitzt.
$\underline{E}_{W_{\text{LF}}}[u_{\text{LF}}(a, \nu_i)]$	Minimaler Wert für den erwarteten Nutzen $\underline{E}_p[u_{\text{LF}}(a, \nu_i)]$ der Aktion $a$ , wenn $\nu_i$ eine Verteilung $p \in W_{\text{LF}}$ besitzt.
$I_{W_{\text{LF}}}[u_{\text{LF}}(a, \nu_i)]$	Intervall für den erwarteten Nutzen $E_p[u_{\text{LF}}(a, \nu_i)]$ der Aktion $a$ , wenn $\nu_i$ eine Verteilung $p \in W_{\text{LF}}$ besitzt.
$p^i$	$i$ -te Komponente des Vektors $p$ , welcher eine diskrete Wahrscheinlichkeitsverteilung beschreibt.
$R(a)$	Regret der Aktion $a$
$R_{\text{LF}}(a)$	Aus den lokalen Berechnungen sich ergebender Wert für den Regret der Aktion $a$
$u(a, z)$	Nutzenfunktion: quantifiziert den aus der Auswahl einer Aktion $a \in A$ resultierenden Nutzen, wenn tatsächlich $z$ der „wahre“ Wert der Größen des Interesses ist.
$u_{\text{LF}}(a, \nu_i)$	Nutzenfunktion bei Abbildung des auf Basis des Ergebnisses der lokalen Bayes'schen Fusion zu lösenden Entscheidungsproblems auf ein Entscheidungsproblem unter linearer partieller Information
$V(W)$	Menge der Ecken von $W$
$V(W_{\text{LF}})$	Menge der Ecken von $W_{\text{LF}}$
$w$	Element (Punkt) aus $W$

$W$	Teilgebiet des Verteilungssimplex $[0,1]^k$ $W$ stellt im betrachteten Fall linearer partieller Information ein konvexes Polyeder dar.
$W_{LF}$	Teilgebiet des Verteilungssimplex $[0,1]^{M+1}$ bei Abbildung des auf Basis des Ergebnisses der lokalen Bayes'schen Fusion zu lösenden Entscheidungsproblems auf ein Entscheidungsproblem unter linearer partieller Information $W$ stellt im betrachteten Fall linearer partieller Information ein konvexes Polyeder dar.
$w_{\max}$	Element (Punkt) aus $V(W)$ , welches eine lineare, reellwertige Funktion über $W$ maximiert.
$w_{\min}$	Element (Punkt) aus $V(W)$ , welches eine lineare, reellwertige Funktion über $W$ minimiert.
$Z_{LF}$	Menge $\{\nu_1, \dots, \nu_M, \nu_{M+1}\}$ , durch die $Z$ bei Abbildung des auf Basis des Ergebnisses der lokalen Bayes'schen Fusion zu lösenden Entscheidungsproblems auf ein Entscheidungsproblem unter linearer partieller Information ersetzt wird.

# 1 Einleitung

## 1.1 Motivation

Die Lösung diverser Aufgabenstellungen aus unterschiedlichsten Anwendungsbereichen profitiert von der Informationsfusion oder setzt diese sogar voraus, da sich nur so eine ausreichend hochwertige Informationsgrundlage schaffen lässt. Als Beispiele für konkrete Anwendungen der Informationsfusion seien Veröffentlichungen aus den Bereichen Sichtprüfung [Hei08, Mey18], Aufklärung und Überwachung [Bou16, Koc05, Bou17, Kuw18], Fahrerassistenz [Kuk18, Stü04], Messtechnik [Som09], Robotik [Bel15, Inc19] und Medizin [Jam17] angeführt. Die zu fusionierenden Informationsbeiträge können sich in vielerlei Hinsicht unterscheiden, z. B. in ihrem Informationsgehalt, in der Natur der zugrundeliegenden, z. B. physikalischen (Mess-)Größen, in ihrem Abstraktionsgrad oder in ihrer Formalisierung. Die Fusion der Beiträge heterogener Informationsquellen bietet oftmals besonderes Potential, da diese sich grob gesagt aufgrund unterschiedlicher Stärken und Schwächen besonders gut gegenseitig ergänzen können.

Wie in der Fachliteratur immer wieder betont wird, hat die (maschinelle) Informationsfusion ihr Vorbild in der realen Welt (siehe z. B. [Das97], [Hal97], [Kha09], [Koc14] und [Mac10]). Mittels ihrer kognitiven Fähigkeiten können Menschen und andere Lebewesen in vielen Fällen auf eine bis heute maschinell nicht vollständig nachahmbare, effiziente Weise unterschiedliche Informationsbeiträge fusionieren. Diese Fähigkeit ist eine wesentliche Voraussetzung dafür, dass sie in der Lage sind, ihre Umwelt ausreichend genau, ausreichend schnell und in einer konsistenten Weise zu erfassen.

Die kognitiven Fähigkeiten des Menschen erlauben es ihm in der Regel, auch beim Vorliegen heterogener Informationsbeiträge recht mühelos die aufgabenrelevanten Aspekte aus diesen zu extrahieren, in Bezug zu setzen, zu einem nützlichen Resultat zu kombinieren und darauf basierend letztlich fundierte Entscheidungen zu treffen. Gerade in Zeiten von Big Data stehen Daten und Information jedoch oft in großen Mengen zur Verfügung. In solchen Situationen ist der Mensch dann schnell überfordert (vgl. z. B. [Bos07a]). Dies gilt auch insbesondere, wenn das Ergebnis innerhalb eines vorgegebenen Zeitrahmens benötigt wird oder wenn kritische Entscheidungen zu treffen sind (vgl. z. B. auch [Aki02], [Cum05] und [Koc14]).

Heutige IT-Systeme besitzen dagegen die Leistungsfähigkeit, auch umfangreiche Daten- und Informationsbestände zeitnah zu überschauen und in kontrollierter Weise zu verarbeiten. Maschinell umsetzbare Modelle und Verfahren, die IT-Systeme und -Anwendungen in die Lage versetzen, gerade auch mit Daten und Information von heterogener Natur dem Menschen ähnlich in „kognitiver“ Weise umzugehen, d. h. sie zu erschließen, zu formalisieren, zu kombinieren und letztlich bestmöglich für die Lösung einer konkreten Aufgabenstellung nutzbar zu machen, besitzen deshalb ein hohes Potential. Hier mangelt es vor allem aber noch an adäquaten Werkzeugen für die Fusion syntaktisch inkompatibler Informationsbeiträge. Auch bedarf es passend ausgestalteter Mensch-Maschine-Schnittstellen, oftmals auch zur Umsetzung interaktiver Assistenzsysteme (vgl. z. B. [Bla08], [Bos07a] und [Sch06]).

Als Basis hierfür notwendig sind mathematische Modelle und Verfahren, um unterschiedliche Informationsbeiträge in geeigneter Weise in eine einheitliche formale Beschreibung zu überführen und sie anschließend zu kombinieren. Dabei muss es methodisch gesehen auch möglich sein, das Fusionsergebnis auf spezielle Einzelfragestellungen zu konzentrieren und es müssen in fundierter und für den Nutzer auch verständlicher Weise Entscheidungen von ihm ableitbar sein. Die Modelle und Verfahren müssen in Software- und Systemarchitekturen sowie in praxisgerechte Algorithmen und Software-Anwendungen umsetzbar sein. Von besonderer Relevanz ist dabei oftmals auch die Skalierbarkeit der Algorithmen bzgl. Rechen- und Speicherbedarf, insbesondere dann, wenn die verfügbaren Ressourcen deutlich begrenzt sind,

z. B. im Fall mobiler Roboter oder beim Vorliegen von Bandbreitenbeschränkungen in Netzwerken. Auch sollte eine modulare Erweiterbarkeit in Bezug auf neue Aufgaben- bzw. Fragestellungen sowie in Bezug auf das Verfügbarwerden zusätzlicher, auch neuartiger Informationsquellen gegeben sein.

Die Welt der Zukunft ist zunehmend digitalisiert und vernetzt. Moderne Kommunikationstechnologien können die Verteilung von Daten und Information in (nahezu) Echtzeit ermöglichen und so die technische Basis schaffen, um Organisationen, Systeme und deren Fähigkeiten über Zeit und Raum zu vernetzen. Die technologische Entwicklung in Bezug auf Sensoren und Plattformen, Datenbanken etc. sowie Netzwerktechnik haben sich zudem in den vergangenen Jahren rasant entwickelt und auch große Mengen an Daten können heute vom technischen Standpunkt aus gesehen oft unkompliziert geteilt werden. Um die sich hieraus ergebenden Möglichkeiten nutzen zu können, besteht ein Bedarf an Systemen bzw. Systemverbänden, die in einer definierten Weise interoperabel miteinander interagieren können. Basis für deren Erstellung sind die Fähigkeit zum übergreifenden Daten- und Informationsaustausch, gegebenenfalls auch unter hoher Zeitkritikalität, auf Basis kompatibler Schnittstellen und die Fähigkeit, die Daten und Information im Verbund korrekt zu interpretieren, zu verarbeiten, sie insbesondere auch zu fusionieren und dem Nutzer seiner Rolle bzw. Funktion entsprechend bedarfsgerecht zur Verfügung zu stellen. Aktuelle Beispiele für den Bedarf an intelligenten Softwaresystemen und Systemverbänden, die die Wertschöpfung von Information und die Umsetzung von Geschäftsprozessen auch über System- und Organisationsgrenzen unterstützen, ergeben sich z. B. aus der Notwendigkeit zur organisations- und oft auch grenzüberschreitenden Zusammenarbeit zum Schutz der Sicherheit einer Nation [Cat17, Ess18a, Zas16] sowie zur Krisenfrüherkennung und zum Krisenmanagement [Ess19, Kuw19, Rot16].

Die IT-gestützte Fusion von sensorieell erfassten Daten, in Datenbanken vorgehaltenen Datenbeständen, kollateral vorliegender Information, Kontext- und Hintergrundwissen etc. stellt insgesamt gesehen eine interdisziplinäre wissenschaftliche Herausforderung dar, welche noch nicht als vollständig gelöst betrachtet werden kann. Das Forschungsgebiet der Informationsfusion mutet oftmals auch recht divers und herausfordernd an, was vor allem einerseits

an der Vielfalt der adressierten Aufgabenstellungen und andererseits an der Vielfalt der zur Problemlösung angewandten, auch interdisziplinären Ansätze liegt (vgl. [Hal04] und [Kla99]). Die Anfänge dieses Forschungsgebiets liegen im militärischen Kontext, was sich u. a. darin widerspiegelt, dass wichtige Grundlagenliteratur von Anwendungen aus dem Bereich der militärischen Aufklärung geprägt ist.

Qualitative Verfahren (z.B. logik- oder regelbasierte Verfahren) erlauben es oftmals, Daten und Information in effektiver Weise gemäß vordefinierter Axiome und Regeln zielgerichtet zu verdichten sowie sie zu neuer Information zu verarbeiten. Die Repräsentation vorliegender Kontext- und Hintergrundinformation mittels formaler Wissensmodelle (z.B. Ontologien) ermöglicht es dabei auch, die Daten und Information gemeinsam in einen semantischen Zusammenhang einzuordnen (vgl. z. B. [Sch06]). Eine grundlegende Anforderung in vielen Anwendungen der Informationsfusion ist darüber hinaus aber auch die Möglichkeit zur Modellierung der Informationsquellen bzw. der von ihnen gelieferten Informationsbeiträge im Hinblick auf die mit diesen korrespondierenden Unsicherheiten und die Propagation dieser Unsicherheiten durch den Fusions- sowie ggf. auch einen anschließenden Entscheidungsprozess. Vgl. hierzu z. B. auch [Das08] und [Kuw18].

Wichtige im Bereich der Informationsfusion eingesetzte quantitative Verfahren beruhen u. a. auf der Bayes'schen Theorie, der klassischen Statistik, der Dempster-Shafer-Theorie, der Fuzzy-Theorie und der Theorie neuronaler Netze (siehe z. B. [Kha09], [Kle04], [Men20], [Mit07]). Mindestens für die meisten anwendungsbezogenen Aufgabenstellungen stellt der Verzicht auf eine Unterscheidung von Unsicherheiten hinsichtlich ihrer Natur (Indeterminismus, Ungenauigkeit, Unwissenheit, Unschärfe, etc.) eine sinnvolle pragmatische Vorgehensweise dar (vgl. [Bey07]). In diesem Fall liefert die Bayes'sche Theorie oftmals ausreichend mächtige Werkzeuge zur Repräsentation von Unsicherheiten und letztlich zur Umsetzung der Informationsfusion auf Basis Bayes'scher Inferenz. Der Informationsstand in Bezug auf alle in eine Fusionsaufgabe involvierten Größen wird dabei probabilistisch mittels Wahrscheinlichkeitsverteilungen in der sog. Degree-of-Belief-Interpretation von Wahrscheinlichkeit repräsentiert. Zusätzlich vorhandenes Vorwissen



wird durch die A-Priori-Verteilung verkörpert und kann dadurch vollständig und in adäquater Weise zusammen mit den von den Informationsquellen gelieferten Informationsbeiträgen in die Lösung der Fusionsaufgabe eingehen. Die Kombination der unterschiedlichen Anteile erfolgt durch die Anwendung des Satzes von Bayes, welcher als vollständiges Fusionsresultat die A-Posteriori-Verteilung liefert. Basierend auf dieser lassen sich dann auch spezielle Einzelfragestellungen beantworten und insbesondere auch Entscheidungen ableiten.

Wie im Verlauf der vorliegenden Arbeit noch genauer dargestellt werden wird, ist die Bayes'sche Fusionsmethodik äußerst fundiert und erfüllt alle essentiellen Anforderungen an eine sinnvolle Methodik zur Fusion der Beiträge heterogener Informationsquellen. Die weitreichende Interpretation von Wahrscheinlichkeit als Degree-of-Belief, welche zudem oftmals dem intuitiven Verständnis von Unsicherheit entspricht, das gerade auch nicht speziell geschulte Personen mitbringen, eröffnet ihr ein breites Anwendungsspektrum. Nicht zuletzt ergibt sich dadurch auch ein durchgängiges Konzept, um mittels Bayes'scher Verfahren Daten und Information unter Berücksichtigung der korrespondierenden Unsicherheiten konsistent weiter zu propagieren (z. B. gerade auch durch die verschiedenen Ebenen des JDL Fusionsmodells), sowie zur Umsetzung brauchbarer Mensch-Maschine-Schnittstellen.

In praktisch relevanten Aufgabenstellungen der Informationsfusion ist jedoch der Zielgrößenraum, d. h. der Wertebereich der Größen des Interesses, welche letztlich den im Rahmen der Fusion interessierenden Sachverhalt mathematisch fassen, oft sehr umfangreich und von recht hoher Dimension. In solchen Situationen verursachen Bayes'sche Verfahren in der Regel hohe, in der Praxis oftmals nicht tragbare Rechenzeiten und immensen Speicherbedarf. Dies begründet sich vor allem aus der Tatsache, dass gemäß dem Bayes'schen Formalismus im Rahmen der Fusion immer die komplette A-Posteriori-Verteilung über dem gesamten Wertebereich der Größen des Interesses berechnet werden muss. Diese verkörpert gerade den sich durch die Fusion im Hinblick auf die Größen des Interesses ergebenden vollständigen Informationsstand.

## 1.2 Zielsetzung

Die vorliegende Arbeit führt neue Ansätze zur Bewältigung Bayes'scher Fusion beim Vorliegen eines umfangreichen, ggf. hochdimensionalen Zielgrößenraums ein. Diese werden unter dem Oberbegriff der lokalen Bayes'schen Fusion zusammengefasst. Das primäre Ziel der Arbeit besteht darin, diese Ansätze formal mathematisch zu formulieren und sie grundlegend zu untersuchen, insbesondere im Hinblick auf ihre Umsetzbarkeit gerade auch für die Fusion der Beiträge heterogener Informationsquellen, ihre Ausdrucksstärke und letztlich ihre Fähigkeit, die Beantwortung von im Rahmen der Informationsfusion vorliegenden Fragestellungen in ausreichender Form zu ermöglichen.

Der Grundgedanke lokaler Bayes'scher Fusionsansätze leitet sich daraus ab, dass in vielen Aufgabenstellungen der Informationsfusion der durch die A-Posteriori-Verteilung in probabilistischer Weise verkörperte vollständige, d. h. maximal umfassende Informationsstand im Hinblick auf die Größen des Interesses eigentlich gar nicht benötigt wird. Vielmehr müssen auf Basis des sich a posteriori ergebenden Informationsstands Antworten auf bestimmte, z. B. für einen Entscheidungsträger relevante Fragestellungen geliefert werden. Beispiele für derartige Fragestellungen sind die Bestimmung einer Schätzung für den tatsächlich vorliegenden Wert der Größen des Interesses oder die bestmögliche Auswahl einer Handlungsoption. Die vollständige A-Posteriori-Verteilung ist in solchen Fällen also letztlich, d. h. als eigentliche Antwort auf diese Fragen, gar nicht nötig. Sie stellt vielmehr nur die Basis zur Ableitung der Antworten dar.

Das Ziel lokaler Bayes'scher Fusion ist es dementsprechend, die Bayes'sche Fusion möglichst konzentriert auf die Anteile des Wertebereichs der Größen des Interesses durchzuführen, in denen sich tatsächlich Aufgabenrelevantes abspielt – in dem Sinne, dass eine detaillierte Betrachtung dieser Anteile des Zielgrößenraums tatsächlich notwendig ist zur Beantwortung der eigentlichen Fragen. Letztlich soll es durch die lokale Bayes'sche Fusion also ermöglicht werden, diese Fragen auf Basis einer nur teilweisen Bestimmung der A-Posteriori-Verteilung zu beantworten und ihre vollständige Berechnung damit zu umgehen.

Zur Präzisierung der Idee, einen derartigen lokalen Standpunkt bei der Bayes'schen Fusion umzusetzen, wurde in [Bey06a] eine Analogie zwischen lokaler Bayes'scher Fusion und kriminalistischen Ermittlungen zur Aufklärung eines Verbrechens hergestellt, also zu einem etablierten Prozess der realen Welt. Diese inspirierte wiederum das Konzept für eine agentenbasierte Fusionsarchitektur, in der unterschiedliche Agenten in kooperativer Weise Spuren im Hinblick auf den tatsächlich vorliegenden („wahren“) Wert der Größen des Interesses in den Informationsbeiträgen identifizieren, auswerten sowie die damit korrespondierenden Hypothesen durch Fusion lokal im Wertebereich der Größen des Interesses weiterentwickeln und letztlich zur Zusammenschau bereitstellen.

In der Arbeit wird angestrebt, die mathematische Basis für die lokale Bayes'sche Fusion derart auszugestalten, dass eine verteilte Realisierung lokaler Bayes'scher Fusion wiederum gerade eine natürliche Entsprechung in der agentenbasierten Fusionsarchitektur hat. Die softwaretechnische Umsetzung der agentenbasierten Fusionsarchitektur stellt dabei kein Thema der Arbeit dar.

Der erste Schritt zur Lösung eines Fusionsproblems ist letztlich dessen formale Modellierung. Zur Ermöglichung der Zielsetzung, dass die Bayes'sche Fusion wie beschrieben möglichst konzentriert auf bestimmte Anteile des Zielgrößenraums erfolgen soll, bedarf es der Formalisierung lokaler Bayes'scher Modelle. Diese müssen die probabilistische Repräsentation und auch die Weiterentwicklung lokaler Informationsstände im Hinblick auf die Größen des Interesses auf Basis Bayes'scher Inferenz in fundierter Weise ermöglichen.

Um die lokalen Bayes'schen Modelle sinnvoll festlegen zu können, d. h. um letztlich die Frage beantworten zu können, welche Anteile des Zielgrößenraums im beschriebenen Sinne aufgabenrelevant sind, bedarf es möglichst aussagekräftiger Kriterien im Hinblick auf die Festlegung des diese umfassenden lokalen Kontexts. Auch sollten Prinzipien und Methoden bereitstehen, auf deren Basis man die Adäquatheit des lokalen Kontexts vorab, d. h. vor Durchführung der Fusion, und möglichst auch ohne vollständige Kenntnis des globalen Bayes'schen Modells wenigstens in einem Mindestmaß auch quantitativ bewerten kann.

Wird die vollständige Bestimmung der A-Posteriori-Verteilung bei der lokalen Bayes'schen Fusion wie beschrieben umgangen, so ist nach Durchführung der Fusion nur ein lokaler (partieller) Informationsstand im Hinblick auf die Größen des Interesses verfügbar. Es muss die Möglichkeit gegeben sein, die lokal vorliegende Information in aufgabengerechter Weise und unter Berücksichtigung der im Hinblick auf die Bildung des lokalen Kontexts vorhandenen (Zusatz)information in objektiver Weise in den eigentlichen, globalen Zusammenhang einzuordnen und darauf basierend auch weiterführende Entscheidungen treffen zu können.

### 1.3 Aufbau der Arbeit

Die eigentliche Arbeit beginnt in Kapitel 2 mit einer Aufarbeitung zentraler Aspekte im Hinblick auf den Stand der Wissenschaft und Technik zum Forschungsgebiet Informationsfusion. Die vorgenommene Darstellung hat vor allem auch das Ziel, einen strukturierenden Überblick über dieses oftmals divers anmutende Forschungsgebiet zu schaffen. Im Zuge dieser Aufarbeitung des Stands der Wissenschaft und Technik werden dann zentrale Begrifflichkeiten im Hinblick auf die Informationsfusion abgeleitet und es wird dediziert auf die Frage nach den Anforderungen, welche eine sinnvolle Methodik zur Fusion von Informationsbeiträgen heterogener Quellen erfüllen muss, eingegangen. Ein besonders Augenmerk in diesem Kapitel wird auch auf die Umsetzung der Informationsfusion in Verbänden gelegt.

In Kapitel 3 werden für die Arbeit wesentliche Grundlagen im Hinblick auf die Bayes'sche Fusion erschlossen. Vorbereitend für den weiteren Verlauf der Arbeit wird dabei zuerst einmal ein allgemeinerer Zusammenhang zwischen der Informationsfusion und der induktiven Statistik hergestellt, auf dessen Basis dann unterschiedliche Inferenzkonzepte eingeführt und im Hinblick auf den Anwendungskontext der Informationsfusion eingeordnet werden. Im Anschluss daran wird detaillierter auf die Bayes'sche Fusionsmethodik eingegangen. Hier werden dann insbesondere auch etablierte Mechanismen eingeführt, anhand derer die Bayes'sche Fusionsmethodik den essentiellen Anfor-

derungen an eine sinnvolle Fusionsmethodik zur Fusion von Informationsbeiträgen heterogener Quellen nachzukommen vermag.

Kapitel 4 adressiert speziell die Herausforderungen, welche sich bei der Bayes'schen Fusion im Hinblick auf die Durchführung der notwendigen Berechnungen ergeben. In diesem Kapitel wird im Rahmen eines Exkurses ein komprimierter Überblick über die wesentlichen in der Fachliteratur vorhandenen Ansätze zur Bewältigung des oft hohen Aufwands Bayes'scher Inferenz vorgenommen. Im Anschluss an diese Darstellung und Diskussion des Stands der Wissenschaft und Technik im Hinblick auf die Berechnung der A-Posteriori-Verteilung bzw. von Kenngrößen derselben werden das Konzept für die agentenbasierte Fusionsarchitektur und dessen Bezugspunkte in Form der Analogie zwischen lokaler Bayes'scher Fusion und kriminalistischen Ermittlungen zur Aufklärung eines Verbrechens sowie wesentliche Grundlagen zur Agententheorie dargestellt und diskutiert.

Aus theoretischer Hinsicht bedarf es zur genaueren Entwicklung lokaler Bayes'scher Fusionsansätze einer Anpassung der Bayes'schen Fusionsmethodik an lokale Betrachtungen. Die notwendigen mathematischen Grundlagen hierfür werden in Kapitel 5 erarbeitet. Hier werden die benötigten lokalen Bayes'schen Modelle, welche gerade die Repräsentation und Propagation lokaler Informationsstände ermöglichen, konzipiert, formalisiert und ausführlich analysiert. Adressiert wird auch die Ausgestaltung dieser Modelle für den Fall, dass die lokale Bayes'sche Fusion kooperativ durch unterschiedliche Agenten, welche jeweils lokale Informationsstände in Bezug auf unterschiedliche Anteile des lokalen Kontexts besitzen, vorgenommen wird. Die so insgesamt formulierten lokalen Bayes'schen Modelle sind aus mathematischer Hinsicht valide, was unabdingbar dafür ist, dass die bei der lokalen Bayes'schen Fusion anvisierte Einnahme eines lokalen Standpunkts konsistent in den Bayes'schen Kalkül integrierbar ist. Ergänzend und in gewissem Sinne auch komplementär zum Vorangegangenen werden weiter auf Basis einer vereinfachten, in Teilen heuristischen Vorgehensweise ausgewählte Teilaspekte im Hinblick auf die lokale Bayes'sche Fusion anhand eines ausführlicheren, möglichst anwendungsnahen Beispiels umgesetzt und untersucht.

Kapitel 6 behandelt die Frage, was aus mathematischer Hinsicht sinnvolle Vorgehensweisen zur Festlegung eines lokalen Kontexts sind und wie sich die Güte lokaler Bayes'scher Modelle generell in Abhängigkeit von der Wahl des lokalen Kontexts bewerten lässt. Hierzu werden zum einen Prinzipien aus dem Kontext der Likelihood-Inferenz und zum anderen Ansätze aus dem Kontext der Informationstheorie aufgegriffen, für die lokale Bayes'sche Fusion angewandt und zueinander in Bezug gesetzt. Der Fokus der Untersuchungen liegt dabei vor allem auf einem Ansatz zur lokalen Bayes'schen Fusion, welcher eine vollständige Einschränkung des probabilistischen Modells auf den lokalen Kontext vornimmt und im vorangegangenen Kapitel als besonders vielversprechend für die lokale Bayes'sche Fusion heterogener Informationsquellen identifiziert wurde.

Die aus der lokalen Bayes'schen Fusion resultierende lokale A-Posteriori-Verteilung stellt partielle Information im Hinblick auf die globale A-Posteriori-Verteilung dar. In Kapitel 7 wird im Hinblick auf den Ansatz zur lokalen Bayes'schen Fusion, welcher eine vollständige Einschränkung des probabilistischen Modells auf den lokalen Kontext vornimmt, untersucht, wie sich das in Form der lokalen A-Posteriori-Verteilung vorliegende lokale Fusionsresultat auf Basis dieses Gedankens und unter zusätzlicher Verwendung von Information darüber, wie der lokale Kontext gebildet wurde, expliziter in Bezug zum in Form der globalen A-Posteriori-Verteilung resultierenden Output der globalen Bayes'schen Fusion setzen und auf dieser Basis auch zur Ableitung weiterführender Entscheidungen nutzen lässt. Hierzu wird ein Wahrscheinlichkeitsintervall-Schema formuliert, anhand dessen sich lokale und globale A-Posteriori-Wahrscheinlichkeiten genauer in Beziehung setzen lassen. Weiter werden Möglichkeiten zur Ausweitung der lokalen A-Posteriori-Verteilung auf eine eindeutige globale A-Posteriori-Verteilung und zur direkten Einbringung der partiellen Information in eine nachfolgende Entscheidungsfindung erarbeitet und in Bezug auf Kernaspekte diskutiert.

Die Arbeit schließt in Kapitel 8 mit Zusammenfassung und Ausblick.

In Bezug auf die vorliegende Arbeit ist der Stand der Wissenschaft und Technik zu weiteren Themengebieten als den oben genannten relevant und be-

stimmte Punkte müssen partiell auch vertieft betrachtet werden. Diese Darstellung erfolgt der Übersichtlichkeit halber jeweils ergänzend in den Kapiteln, wo die Anteile auch benötigt werden.

Im Hinblick auf die praktische Umsetzung ergab sich für die vorliegende Arbeit die Herausforderung, dass die Umsetzung eines Anwendungsbeispiels, das ein ausreichend hohes Maß an Komplexität aufweist, um die im Rahmen der Arbeit erarbeiteten Beiträge vollständig adäquat abzubilden, weder mit vertretbarem Aufwand noch tatsächlich vollständig wirklichkeitsgetreu realisierbar war. Aus diesen Gründen wurde entschieden, die Ergebnisse vor allem anhand illustrativer, dafür aber auch nachvollziehbarer „kleiner“ Beispiele in den jeweiligen Abschnitten darzustellen. Das in Abschnitt 5.3 adressierte, recht ausführliche Beispiel stellt zumindest in gewissem Sinne eine Ausnahme hiervon dar. Dieses Beispiel schien notwendig, da es trotz einer (notwendigerweise) nicht vollständigen Realitätstreue und einer vereinfachten und in Teilen auch heuristischen Umsetzung der lokalen Bayes'schen Fusion weitere relevante Erkenntnisse für die vorliegende Arbeit liefert, welche die aus den theoretischen Untersuchungen resultierenden Erkenntnisse komplettieren.

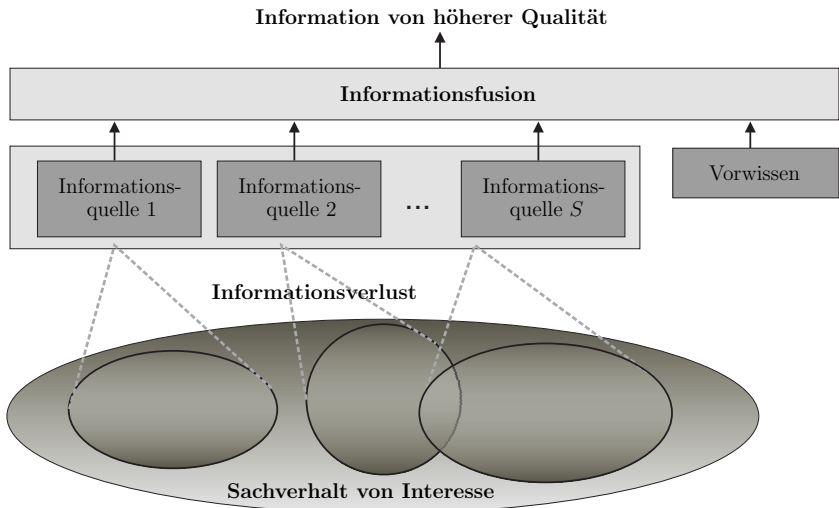




## 2 Informationsfusion

### 2.1 Grundprinzip

In [Hal04] wird die Informationsfusion äußerst treffend charakterisiert als „a discipline [...] to solve a diverse set of problems having common characteristics“. Diese gemeinsamen Charakteristiken ergeben sich direkt aus dem Grundprinzip der Informationsfusion. Dieses ist in Abb. 2.1 illustriert und wird im vorliegenden Abschnitt ausführlich erläutert.



**Abbildung 2.1:** Grundprinzip der Informationsfusion. Abbildung angelehnt an [San10b].

Ein Maß für Information stellt die durch ihr Hinzukommen beseitigte Ungewissheit dar<sup>1</sup> [Bey07]. Der Begriff Ungewissheit ist dabei zu verstehen im Sinne von Unsicherheit in Bezug auf den in einer konkreten Aufgabenstellung interessierenden Sachverhalt. Die Erfassung des interessierenden Sachverhalts durch eine Informationsquelle geht i. A. mit einem Informationsverlust bzw. komplementär gesprochen mit einer Zunahme an Unsicherheit einher. Beispielsweise können Teile der zur Lösung der zugrundeliegenden Aufgabenstellung benötigten Nutzinformation ungenauer werden oder gänzlich verloren gehen.

Dieses Phänomen lässt sich in Anlehnung an [Das97] sehr anschaulich darstellen, indem man Informationsquellen mit spezifisch ausgeprägten Fenstern in die Welt vergleicht. Die konkrete Ausprägung dieser Fenster sowie der Kontext, innerhalb dessen die Informationserfassung geschieht, entscheiden darüber, welche Information in welcher Qualität<sup>2</sup> erfasst wird. Zum Beispiel erfasst ein physikalischer Sensor entsprechend seiner Konstruktionsprinzipien gerade bestimmte physikalische Eigenschaften der Umwelt wie z. B. akustische oder elektromagnetische Wellen und sowohl die Sensorgüte als auch die beim Einsatz des Sensors vorliegenden Rahmenbedingungen (z. B. Umgebungskonditionen wie Wetterbedingungen) beeinflussen weiter die Qualität der erfassten Information.

Erfolgt die Informationserfassung mittels einer (digitalen) sensorischen Informationsquelle, so sind oftmals insbesondere folgende Aspekte im Hinblick auf den Informationsverlust relevant (vgl. [Hei08] und [San10b]):

- Fensterung: Die Informationsquelle erfasst nur einen begrenzten Ausschnitt des interessierenden Sachverhalts. Liefert die Informationsquelle z. B. Bildinformation in Form von Einzelbildern oder Videos, so wäre dieser Ausschnitt zeitlich, örtlich und/oder spektral zu sehen.

---

<sup>1</sup> Siehe [Bey99] für eine informationstheoretische Begründung dieses Sachverhalts.

<sup>2</sup> Der Begriff Informationsqualität lässt sich dabei gemäß [Gil10] auffassen als „measure of how well a piece of information delivered to a user describes a situation or event of interest to that the user.“

- Projektion: Die Erfassung durch die Informationsquelle geht mit einer Projektion einher. Beispielsweise liefert ein herkömmlicher bildgebender Sensor ein zweidimensionales Bild der dreidimensionalen Realität.
- Diskretisierung: Es erfolgt eine Zeit- und/oder Ortsdiskretisierung (Abtastung), z. B. im Pixelraster bei Bildern und bei Videos noch zusätzlich mit der Bildwiederholfrequenz. Weiter erfolgt eine Wertquantisierung, z. B. könnte im Fall eines bildgebenden Sensors dieser ein Grauwertbild mit 8 Bit liefern, welches 256 Grauwerte abbilden kann.
- Störungen beim Erfassungsvorgang: Beispielsweise verursacht das thermische Rauschen eines Sensorchips eine entsprechende nicht-deterministische Störung.

Erfolgt die Informationserfassung durch eine nicht sensorielle Informationsquelle, so könnten z. B. folgende Aspekte relevant sein:

- Unschärfe der Beschreibung: Beispielsweise könnte ein menschlicher Beobachter einer Szene melden, dass sich ein bestimmtes Objekt darin „schnell“ bewegt, ohne die exakte Objektgeschwindigkeit angeben zu können.
- Unzuverlässigkeit bzw. mangelnde Glaubwürdigkeit: Beispielsweise könnte eine menschliche Informationsquelle bewusst oder unbewusst Fehlinformation liefern oder die Einträge in einer Datenbank könnten mutwillig manipuliert, fehlerhaft erfasst oder veraltet sein.
- Fokussierung auf bestimmte Aspekte: Beispielsweise könnte eine Datenbank nur Information in Bezug auf bestimmte Attribute eines interessierenden Objekts vorhalten.

Durch die Fusion der Beiträge unterschiedlicher Informationsquellen lässt sich die in Bezug auf den interessierenden Sachverhalt vorherrschende Unsicherheit gezielt reduzieren. Dies kann sogar bedeuten, dass neue Information verfügbar wird. Letzteres ist insbesondere der Fall, wenn die Nutzinformati-

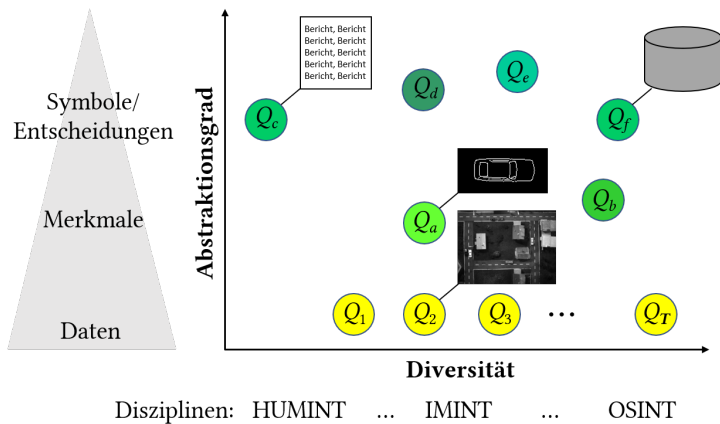
on verteilt über die Informationsbeiträge vorliegt und erst durch die Fusion überhaupt erschlossen werden kann.

Ein besonderes Potential bietet oft die Fusion der Beiträge heterogener Informationsquellen, da sich diese grob gesagt aufgrund unterschiedlicher Stärken und Schwächen optimal ergänzen können, z. B. indem sie im Extremfall komplementäre Eigenschaften des interessierenden Sachverhalts erfassen. Wie in [Ral07] sehr anschaulich dargestellt wird, lassen sich heterogene und homogene Informationsquellen (bzw. die von ihnen gelieferten Beiträge) per se nicht eindeutig voneinander abgrenzen, denn Information (bzw. die dieser zugrundeliegenden Daten) kann auf sehr unterschiedliche Weise heterogen sein bzw. dies nicht sein. Letztlich hängt die Betrachtung von der konkreten Aufgabenstellung und dem zugehörigen Kontext ab.

So werden z. B. im Bereich der militärischen Aufklärung entsprechend der grundlegenden Natur der zur Informationsgewinnung eingesetzten Quellen unterschiedliche Intelligence-Disziplinen unterschieden (siehe z. B. [Das08] oder [NAT19]). Hierzu zählen IMINT (Imagery Intelligence), HUMINT (Human Intelligence), SIGINT (Signals Intelligence), MASINT (Measurement and Signature Intelligence) und OSINT (Open Source Intelligence). Die durch die Verarbeitung und Auswertung erfasster Daten und Information resultierenden Single-Intelligence Ergebnisse können dann in höherwertigen Auswert- und Analyseprozessen zu Multiple-Intelligence Ergebnissen weiter verarbeitet, analysiert und insbesondere auch Intelligence-Disziplin-übergreifend in Bezug gesetzt und fusioniert werden. Bei genauerer Betrachtung arbeiten die einzelnen Intelligence-Disziplinen selbst wieder mit Daten und Information unterschiedlicher Herkunft, z. B. werden im IMINT Bereich unterschiedliche elektrooptische Sensoren, Radar etc. zur Datenerfassung eingesetzt.

Neben den zuvor beispielhaft aufgezeigten Unterscheidungsmöglichkeiten von Informationsquellen hinsichtlich ihrer Diversität ist auch die Unterscheidung von Informationsquellen hinsichtlich der Abstraktionsebene, auf der die von ihnen gelieferten Beiträge vorliegen, wichtig. Eine oft hilfreiche Unterscheidung ist hier die zwischen der Ebene der Daten, der Ebene der Merkmale und der Ebene der Symbole/Entscheidungen.

Abb. 2.2 illustriert die beiden zuvor dargestellten, sich ergänzenden Möglichkeiten zur Unterscheidung heterogener Informationsquellen. Im Hinblick auf die Durchführung der Fusion kann es technisch und/oder organisatorisch vorteilhaft sein, ähnliche Informationsbeiträge zunächst zu Zwischenergebnissen zusammenzuführen und somit deren Verarbeitung und insbesondere auch die Fusion schrittweise durchzuführen. Beispielsweise könnte es im Bereich der militärischen Aufklärung vorteilhaft sein, unterschiedliche Bilder in einem ersten Schritt zu einem höherwertigen Bild zu fusionieren oder unterschiedliche Bilder auszuwerten und die Ergebnisse in einem zusammenfassenden Auswertebereich (vgl. z. B. [San19]) als IMINT Ergebnis festzuhalten.



**Abbildung 2.2:** Unterscheidung heterogener Informationsquellen entsprechend ihrer Diversität (horizontal) und der Abstraktionsebene (vertikal), auf der ihre Beiträge vorliegen, am Beispiel von Intelligence-Diziplinen in der militärischen Aufklärung.

Wie bei der Illustration des Grundprinzips der Informationsfusion in Abb. 2.1 unter Verwendung des Begriffs Vorwissen bereits angedeutet, liegt es auf der Hand, dass bei der Informationsfusion auch Wissen, das zusätzlich zu den von den Informationsquellen gelieferten Beiträgen in Bezug auf den interessie-

renden Sachverhalt vorhanden ist, in adäquater Weise berücksichtigt werden sollte, um so ein möglichst hochwertiges Fusionsergebnis zu erzielen. Im Bereich der militärischen Aufklärung ist in diesem Kontext auch der Begriff Basic Intelligence gebräuchlich [NAT19]. Zur Veranschaulichung seien folgende Beispiele für Vorwissen angeführt:

- Kartenmaterial kann z. B. Information darüber liefern, wo in einem aufzuklärenden Gebiet Straßen oder Wasserwege verlaufen und damit letztlich dazu, wo sich bestimmte Objekte üblicherweise fortbewegen.
- Die grundlegende Kenntnis der Eigenschaften bestimmter Objekttypen kann z. B. Wissen dahingehend liefern, dass sich ein Objekt eines bestimmten Typs nur mit einer bestimmten Maximalgeschwindigkeit fortbewegen kann.

Vorwissen kann insbesondere auch zu früheren Zeitpunkten gewonnene bzw. abgeleitete Information oder allgemein verfügbares Wissen und Regeln (z. B. Naturgesetze) beinhalten. Wie in [Bey08] dargestellt, lässt sich diese Eingangsgröße des Fusionsprozesses prinzipiell auch technisch dahingehend nutzen, um spezielle Anforderungen an das Fusionsresultat einzubringen. Ein konkretes Beispiel hierfür ist das Einbringen von Glattheitsanforderungen an das Ergebnisbild bei der Bildfusion.

## 2.2 Zentrale Begrifflichkeiten

Der Begriff Informationsfusion ist in der Fachliteratur nicht einheitlich definiert. Hinsichtlich einer vertieften Betrachtung dieses Sachverhalts und eines Überblicks über die Vielfalt an vorhandenen Definitionen sei auf [Sni11] und insbesondere auf [Bos07b] verwiesen. Dabei ist zu beachten, dass der Begriff Informationsfusion oftmals synonym zu verwandten Begriffen wie z. B. Datenfusion verwendetet wird. Das Spektrum der Begriffsdefinitionen in der Literatur reicht von sehr allgemein gehaltenen Definitionen bis hin zu auf spezielle Anwendungen bzw. Domänen ausgerichteten Definitionen.

Beispielsweise wird in [Kle04] nachfolgende Definition zugrundegelegt:

„Data fusion is a multilevel, multifaceted process dealing with the automatic detection, association, correlation, estimation, and combination of data and information from single and multiple sources to achieve refined position and identity estimates, and complete and timely assessments of situations and threats and their significance.“

Diese sehr konkrete Festlegung geht im Wesentlichen zurück auf eine durch das amerikanische Joint Directors of Laboratories (JDL) vorgenommene und im sog. Datenfusionslexikon [Whi87] dokumentierte Definition. Um dem fachlichen Fortschritt und der zunehmenden Verbreitung von Methoden der Informationsfusion auch über den militärischen Kontext hinaus gerecht zu werden, wurde die ursprüngliche Definition des JDL in den Folgejahren modifiziert. Erwähnenswert in Bezug auf diese Überarbeitungen ist insbesondere die folgende, in [Ste99] eingeführte Definition:

„Data fusion is the process of combining data to refine state estimates and predictions.“

Parallel dazu wurde im Rahmen einer europäischen Arbeitsgruppe nachfolgende, noch weitreichendere Definition erarbeitet [Wal99]:

„Data fusion is a formal framework in which are expressed means and tools for the alliance of data originating from different sources. It aims at obtaining information of greater quality; the exact definition of ‚greater quality‘ will depend upon the application.“

Diese Definition zeichnet sich dadurch aus, dass sie explizit den Bezug zwischen Informationsfusion und Informationsqualität herstellt. Siehe [San12] für eine ergänzende Betrachtung dieses Bezugs sowie [Ess18b] für weitere Ergebnisse zum Thema Informationsqualität speziell für den Fall von Systemverbänden.

Ein weiteres Beispiel für eine sehr allgemein gehaltene Definition ist die wie folgt in [Das01] getroffene Festlegung:

„Information fusion encompasses the theory, techniques, and tools conceived and employed for exploiting the synergy in the information acquired from multiple sources (sensor, databases, information gathered by human, etc.) such that the resulting decision or action is in some sense better (qualitatively or quantitatively, in terms of accuracy, robustness and etc.) than would be possible, if these sources were used individually without such synergy exploitation.“

Dem in Abschnitt 2.1 eingeführten Grundprinzip der Informationsfusion folgend wird in der vorliegenden Arbeit nachfolgende Definition aus [Bey07] verwendet:

**Definition 2.1.** *Informationsfusion bedeutet das förderliche Zusammenführen, Überlagern und Nutzen der Beiträge mehrerer Informationsquellen im Hinblick auf eine gegebene Aufgabenstellung.*

Diese Definition zeichnet sich dadurch aus, dass sie in hohem Maße allgemeingültig ist und dabei wie auch die ersten beiden der zuvor zitierten Festlegungen den eigentlichen Prozess der Informationsfusion in den Vordergrund stellt. Ebenfalls gemäß [Bey07] ist der Begriff Information dabei in einem abstrakten, informationstheoretisch motivierten Sinne zu verstehen:

**Definition 2.2.** *Information ist alles, das im Kontext der gegebenen Aufgabenstellung relevant ist, in dem Sinne, dass es zur Verminderung einer vorhandenen Unsicherheit in Bezug auf den interessierenden Sachverhalt beitragen kann.*

Weiter ist in der vorliegenden Arbeit unter der Begrifflichkeit *Fusion unterschiedlicher Informationsquellen* die Informationsfusion der von diesen Quellen gelieferten Beiträge zu verstehen und der Begriff *Vorwissen* steht zusammenfassend für alles, das zusätzlich zu den von den Informationsquellen gelieferten Beiträgen hilfreich ist zur Generierung eines möglichst hochwertigen Fusionsresultats.

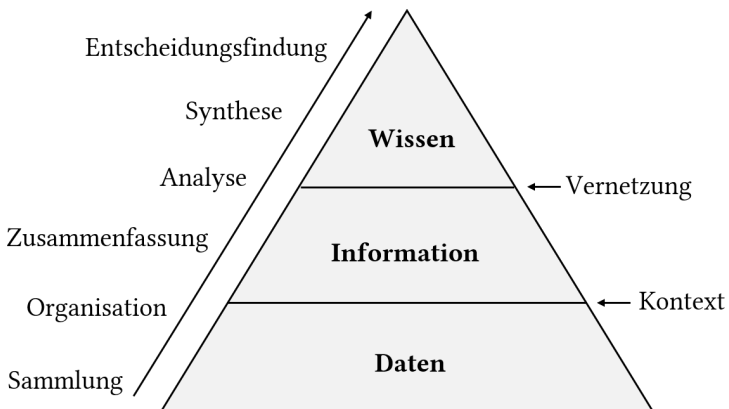
Definition 2.2 impliziert insbesondere, dass in der vorliegenden Arbeit i. A. keine explizite Unterscheidung der Begriffe Daten, Information und Wissen



erfolgt, sondern dass statt dessen der Begriff Information als übergreifender Begriff verwendet wird.

Die Unterscheidung der Begriffe Daten, Information und Wissen wird in der Fachliteratur oftmals anhand des deskriptiven Modells der sog. Wissenspyramide (siehe z. B. [Eng08]) oder anhand verwandter Modelle wie der sog. DIKW (Data-Information-Knowledge-Wisdom) Hierarchie (siehe z. B. [Row07]) vorgenommen. Diese Modelle grenzen die Begriffe Daten, Information und Wissen auf Basis unterschiedlicher, hierarchisch angeordneter Ebenen voneinander ab. Bei der Interpretation der Modelle sollte jedoch beachtet werden, dass es fraglich ist, ob tatsächlich eine derart scharfe Abgrenzung der Realität entspricht (vgl. [Bod06]).

Abb. 2.3 illustriert die entsprechenden drei Hierarchieebenen. Auf der linken Seite sind zusätzlich die Verarbeitungsschritte vermerkt, welche gemäß Auffassung in der Literatur (vgl. [Bre06]) durchlaufen werden müssen, um aus Daten Entscheidungen ableiten zu können.



**Abbildung 2.3:** Zusammenhang zwischen Daten, Information und Wissen. Abbildung mit Ergänzungen angelehnt an [Bre06].

Daten stellen eine digitale Repräsentation von Information dar. Diese Repräsentation erfolgt mittels nach bestimmten bekannten syntaktischen Regeln kodierten Zeichen. Der Begriff Zeichen lässt sich dabei im allgemeinen Sinne auffassen als das kleinste bei einer Programmausführung zugreifbare Element. Zeichen können also z. B. Buchstaben und Zahlen eines Alphabets (bei Daten in Textform) oder Pixelwerte (bei Bilddaten) sein. Daten können aus physikalischen Signalen abgeleitet sein, sie können aber auch auf andere Art zustande kommen.

Daten werden zu Information, indem ihnen eine Bedeutung (Semantik) beigefügt wird (vgl. [Bod06]). Hierzu müssen die Daten in den relevanten Kontext gesetzt werden. In [Reh96] wird diese Notwendigkeit wie folgt auf den Punkt gebracht: „Daten sind das Gegebene zur Verarbeitung ohne Verwendungshinweise“. Die Bedeutung, welche den Daten beigefügt wird, hängt vom jeweiligen Kontext ab. Wenngleich nicht explizit abgebildet in der Wissenspyramide bzw. den Ebenen der DIKW Hierarchie, setzt die Formulierung des Kontexts in der Regel bereits das Vorhandensein und die Einbeziehung von entsprechendem Wissen voraus (vgl. [Dip05]).

Wissen entsteht gemäß [Bod06] nach verbreiteter Auffassung in der Literatur durch die zweckorientierte Verknüpfung von Information unter Berücksichtigung von Kenntnissen hinsichtlich ihres Zusammenhangs und von Möglichkeiten zur sinnvollen Vernetzung.

Die dargestellte Unterscheidung von Daten, Information und Wissen steht nicht im Widerspruch zu der in Definition 2.2 gemachten Festlegung, derzufolge Information (im Kontext der Informationsfusion) als übergreifender Begriff verwendet wird. Gemäß Definition 2.1 erfolgt Informationsfusion immer im Hinblick auf eine gegebene Aufgabenstellung oder, in anderen Worten, in Bezug auf einen interessierenden Sachverhalt (vgl. Abschnitt 2.1). In die Fusion eingehende Daten müssen dafür zwangsläufig entsprechend interpretiert, d. h. in den für das Fusionsproblem relevanten Kontext gesetzt werden. Im Sinne der zuvor vorgestellten Begriffsunterscheidung müssen sie für die Fusion also auf die Informationsebene überführt werden. Definition 2.2 ist ebenso auch zur zuvor eingeführten Auffassung des Begriffs Wissen kompatibel, in dem Sinne, dass Information mit den Bausteinen für Wissen korrespondiert.

## 2.3 Anforderungen an eine Fusionsmethodik

Aus den bisherigen Ausführungen lassen sich drei Basisanforderungen an eine sinnvolle Methodik zur Fusion heterogener Informationsquellen ableiten. Dabei handelt es sich um ihre Fähigkeiten zu Transformation, Fusion und Fokussierung. Diese lassen sich in Anlehnung an [Bey07] wie folgt beschreiben:

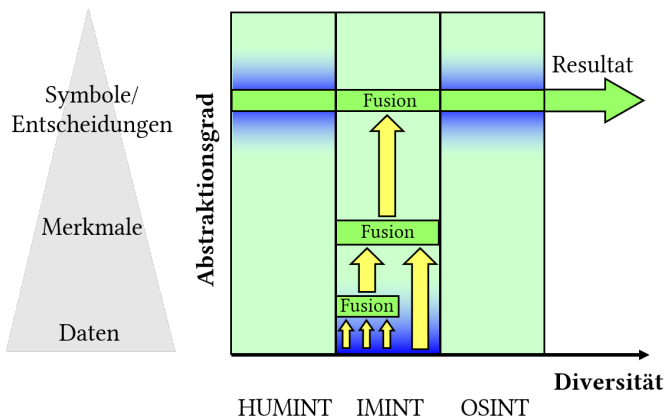
**Transformation:** Die Fähigkeit zur Transformation bedeutet, dass die Fusionsmethodik in der Lage ist, alle in die Informationsfusion eingehenden Informationsanteile möglichst unverfälscht in eine einheitliche formale Beschreibung zu überführen. Dazu müssen heterogene Informationsanteile in Abhängigkeit von ihrem ursprünglichen Abstraktionsgrad entweder verlustarm abstrahiert (bei Erhöhung des Abstraktionsgrads), fehlerfrei gewandelt (bei gleichbleibendem Abstraktionsgrad) oder artefaktarm spezialisiert (bei Absenken des Abstraktionsgrads) werden. Die Transformation in eine einheitliche formale Beschreibung ist nötig, um die zu fusionierenden Informationsanteile mathematisch kompatibel zu machen.

**Fusion:** Die Fähigkeit zur Fusion bedeutet, dass die Fusionsmethodik über Mechanismen verfügt, um die transformierten Informationsanteile auf eine mathematisch fundierte Weise zu einem möglichst hochwertigen Fusionsresultat zu kombinieren.

**Fokussierung:** Die Fähigkeit zur Fokussierung bedeutet, dass die aus der Fusion resultierende umfassende Informationsverkörperung der Gesamtaufgabe bei Bedarf zielgerichtet auf spezielle Einzelfragestellungen konzentriert werden kann. Dies ist wichtig, um sicherzustellen, dass auch ein umfangreiches Fusionsergebnis tatsächlich von Nutzen ist, indem es z. B. möglichst schnell erfassbar ist und den Menschen nicht aufgrund seiner aus dem hohen Umfang resultierenden Komplexität überfordert.

Wie in Abb. 2.4 für ein Beispiel aus der militärischen Aufklärung illustriert, kann es bei der Umsetzung der Fusionsmethodik für eine konkrete Aufgabenstellung vorteilhaft sein, schrittweise vorzugehen. Die Abbildung illus-

triert die Fusion von vier auf der Datenebene vorliegenden Beiträgen aus dem IMINT Bereich, z. B. Bildern, mit je einem auf der Ebene der Symbole/Entscheidungen vorliegenden Beitrag aus den Bereichen OSINT, z. B. Karteninformation, und HUMINT, z. B. die Meldung eines Aufklärungstrupps. Im dargestellten Beispiel wird zuerst ein Teil der Bilder zu einem höherwertigen Bild fusioniert. Das Ergebnis sowie der weitere Beitrag aus dem IMINT Bereich, werden dann auf die Ebene der Merkmale abstrahiert, auf der ein weiterer Fusionsschritt erfolgt. Das Ergebnis wird weiter abstrahiert und anschließend mit den Beiträgen aus dem HUMINT und OSINT Bereich fusioniert, welche hierzu ggf. vorab geeignet gewandelt werden müssen. Das resultierende Fusionsresultat könnten in diesem Beispiel Angaben zu Ort, Position und Typ von sich in einem bestimmten räumlichen Gebiet befindenden Objekten von Interesse sein.



**Abbildung 2.4:** Das Zusammenwirken von Transformation und Fusion am Beispiel von Beiträgen unterschiedlicher Intelligence-Diziplinen in der militärischen Aufklärung. Abbildung angelehnt an [Bey06b].

Eine weitere wichtige Eigenschaft, die eine Fusionsmethodik aufweisen sollte, ist ein ausreichendes Maß an Nachvollziehbarkeit. Gerade in sicherheitskriti-

schen Anwendungen oder in Fällen, in denen geltende Vorgaben und Gesetze wie z. B. die europäische Datenschutzgrundverordnung (DSGVO) fordern, dass maschinell getroffene Entscheidungen angemessen begründet werden müssen, ist diese Eigenschaft unabdingbar. Im Hinblick auf die Nachvollziehbarkeit maschineller Anwendungen lassen sich dabei, wie in [Döb18] dargestellt, die Begriffe Transparenz und Erklärbarkeit unterscheiden. Die Erklärbarkeit einer Anwendung bedeutet, dass die von ihr konkret gelieferten Ergebnisse nachvollziehbar sind – in dem Sinne, dass die wesentlichen Faktoren, die zu diesen Ergebnissen führten, dargelegt werden können. Transparenz bedeutet, dass das Verhalten der Anwendung und damit auch die zugrundeliegenden Modelle und Verfahren vollständig nachvollziehbar sind. Einen grundlegenden Einblick in das dieser Thematik insgesamt zugrundeliegende Forschungsgebiet, welches oft unter dem Begriff der erklärbaren künstlichen Intelligenz (im Englischen: Explainable Artificial Intelligence (Ex-AI, XAI)) gefasst wird, geben z. B. die Referenzen [Du19] und [Hol18].

Eine praxistaugliche Fusionsmethodik muss außerdem auch mit vertretbarem Aufwand anwendbar sein. Es sei hierzu auch angemerkt, dass in vielen praxisrelevanten Anwendungen der Informationsfusion per se nur sehr beschränkte Ressourcen zur Verfügung stehen. Beispielsweise müssen mobile Roboter wie UAVs (Unmanned Aerial Vehicles) oftmals mit der Speicher- und Rechenkapazität auskommen, die sie on board mitführen können. Auch muss z. B. im Bereich der militärischen Aufklärung benötigte Information in der Regel innerhalb eines bestimmten Zeitfensters bereitgestellt werden können, um ihre bestmögliche Nutzbarkeit, z. B. zur Ableitung darauf basierender Entscheidungen, überhaupt zu ermöglichen. In solchen Fällen kann ein vom mathematischen Standpunkt her optimales Fusionsresultat letztlich weniger nützlich sein als ein nicht optimales Fusionsresultat von noch akzeptabler Qualität, wenn die Berechnung des letzteren weniger Ressourcen bedarf.

Eine weitere wünschenswerte Eigenschaft an eine praktikable Fusionsmethodik, welche technische Systeme in die Lage versetzt, gerade auch heterogene Informationsbeiträge zu fusionieren, ist ihre modulare Erweiterbarkeit.

## 2.4 Möglichkeiten zur Einordnung

Im vorliegenden Abschnitt werden gängige Möglichkeiten zur systematischen Einordnung von Fusionsaufgaben, -methoden und -architekturen zusammengefasst und diskutiert.

### 2.4.1 Zusammenhang der Informationsbeiträge

In Abhängigkeit von der Beziehung der von den Informationsquellen gelieferten Beiträge zueinander lassen sich, unter Berücksichtigung des genauen Ziels der Informationsfusion, die folgenden vier elementaren Arten der Fusion unterscheiden (vgl. im Wesentlichen auch [Hei08] und [Rus07]):

**Konkurrierende Fusion:** Die konkurrierende Fusion bezeichnet die Fusion von gleichartigen Informationsbeiträgen mit gleicher Nutzinformation mit dem Ziel, bestimmte Unsicherheiten gezielt zu reduzieren. Ein Beispiel hierfür ist die Überlagerung von unter gleichen Aufnahmebedingungen gemachten Bildern zur Rauschunterdrückung durch die positionsweise Mittelung der Intensitätswerte über die entsprechende Bildserie. Vgl. hierzu z. B. [Rus07].

**Komplementäre Fusion:** Die komplementäre Fusion bezeichnet die Fusion von gleichartigen Informationsbeiträgen mit unterschiedlicher Nutzinformation mit dem Ziel, vollständigere Information zu erlangen. Ein Beispiel für eine komplementäre Fusion ist die Kombination von mittels gleicher Sensorik aufgenommenen Bildern, welche sich überschneidende (jedoch unterschiedliche) Ortsbereiche abdecken, zur Erzeugung eines Ergebnisbilds mit erweitertem örtlichen Abdeckungsbereich. Vgl. z. B. [Mül07] im Hinblick auf die Erzeugung eines Bildteppichs auf Basis von Videodaten.

**Kooperative Fusion:** Bei der kooperativen Fusion liegt die Nutzinformation verteilt über die Informationsbeiträge vor, so dass die gewünschte Information in Bezug auf den interessierenden Sachverhalt überhaupt

erst durch ihre Fusion erschlossen werden kann. Ein Beispiel für eine kooperative Fusion ist die Erzeugung einer Tiefenkarte zur 3D-Rekonstruktion einer Szene durch die Fusion von Stereoserien, d. h. von aus unterschiedlichen Positionen aufgenommenen Bildern. Vgl. hierzu z. B. [Ghe08b].

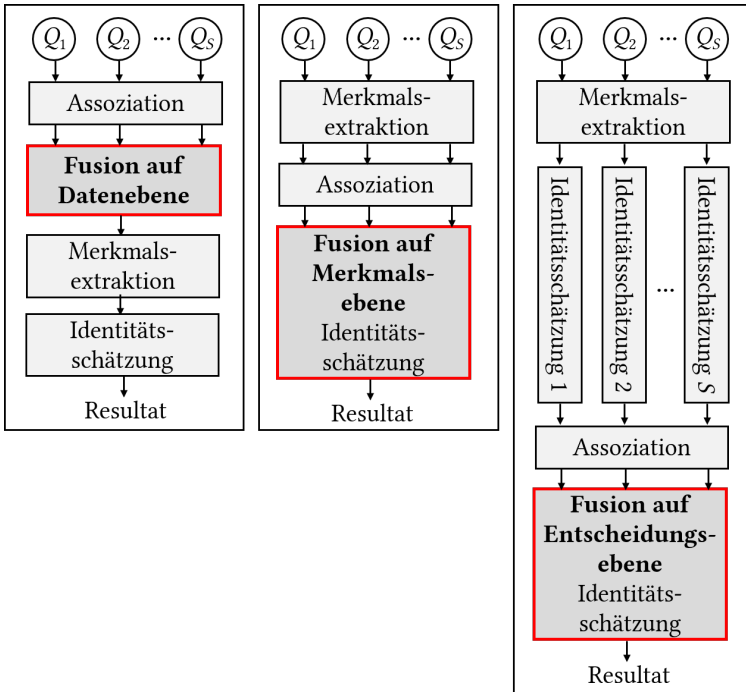
**Orthogonale Fusion:** Bei der orthogonalen Fusion liefern die Informationsquellen Beiträge in Bezug auf disjunkte Aspekte des interessierenden Sachverhalts. Dies ist z. B. der Fall, wenn heterogene Sensoren unterschiedliche physikalische Eigenschaften einer Szene erfassen oder wenn gleichartige Informationsbeiträge vor der eigentlichen Fusion mit unterschiedlichen Verarbeitungsmethoden ausgewertet werden. Vgl. hierzu z. B. [Hei08]. Ein anderes Beispiel für eine orthogonale Fusion ist die Fusion von mittels GMTI (Ground Moving Target Indicator) Sensorik gewonnener Bewegungsinformation eines Objekts und aus militärischen Auswertebereichten stammender Information in Bezug auf den Objekttyp. Vgl. hierzu [Kuw17].

## 2.4.2 Abstraktionsebenen

Die bereits in Abschnitt 2.1 skizzierte Unterscheidung, auf welcher Abstraktionsebene die eigentliche Fusion stattfindet, ist gerade bei datennahen Anwendungen oftmals hilfreich zur Einordnung von Fusionsaufgaben, insbesondere auch im Hinblick auf die Auswahl geeigneter Fusionsmethoden (vgl. hierzu z. B. [Hal97, Hei08, Men20]). In der Fachliteratur werden dabei oftmals die bereits zuvor erwähnten Ebenen der Daten, der Merkmale und der Symbole/Entscheidungen unterschieden. Entsprechende Fusionsaufgaben werden dann in englischsprachigen Veröffentlichungen in die Kategorien Data-Level Fusion, Feature-Level Fusion und Decision-Level Fusion eingeordnet. Im Kontext der Bildfusion wird statt dem Begriff Data-Level Fusion dabei auch der Begriff Pixel-Level Fusion verwendet.

Der Begriff Daten steht bei dieser Unterscheidung für die (maschinell verarbeitbare) digitale Repräsentation von Information, welche z. B. von Sensoren erfasst wird. Merkmale lassen sich auffassen als aus den ursprünglichen Daten

abgeleitete charakteristische Größen (Eigenschaften), die geeignet sein sollen, eine möglichst fehlerfreie Entscheidung hinsichtlich des Fusionsresultats zu ermöglichen. Ein Beispiel für auf der Ebene der Symbole/Entscheidungen angesiedelte Information wären die Ergebnisse einer durchgeführten Klassifikation zum Zwecke der Identifikation von Objekten.



**Abbildung 2.5:** Fusion der Informationsbeiträge von unterschiedlichen Sensoren  $Q_1, \dots, Q_S$  auf den verschiedenen Abstraktionsebenen. Abbildung angelehnt an [Hal97].

Abb. 2.5 illustriert anhand eines in [Hal97] gegebenen Beispiels mögliche Umsetzungen der Fusion auf den verschiedenen Abstraktionsebenen. Anwendungsfall ist dabei die Fusion von Information bezüglich Attributen eines in-



teressierenden Objekts, welche in Daten von unterschiedlichen Sensoren enthalten ist, mit dem Ziel, die Objekt-Identität bestmöglich zu ermitteln.

Die Frage, welche Abstraktionsebene die geeignetste zur Lösung einer bestimmten Aufgabe ist, lässt sich nicht einheitlich beantworten. Wie in Abschnitt 2.3 bereits thematisiert wurde, kann es auch zielführend sein, zur Lösung einer Gesamtausgabe mehrere Fusions Schritte auf unterschiedlichen Abstraktionsebenen zu kombinieren.

Werden Informationsbeiträge für die Durchführung der Fusion auf eine höhere Abstraktionsebene transformiert, so ist die Fusion oft rechnerisch weniger aufwändig. Allerdings können durch die Abstraktion Teile der Nutzinformation verloren gehen. Aus diesem Grund lässt sich durch die Fusion auf einer niedrigeren Abstraktionsebene ggf. ein hochwertigeres Fusionsresultat erzielen (vgl. [Hei08]), wobei in [Das97] darauf hingewiesen wird, dass sich diese Aussage nicht pauschalisieren lässt. Der Grund für Letzteres ist, dass der Umfang an letztlich tatsächlich nicht für die Fusion relevanter Information in der Regel hier am ausgeprägtesten ist und zudem eher die Gefahr besteht, dass mögliche Störungen (z. B. Rauschen) das Ergebnis verfälschen. Unter anderem deshalb kann bei der Durchführung einer Fusion auf einer niedrigeren Abstraktionsebene die Einbeziehung von Information aus höheren Abstraktionsebenen hilfreich oder sogar notwendig sein (vgl. [Bey08] und [Hei08]), z. B. zur Steuerung des Fusionsprozesses. Bei der Auswahl einer geeigneten Abstraktionsebene für die Fusion ist weiter zu berücksichtigen, dass zur Durchführung der Fusion auf einer niedrigeren Abstraktionsebene in der Regel spezifischere, an die Natur der Informationsquellen angepasste Methoden zur Verarbeitung der Informationsbeiträge nötig sind. Dies erschwert dann insbesondere auch die Erweiterung eines Gesamtsystems zur Fusion, wenn neue Informationsquellen eingebunden werden sollen. Je nach vorliegender Diversität der Informationsquellen kann zudem eine Fusion rein auf einer niedrigeren Abstraktionsebene auch gar nicht erst möglich sein. Beispielsweise ist es nicht möglich, Informationsbeiträge, welche durch einen akustischen Sensor und einen bildgebenden Sensor gewonnen werden, direkt auf der Datenebene zu fusionieren. Ein weiteres konkretes Beispiel hierfür stellt die Fusion von mittels eines Multi-Sensor-Systems gewonnenen Informationsbeiträ-

gen und Zusatzinformation (z. B. aus Produktblättern) bei dem in [Mül18a, Mül18b] adressierten System zur Drohnenerkennung und -abwehr dar. Wird die Fusion auf einer höheren Abstraktionsebene umgesetzt, so können zusätzliche Arten von Informationsquellen (z. B. weitere Arten von Sensoren) oder neue Verfahren für die Verarbeitung der Informationsbeiträge oftmals recht unkompliziert und in modularer Weise in ein Gesamtsystem integriert werden. Dabei ist jedoch sicherzustellen, dass die Merkmale bzw. symbolischen Informationen/Entscheidungen, welche aus den einzelnen Informationsbeiträgen abgeleitet werden, unabhängig voneinander sind bzw. dass vorhandene Abhängigkeiten explizit bei der Durchführung der Fusion berücksichtigt werden können.

### 2.4.3 Fusionsmodelle

In der Fachliteratur finden sich zahlreiche unterschiedliche Fusionsmodelle. Einen recht umfassenden Überblick bieten [Bed00], [Elm07], [Das08], [Foo13] und [Mor17]. Während manche Modelle speziell für den Anwendungskontext der Informationsfusion entwickelt wurden, handelt es sich bei vielen Modellen auch um Beschreibungen aus anderen Fachbereichen oder speziellen Anwendungsdomänen, welche dann im Kontext der Informationsfusion als Grundlage herangezogen und ggf. weiterentwickelt werden.

Die fortschreitende Digitalisierung und der technologische Fortschritt in Bezug auf Sensor-, Netzwerktechnologien etc. bewirken heute in vielen Anwendungsbereichen die zunehmende Vernetzung von Systemen, Menschen und Organisationen und ermöglichen dabei insbesondere auch den Zugriff auf weit verteilt vorliegende und umfangreiche Mengen an Information. Im Kontext der Informationsfusion relevante Ereignisse überschreiten dabei auch häufig räumliche und organisatorische, ggf. auch nationale Grenzen. Um diese Gegebenheiten und die daraus resultierenden Anforderungen bestmöglich zu unterstützen, ist auch im Bereich der Informationsfusion die Zusammenarbeit der zur Informationsgewinnung und -verarbeitung zur Verfügung stehenden Systeme und Akteure in system-, ebenen- und/oder organisationsübergreifenden Verbänden notwendig. Vgl. hierzu auch [Cat17], [Ess17] [Ess19]

und [Hal13]. Dies gilt insbesondere für Anwendungen in den Bereichen zivile Sicherheit und militärischen Aufklärung, aber auch darüber hinaus, z. B. im Kontext von Umweltinformationssystemen. Wenngleich sich die nachfolgend dargestellten Fusionsmodelle grundsätzlich auf den Anwendungskontext derartiger Verbände übertragen lassen, so ist anzumerken, dass sie größtenteils per se nicht hierfür entwickelt wurden. Es besteht deshalb gerade hier Potential, ihre Passgenauigkeit und Aussagekraft durch Präzisierung und/oder Weiterentwicklung zu verbessern. Vgl. hierzu auch [Bla17] und [Lig08].

### 2.4.3.1 JDL Modell

Das JDL Modell, das bis heute zu den populärsten Fusionsmodellen zählt, geht in seiner ursprünglichen Form zurück auf die ersten nennenswerten Ansätze zur Strukturierung des Forschungsgebiets Informationsfusion, welche durch das amerikanische Joint Directors of Laboratories (JDL) unternommen wurden. Es sei erwähnt, dass im Rahmen dieser Ansätze auch das in Abschnitt 2.2 erwähnte Datenfusionslexikon entstand. Seit der ursprünglichen Erstellung des JDL Modells im Jahre 1985 gab es vielfältige Ansätze zu seiner Weiterentwicklung und Präzisierung. Hervorzuhebende Beispiele für entsprechende Veröffentlichungen stellen [Bla02], [Bla17], [Ste99], [Ste01] und [Ste08] dar.

Das JDL Modell ist ein funktionales Modell. Wie in Abb. 2.6 dargestellt kategorisiert es auf einem abstrakten Niveau die Fusionsprozesse in verschiedene Ebenen (Levels), welchen sich dann wiederum unterschiedliche, zur Adressierung der in den Ebenen formulierten Zielsetzungen erforderliche Detailfunktionalitäten zuordnen lassen. Da das JDL Modell kein Prozessmodell darstellt, wird die Abfolge der Prozesse dabei nicht festgelegt. Dieser Sachverhalt ist in [Ste99] und [Ste01] anhand von Darstellungen möglicher Informationsflüsse zwischen den Ebenen sehr anschaulich dargestellt.

Dargestellt in Abb. 2.6 ist die in [Ste08] adressierte Version des JDL Modells. Auf eine Übersetzung der in Abb. 2.6 angeführten Begrifflichkeiten wurde zwecks präziser Wiedergabe der originalen Darstellung bewusst verzichtet. Das JDL Modell besteht gemäß Abb. 2.6 aus vier verschiedenen Ebenen (Level

1 bis Level 4) sowie einer zusätzlichen Ebene (Level 0), welche der Vorverarbeitung dient.

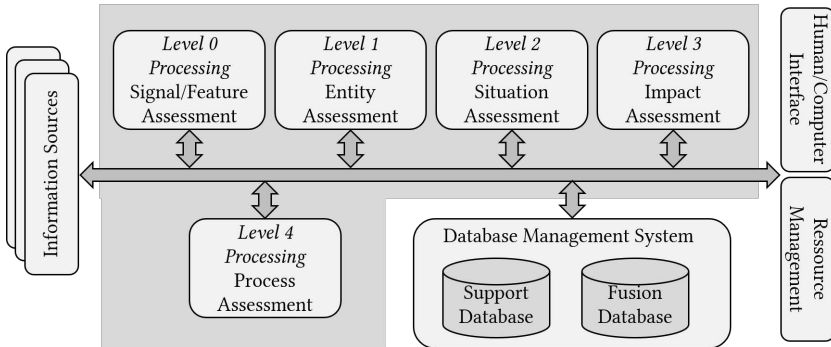


Abbildung 2.6: JDL Modell. Abbildung angelehnt an [Ste08].

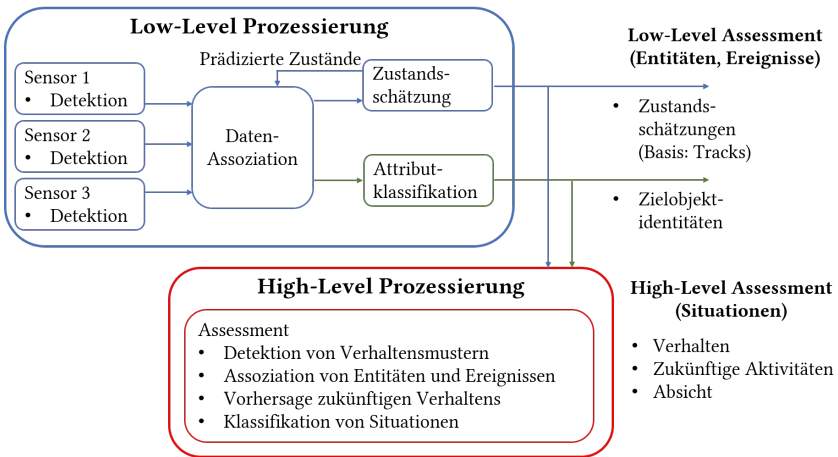
In Anlehnung an [Hal01] und [Ste08] lassen sich Level 1 bis Level 4 wie folgt charakterisieren: Level 1 adressiert die Schätzung der Zustände und Eigenschaften einzelner Entitäten auf Basis der hierzu vorliegenden Information. Anders als in früheren Darstellungen des JDL Modells (wie z. B. in [Ste01]) wird in [Ste08] bei der Benennung von Level 1 bewusst die generische Bezeichnung Entität (statt Objekt) verwendet und es wird darauf hingewiesen, dass die Festlegung der relevanten Typen von Entitäten beim Design eines Fusionssystems explizit adressiert werden muss. Zu den in Level 1 üblicherweise anfallenden Teilaufgaben gehören die Detektion, die Identifikation, die Lokalisierung und das Tracking einzelner physikalischer Objekte. Die in Abb. 2.5 adressierte Teilaufgabe der Objekt-Identifikation ist also in diese Ebene einzuordnen. Entitäten in Level 1 können darüber hinaus auch nicht physikalischer Natur sein und die Schätzung von Entitätszuständen und -eigenschaften kann auch anderweitig, z. B. auf Basis von indirekten Informationen, welche in höheren Ebenen des JDL Modells gewonnen werden, erfolgen. Level 2 adressiert die Ebene der Situationen. Hier geht es darum, Strukturen bestehend aus En-

titäten und Beziehungen zwischen unterschiedlichen Entitäten untereinander sowie deren Umgebung zu verstehen. Level 3 adressiert die Interpretation der gewonnenen Information über Entitäten, Situationen etc. aus Sicht der möglichen Auswirkungen. Ein Beispiel für eine Aufgabenstellung, die unter Level 3 einzuordnen ist, ist die Durchführung einer Bedrohungsanalyse. Diese Aufgabe wird in englischsprachigen Veröffentlichungen oft als Threat Assessment bezeichnet. Bei Level 4 handelt es sich um einen Metaprozess, welcher den resultierenden Gesamtfusionsprozess hinsichtlich Performanz und Effektivität bewertet. In der in [Ste08] vorgestellten Version des JDL Modells wird dabei, anders als bei früheren Versionen, die Funktionalität des eigentlichen Ressourcenmanagements, welches insgesamt gesehen auch die Gewinnung bzw. Sammlung der für die Fusion zur Verfügung stehenden Information umfasst, bewusst aus Level 4 ausgeklammert und als eigenständige Funktionalität notiert.

Eine ursprünglich auf [Wal90] zurückgehende, heute recht gängige Unterscheidung ist die zwischen Low Level Fusion und High Level Fusion. Die Low Level Fusion wird üblicherweise mit Level 0 und Level 1 des JDL Modells assoziiert, während die High Level Fusion üblicherweise in Bezug zu Level 2 und (teilweise) Level 3 gesetzt wird (vgl. z. B. [Bla12], [Das08] und [Tot08]). Die an die Darstellungen in [Wal90] und [Bla12] angelehnte Abb. 2.7 illustriert diese Unterscheidung. Anzumerken ist dabei, dass sich die Darstellung auf den Fall, dass Level 1 des JDL Modells ausschließlich die Zustände und Eigenschaften physikalischer Objekte adressiert, fokussiert und möglichen Rückflüssen von höheren Ebenen des JDL Modells zu niederen Ebenen keine Rechnung trägt.

In der Fachwelt wird überwiegend die Auffassung vertreten, dass gerade im Bereich der High Level Fusion noch deutliche wissenschaftliche Herausforderungen bestehen. Siehe hierzu z. B. [Bla12], [Bla19] und [Tot08]. Zu den Gründen hierfür zählt nach Ansicht vieler Autoren, dass die Low Level Fusion oftmals (zumindest vorwiegend) physikalische Eigenschaften von Objekten der realen Welt adressiert, welche sich überwiegend mittels numerischen Größen repräsentieren lassen, während die High Level Fusion vor allem mit abstrakter Information symbolischer Natur umgehen muss. Weiter erfordert die High Level Fusion in der Regel die Einbeziehung umfangreicheren Zusatzwissens

(z. B. Expertenwissen, Kontextwissen), welches adäquat repräsentiert und eingebunden werden muss. Diese Thematik wird z. B. in den Veröffentlichungen [Hof19], [Kuw13], [Kuw18] und [Kuw19] adressiert. Hervorzuheben ist weiter, dass die High Level Fusion auch ein höheres Maß an Mensch-Maschine-Interaktion erfordert, in dem Sinne, dass der Mensch gerade hier in geeigneter Weise „in-the-Loop“ bzw. „on-the-Loop“ (vgl. z. B. [Not16]) gehalten werden muss. Hieraus folgt insbesondere dann auch, dass der Anforderung nach Nachvollziehbarkeit (vgl. Abschnitt 2.3) der maschinell umgesetzten Funktionalitäten zusätzliche Bedeutung zukommt. Aufgrund der zuvor angeführten Aspekte bestehen auch besondere Herausforderungen hinsichtlich der Evaluation und Bewertung von Verfahren und Systemen zur High Level Fusion.

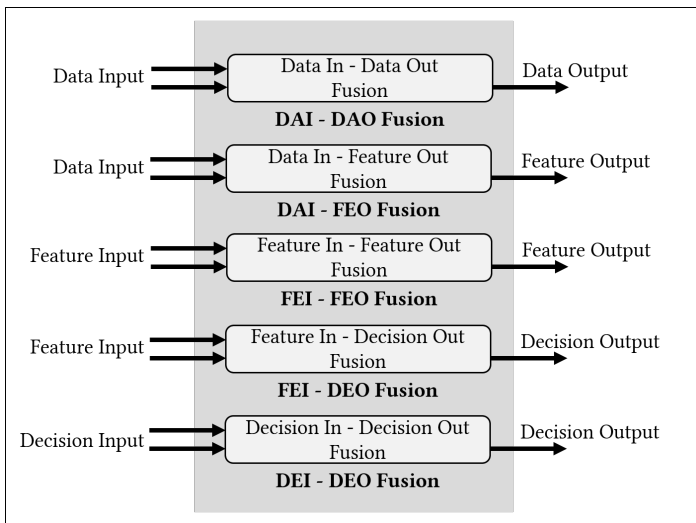


**Abbildung 2.7:** Low Level Fusion versus High Level Fusion gemäß [Wal90]. Abbildung angelehnt an die Darstellungen in [Bla12] und [Wal90].

### 2.4.3.2 I/O Modell nach Dasarathy

Das in [Das97] eingeführte funktionale I/O (Input/Output) Fusionsmodell ordnet einzelne Fusionsprozesse technisch auf Basis eines zweiwertigen Schemas

ein. Die Einordnung erfolgt zum einem in Bezug auf die Abstraktionsebene, auf der die Informationsbeiträge in die eigentliche Fusion eingehen, und zum anderen in Bezug auf die Abstraktionsebene, auf welcher das Fusionsergebnis zu liegen kommt. Das I/O Modell grenzt dabei Daten (im Englischen: Data), Merkmale (im Englischen: Features) und Entscheidungen (im Englischen: Decisions) voneinander ab. Explizit eingeführt werden die in Abb. 2.8 dargestellten Modi sowie zusätzlich noch ein Fusionsprozess über die Zeit, der auf jeder der Abstraktionsebenen zum Einsatz kommen kann.



**Abbildung 2.8:** Fusions-Modi im I/O Modell. Abbildung angelehnt an [Das97].

Die unterschiedlichen Modi dienen als Bausteine für das Design von Fusionsarchitekturen (vgl. [Das97]). Auf Basis der Kategorisierung der anfallenden Teilaufgaben gemäß dem I/O Modell können dabei geeignete Fusionsverfahren und -techniken zu deren Adressierung ausgewählt werden. Die Erweiterung der Kategorisierung auf zusätzliche Modi (z. B. Feature In - Data Out

Fusion, (Decision In + Feature In) - Decision Out Fusion) ist dabei möglich (vgl. [Das97] und [Ste01]).

Wie in [Ste01] und [Ste08] dargestellt wird, kann das I/O Fusionsmodell in einer erweiterten Version zur Verfeinerung des JDL Modells herangezogen werden. Die Erweiterung ist dabei nötig, um auch die Anforderungen der JDL Level 2 bis 4 vollständig abzudecken.

#### 2.4.3.3 Situationsbewusstsein nach Endsley

Während die bisher vorgestellten Modelle das Thema Informationsfusion aus der technischen Perspektive betrachten, kann das Modell von Endsley [End95] verwendet werden, um der kognitiven Ebene verstärkt Rechnung zu tragen. Bei diesem Modell handelt es sich um das erste und bis heute prominenteste Modell zur Beschreibung des Prozesses, anhand dessen menschliches Situationsbewusstsein entsteht.

Endsley definiert den Begriff Situationsbewusstsein dabei wie folgt:

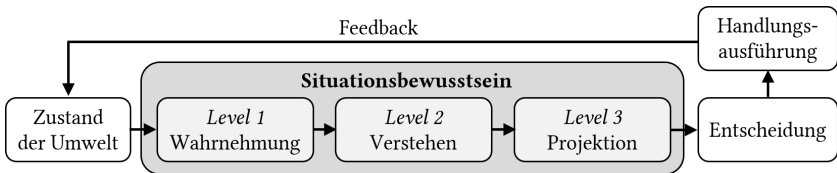
„Situation Awareness is the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future.“

Dieser Definition folgend werden im Modell wie in Abb. 2.9 dargestellt drei aufeinander aufbauende mentale Ebenen (Level) unterschieden. Diese lassen sich wie folgt beschreiben:

- Level 1: Wahrnehmung der relevanten Informationselemente im beobachteten Ausschnitt der Umwelt.
- Level 2: Verstehen der vorliegenden Situation auf Basis der Elemente aus Level 1 und unter Berücksichtigung des relevanten Kontexts.
- Level 3: Projektion zukünftiger Zustände, d. h. Vorhersage der zukünftigen Entwicklung auf Basis der Dynamik der Informationselemente und des Verständnisses der vorliegenden Situation.



Das so erlangte, möglichst hohe Maß an Situationsbewusstsein stellt dann die Basis für das Treffen von Entscheidungen und die nachgelagerte Ausführung von Handlungen dar. Wie durch die Feedback-Schleife in Abb. 2.9 gekennzeichnet, haben die ausgeführten Handlungen dann in der Regel Auswirkungen auf den Zustand des beobachteten Ausschnitts der Umwelt.



**Abbildung 2.9:** Situationsbewusstsein gemäß [End95]; vereinfachte Darstellung.

Bemerkenswert ist, dass sich, wie in [Das08] dargestellt wird, Level 1 bis Level 3 aus dem Modell von Endsley recht direkt den entsprechenden gleichnamigen Ebenen (Level) des JDL Modells zuordnen lassen.

#### 2.4.3.4 OODA Loop

Der OODA (Observe-Orient-Decide-Act) Loop geht zurück auf den amerikanischen Kampfpiloten, Ausbilder und Militärtheoretiker John Richard Boyd (1927–1997). Es handelt sich um ein Modell, welches den Prozess der Entscheidungsfindung anhand unterschiedlicher Phasen beschreibt.

Zu Beginn steht demnach die Erfassung der aktuellen Lage (Observe-Phase). Die dabei gewonnene Information wird dann in Bezug zu weiterer Information und zu übergreifendem Wissen gesetzt, analysiert und eingeordnet (Orient-Phase), um so ein möglichst hohes Maß an Situationsbewusstsein zu erlangen. Auf dieser Basis erfolgt dann, auch unter Berücksichtigung der möglichen Konsequenzen, das eigentliche Entscheiden (Decide-Phase) und die getroffene Entscheidung wird durch entsprechendes praktisches Handeln

umgesetzt (Act-Phase). Das Handeln hat wiederum Auswirkungen auf die Lage, wodurch der OODA Loop erneut angestoßen wird.

Ursprünglich entwickelt zur Beschreibung des Entscheidungsfindungsprozesses von Kampf-Piloten auf der taktischen Ebene, etablierte sich der OODA Loop darüber hinaus als generelles Metamodell zur Beschreibung des Entscheidungsfindungsprozesses von Personen, Organisationen etc. in ihrer Umgebung. Dabei wird in der Regel Bezug zu der in Abb. 2.10 illustrierten, sehr abstrakten Darstellung des OODA Loops genommen. Die ursprünglich von Boyd 1996 entworfene Darstellung, welche z. B. in [Rév17] wiedergegeben ist, ist komplexer und sieht insbesondere auch Interaktionen zwischen den einzelnen Phasen, die nicht durch den in Abb. 2.10 illustrierten Zyklus an sich abgebildet sind, vor. Beispielsweise kontrolliert und leitet demnach die Orient-Phase implizit die Observe-Phase.

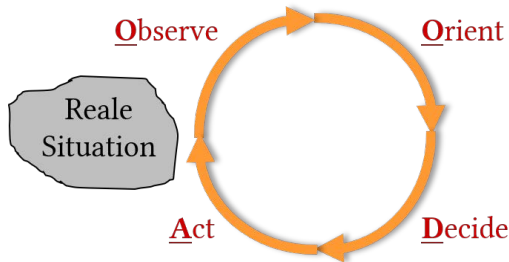


Abbildung 2.10: Die vier Phasen des OODA Loop.

Der OODA Loop hat sich insbesondere auch als grundlegendes Geschäftsprozessmodell im Bereich der militärischen Führung, im Englischen mit C2 (Command and Control) bezeichnet, etabliert (vgl. z. B. [Rév17]). Wie in [Bos07a] angemerkt, wird der C2 Prozess dadurch jedoch stark – möglicherweise zu stark – vereinfacht, insbesondere im Hinblick darauf, dass die Entscheidungsfindung in heutigen militärischen Operationen nicht jeweils in Form einer einzigen und über alle Hierarchiestufen koordinierten Entscheidung erfolgt.

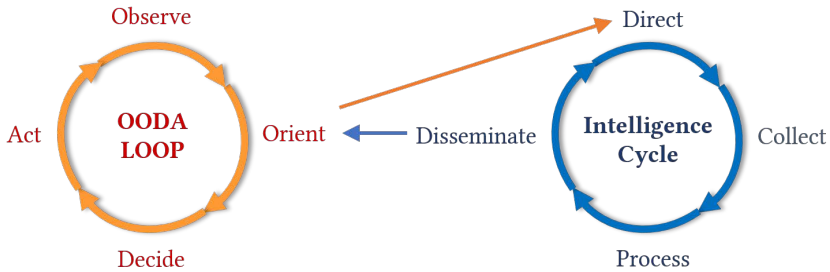
Gemäß [Bry04] lässt sich diese Problematik umgehen, indem man nicht einen, sondern multiple, auf unterschiedlichen Ebenen ablaufende OODA Loops vorsieht. Dies muss dann so ausgestaltet sein, dass die Entscheidungsfindung auf einer bestimmten Ebene die auf niederen Ebenen mitberücksichtigt.

Über diesen spezifischen Punkt hinaus ist jedoch generell anzumerken, dass der OODA Loop nicht selten als zu stark vereinfachend und auch als nicht korrektes Modell für den kognitiven Prozess menschlicher Entscheidungsfindung kritisiert wird. Ein wichtiger Kritikpunkt ist z. B. auch, dass der OODA Loop (zumindest in der abstrakten Version) zu implizieren scheint, dass die Entscheidungsfindung rein reaktiv auf in der Umwelt eingetretene Ereignisse erfolgt – während Entscheidungsträger jedoch tatsächlich auch proaktiv Informationsbedarfe formulieren und diesen folgend Informationen sammeln, um so Pläne und Handlungen abzuleiten (vgl. [Bry04]). Der hohe Abstraktionsgrad des OODA Loops führt zudem gemäß [Bry04] auch dazu, dass er zwar grob beschreibt, wie die Entscheidungsfindung ablaufen kann, dabei aber auch zentrale Aspekte hinsichtlich der Umsetzung nicht adressiert. Ausgeklammert wird z. B. der Aspekt, wie Entscheidungsträger ihre Informationsbedarfe identifizieren und explizit machen. Offen bleibt weiter auch, wie die (in der Observe-Phase) gewonnene Information z. B. wie in [Roy01] dargestellt auf Basis eines mentalen Modells zu Erlangung von Situationsbewusstsein beiträgt. Derartige Kritikpunkte sollten bei der Anwendung dieses Modells berücksichtigt und wo nötig z. B. durch entsprechende Präzisierungen adressiert werden. Eine Präzisierung könnte z. B. auch erfolgen, indem ein Bezug zu anderen Prozessmodellen gesetzt wird.

#### **2.4.3.5 Intelligence Cycle**

Der Intelligence Cycle beschreibt, ebenfalls anhand unterschiedlicher, als Zyklus angeordneter Phasen, die Prozessschritte, die bei der Planung, Gewinnung, Weiterverarbeitung und Verfügbarmachung von entscheidungsrele-

vanter Information<sup>1</sup> anfallen. Die Anzahl und genaue Benennung der Phasen ist nicht einheitlich in der Literatur. Die nachfolgende Beschreibung anhand von vier Phasen, welche die englischsprachigen Bezeichnungen Direction, Collection, Processing und Dissemination tragen, orientiert sich an [NAT19].



**Abbildung 2.11:** Mögliches Zusammenspiel von OODA Loop und Intelligence Cycle.

Zu Beginn eines Durchlaufs des Intelligence Cycle (Direction-Phase) steht die klare Formulierung und ggf. Priorisierung derjenigen Fragestellungen, welche durch die in den nachfolgenden Phasen umgesetzte Informationsgewinnung und -weiterverarbeitung beantwortet werden sollen. Wie in Abb. 2.11 dargestellt könnten entsprechende Fragestellungen aus der Orient-Phase des OODA Loop resultieren, vgl. hierzu auch [Bie06]. Die möglichen Informationsquellen werden identifiziert und die Informationsgewinnung wird geplant. Im nächsten Schritt (Collection-Phase) werden die benötigten Informationen dann gewonnen. Gesammelte Daten werden dabei durch Auswertung/Interpretation in Information überführt. Die gewonnenen Informationsbeiträge

<sup>1</sup> Die genaue Definition des Begriffs Intelligence im NATO Umfeld ist gemäß [NAT19] wie folgt: „The product resulting from the directed collection and processing of information regarding the environment and the capabilities and intentions of actors, in order to identify threats and offer opportunities for exploitation by decision-makers.“ Frei übersetzt handelt es sich bei Intelligence also um im Hinblick auf anstehende Entscheidungen zielgerichtet gewonnene und weiterverarbeitete Information sowie daraus abgeleitetes Wissen.

werden dann in höherwertigen Auswerte- und Analyseprozessen weiterverarbeitet, zueinander in Beziehung gesetzt, analysiert und auch fusioniert, um so die letztlich benötigte entscheidungsrelevante Information zu gewinnen (Processing-Phase). Diese muss anschließend den Personen, Organisationen etc., die sie benötigen, geeignet verfügbar gemacht werden (Dissemination-Phase). Im Erfolgsfall können damit dann die ursprünglichen Fragestellungen als beantwortet betrachtet werden. Es kann aber die Notwendigkeit entstehen, diese in veränderter Form erneut zu stellen oder auch komplett neue Informationsbedarfe zu formulieren, wodurch wiederum ein Durchlauf des Intelligence Cycle angestoßen wird.

Das Konzept des Intelligence Cycle findet vor allem im Bereich der militärischen Aufklärung sowie im nachrichtendienstlichen Kontext Anwendung. Es wird als Metamodell aber auch in anderen Anwendungsbereichen wie z. B. der Marketing Intelligence [The18], d. h. der Sammlung und Analyse von Daten, Informationen und Wissen im Hinblick auf wichtige Schlüsselfaktoren für Marketingunterscheidungen von Unternehmen, eingesetzt.

Für die streitkräftegemeinsame Nachrichtengewinnung und Aufklärung im multinationalen Kontext wird der Intelligence Cycle, auch unter Bezugnahme zu mit ihm interagierenden Prozessen und Funktionalitäten, in [NAT16] beschrieben und detailliert. Die Sammlung der benötigten Daten und Informationsanteile erfolgt gemäß der dort vorgenommenen Präzisierung innerhalb des sog. Joint ISR (Intelligence, Surveillance, and Reconnaissance) Prozesses, eines weiteren zyklischen, aus unterschiedlichen Phasen bestehenden Prozesses. In diesen werden die im Intelligence Cycle formulierten Fragestellungen in geeigneter Weise eingespeist. Sie werden dafür so verfeinert, so dass sie mittels einer bestimmten, passend zu wählenden Intelligence Disziplin und innerhalb dieser mittels einer bestimmten Quelle beantwortet werden können. Jede der entsprechend verfeinerten Fragestellungen initialisiert dann eine Instanz des Joint ISR Prozesses. Hinsichtlich einer detaillierten Beschreibung des Joint ISR Prozesses und der Zusammenhänge zum Intelligence Cycle sei auf [Ess18a] verwiesen. Im Hinblick auf die Zuordnung von Aufgaben und Funktionalitäten der Informationsfusion zu diesen beiden operationellen Prozessen ist zu berücksichtigen, dass der Intelligence Cycle (im Sinne der Begriffsunterschei-

derung gemäß Abschnitt 2.2) die Informations- und die Wissensebene adressiert. Die Gewinnung der relevanten Daten und Informationsanteile sowie die Interpretation gewonnener Daten durch Auswertung erfolgt im Joint ISR Prozess, der dementsprechend die Daten- und die Informationsebene adressiert.

Diese Beobachtung begründet auch die in [Bie06] getätigte Aussage, wonach Fusionsaufgaben, welche Level 2 und Level 3 des JDL Modells zuzuordnen sind, in den Bereich des Intelligence Cycle fallen. Fusionsaufgaben, welche Level 1 des JDL Modells zuzuordnen sind, könnten je nach ihrer konkreten Ausprägung entweder in den Bereich des Joint ISR Prozesses oder des Intelligence Cycle fallen. Grundsätzlich könnte dabei auch die prozessübergreifende Berücksichtigung möglicher Rückflüsse von höheren Ebenen des JDL Modells zu niederen Ebenen gewinnbringend sein. Beispielsweise könnte im Intelligence Cycle gewonnene Information bzw. daraus abgeleitetes Wissen zur Generierung von Hypothesen verwendet werden, welche dann im Joint ISR Prozess zur zielgerichteten Auswertung von Sensordaten beitragen können. Zu berücksichtigen ist dabei dann jedoch, dass dieser Bezug bei der Einspeisung der im Joint ISR Prozess gewonnenen Resultate in den Intelligence Cycle in geeigneter Weise explizit gemacht bzw. erhalten werden muss. Dies ist nötig um zu vermeiden, dass bei einer Fusion im Intelligence Cycle die den Hypothesen zugrundeliegende Information möglicherweise übergewichtet wird, da fälschlicherweise angenommen wird, dass sie von zwei voneinander unabhängigen Quellen stammt.

### 2.4.4 Topologie und Aufgabenverteilung

Hinsichtlich zugrundeliegender Topologie und vorgenommener Aufgabenverteilung lassen sich zentralisierte, verteilte und dezentrale Fusionsarchitekturen unterscheiden. Eine wichtige Mischform dieser Architekturtypen stellen hierarchische Fusionsarchitekturen dar.

Einen Überblick über diese Architekturtypen liefern die Referenzen [Cas13, Hal97, Man94, Kle04, Kha09, Mit07], welche auch als Grundlage für die nachfolgende Aufarbeitung des Themas dienen. Zusätzlich herangezogen wurde

dabei noch die Referenz [Lig08], die sich dadurch auszeichnet, dass die Autoren das Thema speziell im Hinblick auf die Informationsfusion in Systemverbänden beleuchten.

In der Fachliteratur finden sich über die eingangs genannten vier Architekturtypen hinaus noch weitere Unterscheidungen, vgl. hierzu z. B. [Kha09]. Diese sind für den im vorliegenden Abschnitt intendierten Überblick jedoch nicht von zentraler Bedeutung. Weiter ist anzumerken, dass die in der Literatur verwendeten Begrifflichkeiten zur Bezeichnung unterschiedlicher Architekturtypen nicht einheitlich sind.

Wie die nachfolgende Darstellung genauer verdeutlichen wird, hängt die Frage nach dem jeweils geeignetsten Architekturtyp von vielfältigen Faktoren ab. Hierzu zählen die verfügbaren Ressourcen (Rechenleistung und Speicherkapazität in den Komponenten, verfügbare Bandbreiten in einem Netzwerk aus unterschiedlichen Komponenten) ebenso wie die bei den einzelnen Komponenten verfügbaren bzw. realisierbaren Fähigkeiten. Weitere Faktoren können einerseits die Qualität der von den Informationsquellen gelieferten Beiträge und andererseits die konkreten Anforderungen an die mittels der Fusion zu erzielenden Resultate sein.

Je nach Anwendungsdomäne und -bereich können weiter auch operationelle Prozesse, wie sie z. B. in Abschnitt 2.4.3.5 für den Fall der streitkräftegemeinsame Nachrichtengewinnung und Aufklärung im multinationalen Kontext beschrieben wurden, bestimmte Randbedingungen vorgeben. Gerade in organisationsübergreifenden Anwendungskontexten der Informationsfusion muss ggf. auch berücksichtigt werden, welche Information in welcher Form (z. B. in Rohform oder ausgewertet) überhaupt weitergegeben werden darf.

#### **2.4.4.1 Zentralisierte Fusionsarchitekturen**

Beim Vorliegen einer zentralisierten Fusionsarchitektur liefern die Informationsquellen die von ihnen erfassten Beiträge nahezu unverarbeitet an eine zentrale Fusionseinheit, die die eigentliche Fusion samt den ggf. notwendigen Vorverarbeitungsschritten, z. B. im Hinblick auf die Transformation der

Informationsbeiträge in eine einheitliche formale Beschreibung und/oder den Übergang auf eine andere Abstraktionsebene, durchführt.

Da die Informationsbeiträge hier praktisch in Rohform an die Fusionseinheit geliefert werden, kann sichergestellt werden, dass vor Eingang bei der Fusionseinheit zumindest durch Vorverarbeitung keine Nutzinformation verloren ging. Hierdurch ist es theoretisch gesehen einfacher, optimale Fusionsergebnisse zu erzielen (vgl. [Kle04], [Man94] und [Mit07]). Insbesondere können auch fehlerhafte Informationsbeiträge ggf. einfacher in der Zusammenschau identifiziert werden (vgl. [Man94]). In [Cas13] wird allerdings darauf hingewiesen, dass sich die Aussage hinsichtlich der Optimalität zentralisierter Fusionsarchitekturen bei ihrer praktischen Umsetzung oftmals nicht bewahrt. Als ein Grund hierfür ist sicher anzuführen, dass die letztlich erzielte Güte noch von weiteren Faktoren, z. B. im Hinblick auf die Übertragung der Informationsbeiträge, welche unverfälscht und zeitgerecht erfolgen sollte, abhängt.

Bei einer zentralisierten Fusionsarchitektur müssen in den der Fusionseinheit vorgelagerten Komponenten keine Funktionalitäten hinsichtlich der Vorverarbeitung der Informationsbeiträge vorgesehen werden. Dies könnte z. B. die Einbindung kostengünstiger Sensoren ohne eigene Verarbeitungsfunktionalitäten erleichtern. Die Fusionseinheit kann auf den praktisch in Rohform eingehenden Informationsbeiträgen im Prinzip unterschiedliche konkrete Ausprägungen der Fusion (z. B. zur Beantwortung unterschiedlicher Fragestellungen oder zur Präzisierung gewonnener Erkenntnisse) durchführen, ohne dass die Notwendigkeit besteht, zusätzliche Information bei den Informationsquellen anzufordern, was z. B. ein Vorteil sein kann, wenn diese die erfassten Informationen nicht speichern oder wenn sie im Netzwerk nicht durchgehend verfügbar sind. Während der Verarbeitung gewonnene Zwischenergebnisse können bei entsprechender Ausgestaltung der Speichermodalitäten dafür in der Fusionseinheit vorgehalten und ggf. weiterverwendet werden. Zentralisierte Fusionsarchitekturen können auch generell fehlertoleranter sein, z. B. wird in [Kle04] darauf hingewiesen, dass für die Aufgabenstellung des Trackings von Objekten insbesondere im Fall zentralisierter Fusionsarchitekturen Al-



gorithmen, welche das Fehlen einzelner Sensorergebnisse tolerieren, gezielt umsetzbar sind.

Allerdings sind bei einer zentralisierten Fusionsarchitektur, bedingt durch die Übermittlung von Information in nahezu Rohform, die erforderlichen Bandbreiten und die bei der Fusionseinheit anfallenden Verarbeitungskosten im Allgemeinen relativ hoch. Dadurch kann es eher zu zeitlichen Verzögerungen bei der Übermittlung und/oder Verarbeitung der Informationsbeiträge kommen. Liefern die Informationsquellen z. B. umfangreiche Mengen an erfassten Bilddaten, so könnte auch gut der Fall auftreten, dass die im Netzwerk verfügbaren Bandbreiten grundsätzlich nicht ausreichen, um die Übermittlung der entsprechend hohen Datenmengen zu realisieren. Sind in einem Systemverbund abgesetzte Einheiten (z. B. Systeme auf Schiffen) vorhanden, welche nur über sehr begrenzte Bandbreiten angebunden sind und dazu ggf. auch nur zu bestimmten Zeiten Information übermitteln können, so ist zumindest in Bezug auf die entsprechenden Anteile des Verbunds der Ansatz einer zentralisierten Fusionsarchitektur ebenfalls zumindest als schwierig, wenn nicht gar als unrealistisch, zu betrachten.

Werden beim Vorliegen einer zentralisierten Fusionsarchitektur zusätzliche Informationsquellen verfügbar und/oder sind neue Arten von Informationsbeiträgen zu verarbeiten, so muss die Fusionseinheit in der Lage sein, durch Vorhaltung entsprechender Kapazitäten und geeigneter, ggf. neu zu ergänzender, spezifischer Algorithmen auch diese Information adäquat zu handhaben. Die hierfür ggf. erforderliche Modifikation bzw. Erweiterung einer zentralisierten Fusionsarchitektur ist im Vergleich zu alternativen Fusionsarchitekturen in der Regel als eher schwierig anzusehen.

In [Kle04] wird auch auf die im Verhältnis recht hohe Verwundbarkeit zentralisierter Fusionsarchitekturen hingewiesen – in dem Sinne, dass durch einen Ausfall der zentralen Fusionseinheit oder der auf diese ausgerichteten Kommunikationsstruktur das resultierende Gesamtsystem ebenfalls komplett ausfällt, also nicht einmal mehr Teilergebnisse verfügbar sind. In [Man94] wird hinsichtlich dieser Problematik auf die Möglichkeit des Vorsehens einer Back-Up-Komponente für die Fusionseinheit hingewiesen, jedoch verbunden mit

der Kritik, dass hierfür zusätzliche Ressourcen benötigt werden, welche letztlich auch besser genutzt werden könnten.

#### **2.4.4.2 Verteilte Fusionsarchitekturen**

Bei einer verteilten Fusionsarchitektur werden die Informationsbeiträge im Hinblick auf die nachgelagerte, in einer Fusionseinheit durchgeführte Fusion vorverarbeitet, bevor sie dorthin übertragen werden. Beispielsweise könnten aus Bilddaten bereits Positionsschätzungen für sich in den Bildern befindende Objekte gewonnen werden.

Bei einer verteilten Fusionsarchitektur können die zur Übermittlung der Informationsbeiträge erforderlichen Bandbreiten und die bei der Fusionseinheit erforderlichen Ressourcen zu deren Weiterverarbeitung in der Regel geringer gehalten werden als bei einer zentralisierten Fusionsarchitektur. Das Vorhalten dedizierter, jeweils bestmöglich auf die einzelnen Informationsquellen angepasster Verarbeitungsmethoden in den vorgelagerten Komponenten kann dazu beitragen, dass die in den einzelnen Informationsbeiträgen vorliegende Nutzinformation in optimaler Weise erschlossen wird. Ist die Vorverarbeitung bereits so ausgestaltet, dass sie die Informationsbeiträge in eine einheitliche formale Beschreibung überführt, so kann auf das Vorhalten spezifischer Verarbeitungsmethoden für bestimmte Arten von Information in der Fusionseinheit verzichtet werden. Als Konsequenz können bei einer verteilten Fusionsarchitektur Informationsquellen i. A. leichter hinzugefügt oder ausgetauscht werden, als es bei einer zentralisierten Fusionsarchitektur der Fall ist.

Allerdings müssen bei einer verteilten Fusionsarchitektur die entsprechend benötigten Verarbeitungsmethoden bei den der Fusionseinheit vorgelagerten Komponenten auch verfügbar sein und die bei diesen dadurch anfallenden Verarbeitungskosten müssen bei der konkreten System-Ausgestaltung berücksichtigt werden und auch tragbar sein.

Gegebenenfalls kann das Vorsehen eines bidirektionalen Informationsaustausches sinnvoll sein, um so auch Informationsrückflüsse von der Fusionseinheit zu den vorgelagerten Komponenten zu ermöglichen (vgl. [Man94]). Beispiels-

weise könnte es für die Lokalisierung von Objekten zu einem bestimmten Zeitpunkt in den vorgelagerten Komponenten hilfreich sein, wenn die Fusionseinheit diesen durch die Fusion verbesserte Schätzungen der Objektpositionen in Bezug auf frühere Zeitpunkte verfügbar macht.

Die Umsetzung einer verteilten Fusionsarchitektur ist nicht immer möglich, z. B. dann nicht, wenn die in den Informationsbeiträgen vorhandene Nutzinformation nicht ausreichend verlustfrei individuell verarbeitet werden kann oder wenn Abhängigkeiten zwischen den Informationsbeiträgen bestehen, welchen im Rahmen der individuellen Verarbeitung nicht ausreichend Rechnung getragen werden kann.

Die Frage, wie man Abhängigkeiten zwischen den Informationsbeiträgen in geeigneter Weise Rechnung tragen kann, stellt generell oftmals eine Herausforderung im Hinblick auf die Ausgestaltung verteilter Fusionsarchitekturen dar. Eine verteilte Fusionsarchitektur kann dagegen dann prädestiniert sein, wenn die Informationsquellen unterschiedliche z. B. physikalische Prinzipien zur Informationserfassung einsetzen und die ihnen überlagerten Störungen unabhängig sind, so dass man davon ausgehen kann, dass keine derartigen Abhängigkeiten bestehen, vgl. hierzu auch [Kle04].

#### **2.4.4.3 Dezentrale Fusionsarchitekturen**

Bei einer dezentralen Fusionsarchitektur gibt es keine zentrale Fusionseinheit, statt dessen erfolgt die Fusion in unterschiedlichen Fusionsknoten, an welche die Informationsquellen ihre Beiträge liefern. Ein spezieller Fall ist der, dass eine Eins-zu-eins-Entsprechung zwischen Informationsquellen und Fusionsknoten besteht, in dem Sinne, dass die Informationsquellen selbst auch die Fusionsknoten darstellen.

Im Vergleich zu Systemen, die auf einer zentralisierten oder einer verteilten Fusionsarchitektur beruhen, zeichnen sich Systeme, die auf einer dezentralen Fusionsarchitektur beruhen, oftmals durch höhere Skalierbarkeit und Robustheit sowie ein höheres Maß an Modularität aus (vgl. [Man94] und [Mit07]).

Während in zentralisierten und verteilten Fusionsarchitekturen letztlich ausschließlich ein vertikaler Informationsfluss erfolgt, kann mittels dezentraler Fusionsarchitekturen der Anforderung nach einem horizontalen Informationsfluss Rechnung getragen werden. Dies ist u. a. gerade im Hinblick auf die Informationsfusion in system-, ebenen- und/oder organisationsübergreifenden Verbänden, wo der Einsatz von sog. Stove-Pipe-Systemen in der Regel nicht zielführend ist, von Relevanz. Siehe hierzu auch [Lig08].

Bei entsprechender Ausgestaltung einer dezentralen Fusionsarchitektur kann auch ein übergeordneter (gesamthafter) Fusionsprozess iterativ umgesetzt werden. Ein recht anschauliches Beispiel hierfür findet sich in [Ald05], wo der Fall adressiert wird, dass unterschiedliche Sensorknoten in einem iterativen Prozess ihre Informationen austauschen und aktualisieren, bis sie einen Konsens im Hinblick auf die Gültigkeit einer Hypothese erreicht haben. Der Einsatz dezentraler Fusionsarchitekturen bzw. von Architektur-Mischformen, welche dezentrale Architekturbestandteile enthalten, kann darüber hinaus auch erfolgen, um z. B. in Verbänden, in denen unterschiedliche Entscheidungsträger unterschiedliche Information benötigen, dafür zu sorgen, dass in den Fusionsknoten eines entsprechend vernetzten Systems jeweils genau die Information (in der erforderlichen Güte) bereitgestellt wird, welche auch tatsächlich benötigt wird. Siehe hierzu auch [Lig08]. Dies ist insbesondere auch eine wesentliche Voraussetzung, um ein optimales Ressourcenmanagement in solchen Verbänden überhaupt realisieren zu können.

Selbstverständlich muss auch bei der Ausgestaltung dezentraler Fusionsarchitekturen eventuellen Abhängigkeiten zwischen den Beiträgen der Informationsquellen geeignet Rechnung getragen werden. Dies könnte je nach konkretem Anwendungsszenario algorithmisch oder ggf. auch organisatorisch geschehen.

Eine besondere Herausforderung bei der Umsetzung einer dezentralen Fusionsarchitektur besteht in der Ausgestaltung des Informationsflusses zwischen den Knoten untereinander und damit verbunden mit der konkreten Festlegung ihrer Verbindungen. Wird diese Aufgabe nicht adäquat gelöst, so kann zum einen ein unnötig hoher Kommunikationsaufwand resultieren. Man betrachte hierzu z. B. den Extremfall eines Systems, in dem die Fusionsknoten

untereinander komplett vernetzt sind und in dem jeder Fusionsknoten seine Information zu allen anderen Fusionsknoten propagiert. Das Vorsehen zusätzlicher, d. h. über das erforderliche Mindestmaß hinausgehender Verbindungen zwischen den Knoten kann dabei jedoch grundsätzlich die Robustheit des Systems unterstützen, indem Verbindungsausfälle kompensierbar sind. Im Falle eines weiträumig vernetzten Systems sollte auch berücksichtigt werden, dass die Informationsflüsse möglichst so ausgestaltet sein sollten, dass Information nicht unnötigerweise über weite Entfernungen propagiert wird. Gegebenenfalls muss bei der Ausgestaltung der Informationsflüsse auch Bandbreitenbeschränkungen zwischen den einzelnen Knoten Rechnung getragen werden.

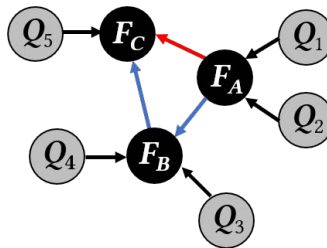
Weiter muss auch verhindert werden, dass bestimmte Information über unterschiedliche Verbindungen und/oder zu unterschiedlichen Zeitpunkten unbeabsichtigt in redundanter Weise in die Fusion eingeht und das Ergebnis verfälscht. Diese Problematik wird im nächsten Abschnitt im Kontext hierarchischer Fusionsarchitekturen noch genauer dargestellt werden.

Abschließend sei angemerkt, dass es bei einer dezentralen Fusionsarchitektur i. A. eher schwierig ist, den Fusionsprozess zu überwachen und zu steuern, da dieses Architekturkonzept letztlich inhärent keine entsprechende globale Struktur vorsieht. Vgl. hierzu auch [Kha09].

#### **2.4.4.4 Hierarchische Fusionsarchitekturen**

Hierarchische Fusionsarchitekturen stellen eine Mischform der vorgenannten drei Architekturtypen dar. In einer hierarchischen Fusionsarchitektur findet die Fusion auf unterschiedlichen Hierarchiestufen statt. In der Regel findet sich auf der obersten Stufe der Hierarchie ein zentraler Fusionsknoten, während auf den niederen Stufen der Hierarchie lokale Fusionsknoten bereits bestimmte Anteile der vorliegenden Information gemäß den Prinzipien zentralisierter oder auch verteilter Fusionsarchitekturen fusionieren. Auf Basis einer hierarchischen Fusionsarchitektur lässt sich dann insbesondere auch eine schrittweise Fusion, wie sie in Abb. 2.4 illustriert wurde, umsetzen.

Wie bei dezentralisierten Fusionsarchitekturen besteht auch bei hierarchischen Fusionsarchitekturen die Gefahr, dass redundant in die Fusion eingehende Information das Ergebnis verfälscht. Abb. 2.12 illustriert diese Problematik anhand eines einfachen Beispiels. In diesem Beispiel liefern fünf Informationsquellen  $Q_s$ ,  $s \in \{1, \dots, 5\}$ , ihre Beiträge an drei Fusionsknoten  $F_A$ ,  $F_B$  und  $F_C$ . Die Verbindungen sind so ausgestaltet, dass der Fusionsknoten  $F_C$  letztlich alle Information empfängt. Die von den Quellen  $Q_1$  und  $Q_2$  gelieferte Information wird im Knoten  $F_A$  fusioniert und muss dann weiter an  $F_B$  geleitet werden, z. B. da sie dort von einem Entscheidungsträger benötigt wird. In diesem Knoten erfolgt auch die Fusion dieser Information mit den von  $Q_3$  und  $Q_4$  gelieferten Beiträgen. Der Knoten  $F_B$  liefert sein Fusionsresultat dann an  $F_C$ . Die in  $F_A$  fusionierte Information wird zusätzlich noch direkt an  $F_C$  geliefert, wodurch sich z. B. ein möglicher Ausfall des Informationsflusses zwischen  $F_A$  und  $F_B$  zumindest im Hinblick auf den Informationsstand in  $F_C$  kompensieren lässt. Wird bei  $F_C$  unabhängig davon, ob der Informationsfluss von  $F_A$  nach  $F_B$  erfolgte oder nicht, stets die direkt von  $F_A$  kommende Information mit der von  $F_B$  und  $Q_5$  gelieferten Information fusioniert, so kann dies zu einer Übergewichtung der von  $F_A$  gelieferten Information und damit zu einer Verfälschung des Fusionsergebnisses führen.



**Abbildung 2.12:** Illustration eines redundanten Informationsflusses (blau, rot) in einer hierarchischen Fusionsarchitektur.

Redundante Informationsflüsse sind oftmals nicht derart direkt aus der Netzwerkstruktur ersichtlich, z. B. beim Vorliegen von sehr umfangreichen Struk-

turen oder bei von über die Zeit ablaufenden Fusionsprozessen. In [Lig08] wird die Problematik nicht direkt aus der Netzwerkstruktur erkennbarer redundanter Informationsflüsse sehr anschaulich anhand sog. Informationsgraphen, die die Informationsflüsse über die Zeit verdeutlichen, aufgezeigt.





## 3 Bayes'sche Fusion

### 3.1 Festlegung der Basisgrößen

Wie in Abschnitt 2.1 dargestellt besteht das Ziel der Informationsfusion letztlich darin, die in Bezug auf den in einer konkreten Aufgabenstellung interessierenden Sachverhalt vorherrschende Unsicherheit gezielt zu reduzieren.

Vor der eigentlichen Durchführung einer Fusionsaufgabe mit mathematischen Methoden ist es notwendig, den interessierenden Sachverhalt formal zu beschreiben. Hierzu wird in der vorliegenden Arbeit der Vektor

$$z := (z_1, \dots, z_N) \in Z := Z_1 \times \dots \times Z_N, \quad (3.1)$$

eingeführt. Dieser Vektor fasst die Größen des Interesses, d. h. die Parameter des nicht direkt beobachtbaren „State of Nature“<sup>1</sup>, über welche in der konkreten Aufgabenstellung Information erlangt werden soll, zusammen. Gemäß der Festlegung in Gleichung (3.1) wird davon ausgegangen, dass  $z$  prinzipiell Werte aus einer Menge  $Z$  annehmen kann und es bezeichnet  $Z_n$  den Wertebereich der  $n$ -ten Komponente  $z_n$  von  $z$ ,  $n \in \{1, \dots, N\}$ ,  $N \in \mathbb{N}$ .

Weiter ist es notwendig, vor der eigentlichen Durchführung der Fusion die zur Lösung der Fusionsaufgabe vorliegenden Informationsbeiträge ebenfalls formal zu beschreiben. Bezeichnet  $d_s$  den von der Informationsquelle Nummer  $s$  gelieferten Informationsbeitrag,  $s \in \{1, \dots, S\}$ , welcher grundsätzlich Werte aus der Menge  $D_s$  annehmen kann, so fasst der Vektor

$$d := (d_1, \dots, d_S) \in D := D_1 \times \dots \times D_S \quad (3.2)$$

---

<sup>1</sup> Diese Terminologie ist angelehnt an die statische Entscheidungstheorie, vgl. z. B. [Ber85].

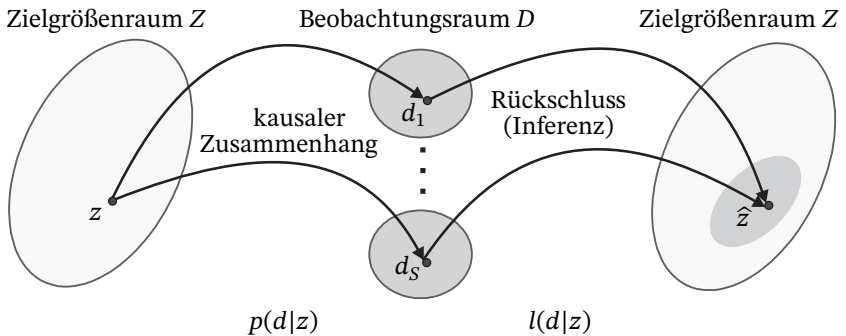
die von den  $S \in \mathbb{N}$  Informationsquellen insgesamt gelieferte Information zusammen. Die Menge  $D$  bezeichnet dabei den Wertebereich des Vektors  $d$ , d. h. die Menge der grundsätzlich möglichen Werte von  $d$ .

In einer konkret vorliegenden Fusionsaufgabe ist der Wert von  $d$  bekannt. Auf dieser Basis soll dann im Rahmen der Fusion ein Rückschluss auf den in der konkreten Aufgabe vorliegenden Wert von  $z$ , welcher nicht direkt erfassbar ist, erfolgen. Bei der Bayes'schen Fusionsmethodik geschieht dies auf Basis Bayes'scher Inferenz, einem Ansatz aus dem Bereich der induktiven Statistik. Die in den nachfolgenden Abschnitten dargestellte Rückführung eines Fusionsproblems auf ein Problem der induktiven (Bayes'schen) Statistik wurde in den Grundzügen zuvor erstmals in [Bey06a] und [Bey06b] veröffentlicht.

## 3.2 Fusion als statistische Inferenz

Induktive statistische Verfahren haben das Ziel, ausgehend von bestimmten Beobachtungen einen Rückschluss auf die zu den Beobachtungen führende Ursache zu ziehen. Dabei wird angenommen, dass der Rückschluss i. A. unsicherheitsbehaftet, also insbesondere nicht eindeutig möglich ist.

Im Kontext der Informationsfusion entsprechen die Beobachtungen gerade den von den Informationsquellen gelieferten Informationsbeiträgen, welche im Vektor  $d$  zusammengefasst sind. Die Ursache, auf die zurück geschlossen werden soll, ist der durch den Vektor  $z$  verkörperte, in der konkreten Aufgabe interessierende Sachverhalt.  $z$  nimmt in der konkret vorliegenden Fusionsaufgabe zwar einen festen „wahren“ Wert an, dieser ist jedoch nicht direkt beobachtbar. Er äußert sich aber gerade in der konkreten Ausprägung der erfassten Informationsbeiträge, d. h. im Wert von  $d$ . Das Ziel der Informationsfusion ist es nun, unter bestmöglicher Ausnutzung der durch  $d$  verkörperten Information einen Rückschluss bezüglich des „wahren“ Werts von  $z$  zu ziehen, vgl. Abb. 3.1. Hierfür ist es notwendig, den Zusammenhang zwischen  $z$  und  $d$  und die korrespondierenden Unsicherheiten adäquat mathematisch zu beschreiben.



**Abbildung 3.1:** Informationsfusion aufgefasst als statistische Inferenz. Abbildung angelehnt an [Bey06b].

Der sog. Grundannahme der induktiven Statistik (vgl. [Rüg99]) folgend wird  $d$  hierfür als Realisierung einer Zufallsvariablen angesehen, deren Verteilung nicht vollständig bekannt ist. Auf Basis des Wissens, welches unabhängig von der konkreten Beobachtung des Werts von  $d$  vorliegt, wird eine Verteilungsannahme formuliert. Diese besteht in der Angabe einer Menge von Verteilungen verbunden mit der Annahme, dass die Verteilung der Zufallsvariablen  $d$  in dieser Verteilungsklasse enthalten ist. Zur Formulierung der Verteilungsannahme wird konkret das Wissen bezüglich des Übergangs von  $z$  zu  $d$  verwendet. Der Rückschluss von  $d$  auf  $z$  korrespondiert dann mit einer genaueren Aussage über die Verteilung von  $d$  oder über bestimmte Kennwerte dieser Verteilung. Die Erarbeitung entsprechender Verfahren stellt das sog. Grundproblem der statistischen Inferenz (vgl. [Rüg99]) dar.

In der vorliegenden Arbeit wird angenommen, dass alle Zufallsvariablen, welche in die Modellierung eines Fusionsproblems eingehen, diskret, absolut stetig oder gemischt diskret absolut stetig sind. Mit der Bezeichnung gemischt diskret absolut stetig ist im Fall mehrdimensionaler Zufallsvariablen gemeint, dass ihre Komponenten teilweise diskret und teilweise absolut stetig sind. Die Verteilung einer solchen Zufallsvariablen lässt sich dann eindeutig durch eine Wahrscheinlichkeitsfunktion (diskreter Fall), durch eine Wahrscheinlichkeitsdichtefunktion (kontinuierlicher, absolut stetiger Fall) bzw. durch

eine Mischung derselben (gemischter Fall) beschreiben (vgl. hierzu z. B. auch [Hen10] sowie [DeG70] für den gemischt diskret absolut stetigen Fall). Im Folgenden wird für diese Wahrscheinlichkeitstheoretischen Funktionen einheitlich der Begriff Wahrscheinlichkeitsverteilung bzw. (verkürzt) Verteilung verwendet. Um unnötige Fallunterscheidungen zu vermeiden, wird außerdem im Folgenden meist nur die kontinuierliche Schreibweise verwendet.

Die Grundannahme der induktiven Statistik sagt damit dann gerade aus, dass der Zusammenhang zwischen  $d$  und  $z$  durch die Angabe einer durch  $z$  parametrisierten Verteilungsklasse der Form  $\{p(d|z)|z \in Z\}$  beschreibbar ist. Basis des Rückschlusses von  $d$  auf  $z$  in einer konkreten Fusionsaufgabe ist dann die sog. Likelihood-Funktion  $l(d|z)$ , die im Wesentlichen mit  $p(d|z)$  korrespondiert, wobei der Wert von  $d$  fixiert und der Wert von  $z$  variabel ist.

In der Regel kann aufgrund der auch nach der Durchführung der Informationsfusion noch vorhandenen Restunsicherheit nicht zweifelsfrei von den Informationsbeiträgen auf den „wahren“ Wert der Größen des Interesses zurückgeschlossen werden, so dass unterschiedliche mögliche Werte von  $z$  zu einem gewissen Mindestmaß mit dem vorliegenden Wert von  $d$  kompatibel sind. Dieser Sachverhalt ist in Abb. 3.1 durch die dunkler eingefärbte Ellipse in der rechtsseitigen Darstellung des Wertebereichs  $Z$  von  $z$  illustriert. Eine Auswahl bezüglich eines eindeutigen Werts  $\hat{z}$  oder auch einer Menge von Werten für  $z$  könnte durch Anwendung einer geeigneten Methode zur Punkt- oder Bereichsschätzung erfolgen. Durch Einbeziehung von Ansätzen aus dem Bereich der statistischen Entscheidungstheorie (vgl. z. B. [Ber85]) lassen sich auch anderweitige Entscheidungen bzw. Aktionen ableiten.

Verfahren der statistischen Inferenz ist gemeinsam, dass Fakten und damit verbundene Unsicherheiten probabilistisch, d. h. auf Basis der Wahrscheinlichkeitstheorie, modelliert und propagiert werden. Sie erfordern deshalb ein grundsätzliches Verständnis des Wahrscheinlichkeitsbegriffs.

## 3.3 Wahrscheinlichkeit

### 3.3.1 Mathematische Festlegung

Die Basis der modernen Wahrscheinlichkeitstheorie stellt die Kolmogorov'sche Axiomatik [Kol33] dar. Diese legt fest, welche Eigenschaften (absolute) Wahrscheinlichkeit als abstrakte mathematische Größe besitzen muss. Aus nachfolgenden beiden Definitionen wird deutlich, dass Wahrscheinlichkeit dabei mathematisch als normiertes Maß, d. h. als eine spezielle, über einem geeigneten Mengensystem definierte Funktion aufgefasst wird.

Definition 3.1 spezifiziert im Wesentlichen die notwendige Struktur dieses Mengensystems:

**Definition 3.1.** Sei  $\Omega$  eine nichtleere Menge. Ein Mengensystem  $\mathcal{A}$  ist eine  $\sigma$ -Algebra über  $\Omega$ , wenn  $\mathcal{A}$  eine Teilmenge der Potenzmenge  $\mathcal{P}(\Omega)$  von  $\Omega$  ist, welche folgende Anforderungen erfüllt:

- (a)  $\emptyset \in \mathcal{A}$ .
- (b) Aus  $A \in \mathcal{A}$  folgt  $\bar{A} := \Omega \setminus A \in \mathcal{A}$ .
- (c) Aus  $A_i \in \mathcal{A}$ ,  $i \in \mathbb{N}$ , folgt  $\bigcup_{i \in \mathbb{N}} A_i \in \mathcal{A}$ .

Das Tupel  $(\Omega, \mathcal{A})$  heißt dann messbarer Raum und jede in  $\mathcal{A}$  enthaltene Menge  $A$  heißt messbar.

Definition 3.2 legt darauf aufbauend anhand der sog. Kolmogorov'schen Axiome fest, welche Eigenschaften Wahrscheinlichkeit aus mathematischer Sicht erfüllen muss:

**Definition 3.2.** Sei  $(\Omega, \mathcal{A})$  ein messbarer Raum und  $P$  eine auf  $\mathcal{A}$  definierte reellwertige Funktion. Das Tripel  $(\Omega, \mathcal{A}, P)$  heißt Wahrscheinlichkeitsraum, falls  $P$  die Kolmogorov'schen Axiome erfüllt:

- (a) Nichtnegativität:  $P(A) \geq 0$  für alle  $A \in \mathcal{A}$ .
- (b) Normiertheit:  $P(\Omega) = 1$ .

(c)  $\sigma$ -Additivität: Ist  $(A_i)_{i \in \mathbb{N}}$  eine Folge disjunkter Mengen aus  $\mathcal{A}$ , so gilt

$$P\left(\bigcup_{i \in \mathbb{N}} A_i\right) = \sum_{i \in \mathbb{N}} P(A_i).$$

$P$  heißt dann Wahrscheinlichkeitsmaß auf  $(\Omega, \mathcal{A})$  und  $\Omega$  heißt Grundraum. Jede in  $\mathcal{A}$  enthaltene (d. h. messbare) Menge  $A$  wird Ereignis genannt.

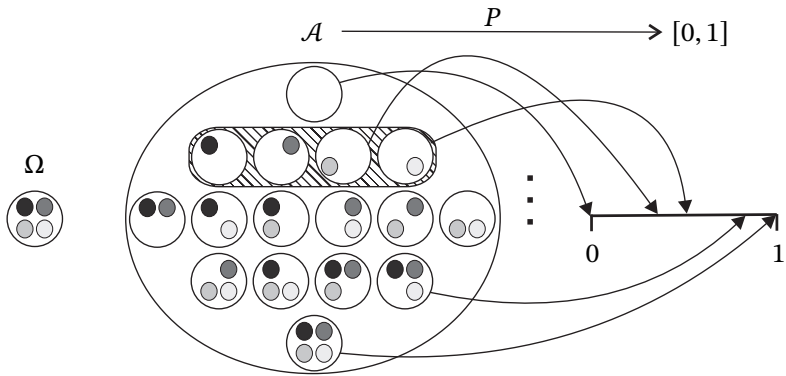
Das Wahrscheinlichkeitsmaß  $P$  weist also gerade den Ereignissen  $A \in \mathcal{A}$  ihre Wahrscheinlichkeit  $P(A)$  zu. Ist eine Teilmenge des Grundraums  $\Omega$  dagegen nicht in  $\mathcal{A}$  enthalten, so stellt sie kein Ereignis dar und ihre Wahrscheinlichkeit ist nicht definiert.

Gemäß der Festlegung eines Wahrscheinlichkeitsraums  $(\Omega, \mathcal{A}, P)$  enthält der zugehörige Grundraum  $\Omega$  alle Ergebnisse, welche grundsätzlich eintreten können. Während das Ereignis  $\Omega$  deshalb sicher eintritt, tritt das Ereignis  $\emptyset$  sicher nicht ein.

Ist  $\Omega$  abzählbar, so ist die Festlegung  $\mathcal{A} = \mathcal{P}(\Omega)$  üblich und wird oft nicht explizit angegeben. In nachfolgender Abbildung 3.2 ist die entsprechende wahrscheinlichkeitstheoretische Modellierung für den Fall eines Grundraums  $\Omega$  der Kardinalität vier graphisch illustriert. Das Wahrscheinlichkeitsmaß  $P$  ist hier eindeutig festgelegt durch seine Werte auf den Elementarereignissen, d. h. den einelementigen Teilmengen von  $\Omega$ .

Im allgemeinen Fall muss die zu  $\Omega$  gehörige  $\sigma$ -Algebra  $\mathcal{A}$  jedoch aus mathematischen Gründen eine echte Teilmenge von  $\mathcal{P}(\Omega)$  darstellen. Gilt  $\Omega = \mathbb{R}^k$ ,  $k \in \mathbb{N}$ , so wählt man als zu  $\Omega$  gehörige  $\sigma$ -Algebra oft ebenfalls ohne explizite Angabe derselben die sog. Borel'sche  $\sigma$ -Algebra  $\mathcal{B}(\mathbb{R}^k)$ . Dies ist die kleinste  $\sigma$ -Algebra, die alle offenen Teilmengen des  $\mathbb{R}^k$  enthält. Es gilt  $\mathcal{B}(\mathbb{R}^k) \neq \mathcal{P}(\mathbb{R}^k)$ . Dabei umfasst  $\mathcal{B}(\mathbb{R}^k)$  jedoch alle in der Praxis relevanten Teilmengen des  $\mathbb{R}^k$ . Vgl. hierzu z. B. [Geo07] oder [Hen10].

Aus den Kolmogorov'schen Axiomen lassen sich die gängigen Regeln für das Rechnen mit Wahrscheinlichkeitsmaßen ableiten. Die Forderung, dass  $(\Omega, \mathcal{A})$  einen messbaren Raum darstellen muss, garantiert dabei, dass das Rechnen mit Wahrscheinlichkeitsmaßen wohldefiniert ist.



**Abbildung 3.2:** Endlicher Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  mit  $|\Omega| = 4$  und  $\mathcal{A} = \mathcal{P}(\Omega)$ . Das Wahrscheinlichkeitsmaß  $P$  weist hier jeder Teilmenge  $A$  von  $\Omega$  ihre Wahrscheinlichkeit  $P(A) \in [0, 1]$  zu. Die Menge der Elementarereignisse, mittels der  $P$  hier eindeutig festgelegt ist, ist in der Abbildung schraffiert markiert. Die Elementarereignisse haben die Form  $\{\omega_i\}$  mit  $\omega_i \in \Omega$ ,  $i \in \{1, 2, 3, 4\}$ .

Die in Definition 3.2 vorgenommene Festlegung absoluter Wahrscheinlichkeit wird komplettiert durch die Festlegung bedingter Wahrscheinlichkeit (siehe z. B. [Geo07] oder [Irl05]). Diese legt fest, wie sich die Wahrscheinlichkeiten von Ereignissen ändern, wenn als Teilinformation bekannt wird, dass ein bestimmtes Ereignis sicher eingetreten ist:

**Definition 3.3.** Seien  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum und  $B \in \mathcal{A}$  ein Ereignis mit  $P(B) > 0$ . Dann ist für jedes Ereignis  $A \in \mathcal{A}$  die bedingte Wahrscheinlichkeit von  $A$  gegeben  $B$  definiert durch

$$P(A|B) := \frac{P(A \cap B)}{P(B)}. \quad (3.3)$$

Bei  $P(\cdot|B)$  handelt es sich wieder um ein Wahrscheinlichkeitsmaß auf  $(\Omega, \mathcal{A})$ , vgl. z. B. [Geo07] für einen mathematischen Beweis.

Die Wahrscheinlichkeitstheorie liefert mittels der in diesem Abschnitt eingeführten Festlegungen eine Antwort auf die Frage, wie man Unsicherhei-

ten mittels Wahrscheinlichkeiten mathematisch beschreiben und verrechnen kann. Sie lässt dabei bewusst offen, wie Wahrscheinlichkeiten zu interpretieren sind. Als Konsequenz sind unterschiedliche Interpretationsmöglichkeiten zulässig. Dies ist ein wesentlicher Grund dafür, dass der Bereich der induktiven Statistik unterschiedliche Inferenzkonzepte umfasst.

### 3.3.2 Interpretationsmöglichkeiten

In der Fachliteratur finden sich diverse Möglichkeiten, Wahrscheinlichkeit zu interpretieren, vgl. z. B. [Gal17], [Goo59] und [Háj19]. Für den Kontext der vorliegenden Arbeit ist die Unterscheidung zwischen im frequentistischen Sinne interpretierter Wahrscheinlichkeit und Wahrscheinlichkeit, welche als sog. Degree-of-Belief (im Deutschen: Grad des Dafürhaltens) aufgefasst wird, wesentlich. Vgl. z. B. auch [Ait04] und [Bey99] zu dieser Unterscheidungsmöglichkeit.

#### 3.3.2.1 Frequentistische Interpretation

Der frequentistischen Interpretation von Wahrscheinlichkeit liegt der Gedanke eines unendlich oft wiederholbaren Zufallsexperiments zugrunde. Ein Zufallsexperiment ist gemäß [Bei03] ein Experiment mit vorab bekannter Ergebnismenge und nichtdeterministischem Ausgang, das unter genau festgelegten Bedingungen zumindest gedanklich beliebig oft wiederholbar ist.

Notiert  $h_k(A)$  die Anzahl des Eintretens eines Ereignisses  $A$  bei  $k$ -maliger Wiederholung des Zufallsexperiments,  $k \in \mathbb{N}$ , so ist die relative Häufigkeit des Ereignisses  $A$  in den  $k$  Wiederholungen gegeben durch

$$r_k(A) := \frac{h_k(A)}{k}. \quad (3.4)$$

Erfahrungsgemäß stabilisiert sich der Wert von  $r_k(A)$  bei einer zunehmenden Anzahl  $k$  von Wiederholungen des Zufallsexperiments. Dieser Sachverhalt wird als empirisches Gesetz über die Stabilisierung relativer Häufigkeiten bezeichnet (vgl. [Hen10]). Bei der frequentistischen Interpretation von Wahr-



scheinlichkeit wird die Wahrscheinlichkeit  $P(A)$  des Ereignisses  $A$  aufgefasst als der Wert, gegen den sich die relative Häufigkeit  $\eta_k(A)$  dementsprechend erfahrungsgemäß stabilisiert.

Es sei betont, dass es sich hierbei um eine Interpretation der Wahrscheinlichkeit  $P(A)$  des Ereignisses  $A$ , nicht um eine Definition im mathematischen Sinne handelt. Wie in [Hen10] dargestellt wäre der Versuch einer Definition von  $P(A)$  als Grenzwert der relativen Häufigkeit  $\eta_k(A)$  für  $k \rightarrow \infty$  als höchst problematisch anzusehen. Das empirische Gesetz über die Stabilisierung relativer Häufigkeiten stellt nämlich keinen mathematischen Beweis dafür dar, dass die Folge  $(\eta_k(A))_{k \in \mathbb{N}}$  tatsächlich gegen einen festen Grenzwert konvergiert, es drückt nur die auf einer endlichen Anzahl von Wiederholungen basierende Erfahrung aus, dass der Wert  $\eta_k(A)$  sich mit zunehmender Anzahl der Wiederholungen  $k$  stabilisiert. Zudem wäre selbst im Falle der Konvergenz die exakte Ermittlung des Grenzwerts nicht möglich. Eine solche würde ja bedeuten, dass eine unendlich große Anzahl von Wiederholungen des entsprechenden Zufallsexperiments tatsächlich durchführbar sein müsste.

### 3.3.2.2 Degree-of-Belief-Interpretation

Bei der Degree-of-Belief-Interpretation von Wahrscheinlichkeit stellt die Wahrscheinlichkeit  $P(A)$  eines Ereignisses  $A$  ein quantitatives Maß für den bezüglich  $A$  vorliegenden Informationsstand bzw. komplementär ausgedrückt für die diesbezüglich vorhandene Unsicherheit dar. Wahrscheinlichkeiten im Sinne der Degree-of-Belief-Interpretation haben stets bedingten Charakter. Es gilt genau genommen  $P(A) = P(A|I)$ , wobei  $I$  für die bezüglich  $A$  vorliegende Information steht. Die beiden Extremwerte Null und Eins korrespondieren damit, dass völlige Ungewissheit (im Fall  $P(A) = 0$ ) bzw. vollständige Information (im Fall  $P(A) = 1$ ) in Bezug auf den durch  $A$  verkörperten Sachverhalt vorliegt.

Wahrscheinlichkeiten im Sinne der Degree-of-Belief-Interpretation von Wahrscheinlichkeit können objektiver oder subjektiver Natur sein. Bei der subjektiven Auffassung der Wahrscheinlichkeit  $P(A)$  als Degree-of-Belief verkörpert der Wert von  $P(A)$  sowohl die bezüglich  $A$  vorhandene Infor-

mation als auch die Überzeugung der diesen Wert festlegenden Person. Bei der objektiven Auffassung der Wahrscheinlichkeit  $P(A)$  als Degree-of-Belief müssen dagegen exakt die in Bezug auf  $A$  vorhandenen Fakten (und nur diese) bei der Festlegung des Wertes von  $P(A)$  berücksichtigt werden. Unterschiedliche Personen, welchen die gleiche Information  $I$  bezüglich des durch  $A$  verkörperten Sachverhalts vorliegt, müssen in diesem Fall also auf den gleichen Wert für  $P(A)$  kommen.

Wie z. B. in [Ber03] angeführt stellt die Degree-of-Belief-Interpretation von Wahrscheinlichkeit die weitreichendste Möglichkeit, Wahrscheinlichkeit zu deuten, dar. Insbesondere beschränkt sie sich hinsichtlich ihrer Anwendbarkeit nicht auf Wahrscheinlichkeiten von Ereignissen, welche in Bezug zu Zufallsexperimenten gesetzt werden können, wie es bei der frequentistischen Interpretation von Wahrscheinlichkeit der Fall ist.

In der Fachliteratur finden sich sogar unterschiedliche Veröffentlichungen, in denen recht überzeugend formal begründet wird, dass Wahrscheinlichkeit im Sinne der Degree-of-Belief-Interpretation ein adäquates und generell ausreichendes Maß darstellt, um die in Bezug auf einen interessierenden Sachverhalt vorliegende Information oder komplementär ausgedrückt die diesbezüglich vorhandene Unsicherheit zu beschreiben. Konkret angeführt hierzu seien die Referenzen [Ber04], [Cox61] und [Lin87]. Dabei muss jedoch auch erwähnt werden, dass hinsichtlich dieser Aussage keine Einigkeit in der Fachliteratur besteht, sie von anderen Autoren also wiederum in Zweifel gezogen wird. Ein häufiger Einwand (vgl. z. B. [Edw92] und [Roy97]) betrifft den Fall, dass keinerlei Information in Bezug auf den interessierenden Sachverhalt vorliegt, und damit verbunden die Frage, ob es in einer solchen Situation überhaupt angemessen ist, eine Wahrscheinlichkeitsaussage zur Beschreibung des Informationsstands zu verwenden. Hintergrund ist letztlich, dass eine solche Wahrscheinlichkeitsaussage zwangsläufig immer eine bestimmte Art von Festlegung macht und somit unvermeidbar doch bestimmte Information transportiert.

Dieser Sachverhalt äußert sich gerade im Kontext der Bayes'schen Theorie dann dadurch, dass man Fälle konstruieren kann, in denen die sog. nichtinformativen A-Priori-Verteilungen, welche das vollständige Nichtvorhandensein

von Information vor Bekanntwerden zusätzlicher Informationsbeiträge probabilistisch repräsentieren sollen, bei der Durchführung statistischer Analysen letztlich dann doch in bedeutendem Umfang Information in das Ergebnis einbringen bzw. zu Widersprüchen führen. Weiter sind diese nichtinformativen A-Priori-Verteilungen in der Regel auch nicht eindeutig und ggf. nicht normierbar. Vgl. hierzu z. B. auch [Ber04] und [Rüg99] sowie anteilig Abschnitt 3.5.1.

Wenngleich man entsprechende Einwände nicht grundsätzlich ignorieren sollte, so lässt sich dennoch feststellen, dass für viele reale Aufgabenstellungen der Informationsfusion die Degree-of-Belief-Interpretation von Wahrscheinlichkeit eine hinreichend mächtige und gut geeignete Möglichkeit zur Beschreibung und Weiterentwicklung der mit dem interessierenden Sachverhalt verbundenen Unsicherheit darstellt. Ein Verzicht auf eine Differenzierung von Unsicherheiten hinsichtlich ihrer Ursachen und Natur (Nichtdeterminismus, Unwissen, Unschärfe etc.) entspricht hier oftmals einem sinnvollen pragmatischen Standpunkt (vgl. [Bey07]). Die Verwendung der Degree-of-Belief-Interpretation von Wahrscheinlichkeit hat insbesondere auch den Vorteil, dass sie oftmals dem intuitiven Verständnis von Unsicherheit entspricht, das auch nicht speziell geschulte Personen mitbringen. Sie stellt damit insbesondere eine gute Basis im Hinblick auf die Umsetzung probabilistischer Verfahren, welche im Sinne von Abschnitt 2.3 ausreichend nachvollziehbar sind, dar.

Im Hinblick auf die konkrete Problemstellung, das vollständige Nichtvorhandensein von Information im Kontext der Wahrscheinlichkeitstheorie exakt mathematisch zu verkörpern, lässt sich festhalten, dass diese Problemstellung sicher noch nicht vollständig gelöst ist bzw. wohl auch gar nicht vollständig lösbar sein wird (vgl. hierzu auch [Bey99] und [Rüg99]). Jedoch stellen die im Kontext der Bayes'schen Theorie entwickelten nichtinformativen A-Priori-Verteilungen in der Praxis oftmals gute Werkzeuge zur Lösung entsprechender Aufgabenstellungen dar. Es ist dabei allerdings wichtig, dass diesen Wahrscheinlichkeitsverteilungen zugrundeliegende Konzept nicht misszuverstehen und sie dementsprechend auch korrekt zu verwenden. Wie in [Iro97] dargestellt, müssen solche nichtinformativen A-Priori-Verteilungen

in dem Sinne aufgefasst werden, dass sie tatsächlich sehr wohl Information transportieren, diese jedoch im Hinblick auf ein bestimmtes statistisches Modell und eine bestimmte Problemstellung derart ausgestaltet ist, dass die in Form von zusätzlichen Informationsbeiträgen bekannt werdende Information die resultierenden Schlüsse dominiert.

### 3.3.2.3 Beispiele

In diesem Abschnitt werden die zwei zuvor eingeführten Möglichkeiten, Wahrscheinlichkeit zu interpretieren, anhand der nachfolgenden beiden Beispiele noch einmal verdeutlicht.

Beispiel A: Eine faire Münze werde geworfen. Betrachtet wird die Aussage „Die Wahrscheinlichkeit, dass die Münze Kopf zeigt, beträgt  $1/2$ .“

Beispiel B: Betrachtet wird die Aussage: „Die Wahrscheinlichkeit, dass Thomas Bayes im Jahr 1701 geboren wurde, beträgt  $8/10$ .“

Beispiel A ist ein gängiges Beispiel zur Verdeutlichung der frequentistischen Interpretation von Wahrscheinlichkeit. Da die Münze fair ist, wird üblicherweise bei einer sehr großen Anzahl von Wiederholungen des Münzwurfs in circa der Hälfte der Fälle das Ergebnis Kopf eintreten, siehe z. B. [Bei03] für eine Aufzeichnung entsprechender Versuche. Es sei jedoch angemerkt, dass sich die tatsächliche Messung des exakten Werts  $1/2$  auch mittels einer noch so großen Anzahl von Münzwürfen nicht in valider Form durchführen lässt (vgl. Abschnitt 3.3.2.1). Die Aussage in Beispiel A ist natürlich auch konsistent mit der umfassenderen Degree-of-Belief-Interpretation von Wahrscheinlichkeit. Die angegebene Wahrscheinlichkeit ist objektiv, denn der Wert  $1/2$  repräsentiert gerade die Information, dass die Münze fair ist und sich somit völlige Indifferenz hinsichtlich des Ergebnisses des Münzwurfs ergibt.

Die Aussage in Beispiel B macht beim Zugrundelegen der frequentistischen Interpretation von Wahrscheinlichkeit keinen Sinn. Die Vorstellung eines wiederholbaren Zufallsexperiments im Zusammenhang mit dem Ereignis „Thomas Bayes wurde im Jahr 1701 geboren“ ist abwegig. Die Aussage in Beispiel B lässt sich jedoch ohne Weiteres als Grad des Dafürhaltens, dafür, dass

Thomas Bayes im Jahr 1701 geboren wurde, verstehen. Sie ist also durchaus sinnvoll im Sinne der Degree-of-Belief-Interpretation von Wahrscheinlichkeit. Das Geburtsjahr von Thomas Bayes, des Namensgebers der Bayes'schen Theorie, ist tatsächlich nicht zweifelsfrei auf das Jahr 1701 datierbar (siehe [Ber04]). Der angegebene konkrete Wahrscheinlichkeitswert  $8/10$  wurde von der Autorin nach der Lektüre von [Ber04] gefühlsmäßig festgesetzt, so dass davon auszugehen ist, dass eine subjektive Wahrscheinlichkeit im Sinne der Degree-of-Belief-Interpretation vorliegt.

## 3.4 Ansätze zur statistischen Inferenz

In Abschnitt 3.2 wurden bereits die Grundannahme der induktiven Statistik und darauf basierend das Grundproblem statistischer Inferenz eingeführt. Die unterschiedlichen Möglichkeiten zur Interpretation von Wahrscheinlichkeit sind ein wesentlicher Grund dafür, dass es darauf basierend keine einheitliche Theorie statistischer Inferenz gibt, sondern dass statt dessen unterschiedliche Inferenzkonzepte nebeneinander bestehen. Die vollständige wahrscheinlichkeitstheoretische Modellierung eines Inferenz-Problems, die genaue Art und Weise, wie Rückschlüsse getroffen werden und wie diese bewertet werden, hängen entscheidend vom konkreten Inferenzkonzept ab.

Im Folgenden wird ein kurzer (ausschnittsweiser) Überblick über die drei bedeutendsten Inferenzkonzepte gegeben mit dem Ziel, diese voneinander abzugrenzen und zugleich wichtige Grundlagen im Hinblick auf die vorliegende Arbeit einzuführen.

### 3.4.1 Klassische Inferenz

Der klassischen Inferenz liegt die frequentistische Interpretation von Wahrscheinlichkeit zugrunde. Die wahrscheinlichkeitstheoretische Modellierung eines Inferenzproblems beschränkt sich auf die in Abschnitt 3.2 beschriebene Einführung der Verteilungsklasse  $\{p(d|z)|z \in Z\}$ . Während  $d$  dabei als

Realisierung einer Zufallsvariablen aufgefasst wird, nimmt  $z$  die Rolle einer deterministischen Konstanten ein.

Der Rückschluss von  $d$  auf  $z$  erfolgt indirekt und wird von einem pre-experimentellen Standpunkt aus betrachtet. Genauer basiert er auf im frequentistischen Sinne zu interpretierenden Wahrscheinlichkeitsaussagen bezüglich  $d$ . Dabei werden dann notwendigerweise auch die möglichen Werte für  $d$ , die in der konkreten Aufgabe überhaupt nicht vorliegen, berücksichtigt.

Beispielsweise könnte eine Bereichsschätzung für eine kontinuierliche, ein-dimensionale Größe des Interesses  $z$  durch Angabe eines sog. Konfidenzintervalls zu einem bestimmten Vertrauensgrad  $\alpha$  erfolgen. Ein derartiges Konfidenzintervall zeichnet sich dadurch aus, dass bei einer großen Anzahl von Durchführungen des Zufallsexperiments, das zum Eintreten von  $d$  führt, erfahrungsgemäß mindestens  $(1 - \alpha) \cdot 100$  Prozent der resultierenden Konfidenzintervalle den „wahren“ Wert von  $z$  enthalten. Diese Aussage gleichzusetzen mit der Aussage, dass der „wahre“ Wert von  $z$  mit Wahrscheinlichkeit  $1 - \alpha$  im Konfidenzintervall enthalten ist, ist nicht zulässig. In einer konkreten Aufgabenstellung ist der „wahre“ Wert von  $z$  im aufgrund der konkreten Ausprägung von  $d$  resultierenden Konfidenzintervall enthalten oder nicht.

Die Auffassung eines Fusionsproblems als Inferenzproblem im Sinne der klassischen Inferenz ist in Anbetracht der zugrundegelegten Interpretation von Wahrscheinlichkeit nur dann zulässig, wenn die im Vektor  $d$  zusammengefassten Informationsbeiträge tatsächlich als zufällig im frequentistischen Sinne angesehen werden können und der Übergang von  $z$  zu  $d$  tatsächlich (rein) als Nichtdeterminismus modelliert werden kann. Ein Fall, in dem dies z. B. möglich wäre, ist der, dass  $d$  die Werte, die sich bei mehrfacher Bestimmung der Position  $z$  eines Objekts ergeben, spezifiziert, und dass die erfassten Positionangaben aufgrund von nicht systematischen Messabweichungen unsicher im frequentistischen Sinne sind.

### 3.4.2 Bayes'sche Inferenz

Der Bayes'schen Inferenz liegt die Degree-of-Belief-Interpretation von Wahrscheinlichkeit zugrunde. Dabei wird der Informationsstand in Bezug auf jede der in die Fusionsaufgabe involvierten Größen durch geeignete Wahrscheinlichkeitsverteilungen ausgedrückt, d. h. sowohl  $d$  als auch  $z$  werden als Zufallsvariablen aufgefasst.

Ein Bayes'sches statistisches Modell ergänzt dementsprechend die mit der Verteilungsklasse  $\{p(d|z)|z \in Z\}$  korrespondierende probabilistische Modellierung durch die Einführung der sog. A-Priori-Verteilung  $p(z)$ . Arbeitet man mit objektiven Degree-of-Belief-Verteilungen<sup>1</sup>, wovon im Kontext der Informationsfusion in der Regel auszugehen sein wird, so repräsentiert die A-Priori-Verteilung gerade das Vorwissen  $d_0$ , welches bezüglich des „wahren“ Werts von  $z$  zusätzlich zu den von den Informationsquellen gelieferten Beiträgen verfügbar ist. Wie jede wahrscheinlichkeitstheoretische Aussage im Sinne der Degree-of-Belief-Interpretation von Wahrscheinlichkeit hat also auch die A-Priori-Verteilung bedingten Charakter, d. h. die Notation  $p(z)$  steht abkürzend für  $p(z|d_0)$ . Anders als die klassische Inferenz bietet die Bayes'sche Inferenz mittels der A-Priori-Verteilung somit einen Weg, zusätzlich zu der von  $d$  gelieferten Information verfügbares Wissen bezüglich  $z$  in koordinierter Weise in die Inferenz einzubringen.

Gemäß der Degree-of-Belief-Interpretation von Wahrscheinlichkeit stellt dann die sog. A-Posteriori-Verteilung  $p(z|d) := p(z|d, d_0)$  die vollständige Verkörperung des Informationsstands bezüglich  $z$  unter Berücksichtigung der von  $d_0$  und  $d$  gelieferten Information dar. Sie berechnet sich gemäß dem Satz von Bayes<sup>2</sup>, einer Rechenregel aus der Wahrscheinlichkeitstheorie:

$$p(z|d) = \frac{p(d|z)p(z)}{\int_Z p(d|z)p(z) dz} . \quad (3.5)$$

<sup>1</sup> Unter einer Degree-of-Belief-Verteilung ist eine Wahrscheinlichkeitsverteilung im Sinne der Degree-of-Belief-Interpretation von Wahrscheinlichkeit zu verstehen.

<sup>2</sup> Ein Beweis des Satzes von Bayes findet sich z. B. in [Ber04].

Der Nenner in Gleichung (3.5) korrespondiert gerade mit der Marginalverteilung von  $p(d)$  von  $d$ . Da in einer konkreten Aufgabenstellung der Wert von  $d$  (wie auch der Wert von  $d_0$ ) fest ist, alle involvierten Größen also rein als Funktionen von  $z$  betrachtet werden, lässt sich Gleichung (3.5) abkürzend schreiben als

$$p(z|d) \propto p(d|z)p(z) . \quad (3.6)$$

Wie bereits in Abschnitt 3.2 erwähnt wird  $p(d|z)$  bei Zugrundelegung des in einer konkreten Aufgabe festen Werts von  $d$  als Likelihood-Funktion bezeichnet und im Folgenden explizit als  $l(d|z)$  notiert. Aufgefasst als Funktion von  $z$  stellt diese Größe i. A. keine Wahrscheinlichkeitsverteilung dar. Wie aus Gleichung (3.6) erkennbar ist, genügt es zur Durchführung der Bayes'schen Inferenz letztlich, wenn  $l(d|z)$  bis auf einen (von  $z$  unabhängigen) Proportionalitätsfaktor bekannt ist. In diesem Sinne lässt sich die Likelihood-Funktion auch allgemeiner definieren als die Klasse von Funktionen, welche für festes  $d$  proportional zu  $p(d|z)$  sind (vgl. hierzu z. B. auch [Edw92], [Gho06] und [Koc14]). Es lässt sich also verallgemeinert festlegen:

$$l(d|z) := c_d \cdot p(d|z) , \quad (3.7)$$

wobei  $c_d$  eine beliebige, jedoch für alle Werte  $z \in Z$  gleich zu wählende Konstante ist.

Im Kontext der Bayes'schen Inferenz kommt der A-Posteriori-Verteilung  $p(z|d)$  fundamentale Bedeutung zu. Unter Zugrundelegung der Rechenregeln der Wahrscheinlichkeitstheorie verbunden mit der Interpretation von Wahrscheinlichkeit als Degree-of-Belief stellt sie letztlich die einzig stimmige probabilistische Verkörperung des gesamten Informationsstands bezüglich des „wahren“ Werts von  $z$  dar. Anders als bei der klassischen Inferenz kann dadurch, dass auch  $z$  als Zufallsgröße aufgefasst wird, hier also tatsächlich direkt probabilistisch auf den „wahren“ Wert von  $z$  zurückgeschlossen werden und die Bewertung der Güte des Rückschlusses sowie die Ableitung spezifischerer Aussagen bzw. Schlussfolgerungen können darauf basierend



von einen post-experimentellen Standpunkt, d. h. auf Basis des tatsächlich für  $d$  vorliegenden Werts, erfolgen.

Beispielsweise könnte eine Bereichsschätzung für eine kontinuierliche, eindimensionale Größe des Interesses  $z$  durch Angabe eines Intervalls erfolgen, welches mindestens  $(1 - \gamma) \cdot 100$  Prozent der A-Posteriori-Wahrscheinlichkeit bezüglich des „wahren“ Werts von  $z$  enthält. Um dieses Intervall möglichst aussagekräftig zu wählen, kann dabei z. B. zusätzlich gefordert werden, dass die A-Posteriori-Verteilung von  $z$  innerhalb dieses Intervalls einen gewissen, geeignet festzulegenden Minimalwert  $\delta$  nicht unterschreitet. In diesem Fall wird das resultierende Intervall dann als HPD (Highest Posterior Density) Intervall zum Glaubwürdigkeitsgrad  $\gamma$  bezeichnet. Vgl. hierzu [Rüg99].

Aufgrund der weitreichenden Interpretation von Wahrscheinlichkeit als Degree-of-Belief ist die Auffassung eines Fusionsproblems als Inferenzproblem im Sinne der Bayes'schen Inferenz nicht nur auf diejenigen Fälle beschränkt, in denen die für den Rückschluss von  $d$  auf  $z$  zu modellierende Unsicherheit im Sinne der frequentistischen Interpretation von Wahrscheinlichkeit auffassbar ist. Während eine Interpretation im frequentistischen Sinne zumindest bei solchen Low Level Fusionsaufgaben (vgl. Abschnitt 2.4.3.1), bei denen physikalische Eigenschaften von Objekten der realen Welt mittels Sensoren, welche bestimmte physikalische Prinzipien umsetzen, erfasst werden, oft noch sinnvoll möglich sein mag, ist dies insbesondere bei Aufgaben im Bereich der High Level Fusion in der Regel nicht möglich. Vgl. hierzu auch [Tot08].

Prominente Beispiele in der Fachliteratur wie Bayes'sche Tracking Verfahren (vgl. z. B. [Koc14]) oder der in der STANAG<sup>1</sup> 4162 [NAT09] standardisierte IDCPC (Identification Data Combining Process) zur Charakterisierung der Eigenschaften aufklärungsrelevanter Objekte auf Basis Bayes'scher Inferenz (vgl. z. B. [Krü08]) verdeutlichen darüber hinaus auch die gute Anwendbarkeit Bayes'scher Verfahren in Low Level Fusionsaufgaben. Bayes'sche Inferenz kann in diesem Sinne ein durchgängiges Konzept zur Modellierung von

---

<sup>1</sup> Die Abkürzung STANAG steht für (NATO) Standardization Agreement, ein NATO Standardisierungsübereinkommen.

Unsicherheiten und zu ihrer Propagierung durch verschiedene Ebenen des JDL Modells liefern<sup>1</sup>. Anzumerken ist dabei, dass (bei Zugrundelegung von objektiven Degree-of-Belief-Verteilungen) mittels der Bayes'schen Inferenz auch Ergebnisse abgeleitet werden können, welche aus Sicht der klassischen Inferenz ebenfalls sinnvoll sind (vgl. z. B. [Gho06]).

Wesentlich für die breite Anwendbarkeit der Bayes'schen Inferenz im Kontext der Informationsfusion ist insbesondere auch die Tatsache, dass sie die gleichzeitige Handhabung kontinuierlicher und diskreter Größen erlaubt, diese verläuft praktisch analog. Weiter lassen sich auch unterschiedlich skalierte Größen unkompliziert gemeinsam handhaben. Diese wäre im Kontext der Informationsfusion z. B. nötig, wenn Objekte gleichzeitig lokalisiert und klassifiziert werden sollen. Während es sich bei der Position um eine intervall-skalierte Größe handelt, stellt der Objekttyp eine nominal-skalierte Größe dar.

### 3.4.3 Likelihood-Inferenz

Die Likelihood-Inferenz beschränkt sich wie die klassische Inferenz auf die Einführung der Verteilungsklasse  $\{p(d|z)|z \in Z\}$ . Anders als in der klassischen Inferenz wird bei der Likelihood-Inferenz jedoch allein der in der konkreten Aufgabe tatsächlich vorliegende Wert von  $d$  für den Rückschluss auf  $z$  herangezogen. Das grundsätzliche Vorgehen dabei ist unabhängig davon, ob  $z$  als deterministische Konstante (wie bei der klassischen Inferenz der Fall) oder als Zufallsvariable im Sinne der Degree-of-Belief-Interpretation von Wahrscheinlichkeit (wie bei der Bayes'schen Inferenz der Fall) aufgefasst wird, anwendbar.

Wesentlicher Ausgangspunkt der Likelihood-Inferenz ist die Feststellung, dass die Likelihood-Funktion  $l(d|z)$  die gesamte Information, welche von den Beobachtungen  $d$  über den „wahren“ Wert von  $z$  geliefert wird, enthält. Das sog. Likelihood-Korollar (vgl. [Rüg99]) sagt dementsprechend aus, dass der Schluss von  $d$  auf  $z$  allein auf der Likelihood-Funktion zu basieren hat.

---

<sup>1</sup> Diesen Sachverhalt illustriert im Kleinen bereits der IDCP. Er unterstützt nämlich letztlich die Schnittstelle zwischen JDL Level 1 und JDL Level 3, siehe hierzu auch [Bla14].

Eine eindeutige Leitlinie, wie dies erfolgen soll, ergibt sich aus einem Postulat, das, der z. B. in [Hac65] und [Roy97] hierfür verwendeten Terminologie folgend, in der vorliegenden Arbeit als Law of Likelihood bezeichnet wird. Das Law of Likelihood lässt sich wie folgt fassen:

Es seien  $z^*, z^{**} \in Z$ . Die (gemachten) Beobachtungen  $d$  liefern genau dann statistische Evidenz für  $z^*$  vis-à-vis  $z^{**}$  wenn  $l(d|z^*) > l(d|z^{**})$  ist. Ein quantitatives Maß für die Stärke der statistischen Evidenz stellt die Likelihood-Ratio dar, d. h. der Quotient  $l(d|z^*)/l(d|z^{**})$  gibt an, wie stark  $d$  den Wert  $z^*$  relativ zum Wert  $z^{**}$  stützt.

Die auf Basis des Law of Likelihood angestellte Betrachtung hinsichtlich der von den Beobachtungen  $d$  gelieferten statistischen Evidenz ist stets relativ. Beantwortet wird also die Frage, inwieweit die gemachten Beobachtungen  $d$  eine Hypothese bezüglich des „wahren“ Werts von  $z$  im Vergleich zu einer anderen stützen – nicht die Frage, ob die gemachten Beobachtungen für oder gegen eine bestimmte (einzelne) Hypothese sprechen. Diese Betrachtungsweise ist von daher sinnvoll, da ein niedriger Wert von  $p(d|z)$  für ein bestimmtes  $z^* \in Z$  nicht bedeutet, dass die gemachten Beobachtungen  $d$  gegen  $z^*$  sprechen. Für festes  $d$  als Funktion von  $z$  betrachtet stellt  $p(d|z)$  i. A. keine Wahrscheinlichkeitsverteilung dar. Der Wert von  $p(d|z)$  kann als Konsequenz recht gering für alle  $z \in Z$  ausfallen und  $d$  kann in diesem Fall (relativ zu allen anderen möglichen Werten aus  $Z$ ) sogar sehr starke statistische Evidenz für einen Wert  $z^*$ , für den  $p(d|z)$  absolut betrachtet recht niedrig ist, liefern. Vgl. auch [Bar04] und [Roy97] für vertiefte Diskussionen dieses Sachverhalts. Weiter sei ergänzend auf die Referenz [Edw92] verwiesen, in der vor allem auch die Sinnhaftigkeit der Likelihood-Ratio für den kontinuierlichen Fall (sowohl im Hinblick auf  $d$  als auch auf  $z$ ) dargelegt wird.

Um einen besseren Überblick über die von  $d$  gelieferte statistische Evidenz zu erhalten, bietet sich oftmals die Betrachtung der sog. standardisierten Likelihood-Funktion  $l_{\text{st}}(d|z)$  an (vgl. z. B. [Blu02] oder [Roy97]). Diese resultiert, indem man die Likelihood-Funktion  $l(d|z)$  auf einen Maximalwert von Eins skaliert. Die standardisierte Likelihood-Funktion  $l_{\text{st}}(d|z)$  gibt für jedes  $z \in Z$  die statistische Evidenz, welche die Beobachtungen  $d$  für  $z$  relativ

zum Maximum-Likelihood-Schätzwert, dem im Vergleich zu allen anderen möglichen Werten aus  $Z$  am besten gestützten Wert, liefern, an. Um die Darstellung nicht unnötig zu verkomplizieren, wird in der vorliegenden Arbeit jeweils angenommen, dass die betrachteten Likelihood-Funktionen stets ein entsprechendes globales Maximum besitzen, auch wenn diese Voraussetzung tatsächlich, wie z. B. in [Rüg99] thematisiert wird, nicht immer erfüllt sein muss.

Gemäß den Leitlinien der Likelihood-Inferenz sollte beispielsweise eine Bereichsschätzung für eine kontinuierliche, eindimensionale Größe des Interesses  $z$  diejenigen Werte aus  $Z$  enthalten, welche gemäß den gemachten Beobachtungen  $d$  besonders schlüssig sind, in dem Sinne, dass die Beobachtungen für sie ein besonders hohes Maß an statistischer Evidenz (im Vergleich zu den anderen Werten aus  $Z$ ) liefern. Diese Aufgabe lässt sich oftmals am Besten umsetzen, indem man eine untere Schranke  $\epsilon \in (0,1)$  an die standardisierte Likelihood-Funktion ansetzt<sup>1</sup>.

Wie in [Roy97] sehr anschaulich erläutert wird, besteht das Ziel des entsprechenden, bei der Likelihood-Inferenz angestellten Schlusses nicht darin, den bezüglich des „wahren“ Werts von  $z$  insgesamt vorliegenden Informationsstand zu ermitteln. Wie weiter erläutert wird, wäre diese Zielsetzung, den vollständigen Informationsstand zu ermitteln, durch die Berechnung der A-Posteriori-Verteilung  $p(z|d)$ , wie sie im Kontext der Bayes'schen Inferenz eingeführt wurde, zu adressieren. Das Ziel der Likelihood-Inferenz ist dementsprechend die objektive Evaluation allein der Beobachtungen  $d$  in Bezug auf die Frage nach dem „wahren“ Wert von  $z$ . Vorwissen, welches zusätzlich zur Verfügung stehen mag, wird dabei bewusst ausgeklammert, es darf in Bezug auf diese spezifische Frage ja auch letztlich überhaupt nicht berücksichtigt werden.

Dies erklärt sehr gut die Tatsache, dass die Anwendung der Likelihood-Inferenz heute z. B. im Kontext der statistischen Auswertung klinischer Stu-

---

<sup>1</sup> Dieser Sachverhalt wird in Abschnitt 6.2 der vorliegenden Arbeit noch vertieft betrachtet und genutzt werden, so dass im Rahmen des hier angeführten Beispiels nicht weiter darauf eingegangen wird.

dien (vgl. [Blu02]) und der forensischen Statistik (vgl. [Ait04] und [Luc05]) propagiert wird. Beispielsweise hat ein forensischer Experte, der bestimmte Daten in Bezug auf eine mögliche Straftat auswerten soll, letztlich die Aufgabe, ein allein auf den ihm vorliegenden Daten beruhendes Ergebnis zu liefern. Die Kombination dieses Ergebnisses mit den Ergebnissen der Auswertung weiterer Daten und eine Aussage hinsichtlich des Grad des Dafürhaltens dafür, ob eine Straftat vorliegt bzw. wer der Verursacher ist, sollte zwar möglich sein, hat aber nicht durch den forensischen Experten, sondern durch die ermittelnden Polizeibeamten bzw. später final durch ein Gericht zu erfolgen.

Gemäß dem in Abschnitt 2.1 ausführlich erläuterten Grundprinzip der Informationsfusion macht es bei der Informationsfusion nun aber i. A. gerade Sinn, Vorwissen, welches zusätzlich zu den von den Informationsquellen gelieferten Beiträgen verfügbar ist, gezielt zu berücksichtigen, um so ein möglichst hochwertiges Fusionsergebnis zu erzielen. In anderen Worten hat die Informationsfusion also i. A. gerade das übergeordnete Ziel, den vollständigen Informationsstand im Hinblick auf  $z$  zu erschließen – so wie es die Bayes'sche Inferenz zu leisten vermag. Die Zugrundelegung der Likelihood-Inferenz als allgemeingültiges Inferenzkonzept zur Adressierung von Aufgabenstellungen der Informationsfusion ist in diesem Sinne also als nicht sinnvoll zu erachten.

Wie im weiteren Verlauf der vorliegenden Arbeit erkennbar werden wird, eignen sich Prinzipien aus dem Bereich der Likelihood-Inferenz jedoch sehr gut, um bestimmte spezielle Betrachtungen im Kontext der Informationsfusion, insbesondere auch im Hinblick auf die lokale Bayes'sche Fusion, wesentlich zu unterstützen. Voraussetzung dafür, dass dieses Vorgehen möglich ist, ist, dass sich die der Likelihood-Inferenz und der Bayes'schen Inferenz zugrundeliegenden Prinzipien grundsätzlich nicht widersprechen. Das dem tatsächlich so ist, soll im Folgenden kurz skizziert werden.

Gemäß dem Satz von Bayes (vgl. Gleichung (3.5)) gilt für  $z^*, z^{**} \in Z$  die folgende Beziehung:

$$\underbrace{\frac{p(z^*|d)}{p(z^{**}|d)}}_{\text{Posterior-Odds}} = \underbrace{\frac{l(d|z^*)}{l(d|z^{**})}}_{\text{Likelihood-Ratio}} \cdot \underbrace{\frac{p(z^*)}{p(z^{**})}}_{\text{Prior-Odds}}. \quad (3.8)$$

Ist  $Z$  diskret, so geben die sog. Prior-Odds in Gleichung (3.8) an, um wie viel wahrscheinlicher (oder unwahrscheinlicher) die mit  $z^*$  korrespondierende Hypothese bezüglich des „wahren“ Werts von  $z$  gegenüber der mit  $z^{**}$  korrespondierenden Hypothese ist, wenn man allein das vorhandene Vorwissen berücksichtigt. Nachdem der Informationsstand in Bezug auf den „wahren“ Wert von  $z$  durch Einbeziehung der von  $d$  gelieferten Information aktualisiert wurde, stellen die sog. Posterior-Odds die Revision dieser quantitativen (relativen) Aussage dar. Wie aus Gleichung (3.8) erkennbar ist, stellt der Mechanismus, durch welchen die Prior-Odds in die Posterior-Odds überführt werden, eine Multiplikation mit der Likelihood-Ratio dar. Die Likelihood-Ratio quantifiziert also den Faktor, um den diese Revision erfolgen sollte, und stellt in diesem Sinne damit auch aus Sicht der Bayes'schen Theorie ein adäquates (relatives) Maß für die von den Beobachtungen gelieferte statistische Evidenz im Hinblick auf den „wahren“ Wert von  $z$  dar. Wie in [Edw92] im Kontext der Herleitung des Satzes von Bayes für den Fall, dass es sich bei  $z$  um eine kontinuierliche Größe handelt, dargestellt, lässt sich diese Überlegung letztlich so auch auf den kontinuierlichen Fall übertragen. Die beiden bezüglich des „wahren“ Werts von  $z$  betrachteten Hypothesen korrespondieren gemäß [Edw92] dann mit infinitesimal kleinen Bereichen aus  $Z$  der Form  $z^* < z < z^* + dz^*$  und  $z^{**} < z < z^{**} + dz^{**}$ , wobei sich die infinitesimalen Elemente  $dz^*$  und  $dz^{**}$  aus Gleichung (3.8) herauskürzen.

In diesem Sinne könnte man die zuvor dargestellten Prinzipien aus dem Bereich der Likelihood-Inferenz als auch inhärenten Bestandteil der der Bayes'schen Inferenz zugrundeliegenden Bayes'schen Theorie betrachten. Voraussetzung ist dabei jedoch, dass ein Standpunkt eingenommen wird, demzufolge Wahrscheinlichkeitsaussagen tatsächlich im Sinne der Degree-of-Belief-Interpretation von Wahrscheinlichkeit aufgefasst werden können. Hierzu ist anzumerken, dass prominente Vertreter der Likelihood-Inferenz diesen Standpunkt jedoch mindestens in dem Fall, in dem keinerlei Vorwissen bezüglich  $z$  vorliegt, für nicht sinnvoll erachten (vgl. hierzu auch Abschnitt 3.3.2.2).

Abschließend sei noch angemerkt, dass sich manche Autoren wie z.B. [Edw92] dafür aussprechen, den Grad, zu dem die Beobachtungen  $d$  be-

stimmte Hypothesen (relativ zueinander betrachtet) stützen, auf einer logarithmischen Skala anzugeben. In diesem Fall lässt sich die in Gleichung (3.8) dargestellte Beziehung auch wie folgt sehr anschaulich als Differenz interpretieren:

$$\underbrace{\log\left(\frac{l(d|z^*)}{l(d|z^{**})}\right)}_{\text{Unterstützung durch } d} = \underbrace{\log\left(\frac{p(z^*|d)}{p(z^{**}|d)}\right)}_{\text{Posterior-Log-Odds}} - \underbrace{\log\left(\frac{p(z^*)}{p(z^{**})}\right)}_{\text{Prior-Log-Odds}}. \quad (3.9)$$

## 3.5 Erfüllung der Basisanforderungen

Gemäß Abschnitt 2.3 muss eine sinnvolle Methodik zur Fusion heterogener Informationsquellen die Fähigkeiten zur Transformation, Fusion und Fokussierung besitzen. Im vorliegenden Abschnitt werden Mechanismen aufgezeigt, anhand deren die Bayes'sche Fusionsmethodik diesen drei Basisanforderungen nachzukommen vermag. Diese Mechanismen gründen sich auf fundierte Konzepte und Verfahren aus der Wahrscheinlichkeitstheorie, der Informationstheorie und der statistischen Entscheidungstheorie. Der vorliegende Abschnitt stellt eine hinsichtlich der erforderlichen Grundlagen deutlich erweiterte Darstellung der diesbezüglich in [Bey06b] erstmalig veröffentlichten Ergebnisse im Hinblick auf die Frage nach der Adäquatheit der Bayes'schen Fusionsmethodik bezüglich der vorab formulierten Basisanforderungen dar.

### 3.5.1 Fähigkeit zur Transformation

Wie in Abschnitt 3.4.2 bereits dargelegt wird bei der Bayes'schen Fusion der Informationsstand in Bezug auf alle in die Fusionsaufgabe involvierten Größen mittels Wahrscheinlichkeitsverteilungen im Sinne der Degree-of-Belief-Interpretation von Wahrscheinlichkeit repräsentiert. Üblicherweise enthält die Menge der Wahrscheinlichkeitsverteilungen, welche mit der in Bezug auf eine involvierte Größe vorliegenden Information konsistent sind, mehr als ein

Element. Als Konsequenz sind Mechanismen notwendig, um aus dieser Menge bestmöglich eine eindeutige Wahrscheinlichkeitsverteilung auszuwählen. In der Regel strebt man dabei an, dass die gewählte Wahrscheinlichkeitsverteilung objektiv ist, d. h. dass sie einerseits konsistent mit der vorliegenden Information ist und dass sie andererseits keine zusätzliche Information (in Form von Artefakten) in das Fusionsergebnis einbringt.

Die Bayes'sche Theorie stellt unterschiedliche, an spezifische Anforderungen angepasste Mechanismen zu Verfügung, um einen vorliegenden Informationsstand dementsprechend in eine objektive probabilistische Repräsentation im Sinne der Degree-of-Belief-Interpretation von Wahrscheinlichkeit zu überführen. Einen Überblick hierzu liefern z. B. die Referenzen [Ber85, Ber04, Kas96, Rob01]. Einen konzeptionell einfachen und gleichzeitig theoretisch gut begründeten Mechanismus zur Transformation, welcher oft erfolgreich anwendbar ist, liefert das erstmals in [Jay68] veröffentlichte Prinzip der Maximalen Entropie. Seine Anwendung wird heute u. a. im Kontext der Messtechnik (siehe z. B. [Bey99], [Som09] und [Wei99a]) als zielführend angesehen. Darüber hinaus hat es auch in anderen Anwendungsbereichen (siehe z. B. [Cre04], [De 18], [Har14], [Kap89] und [Moh90]) seine Praxistauglichkeit bewiesen.

Das Prinzip der Maximalen Entropie lässt sich auf grundlegende Konzepte aus der Informationstheorie [Sha48] zurückführen. Im Folgenden wird es zuerst für den Fall diskreter Wahrscheinlichkeitsverteilungen eingeführt, anschließend wird es dann auf den allgemeinen Fall erweitert.

**Definition 3.4.** *Es sei  $p(z)$  die Wahrscheinlichkeitsverteilung einer diskreten Zufallsgröße  $z$ , welche Werte aus der Menge  $Z$  annehmen kann. Die (Shannon'sche) Entropie  $H[p(z)]$  von  $p(z)$  ist gegeben durch*

$$H[p(z)] := -E_{p(z)}[\log p(z)] = - \sum_{z \in Z} p(z) \log p(z) . \quad (3.10)$$

Die (Shannon'sche) Entropie stellt ein quantitatives Maß für die durch die Wahrscheinlichkeitsverteilung  $p(z)$  verkörperte Unsicherheit, oder in anderen Worten, ein inverses Maß für den Informationsgehalt von  $p(z)$  dar. Die-



ser Sachverhalt lässt sich direkt aus Axiomen für ein entsprechendes Unsicherheitsmaß ableiten, siehe hierzu z. B. [Gui77] oder [Kli06]. Man kann ihn sich wie folgt verdeutlichen: Die Größe  $-\log p(z)$ , deren Erwartungswert  $E_{p(z)}[-\log p(z)]$  in Gleichung (3.10) unter Zugrundelegung der Verteilungsannahme  $p(z)$  gebildet wird, lässt sich auffassen als Maß für den Informationsgewinn, den es zu Folge hat, wenn bekannt wird, dass der entsprechende Wert  $z$  der Zufallsgröße der „wahre“ Wert ist. Sie spiegelt dabei u. a. die Tatsache wieder, dass der Informationsgewinn beim Bekanntwerden eines Ereignisses umso höher bewertet werden sollte, je unwahrscheinlicher das Ereignis ist. Die (Shannon'sche) Entropie  $H[p(z)]$  lässt sich in diesem Sinne auffassen als der mittlere Informationsgewinn, der unter Zugrundelegung der Verteilungsannahme  $p(z)$  beim Bekanntwerden des „wahren“ Werts von  $z$  auftritt. Je geringer der Wert von  $H[p(z)]$  ausfällt, umso geringer ist die verbleibende Unsicherheit bzw. (in anderen Worten) umso weniger Information fehlt noch im Hinblick auf die vollständige Kenntnis des „wahren“ Werts von  $z$ .

Gemäß dem Prinzip der Maximalen Entropie ist aus der Menge aller Wahrscheinlichkeitsverteilungen auf  $Z$ , welche konsistent mit der bezüglich  $z$  vorliegenden Information  $I$  sind, diejenige Wahrscheinlichkeitsverteilung auszuwählen, welche maximale (Shannon'sche) Entropie aufweist. Dadurch wird die Unsicherheit bezüglich des „wahren“ Werts von  $z$  also nur in dem Umfang verringert, in dem dies notwendig ist, um die Konsistenz mit  $I$  sicherzustellen. Die Maximum-Entropie-Verteilung  $p_{\text{ME}}(z)$  ergibt sich damit als Lösung des folgenden Optimierungsproblems:

$$p_{\text{ME}}(z) = \arg \max_{p(z) \in \mathfrak{P}_z} H[p(z)] . \quad (3.11)$$

Dabei bezeichnet  $\mathfrak{P}_z$  die Menge aller Wahrscheinlichkeitsverteilungen auf  $Z$ , welche konsistent mit der vorliegenden Information  $I$  sind.

Als Basis für die Erweiterung des Prinzips der Maximalen Entropie auf den Fall kontinuierlicher Zufallsgrößen dient in der Fachliteratur oft die direkte, in nachfolgender Definition 3.5 gefasste Erweiterung des Entropie-Begriffs. Die Maximum-Entropie-Verteilung  $p_{\text{ME}}(z)$  ergibt sich dann analog zum diskreten Fall als Lösung des in Gleichung (3.11) formulierten Optimierungsproblems.

**Definition 3.5.** *Es sei  $p(z)$  die Wahrscheinlichkeitsverteilung einer kontinuierlichen Zufallsgröße  $z$ , welche Werte aus der Menge  $Z$  annehmen kann. Die (Boltzmann) Entropie  $H[p(z)]$  von  $p(z)$  ist gegeben durch*

$$H[p(z)] := -E_{p(z)}[\log p(z)] = - \int_Z p(z) \log p(z) dz . \quad (3.12)$$

Die (Boltzmann) Entropie lässt sich grundsätzlich analytisch sowie (wie auch im diskreten Fall) auf Basis entsprechender Axiome (hier dann speziell für den kontinuierlichen Fall) betrachten (vgl. z. B. [Gui77]). Ihre Verwendung als quantitatives Maß für die durch die Wahrscheinlichkeitsverteilung  $p(z)$  verkörperte Unsicherheit ist dennoch zumindest manchmal mit gewissen konzeptionellen Schwierigkeiten verbunden.

Diese Schwierigkeiten resultieren letztlich daraus, dass – wenn auch nicht direkt ersichtlich – in Definition 3.4 und Definition 3.5 letztlich in folgendem Sinne unterschiedliche mathematische Größen eingeführt wurden: Zwar resultiert die rechte Seite von Gleichung (3.12) auf den ersten Blick direkt aus der rechten Seite von Gleichung (3.10), indem die Summation durch eine Integration ersetzt wird. Allerdings lässt sich zeigen, dass die rechte Seite von Gleichung (3.12) gerade nicht als Grenzwert der rechten Seite von Gleichung (3.10) bei einer immer feiner werdenden Diskretisierung des Wertebereichs  $Z$  von  $z$  resultiert (siehe hierzu z. B. [Bey99], [Gui77] oder [Kli06]). Anders als die (Shannon'sche) Entropie diskreter Zufallsgrößen kann die (Boltzmann) Entropie kontinuierlicher Zufallsgrößen auch negative Werte annehmen und ist nicht invariant unter Koordinatentransformationen (siehe hierzu insbesondere auch [Jay68]).

Eine vollständig einheitliche Herleitung eines quantitativen Maßes für die durch eine Wahrscheinlichkeitsverteilung  $p(z)$  verkörperte Unsicherheit für den diskreten und den kontinuierlichen Fall, welches keine derartigen Schwierigkeiten mit sich bringt, lässt sich jedoch bewerkstelligen (vgl. hierzu z. B. [Gui77], [Kli06] sowie insbesondere [Hob73]). Hierfür setzt man statt eines absoluten Unsicherheitsmaßes, wie es die Entropie darstellt, ein relatives Unsicherheitsmaß an. Konkret herangezogen wird die in [Kul59] festgelegte

Kullback-Leibler-Divergenz, welche auch als relative Entropie bezeichnet wird. Diese ist wie folgt definiert:

**Definition 3.6.** *Es seien  $p(z)$  und  $q(z)$  Wahrscheinlichkeitsverteilungen auf  $Z$ . Die Kullback-Leibler-Divergenz  $KD[p(z),q(z)]$  von  $p(z)$  relativ zu  $q(z)$  ist gegeben durch*

$$KD[p(z),q(z)] := E_{p(z)} \left[ \log \frac{p(z)}{q(z)} \right] = \int_Z p(z) \log \frac{p(z)}{q(z)} dz . \quad (3.13)$$

*In Fällen, in denen  $z$  anteilig oder vollständig aus diskreten Komponenten besteht, ist die Integration in Gleichung (3.13) dementsprechend (anteilig oder vollständig) durch eine Summation zu ersetzen.*

Die Kullback-Leibler-Divergenz  $KD[p(z),q(z)]$  lässt sich in folgendem Sinne als relatives Unsicherheitsmaß auffassen (vgl. [Kul59] zur Herleitung der mathematischen Beziehungen<sup>1</sup>): Es bezeichne  $A$  die Hypothese, dass  $z$  die Verteilung  $p(z)$  besitzt. Weiter bezeichne  $B$  die Hypothese, dass  $z$  die Verteilung  $q(z)$  besitzt. Durch Ansetzen der Odds-Form des Satzes von Bayes (ähnlich wie in Gleichung (3.8)) lässt sich zeigen, dass  $p(z)/q(z)$  gerade die durch das Bekanntwerden eines Werts von  $z$  gelieferte statistische Evidenz (vgl. Abschnitt 3.4.3) für die Hypothese  $A$  relativ zur Hypothese  $B$  darstellt, wenn man davon ausgeht, dass Hypothese  $A$  wahr ist,  $z$  also die Verteilung  $p(z)$  besitzt. Die Kullback-Leibler-Divergenz  $KD[p(z),q(z)]$  gibt somit also die mittlere statistische Evidenz (im Sinne einer Erwartungswertbildung gemäß der tatsächlich vorliegenden Verteilung  $p(z)$ ) für die Hypothese  $A$  relativ zur Hypothese  $B$  an, wenn man diese auf einer logarithmischen Skala (vgl. hierzu auch Gleichung (3.9)) misst.

Die Kullback-Leibler-Divergenz stellt die Basis des sog. Prinzips der Minimalen Information (siehe [Par06, Van81, Wil80]) dar, welches sich als Erweiterung des Prinzips der Maximalen Entropie auffassen lässt (vgl. auch [Hob73]).

<sup>1</sup> Die Terminologie in [Kul59] ist anders als die nachfolgend verwendete, insbesondere wird dort der Begriff der statistischen Evidenz nicht verwendet.

Zur Erläuterung des Prinzips der Minimalen Information sei Folgendes angenommen: Es stelle  $q(z)$  eine objektive Wahrscheinlichkeitsverteilung für  $z$  im Sinne der Degree-of-Belief-Interpretation von Wahrscheinlichkeit dar. Es werde zusätzliche Information  $I$  in Bezug auf  $z$  verfügbar und die Aufgabe sei nun,  $q(z)$  durch eine neue objektive Wahrscheinlichkeitsverteilung zu ersetzen, welche zusätzlich auch  $I$  berücksichtigt. Basierend auf der zuvor erläuterten Interpretation der Kullback-Leibler-Divergenz als relatives Unsicherheitsmaß besagt das Prinzip der Minimalen Information, dass die neue Verteilung  $p_{\text{MI}}(z)$  so aus der Menge aller Wahrscheinlichkeitsverteilungen auf  $Z$ , welche konsistent mit  $I$  sind, zu wählen ist, dass die Kullback-Leibler-Divergenz von  $p_{\text{MI}}(z)$  relativ zu  $q(z)$  minimal wird. Die Verteilung  $p_{\text{MI}}(z)$  ergibt sich somit als Lösung des folgenden Minimierungsproblems:

$$p_{\text{MI}}(z) = \arg \min_{p(z) \in \mathfrak{P}_z} \text{KD}[p(z), q(z)], \quad (3.14)$$

wobei  $\mathfrak{P}_z$  die Menge aller Wahrscheinlichkeitsverteilungen auf  $Z$  bezeichnet, welche konsistent mit der neu hinzugekommenen Information  $I$  sind.

Das Prinzip der Minimalen Information lässt sich vereinfacht<sup>1</sup> wie folgt als Verallgemeinerung des Prinzips der Maximalen Entropie veranschaulichen: Es gilt

$$\text{KD}[p(z), q(z)] = - \int_Z p(z) \log q(z) dz - \underbrace{\left( - \int_Z p(z) \log p(z) dz \right)}_{H[p(z)]}. \quad (3.15)$$

Ist vor dem Bekanntwerden von  $I$  keinerlei Information bezüglich  $z$  verfügbar, so sollte  $q(z)$  keinen Wert  $z \in Z$  favorisieren. In vielen praktischen Aufgabenstellungen lässt sich dies umsetzen, indem man annimmt, dass  $q(z) \propto 1$  gilt,  $q(z)$  also einen konstanten Wert  $c$  über ganz  $Z$  annimmt. Der erste Term

<sup>1</sup> Eine detaillierte mathematische Betrachtung des Zusammenhangs zwischen Entropie und Kullback-Leibler-Divergenz findet man z. B. in [Hob73]. Auf eine derartige, komplexere Darstellung wird in der vorliegenden Arbeit bewusst verzichtet, da es an dieser Stelle vor allem darum geht, die grundsätzlichen Prinzipien zu verdeutlichen.

auf der rechten Seite von Gleichung (3.15) entspricht dann gerade  $-\log c$ . Das bei Anwendung des Prinzips der Minimalen Information zu lösende Minimierungsproblem gemäß Gleichung (3.14) und das bei Anwendung des Prinzips der Maximalen Entropie zu lösende Maximierungsproblem gemäß Gleichung (3.11) sind dann äquivalent.

Eine Festlegung der Form  $q(z) \propto 1$  für eine nichtinformative A-Priori-Verteilung ist innerhalb der Bayes'schen Theorie oftmals auch dann zulässig und vielversprechend, wenn  $Z$  so beschaffen ist, dass keine dieser Festlegung entsprechende Gleichverteilung auf  $Z$  existiert. In diesem Fall stellt  $q(z)$  dann eine nicht normierbare sog. uneigentliche Wahrscheinlichkeitsverteilung<sup>1</sup> dar. Der Einsatz von uneigentlichen A-Priori-Verteilungen ist innerhalb der Bayes'schen Theorie i. A. problemlos möglich, wenn bei der Bayes'schen Inferenz als A-Posteriori-Verteilung wieder eine wohldefinierte (normierbare) Wahrscheinlichkeitsverteilung resultiert. Vgl. zu dieser Thematik z. B. auch [Car09], [Ntz09] oder [Rüg99].

Das Prinzip der Minimalen Information liefert in diesem Sinne eine einheitliche theoretische Rechtfertigung für das Prinzip der Maximalen Entropie sowohl für den diskreten als auch für den kontinuierlichen Fall. Dementsprechend wird im weiteren Verlauf der vorliegenden Arbeit nicht mehr zwischen diesen beiden Fällen unterschieden.

In vielen praktisch relevanten Fällen ist eine analytische Bestimmung der Maximum-Entropie-Verteilung  $p_{ME}(z)$  möglich, vgl. hierzu auch [Kap89]. Ein wichtiger Fall ist der, dass sich die bezüglich  $z$  vorliegende Information  $I$  anhand von Erwartungswerten der Form

$$E_{p(z)}[f_j(z)] = \kappa_j, \quad j \in \{1, \dots, J\}, \quad (3.16)$$

mit Funktionen  $f_1(z), \dots, f_J(z)$  ausdrücken lässt,  $J \in \mathbb{N}$ . Wie u. a. in [Cov91] und [Jay57] aufgezeigt, liefert ein Lagrange-Ansatz zur Lösung des korrespon-

---

<sup>1</sup> Uneigentliche A-Priori-Verteilungen werden in englischsprachigen Veröffentlichungen als *improper priors* bezeichnet.

dierenden Optimierungsproblems (gemäß Gleichung (3.11)) hier die geschlossene Lösung

$$p_{\text{ME}}(z) = \lambda_0 \exp \left\{ \sum_{j=1}^J \lambda_j f_j(z) \right\}, \quad (3.17)$$

wobei die Konstanten  $\lambda_0, \dots, \lambda_J$  anhand der Nebenbedingungen gemäß Gleichung (3.16) und der Nebenbedingung, dass  $p_{\text{ME}}(z)$  eine Wahrscheinlichkeitsverteilung darstellen muss, zu bestimmen sind.

Beispielsweise lassen sich durch die Festlegungen

$$f_1(z) = z, \quad (3.18)$$

$$f_j(z) = (z - E_{p(z)}[z])^j, \quad j \in \{2, \dots, J\}, \quad (3.19)$$

die ersten  $J$  zentralen Momente einer reellwertigen Zufallsvariablen  $z$  als Nebenbedingungen in der durch Gleichung (3.16) vorgegebenen Form formulieren. Im Fall  $j = 2$  entspricht dies gerade der Vorgabe von Erwartungswert  $\mu$  und Varianz  $\sigma^2$ . In diesem Fall ergibt sich als Maximum-Entropie-Verteilung die Normalverteilung  $\mathcal{N}(\mu, \sigma^2)$ .

Weiter entspricht beispielsweise eine Festlegung der Form

$$f_j(z) = \mathbf{1}_{(-\infty, \theta_j]}(z) := \begin{cases} 1, & \text{falls } z \leq \theta_j \\ 0, & \text{falls } z > \theta_j \end{cases}, \quad j \in \{1, \dots, J\}, \quad (3.20)$$

der Vorgabe bestimmter Quantile der gesuchten Wahrscheinlichkeitsverteilung, denn es gilt dann

$$E_{p(z)}[f_j(z)] = \int_{-\infty}^{\theta_j} p(z) dz. \quad (3.21)$$

Die Erhebung derartiger Quantile könnte sich z. B. anbieten, wenn die Information bezüglich  $z$  bei menschlichen Experten abgefragt wird.

Mathematische Mechanismen zur Transformation von Information in eine objektive probabilistische Repräsentation im Sinne der Degree-of-Belief-Interpretation von Wahrscheinlichkeit wie das Prinzip der Maximalen Entropie werden innerhalb der Bayes'schen Theorie üblicherweise zur Festlegung der A-Priori-Verteilung  $p(z)$  eingesetzt. Für eine Festlegung der Likelihood-Funktion  $l(d|z)$  ist dagegen oftmals das Wissen über den Prozess, durch den  $d$  in Abhängigkeit von  $z$  generiert wurde, direkt nutzbar, siehe hierzu z. B. [Edw63] oder [Wei99a]. Konkret können hierzu in der Praxis oftmals physikalische Modelle oder empirische Erhebungen genutzt werden. Ist eine derartige Festlegung der Likelihood-Funktion  $l(d|z)$  nicht möglich, so können jedoch auch hierzu entsprechende mathematische Mechanismen zur Transformation eingesetzt werden, siehe hierzu z. B. [Bey99] oder [Iuc06].

### 3.5.2 Fähigkeit zur Fusion

Wie bereits in Abschnitt 3.4.2 dargelegt wurde, stellt beim Zugrundelegen der Degree-of-Belief-Interpretation von Wahrscheinlichkeit die A-Posteriori-Verteilung  $p(z|d)$  die vollständige und vom mathematischen Standpunkt aus einzig stimmige Verkörperung des Informationsstands bezüglich  $z$  dar. Sie berücksichtigt gerade das Vorwissen und die von den Informationsquellen in Form von  $d$  gelieferte Information.

Bei der Bayes'schen Fusion entspricht in diesem Sinne dann also die A-Posteriori-Verteilung  $p(z|d)$  der vollständigen Verkörperung des Fusionsresultats und der in Abschnitt 2.3 geforderte Mechanismus, mittels dem die transformierten Fusionsanteile kombiniert werden, ist gerade durch den Satz von Bayes (vgl. Gleichung (3.5) bzw. Gleichung (3.6) sowie Gleichung (3.7)) vorgegeben:

$$p(z|d) = \frac{l(d|z)p(z)}{\int_Z l(d|z)p(z) dz} \propto l(d|z)p(z) . \quad (3.22)$$

Es lässt sich zeigen, dass diese Kombination nicht nur aus Sicht der Wahrscheinlichkeitstheorie, sondern auch aus informationstheoretischer Sicht optimal ist. Dieser Sachverhalt wird in Abschnitt 6.3 noch genauer betrachtet

und genutzt werden. An dieser Stelle sei weiter auch noch einmal hervorgehoben, dass die Bayes'sche Fusionsmethodik in Form der A-Priori-Verteilung  $p(z)$  eine explizite Eingangsgröße liefert, mittels der das hinsichtlich der Größen des Interesses vorhandene Vorwissen adäquat in den Fusionsvorgang eingespeist werden kann.

Gemäß der Kettenregel aus der Wahrscheinlichkeitstheorie gilt

$$p(d|z) = p(d_1|z)p(d_2|z,d_1) \dots p(d_S|z,d_1, \dots, d_{S-1}), \quad (3.23)$$

d. h. die von den Informationsquellen gelieferten Beiträge  $d_s$ ,  $s \in \{1, \dots, S\}$ , können i. A. nicht individuell in die Bayes'sche Fusion eingehen. Gerade beim Vorliegen heterogener Informationsquellen ist dies aber dennoch oft der Fall, nämlich dann, wenn die Informationsbeiträge vom mathematischen Standpunkt aus als stochastisch bedingt unabhängig gegeben  $z$  angesehen werden können. Diese Annahme lässt sich z. B. rechtfertigen, wenn die Informationsquellen Sensoren, welche auf unterschiedlichen physikalischen Prinzipien aufbauen, entsprechen und die ihnen überlagerten Störungen unabhängig sind (vgl. [Bey06b] und [Kle04]).

Beim Vorliegen der stochastisch bedingten Unabhängigkeit vereinfacht sich Gleichung (3.23) zu

$$p(d|z) = \prod_{s=1}^S p(d_s|z), \quad (3.24)$$

d. h. bei festgehaltenem Wert von  $z$  (ihrer Ursache) enthalten die Informationsbeiträge keine Information übereinander. Gemäß Gleichung (3.24) ergibt sich in diesem Fall die Likelihood-Funktion  $l(d|z)$  als Produkt individueller Likelihood-Funktionen  $l(d_s|z)$  für die einzelnen Informationsquellen,  $s \in \{1, \dots, S\}$ , was bereits eine deutliche Vereinfachung im Hinblick auf ihre Bestimmung darstellt.

Außerdem lässt sich in diesem Fall durch rekursive Anwendung des durch den Satz von Bayes vorgegebenen Fusionschemas (vgl. Gleichung (3.22)) un-



kompliziert ein sequentielles Fusionsschema realisieren:

$$p(z|d_1, \dots, d_{s+1}) \propto l(d_{s+1}|z)p(z|d_1, \dots, d_s) \quad \forall s \in \{1, \dots, S-1\}. \quad (3.25)$$

Die von den Informationsquellen gelieferten Beiträge tragen hier sukzessive zum Informationsstand bezüglich  $z$  bei. Letztlich stellt dabei die A-Posteriori-Verteilung eines bestimmten Fusionsschritts die A-Priori-Verteilung für den nächsten Schritt dar. Es sei angemerkt, dass die Reihenfolge, in der die Informationsbeiträge in diese schrittweise Fusion eingehen, keine Rolle spielt.

Prinzipiell lässt sich durch die Umsetzung eines derartigen sequentiellen Fusionsschemas beim Vorliegen von verschiedenen, u. a. auch nach und nach verfügbar werdenden Informationsquellen insbesondere auch solange Information ansammeln, bis der Informationsstand in Bezug auf die durch  $z$  verkörperten Größen des Interesses eine ausreichend gute Entscheidungsgrundlage darstellt. Es besteht also erst einmal kein Zwang, nach einer vorab fixierten Anzahl von Fusionsschritten eine „harte“ Entscheidung bezüglich des „wahren“ Werts von  $z$  zu treffen.

Bayes'sche Verfahren, welche sich auf die beschriebene bedingte Unabhängigkeitsannahme stützen, werden in der Fachliteratur oft als naive Bayes'sche Verfahren bezeichnet. Sie liefern gerade bei Aufgabenstellungen der Klassifikation oft auch gute Ergebnisse, obwohl die bedingte Unabhängigkeitsannahme tatsächlich nicht erfüllt ist (siehe hierzu [Dom97, Ekd06, Han01, Ris01]). Da bei einer Klassifikation nur bestimmte Aspekte der A-Posteriori-Verteilung relevant sind (siehe hierzu auch Abschnitt 3.5.3), impliziert diese Feststellung aber nicht unbedingt, dass in diesen Fällen die beim Zugrundelegen einer (nicht geltenden) bedingten Unabhängigkeitsannahme resultierende A-Posteriori-Verteilung insgesamt eine gute Näherung für die tatsächliche A-Posteriori-Verteilung darstellt.

In praktischen Aufgabenstellungen der Informationsfusion müssen möglicherweise für eine realistische bzw. praktikable Modellbildung neben  $d$  und  $z$  (und dem Vorwissen) noch weitere Einflussgrößen explizit berücksichtigt werden. Die Bayes'sche Fusionsmethodik zeichnet sich in diesen Fällen durch einen zwanglosen Umgang mit den entsprechenden sog. Nuisance-

Parametern (vgl. [Ber04]) aus. Sie können z. B. als weitere Zufallsgrößen in die Berechnung der A-Posteriori-Verteilung miteinbezogen werden und dann nachgelagert mittels Mechanismen zur Fokussierung wieder herausgerechnet werden. Derartige Mechanismen werden im nachfolgenden Abschnitt genauer aufgegriffen.

### 3.5.3 Fähigkeit zur Fokussierung

Die Bayes'sche Theorie stellt unterschiedliche Mechanismen zur Verfügung, mittels derer sich die in Form der A-Posteriori-Verteilung  $p(z|d)$  vorliegende umfassende Informationsverkörperung des Fusionsresultats bei Bedarf zielgerichtet auf spezielle Einzelfragestellungen konzentrieren lässt. Die Fokussierung geht dabei aber natürlich (unvermeidbar) mit einem Informationsverlust einher, d. h. es ist wichtig, dass dabei möglichst keine Information verloren geht, die in Bezug auf die jeweilige Einzelfragestellung relevant ist.

Sind im Hinblick auf eine Einzelfragestellung nur bestimmte der im Vektor  $z$  zusammengefassten Größen des Interesses  $z_n$ ,  $n \in \{1, \dots, N\}$ , relevant, so nehmen die übrigen Komponenten von  $z$  die Rolle von Nuisance-Parametern ein. Eine Reduktion der Dimension der A-Posteriori-Verteilung lässt sich in diesem Fall durch Marginalisierung umsetzen.

Entsprechen z. B. in einer konkreten Aufgabenstellung  $(z_1, z_2)$  der Position,  $z_3$  dem Typ und  $z_4$  der (als kontinuierliche Größe modellierten) Bewegungsrichtung eines Objekts in einer Szene und ist in einer speziellen Einzelfragestellung nur die Position von Interesse, so lässt sich die A-Posteriori-Verteilung  $p(z|d) = p(z_1, z_2, z_3, z_4|d)$  wie folgt durch Bildung der entsprechenden Marginalverteilung  $p(z_1, z_2|d)$  auf diesen Aspekt konzentrieren:

$$p(z_1, z_2|d) = \sum_{z_3 \in Z_3} \int_{Z_4} p(z_1, z_2, z_3, z_4|d) dz_4. \quad (3.26)$$

Einen weiteren Mechanismus zur Fokussierung stellt die Berechnung spezifischer A-Posteriori-Erwartungswerte  $E_{p(z|d)} [f(z)]$  dar. Dabei ist  $f(z)$  eine für die spezielle Einzelfragestellung passend gewählte Funktion. Hängt diese

nicht von allen Komponenten von  $z$  ab, so beinhaltet die Erwartungswertbildung auch eine Marginalisierung. Im vorgenannten Beispiel resultiert z. B. aus der Festlegung  $f(z) = z_4$  gerade die erwartete Bewegungsrichtung des Objekts in Form des marginalen A-Posteriori-Erwartungswerts  $E_{p(z|d)} [z_4]$ :

$$E_{p(z|d)} [z_4] = \int_{Z_1} \int_{Z_2} \sum_{z_3 \in Z_3} \int_{Z_4} z_4 p(z_1, z_2, z_3, z_4 | d) dz_1 dz_2 dz_4 \quad (3.27)$$

$$= \int_{Z_4} z_4 p(z_4 | d) dz_4 . \quad (3.28)$$

Mit dem dargestellten Mechanismus zur Fokussierung durch Erwartungswertbildung eng verwandt sind die in Abschnitt 3.2 bereits kurz erwähnten Verfahren zur Schätzung des „wahren“ Werts von  $z$ . So kann z. B. mittels einer Punktschätzung ein Übergang von der A-Posteriori-Verteilung  $p(z|d)$  zu einem gemäß eines bestimmten Bewertungskriteriums optimalen Schätzwert  $\hat{z}$  für den „wahren“ Wert von  $z$  realisiert werden. Das der Auswahl des Schätzwerts zugrundegelegte Bewertungskriterium kann dabei z. B. in Form einer Verlustfunktion  $v(e, z)$  eingebracht werden. Diese setzt die möglichen Schätzwerte  $e$  und den (unbekannten) „wahren“ Wert von  $z$  durch die Zuweisung von dem mit der Auswahl verbundenen Verlust in Beziehung. Die Bestimmung eines optimalen Schätzwerts erfolgt dann durch die Minimierung des A-Posteriori-Erwartungswerts  $E_{p(z|d)} [v(e, z)]$  der Verlustfunktion. Dieser stellt eine Funktion von  $e$  dar und gibt unter Berücksichtigung der bezüglich  $z$  a posteriori vorhandenen Information den mit der Auswahl erwartungsgemäß verbundenen Verlust an.

Grundsätzlich kann auch eine derartige Schätzung mit einer Marginalisierung gekoppelt sein. Im Beispiel oben würde beim Zugrundelegen der auf den Fahrzeugtyp  $z_3$  konzentrierten 0-1-Verlustfunktion

$$v(e, z_3) = \begin{cases} 1 & \text{falls } e \neq z_3 \\ 0 & \text{falls } e = z_3 \end{cases} \quad (3.29)$$

eine Punktschätzung gerade den Maximum-A-Posteriori-Schätzwert der Mar-

ginalverteilung  $p(z_3|d)$  liefern, denn es gilt:

$$\arg \min_{e \in Z_3} E_{p(z|d)} [v(e, z_3)] \quad (3.30)$$

$$= \arg \min_{e \in Z_3} \int_{Z_1} \int_{Z_2} \sum_{z_3 \in Z_3} \int_{Z_4} v(e, z_3) p(z_1, z_2, z_3, z_4 | d) dz_1 dz_2 dz_4 \quad (3.31)$$

$$= \arg \min_{e \in Z_3} \sum_{z_3 \in Z_3} v(e, z_3) p(z_3 | d) \quad (3.32)$$

$$= \arg \min_{e \in Z_3} (1 - p(e|d)) \quad (3.33)$$

$$= \arg \max_{e \in Z_3} p(e|d) . \quad (3.34)$$

Eine den vorliegenden Abschnitt in dieser Hinsicht ergänzende Darstellung von Verfahren zu Schätzung des „wahren“ Werts von  $z$  auf Basis unterschiedlicher Verlustfunktionen wurde in [Bey08] mit Bezug auf den Anwendungskontext der Bildfusion veröffentlicht.

## 4 Durchführung der Fusion

### 4.1 Herausforderung im Hinblick auf die Berechnungen

Aus Kapitel 3 geht hervor, dass es sich bei der Bayes'schen Fusionsmethodik grundsätzlich gesehen um eine geeignete und zudem mathematisch äußerst fundierte Methodik zur Fusion heterogener Informationsquellen handelt, welche sich zusätzlich noch durch weitere Vorzüge auszeichnet. Hervorgehoben werden sollen an dieser Stelle insbesondere noch einmal ihre Durchgängigkeit, ihre modulare Erweiterbarkeit und die Möglichkeit, Vorwissen adäquat in die Fusion einzubringen. Auch entspricht die zur Repräsentation des jeweiligen Informationsstands verwendete probabilistische Modellierung im Sinne der Degree-of-Belief-Interpretation von Wahrscheinlichkeit oftmals dem intuitiven Verständnis auch nicht speziell geschulter Personen. In der Praxis kann die Bayes'sche Fusionsmethodik jedoch schnell an ihre Grenzen stoßen, indem sie einen nicht tragbaren Aufwand verursacht.

Auch bei einer Modellierung von A-Priori-Verteilung  $p(z)$  und Likelihood-Funktion  $l(d|z)$  unter Zugrundelegung nicht allzu komplizierter parametrischer Verteilungsannahmen lässt sich die A-Posteriori-Verteilung  $p(z|d)$  in der Regel nicht analytisch bestimmen. Liegt z. B. als A-Priori-Verteilung eine Cauchy-Verteilung  $\mathcal{C}(\alpha, \lambda)$  mit bekannten (Hyper)parametern  $\alpha \in \mathbb{R}$  und  $\lambda > 0$  vor und korrespondiert die Likelihood-Funktion mit einer Normalverteilung  $\mathcal{N}(z, \sigma^2)$  mit (unbekanntem) Erwartungswert  $z \in \mathbb{R}$  und bekannter Varianz  $\sigma^2 > 0$ , so ist eine analytische Bestimmung der A-Posteriori-Verteilung nicht möglich (vgl. [Ber85]).

Schon beim Vorliegen eines eindimensionalen Wertebereichs  $Z = Z_1$  der Größen des Interesses kann die praktische Anwendung der Bayes'schen Fusionsmethodik also in diesem Sinne Probleme bereiten. In praktisch relevanten Aufgabenstellungen können nun aber sowohl die Dimension  $N \in \mathbb{N}$  des Wertebereichs  $Z$  der Größen des Interesses  $z = (z_1, \dots, z_N)$  als auch die individuellen Wertebereiche  $Z_n$  der einzelnen Komponenten  $z_n$ ,  $n \in \{1, \dots, N\}$ , recht umfangreich sein. Die vorgenannte Problematik verschärft sich dann noch weiter. Gilt  $Z = Z_1 \times \dots \times Z_N$  und besitzt  $Z_n$  (ggf. nach Diskretisierung) die Kardinalität  $|Z_n|$ , so liegt der Aufwand zur numerischen („punktweisen“) Berechnung der A-Posteriori-Verteilung  $p(z|d)$  in

$$O(|Z|) = O\left(\prod_{n=1}^N |Z_n|\right) = O(\eta^N) \quad (4.1)$$

wobei

$$\eta := \sqrt[N]{\prod_{n=1}^N |Z_n|} \quad (4.2)$$

das geometrische Mittel von  $Z_1, \dots, Z_n$  notiert. Gemäß Gleichung (4.1) wächst der Aufwand zur Berechnung der A-Posteriori-Verteilung  $p(z|d)$  also i. A. exponentiell mit der mittleren Dimension des Wertebereichs  $Z$  der Größen des Interesses (vgl. [San09c]). Dieser Sachverhalt ist darauf zurückzuführen, dass sich die Berechnung der A-Posteriori-Verteilung stets auf ganz  $Z$  bezieht, in diesem Sinne also globalen Charakter hat.

Gemäß dem Bayes'schen Formalismus erfolgt die Berechnung der A-Posteriori-Verteilung  $p(z|d)$  auch immer komplett – unabhängig davon, ob in einer konkreten Aufgabenstellung tatsächlich der vollständige a posteriori verfügbare Informationsstand bezüglich des „wahren“ Werts von  $z$  benötigt wird oder nicht. Beispielsweise wäre wie in Abschnitt 3.5.3 skizziert auch ein Schätzwert  $\hat{z}$  für den „wahren“ Wert der Größen des Interesses auf Basis der gesamten A-Posteriori-Verteilung  $p(z|d)$  zu bestimmen.

Die in der vorliegenden Arbeit nachfolgend formulierten lokalen Ansätze zur Bayes'schen Fusion adressieren diese Problematik grob gesagt durch die Ermöglichung der Einnahme eines lokalen Standpunkts. Das Ziel besteht darin, die eigentliche Bayes'sche Fusion möglichst nur dort lokal in  $Z$  vorzunehmen, wo sich auch tatsächlich Aufgabenrelevantes abspielt. Bevor dies ausgeführt wird, werden zuvor im nachfolgenden Abschnitt im Rahmen eines Exkurses in der Fachliteratur vorhandene Ansätze zur Adressierung der dargestellten Problematik kurz vorgestellt und diskutiert.

## 4.2 Ansätze in der Fachliteratur

Im vorliegenden Abschnitt wird ein komprimierter Überblick über die wesentlichen in der Fachliteratur vorhandenen Ansätze zur Adressierung der zuvor dargestellten Problematik im Hinblick auf die Berechnung der A-Posteriori-Verteilung  $p(z|d)$  bzw. von Kenngrößen derselben gegeben. Der Fokus dieser Aufbereitung und Diskussion des Stands der Wissenschaft und Technik liegt dabei auf den Ansätzen an sich, nicht auf den speziellen, i. A. recht vielfältigen Aufgabenstellungen, in denen sie eingesetzt werden. Um die Darstellung überschaubar zu halten, wird wo möglich auch auf die Angabe und Herleitung umfangreicher mathematischer Details bewusst verzichtet und statt dessen auf entsprechende Referenzen verwiesen. Die nachfolgenden Unterabschnitte stellen eine überarbeitete und deutlich erweiterte Version der hierzu zuvor in [San13] veröffentlichten Anteile dar.

### 4.2.1 Konjugierte Verteilungsfamilien

Wie bereits in Abschnitt 4.1 dargestellt wurde, lässt sich auch bei einer Modellierung der in die Bayes'sche Fusion involvierten Größen unter Zugrundelegung nicht allzu komplizierter parametrischer Verteilungsannahmen die A-Posteriori-Verteilung  $p(z|d)$  in der Regel nicht analytisch bestimmen. Es gibt jedoch Ausnahmen. Dem Prinzip konjugierter Verteilungsfamilien liegt der Gedanke zugrunde, diese gezielt aufzugreifen. Die hierfür notwendigen, recht

komplexen mathematischen Grundlagen werden in der Standardliteratur zur Bayes'schen Theorie adressiert, siehe z. B. [Ber04], [Ber85] oder [Rob01].

**Definition 4.1.** *Eine Menge von Wahrscheinlichkeitsverteilungen  $\mathfrak{P}_z$  ist konjugiert zu einer Likelihood-Funktion  $l(d|z)$ , wenn aus  $p(z) \in \mathfrak{P}_z$  folgt, dass auch  $p(z|d) \in \mathfrak{P}_z$  gilt.*

Ein triviales, allerdings nicht hilfreiches Beispiel ergibt sich (für jede beliebige Likelihood-Funktion  $l(d|z)$ ), wenn man  $\mathfrak{P}_z$  als die Menge aller Wahrscheinlichkeitsverteilungen auf  $Z$  definiert.

Rechnerisch interessant sind vor allem konjugierte Verteilungsfamilien, welche sich durch eine überschaubare Anzahl an Parametern beschreiben lassen. Beschränkt man sich bei der Modellierung der in die Bayes'sche Fusion involvierten Größen nämlich dann auf einen Fall, in dem das Prinzip konjugierter Verteilungsfamilien anwendbar ist, so lässt sich die Berechnung der A-Posteriori-Verteilung wesentlich vereinfachen. Werden Likelihood-Funktion  $l(d|z)$  und A-Priori-Verteilung  $p(z)$  so gewählt, dass  $p(z)$  zu einer entsprechenden, zu  $l(d|z)$  konjugierten Verteilungsfamilie  $\mathfrak{P}_z$  gehört, so weiß man, dass auch die A-Posteriori-Verteilung  $p(z|d)$  wieder in  $\mathfrak{P}_z$  enthalten ist. Lassen sich weiter die in  $\mathfrak{P}_z$  enthaltenen Wahrscheinlichkeitsverteilungen durch eine überschaubare Anzahl an Parametern beschreiben, so lässt sich die Berechnung der A-Posteriori-Verteilung im Erfolgsfall auf die analytische Berechnung dieser Parameter reduzieren. Der Aufwand, den eine numerische („punktweise“) Berechnung der A-Posteriori-Verteilung bedeutet hätte, fällt dann also nicht an und die A-Posteriori-Verteilung lässt sich zudem auch exakt (in parametrischer Form) angeben.

Den in der Praxis wichtigsten Fall hinsichtlich der Anwendung des Prinzips konjugierter Verteilungsfamilien stellt der von Normalverteilungsannahmen dar. Gelten beispielsweise  $z \in \mathbb{R}$ ,  $d \in \mathbb{R}$  und liegt als Likelihood-Funktion eine Normalverteilung  $\mathcal{N}(z, \sigma^2)$  mit (unbekanntem) Erwartungswert  $z$  und bekannter Varianz  $\sigma^2 \in \mathbb{R}_{>0}$  vor, so bildet die Menge aller Normalverteilungen  $\{\mathcal{N}(\mu, \tau^2) | \mu \in \mathbb{R}, \tau^2 \in \mathbb{R}_{>0}\}$  eine zu dieser konjugierte Verteilungsfamilie. Liegt als A-Priori-Verteilung die  $\mathcal{N}(\mu, \tau^2)$  Normalverteilung vor, so ergibt sich



als A-Posteriori-Verteilung gerade die  $\mathcal{N}\left(\frac{\sigma^2\mu+\tau^2d}{\sigma^2+\tau^2}, \frac{\sigma^2\tau^2}{\sigma^2+\tau^2}\right)$  Normalverteilung. In diesem Fall vereinfacht sich die Berechnung der A-Posteriori-Verteilung also signifikant, es sind lediglich ihr Erwartungswert und ihre Varianz durch Verknüpfung der entsprechenden Parameter von A-Priori-Verteilung und Likelihood-Funktion zu berechnen. Vergleichbare Aussagen gelten auch für den multivariaten Fall. Vgl. z. B. [Car09] und [Rob01] für weitere Details zu diesen Sachverhalten.

Die dem Prinzip konjugierter Verteilungsfamilien zugrundeliegende Theorie adressiert letztlich die Frage, wann sich die Berechnung der A-Posteriori-Verteilung  $p(z|d)$  wie beschrieben auf die Verknüpfung einer überschaubaren Anzahl an Verteilungsparametern vereinfachen lässt. Dieser Theorie folgend ist der Einsatz des Prinzips konjugierter Verteilungsfamilien besonders dann vielversprechend, wenn man die Likelihood-Funktion so wählen kann, dass sie mit einem Element der sog. natürlichen Exponentialfamilie<sup>1</sup> korrespondiert. In diesem Fall lässt sich nämlich eine (nicht triviale) zugehörige konjugierte Verteilungsfamilie angeben und die Parameter der A-Posteriori-Verteilung, welche resultiert, wenn man ein Element dieser Verteilungsfamilie als A-Priori-Verteilung wählt, lassen sich analytisch ableiten. Siehe z. B. [Rob01] für eine formal mathematische Darstellung dieser Sachverhalte, die zugehörigen mathematischen Beweise und weitere Beispiele für solche Kombinationen aus Likelihood-Funktion und konjugierter Verteilungsfamilie.

Grundsätzlich bietet das Prinzip konjugierter Verteilungsfamilien in vielen praktischen Aufgabenstellungen einen Weg, um den hohen Aufwand, den eine anderweitige Berechnung der A-Posteriori-Verteilung  $p(z|d)$  bedeuten würde, zu umgehen. Man darf dabei allerdings nicht vergessen, dass das Hauptargument, das für die Nutzung konjugierter Verteilungsfamilien spricht, oftmals rein technisch ist.

Die Sinnhaftigkeit des dahinterliegenden Konzepts kann zwar allgemein in gewissen Sinne damit begründet werden, dass die Information in  $d$  bezüglich  $z$  stets als begrenzt angenommen werden sollte, ein Übergang von der

---

<sup>1</sup> Diese ist nicht zu verwechseln mit der vom Namen her ähnlichen Exponentialverteilung.

A-Priori-Verteilung  $p(z)$  zur A-Posteriori-Verteilung  $p(z|d)$  demnach nur die Parameter der Verteilung von  $z$ , nicht aber ihre vollständige Struktur verändern sollte (vgl. [Rai61]). Allerdings stellen die bei der Festlegung von Likelihood-Funktion und A-Priori-Verteilung gemachten Annahmen oftmals dennoch Einschränkungen dar, die nicht notwendigerweise der Realität entsprechen. So wird z. B. die A-Priori-Verteilung aus der konjugierten Verteilungsfamilie  $\mathfrak{P}_z$  oft so ausgewählt, dass sie die tatsächliche A-Priori-Verteilung, welche nicht in  $\mathfrak{P}_z$  enthalten ist, möglichst gut annähert. Daraus folgt dann allerdings nicht zwangsläufig, dass die resultierende A-Posteriori-Verteilung die tatsächliche A-Posteriori-Verteilung, welche bei Verwendung der tatsächlichen A-Priori-Verteilung resultieren würde, auch entsprechend gut annähert (vgl. hierzu [Ber85] und [Rob01]). Dieser Sachverhalt muss dann also ggf. separat validiert werden, z. B. durch eine Bayes'sche Robustheitsanalyse (vgl. hierzu auch [Ins00]), welche selbst wiederum mathematisch gesehen recht komplex sein kann.

Das Prinzip konjugierter Verteilungsfamilien lässt sich auf Mixturen erweitern (vgl. z. B. [Rob01]). Damit ist gemeint, dass eine Menge von Mixturen von Wahrscheinlichkeitsverteilungen, welche in einer konjugierten Verteilungsfamilie enthalten sind, selbst wieder eine konjugierte Verteilungsfamilie bildet (für die entsprechende Likelihood-Funktion). Dies liefert einen Ansatzpunkt, um die tatsächliche A-Priori-Verteilung theoretisch gesehen beliebig genau zu approximieren (vgl. z. B. [Rob01]). Beispielsweise kann man gerade auch mittels Mixturen von Normalverteilungen jede beliebige Wahrscheinlichkeitsverteilung beliebig gut annähern (vgl. z. B. [Rob01] und [Hub11]). Voraussetzung ist dabei jedoch, dass man die Anzahl der in den Mixturen enthaltenen Komponenten nicht beschränkt.

Da die erforderlichen Rechenoperationen, welche man bei Anwendung des Prinzips konjugierter Verteilungsfamilien durchführen muss, um die Parameter der A-Posteriori-Verteilung zu berechnen, von der Anzahl der in den Mixturen enthaltenen Komponenten abhängen, bieten konjugierte Mixturen mit recht vielen Komponenten schnell keine ausreichenden rechnerischen Vorteile mehr. Als Konsequenz besteht i. A. ein Trade-Off zwischen der Genauigkeit der Approximation (der A-Priori-Verteilung) und dem (für die Berechnung der

A-Posteriori-Verteilung) resultierenden Rechenaufwand, welcher sich oftmals nicht so einfach lösen lässt. Dies gilt gerade auch bei sukzessiver Anwendung des Satzes von Bayes im Rahmen einer sequentiellen Fusion, da in diesem Fall die Anzahl der Parameter der resultierenden A-Posteriori-Verteilungen i. A. deutlich mit der Anzahl der durchgeführten Fusionschritte wächst. Ein Beispiel aus der Fachliteratur, um wiederum speziell diese Problematik zu adressieren, ist der in [Hub11] vorgestellte Ansatz einer dynamischen Reduktion von Mixtur-Komponenten.

### 4.2.2 Deterministische Approximation

Eine deterministische Approximation der A-Posteriori-Verteilung  $p(z|d)$  bzw. von Kenngrößen derselben kann mittels sog. analytischer Approximationsverfahren durchgeführt werden. Diese basieren auf bestimmten Annahmen hinsichtlich der Struktur der A-Posteriori-Verteilung. In der Regel wird konkret die Annahme gemacht, dass die A-Posteriori-Verteilung sich ausreichend gut durch eine Normalverteilung approximieren lässt und es wird eine entsprechende Approximation vorgenommen.

In [Car09] wird dargestellt, dass sich eine Approximation der A-Posteriori-Verteilung durch eine Normalverteilung unter bestimmten Bedingungen in Aufgabenstellungen, in denen sehr viele, unabhängig und identisch verteilte Beobachtungen  $d_s$  vorliegen, asymptotisch rechtfertigen lässt. Konkret lässt sich dann zeigen, dass die A-Posteriori-Verteilung durch eine Normalverteilung um ihren Modalwert angenähert werden kann und dass sich deren Kovarianzmatrix durch Auswertung der Hesse-Matrix von  $\log p(z|d)$  im Modalwert berechnen lässt (vgl. [Car09]). Bei der sog. Normal-Approximation wird die entsprechende Approximation dann mittels einer Taylorentwicklung bis zur zweiten Ordnung, welche angewandt wird auf  $\log(p(d|z)p(z))$ , vorgenommen (vgl. [Car09]). Es sei noch angemerkt, dass sich eine Verbesserung der Approximationsgüte durch die sog. Laplace-Approximation, welche z. B. in [Ber04] und [Gam06] recht ausführlich beschrieben ist, erzielen lässt.

Falls bei der Anwendung analytischer Approximationsverfahren die hinsichtlich der Struktur der A-Posteriori-Verteilung gemachten Annahmen gut zu-

treffen, die A-Posteriori-Verteilung strukturell also tatsächlich recht ähnlich zu (in der Regel) einer Normalverteilung ist, so lassen sich mit diesen sehr gute Ergebnisse erzielen (vgl. z. B. [Ber04]). U. a. aufgrund der zunehmenden Komplexität der zur Durchführung der Approximation notwendigen Berechnungen und den damit verbundenen Schwierigkeiten im Hinblick auf ihre praktische Durchführung ist die Anwendbarkeit analytischer Approximationsmethoden jedoch begrenzt auf Aufgabenstellungen, in denen die Dimension von  $Z$  eher gering ist (vgl. z. B. [Ber04], [Car09] oder [Gam06]).

Falls keine analytische Approximation möglich ist, so kann man versuchen, alternativ auf numerische Approximationsverfahren, d. h. auf entsprechende Verfahren aus der numerischen Mathematik, zurückzugreifen. Deren Anwendbarkeit wird im Bayes'schen Kontext allerdings i. A. auch als äußerst begrenzt gesehen. Der Grund ist, dass sie letztlich dem sog. Fluch der Dimensionalität unterliegen. Dies bedeutet, dass ihre Approximationsgüte mit zunehmender Dimension von  $Z$  drastisch abnimmt (vgl. [Rob01]). Versucht man, eine bestimmte Approximationsgüte sicherstellen, indem man die Anzahl der numerischen Funktionsauswertungen erhöht, so führt dies dazu, dass diese exponentiell mit der Anzahl der Dimensionen wächst (vgl. [Car09]).

## 4.2.3 Nichtdeterministische Approximation

### 4.2.3.1 Grundprinzip

Angenommen, man kann auf Basis einer unabhängig und identisch verteilten Folge von Zufallsvariablen eine Folge von Samples  $z^{(1)}, z^{(2)}, \dots$  der A-Posteriori-Verteilung  $p(z|d)$  erzeugen. Gemäß dem sog. starken Gesetz der großen Zahlen aus der Wahrscheinlichkeitstheorie (vgl. z. B. [Geo07]) gilt dann fast sicher<sup>1</sup>

$$\lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=1}^m f(z^{(i)}) = E_{p(z|d)}[f(z)]. \quad (4.3)$$

---

<sup>1</sup> Eine Aussage in der Wahrscheinlichkeitstheorie gilt fast sicher, wenn die Wahrscheinlichkeit dafür, dass sie zutrifft, Eins beträgt.

Gleichung (4.3) liefert eine Basis, um (bei entsprechender Wahl der Funktion  $f(z)$ ) Kenngrößen der A-Posteriori-Verteilung  $p(z|d)$  mittels einer ausreichenden Anzahl  $m \in \mathbb{N}$  von Samples anzunähern (vgl. z. B. [Car09]). Beispielsweise lassen sich so die (theoretischen) Momente und Quantile der A-Posteriori-Verteilung durch die empirischen, auf Basis der Samples berechneten Momente und Quantile approximieren. Insbesondere kann auch das durch Auswertung der Samples insgesamt resultierende Histogramm als Approximation der gesamten A-Posteriori-Verteilung aufgefasst werden. Da Gleichung (4.3) fast sicher gilt, erfolgt die Approximation in der Regel korrekt.

Im Hinblick auf die Verfahren, welche für die nichtdeterministische Approximation der A-Posteriori-Verteilung  $p(z|d)$  bzw. von Kenngrößen derselben zur Verfügung stehen, muss zwischen nicht-iterativen und iterativen Verfahren unterschieden werden. Nicht-iterative Verfahren erzeugen eine Menge von  $m \in \mathbb{N}$  Samples, wobei  $m$  im Voraus geeignet festzulegen ist. Bei iterativen Verfahren wird  $m$  dagegen nicht im Voraus fixiert.

Die Anwendbarkeit nicht-iterativer Verfahren ist gerade in Fällen, in denen  $Z$  eine recht umfangreiche Dimension besitzt, i. A. als äußerst limitiert anzusehen. Im nachfolgenden Abschnitt wird dieser Sachverhalt exemplarisch anhand zweier konkreter nicht-iterativer Verfahren genauer erläutert.

#### 4.2.3.2 Nicht-iterative Verfahren

Gängige Beispiele in der Standardliteratur für nicht-iterative Verfahren sind Importance Sampling, Weighted Resampling sowie Rejection Sampling. Im Folgenden wird exemplarisch genauer auf Importance Sampling und Weighted Resampling eingegangen. Einen Einblick in das Rejection Sampling bieten z. B. [Car09] und [Gam06]. Ergänzend sei noch auf die Referenz [Cap07] verwiesen, in welcher die (nachfolgend nicht adressierte) sequentielle Anwendung derartiger nicht-iterativer Verfahren ausführlicher betrachtet wird.

Beim Importance Sampling und beim Weighted Resampling wird eine Menge  $\{z^{(1)}, \dots, z^{(m)}\}$  von Samples mittels einer Hilfsverteilung  $h(z)$  auf  $Z$ , welche

zumindest prinzipiell gesehen weitgehend beliebig gewählt werden kann, erzeugt. Weiter werden Gewichte  $\omega_i$  wie folgt festlegt:

$$\omega_i := \frac{p(d|z^{(i)})p(z^{(i)})}{h(z^{(i)})}, i \in \{1, \dots, m\}. \quad (4.4)$$

Beim Importance Sampling werden die mittels der Hilfsverteilung  $h(z)$  erzeugten Samples dann direkt verwendet, um die nachfolgende Approximation umzusetzen (vgl. [Mar07]):

$$E_{p(z|d)}[f(z)] \approx \frac{\sum_{i=1}^m f(z^{(i)})\omega_i}{\sum_{i=1}^m \omega_i}. \quad (4.5)$$

Die Tatsache, dass sich dadurch bei ausreichend großer Wahl von  $m$  in der Regel eine korrekte Approximation ergibt, lässt sich im Wesentlichen anhand der folgenden Beziehungen erkennen:

$$E_{p(z|d)}[f(z)] = E_{h(z)}\left[\frac{f(z)p(z|d)}{h(z)}\right] \quad (4.6)$$

$$= \frac{1}{p(d)} E_{h(z)}\left[\frac{f(z)p(d|z)p(z)}{h(z)}\right], \quad (4.7)$$

$$p(d) = E_{h(z)}\left[\frac{p(d|z)p(z)}{h(z)}\right]. \quad (4.8)$$

Anders als beim Importance Sampling wird beim Weighted Resampling Gleichung (4.5) nicht direkt anhand der mittels  $h(z)$  erzeugten Samples ausgewertet. Statt dessen erfolgt noch ein zweiter Sampling-Schritt (vgl. z. B. [Gam06]). Nachdem die Gewichte  $\omega_i$  auf Basis der mittels  $h(z)$  erzeugten Samples  $z^{(i)}$  gemäß Gleichung (4.4) berechnet wurden, werden beim Weighted Resampling neue Samples aus der Wertemenge  $\{z^{(1)}, \dots, z^{(m)}\}$  erzeugt. Dabei wird  $z^{(i)}$  mit Wahrscheinlichkeit  $\omega_i / \sum_{k=1}^m \omega_k$  gewählt. Die so erzeugten Samples entsprechen dann näherungsweise Samples der A-Posteriori-Verteilung  $p(z|d)$  und

können direkt zur näherungsweisen Approximation von  $E_{p(z|d)}[f(z)]$  gemäß Gleichung (4.3) verwendet werden.

Der Hintergrund des beim Weighted Resampling, das auch als Sampling/Importance Resampling (SIR) bezeichnet wird, durchgeführten zusätzlichen Sampling-Schritts ist wie folgt: Bildet die Hilfsverteilung  $h(z)$  die A-Posteriori-Verteilung  $p(z|d)$  nicht gut ab, so kann es sein, dass sie recht viele Samples liefert, die in Bereichen von  $Z$  liegen, in denen der Wert von  $p(z|d)$  sehr gering ist und nur wenige Samples, für die dies nicht gilt. In diesem Fall kann dann die in Gleichung (4.5) vorgenommene Gewichtung dazu führen, dass im Wesentlichen nur sehr wenige Samples mit recht hohem Gewicht zur Approximation von  $E_{p(z|d)}[f(z)]$  beitragen und als Ergebnis eine Approximation von schlechter Qualität resultiert (vgl. z. B. auch [Car09]).

Beim Weighted Resampling wird der dargestellten Problematik, die entstehen kann, wenn  $h(z)$  die A-Posteriori-Verteilung  $p(z|d)$  nicht gut abbildet, also entgegengewirkt. Da der Resampling-Schritt dabei allein auf Basis der zuvor mittels  $h(z)$  erzeugten Samples und unter Betrachtung von deren Gewichtung relativ zueinander erfolgt, bietet das Weighted Resampling allerdings z. B. in Fällen, in denen alle mittels  $h(z)$  erzeugten Samples in Bereichen von  $Z$  liegen, in denen der Wert von  $p(z|d)$  gering ist, auch keine Lösung, um eine schlechte Approximation von  $E_{p(z|d)}[f(z)]$  zu verhindern (vgl. z. B. auch [Car09]).

Die vorgestellten Beispiele des Importance Sampling und des Weighted Resampling verdeutlichen, dass die Schwierigkeit bei nicht-iterativen Verfahren in der Regel darin besteht, eine Hilfsverteilung zu bestimmen, welche einerseits ein einfaches Sampling ermöglicht, andererseits bereits die A-Posteriori-Verteilung  $p(z|d)$  gut genug abbildet. Gelingt dies nicht, so kann die resultierende Approximation von  $E_{p(z|d)}[f(z)]$  von unzureichender Qualität sein. Ein Ausweg könnte darin bestehen, die Anzahl  $m$  der Samples deutlich höher anzusetzen, allerdings wird dies in der Praxis kaum wünschenswert bzw. unter Umständen auch überhaupt nicht möglich sein. Die beschriebene Problematik verschärft sich mit zunehmender Dimension von  $Z$ . Aus diesen Grund wird in hochdimensionalen Problemstellungen heute üblicherweise auf iterative Verfahren, konkret sog. Markov Chain Monte Carlo (MCMC) Verfahren, zurückgegriffen (vgl. hierzu z. B. auch [Car09] und [Gam06]).

Abschließend sei noch angemerkt, dass bei der Umsetzung nicht-iterativer Verfahren oftmals auf die A-Priori-Verteilung  $p(z)$  als Hilfsverteilung zurückgegriffen wird (vgl. z.B. die Darstellung in [Gam06]). In diesem Fall reduziert sich Gleichung (4.4) gerade auf die Festlegung  $\omega_i = p(d|z^{(i)})$ . Die vorab für das Importance Sampling und das Weighted Resampling dargestellten Schwierigkeiten resultieren in diesem Fall also grob gesagt, wenn sich A-Priori-Verteilung und Likelihood-Funktion deutlich widersprechen.

### 4.2.3.3 Markov Chain Monte Carlo Verfahren

Bayes'sche Verfahren konnten sich in der Fachwelt erst dann als tatsächliche Alternative zu Verfahren der klassischen Statistik etablieren, als man das Potential von MCMC Verfahren zur approximativen Berechnung der A-Posteriori-Verteilung bzw. von Kenngrößen derselben erkannte und sie für diesen Zweck einzusetzen begann (vgl. z.B. auch [Ntz09]). Interessant dabei ist, dass MCMC Verfahren zu diesem Zeitpunkt bereits schon recht lange in der Physik bekannt waren, ihre Einsatzmöglichkeit im Kontext der Bayes'schen Theorie aber lange nicht gesehen wurde, siehe [Hit03] für weitere Details zur Historie. MCMC Verfahren lassen sich vor allem auch oft zur Lösung hochdimensionaler Problemstellungen erfolgreich anwenden. Durch das Verfügbarwerden entsprechender Methoden und Werkzeuge für die Bayes'sche Statistik verbesserte sich die Möglichkeit, Bayes'sche Verfahren in der Praxis überhaupt erst adäquat, d. h. vor allem auch auf ausreichend komplexe Modelle, anwenden zu können, deutlich.

Im Kontext der Bayes'schen Theorie wird mittels MCMC Verfahren grob gesagt ein zeitdiskreter stochastischer Prozess mit Zustandsraum  $Z$  in Form einer sog. Markovkette erzeugt, welcher gegen die A-Posteriori-Verteilung  $p(z|d)$  konvergiert. Der zugrundeliegende Gedanke ist, dass man die Markovkette, wenn sie sich lange genug entwickelt hat, nutzen kann, um näherungsweise Samples der A-Posteriori-Verteilung zu generieren.

Ein Markovkette mit Zustandsraum  $Z$  ist definiert als stochastischer Prozess  $(X_t)_{t \in \mathbb{N}}$ , welcher für alle  $A, A_1, \dots, A_{t-1} \subset Z$ ,  $x \in Z$ ,  $t \in \mathbb{N}$  folgende Eigenschaft besitzt:



$$\begin{aligned} & P(X_{t+1} \in A | X_t = x, X_{t-1} \in A_{t-1}, \dots, X_1 \in A_1) \\ &= P(X_{t+1} \in A | X_t = x). \end{aligned} \quad (4.9)$$

Ist  $Z$  diskret, so lässt sich Gleichung (4.9) auch wie folgt äquivalent schreiben (vgl. z. B. [Gam06]): Für alle  $y_1, \dots, y_{t+1} \in Z$ ,  $t \in \mathbb{N}$  muss gelten

$$\begin{aligned} & P(X_{t+1} = y_{t+1} | X_t = y_t, X_{t-1} = y_{t-1}, \dots, X_1 = y_1) \\ &= P(X_{t+1} = y_{t+1} | X_t = y_t). \end{aligned} \quad (4.10)$$

Die Gleichungen (4.9) und (4.10) besagen gerade, dass zu jedem Zeitpunkt  $t \in \mathbb{N}$  der Zustand, in dem sich die Markovkette zum nächsten Zeitpunkt  $t + 1$  befindet, nur vom aktuellen Zustand und nicht von der darüber hinausgehenden Historie abhängt. Die Theorie der Markovketten ist im allgemeinen Fall komplexer als im Fall, dass  $Z$  diskret ist. Da die in Bezug auf MCMC Verfahren wichtigen Aussagen im Wesentlichen gleich sind in beiden Fällen, konzentriert sich die nachfolgende, bewusst kurz gehaltene Darstellung der mathematischen Grundlagen auf den diskreten Fall. Für den allgemeinen Fall sei z. B. auf [Gam06] und [Mey93] verwiesen.

Eine homogene Markovkette  $(X_t)_{t \in \mathbb{N}}$  liegt vor, wenn die Übergangswahrscheinlichkeit  $P(X_{t+1} = x | X_t = y)$  für alle  $x, y \in Z$  unabhängig von  $t$  ist. Diese wird im Nachfolgenden dann abkürzend mit  $P(y, x)$  notiert, d. h.

$$P(y, x) := P(X_{t+1} = x | X_t = y). \quad (4.11)$$

Eine Wahrscheinlichkeitsverteilung  $q(z)$  auf  $Z$  ist dann eine sog. stationäre Verteilung der Markovkette wenn für alle  $x \in Z$  die folgende Bedingung erfüllt ist:

$$q(x) = \sum_{y \in Z} q(y)P(y, x). \quad (4.12)$$

Wenn die Markovkette also zu einem gewissen Zeitpunkt die stationäre Verteilung  $q(z)$  besitzt, so behält sie gemäß Gleichung (4.12) diese zu allen späteren Zeitpunkten bei.

Es lässt sich zeigen, dass eine Markovkette eine stationäre Verteilung besitzt und auch unabhängig von ihrem Startpunkt, d. h. dem Punkt, in dem sie sich zum Zeitpunkt  $t = 1$  befindet, gegen diese konvergiert, wenn sie bestimmte Eigenschaften erfüllt. Konkret muss die Markovkette dafür ergodisch, irreduzibel und positiv rekurrent sein. Um den Umfang der vorliegenden Darstellung überschaubar zu halten, wird hinsichtlich der genauen Bedeutung dieser Eigenschaften sowie auf den formalen Beweis der zuvor gemachten Aussagen auf die Fachliteratur, z. B. [Gam06] oder [Bré99], verwiesen.

Gelingt es, eine derartige Markovkette auf  $Z$  zu erzeugen, welche die A-Posteriori-Verteilung  $p(z|d)$  als stationäre Verteilung besitzt, und lässt man diese lange genug laufen, bis die Konvergenz zumindest näherungsweise eingetreten ist, so lassen sich anhand dieser iterativ näherungsweise Samples der A-Posteriori-Verteilung erzeugen. Damit dies in der Praxis auch gut durchführbar ist, besteht natürlich weiter der Wunsch, dass die Samples möglichst einfach erzeugt werden können. Üblicherweise können der sog. Metropolis-Hastings-Algorithmus oder der sog. Gibbs Sampler, welcher letztlich einen Spezialfall des Metropolis-Hastings-Algorithmus darstellt, benutzt werden, um eine entsprechende Markovkette zu erzeugen. Eine genauere Darstellung dieser Algorithmen würde den Umfang der Darstellung sprengen und vor allem auch zu weit weg vom eigentlichen Thema der vorliegenden Arbeit führen, weshalb darauf verzichtet wird. Beide Algorithmen werden in der (neueren) Standardliteratur zur Bayes'schen Theorie jedoch ausführlich beschrieben. Der Leser sei diesbezüglich konkret z. B. auf die Referenzen [Car09, Gam06, Mar07] sowie auf [Ntz09] für eine recht praxis- und Werkzeug-orientierte Darstellung verwiesen.

Trotz ihrer Erfolge kann der Einsatz von MCMC Verfahren in der Praxis mit gewissen Schwierigkeiten verbunden sein. Diese sollen nachfolgend kurz skizziert werden.

Wie aus Gleichung (4.9) und Gleichung (4.10) erkennbar ist, liefert das iterative Sampling mittels einer Markovkette keine tatsächlich voneinander unabhängigen Samples. Diese Problematik kann in der Regel z. B. dadurch adressiert werden, indem man nicht in jedem Schritt der Markovkette ein Sample zieht, sondern nur alle  $k \in \mathbb{N}$  Schritte, um so zu erreichen, dass die Samples

zumindest näherungsweise voneinander unabhängig sind. Um  $k$  passend zu wählen, müssen die sog. Autokorrelationen der Samples dabei separat ermittelt werden (vgl. z. B. [Ntz09]).

Eine wesentliche Schwierigkeit kann darüber hinaus die grundsätzliche Frage darstellen, wie lange man die Markovkette laufen lassen muss, bis anhand von ihr tatsächlich Samples generiert werden können, welche näherungsweise gemäß der stationären Verteilung verteilt sind. Da die verfügbaren theoretischen Schranken hierzu i. A. eher schwach sind, müssen üblicherweise (auch verschiedene) empirische Verfahren zur Konvergenz-Diagnostik (vgl. z. B. [Ntz09]) angewandt werden, um hierzu ausreichend präzise Aussagen zu bekommen.

Abschließend muss noch darauf eingegangen werden, dass der Einsatz von MCMC Verfahren auch in komplexen und hochdimensionalen Problemstellungen üblicherweise gut funktioniert, es aber durchaus auch Einschränkungen gibt. Dies kann insbesondere dann der Fall sein, wenn  $Z$  umfangreich und von heterogener Struktur ist – wie es gerade auch in vielen praktischen Aufgabenstellungen der Informationsfusion der Fall sein wird. In [Zhu00] findet sich, dort konkret im Hinblick auf die Anwendung von MCMC Verfahren zur Objekterkennung in Bildern und zur Bildsegmentierung, die folgende, recht prägnante Aussage

„Intuitively, in a complicated solution space, an arbitrarily designed MCMC is like a drunk man walking in a cluttered building in the dark.“

Die als Antwort auf diese Problematik entwickelten DDMCMC (Data-Driven Markov Chain Monte Carlo) Verfahren (siehe z. B. auch [Tu02] und [Bar20]) nutzen im Gegensatz zu normalen MCMC Verfahren zusätzlich sehr datennahe Information, z. B. aus einem Kanten- oder Eckendetektor, um die Konvergenz der dann umgesetzten Markovkette zu beschleunigen. Zum Design der entsprechenden Verfahren wird recht viel spezifisches Expertenwissen im Hinblick auf die konkret adressierte Aufgabenstellung bzw. Anwendung benötigt, welches dann auch in die konkrete Ausgestaltung des anvisierten DDMCMC Algorithmus einfließen muss.

## 4.3 Ansatzpunkt der Arbeit

Die erstmalig in [Bey06a] veröffentlichte Idee lokaler Bayes'scher Fusionsansätze besteht darin, die in Abschnitt 4.1 dargestellte Problematik durch die Ermöglichung der Einnahme eines lokalen Standpunkts bei der Bayes'schen Fusion zu adressieren. Das Ziel ist es, die Fusion nur lokal, d. h. möglichst konzentriert auf Bereiche von  $Z$  vorzunehmen, in denen sich auch tatsächlich Aufgabenrelevantes abspielt.

Zur genaueren Formulierung dieses Gedankens wurde in [Bey06a] eine Analogie zu kriminalistischen Ermittlungen zur Aufklärung eines Verbrechens hergestellt, also zu einem etablierten Prozess der realen Welt. Dies inspirierte wiederum das Konzept für eine agentenbasierte Fusionsarchitektur, in der unterschiedliche Agenten in kooperativer Weise Spuren in Informationsbeiträgen identifizieren, auswerten sowie die damit korrespondierenden Hypothesen lokal durch die Einbeziehung zusätzlicher Informationsbeiträge weiterentwickeln und letztlich zur Zusammenschau bereitstellen. Im vorliegenden Abschnitt, welcher eine aufbereitete und erweiterte Version der zuvor u. a. auch weiter in [Bey07] veröffentlichten Ergebnisse zur lokalen Bayes'schen Fusion darstellt, werden dieses Konzept sowie die erforderlichen Bezugspunkte vorgestellt.

### 4.3.1 Bezugspunkte

#### 4.3.1.1 Kriminalistische Ermittlungen

Im Folgenden wird ein vereinfachter Blick auf kriminalistische Ermittlungen zur Aufklärung eines Verbrechens geworfen, mit dem Ziel, die für die Idee lokaler Bayes'scher Fusionsansätze relevanten Aspekte darzustellen. Es sei vorausgeschickt, dass die vorliegende Arbeit nicht den Anspruch erhebt, den Prozess kriminalistischer Ermittlungen im Detail zu eruieren und abzubilden. Ein derartiges Unterfangen ist weder notwendig noch sinnvoll, um die im Rahmen der Arbeit gesteckten Ziele zu erreichen. Angemerkt sei hierzu auch, dass sich in der Fachliteratur zwar unterschiedliche Modelle zur Beschreibung

des Prozesses kriminalistischer Ermittlungen finden, die Forschung diesbezüglich aber längst nicht als abgeschlossen erachtet werden kann (vgl. z. B. [And19]). Die nachfolgend dargestellten Aspekte sind recht offensichtliche Erkenntnisse im Hinblick auf die Funktionsweise kriminalistischer Ermittlungen, welche aber auch durch entsprechende Aussagen in der Fachliteratur gestützt werden. Konkret hierzu angeführt seien die Referenzen [And19] und [Kuy82].

Kriminalistische Ermittlungen stellen insgesamt einen komplexen Prozess dar. Dieser kann in unterschiedlichen Ausprägungen (z. B. proaktiv oder reaktiv) erfolgen. Er umfasst in der Regel die Sammlung, Verarbeitung bzw. Analyse und Kombination relevanter Daten und Information.

Wurde konkret festgestellt, dass sich ein Verbrechen ereignet hat, so wird eine Untersuchung dieses Verbrechens eingeleitet. Das Ziel der Untersuchung besteht darin, ausreichend sicher feststellen zu können, ob gegen jemanden eine Anklage erhoben werden muss und diese dann gegebenenfalls vorzubereiten. Hierfür müssen unterschiedliche mögliche Erklärungen des Vorfalls (auch bezeichnet als Untersuchungslinien) erkundet werden. Die für die Untersuchung eingesetzten Ressourcen, z. B. die Anzahl der Ermittler, werden in der Regel von der Wichtigkeit bzw. Schwere des Falls mitbeeinflusst.

Die Durchführung der Untersuchung beinhaltet oft zahlreiche, aneinander gekoppelte Ermittlungsschritte. Dabei können Spuren, d. h. aus der verfügbaren Information generierte Hinweise, welche sich z. B. auf die Identität oder den Aufenthaltsort eines Verdächtigen beziehen, genutzt werden, um Ansatzpunkte und Pfade zu identifizieren, an denen die weiteren Ermittlungen dann ausgerichtet werden können. Die vorliegende Information bzw. die daraus generierten Spuren werden ggf. entsprechend ihrer Relevanz priorisiert.

Der erste Schritt der Untersuchung besteht i. A. darin, die (möglichen und realistischen) Hypothesen zu formulieren, welche den Vorfall erklären können. Um festzustellen, ob eine der Hypothesen ausreichend gestützt wird, muss durch Einbeziehung der vorliegenden bzw. der im Rahmen der Ermittlungen gewonnenen Information der Informationsstand in Bezug auf die Hypothesen aktualisiert werden. Im Erfolgsfall kann mindestens am Ende der Untersu-

chung der Grad des Dafürhaltens bezüglich der Gültigkeit bestimmter Hypothesen als so hoch eingeschätzt werden, dass sie als bewiesen gelten können. Ist dies nicht der Fall, so muss die Untersuchung fortgesetzt oder ohne Ergebnis beendet werden.

Während der Untersuchung werden die Ermittler mit heterogenen Informationsbeiträgen aus unterschiedlichen Quellen konfrontiert. Hierzu können z.B. Zeugenaussagen, Fingerabdrücke, DNA-Spuren, Information aus Datenbanken zählen. Diese müssen letztlich von Personen interpretiert und in Zusammenschau betrachtet werden, welche über ausreichend Kompetenz und Kenntnis verfügen. Hierzu ist i. A. auch Vorwissen nötig. Auch kann es einer Abstimmung unterschiedlicher Ermittler erfordern. Weiter ist zur Erfassung, Verarbeitung und Auswertung spezifischer Daten oftmals die Unterstützung der Ermittler durch Sachverständige nötig – nämlich dann, wenn die Ermittler, welche die eigentliche Untersuchung durchführen, nicht über ausreichend Fachexpertise verfügen, um diese Aufgabe selbst zu leisten. Die Sachverständigen müssen den Ermittlern dann ein für sie verständliches, rein auf die Auswertung der ihnen vorgelegten Daten bezogenes Auswertergebnis liefern. Beispiele für derartige Sachverständige sind Gerichtsmediziner, DNA-Analysten, IT-Experten.

Charakteristisch für die kriminalistischen Ermittlungen sind insbesondere die heterogenen Informationsquellen und die von den Ermittlern eingenommene lokale Sicht. Die zur Aufklärung eines Verbrechens angestoßene Untersuchung konzentriert sich auf bestimmte verdächtige Ereignisse bzw. Personen.

Für eine finale Entscheidung bezüglich der Schuldigkeit eines Tatverdächtigen müssen nach Abschluss der Ermittlungen alle relevanten Teilergebnisse zu einem Ermittlungsergebnis zusammengefasst werden. Sie können dann so an ein Organ der Judikative übergeben werden.

#### **4.3.1.2 Agenten**

In der agentenbasierten Fusionsarchitektur zur lokalen Bayes'schen Fusion sollen auf Basis eines Multi-Agenten-Systems in Analogie zu kriminalisti-

schen Ermittlern sog. Fusionsagenten zur Fusion heterogener Informationsquellen eingesetzt werden. Bevor dies in Abschnitt 4.3.2 genauer dargestellt wird, gibt der vorliegende Abschnitt einen kurzen Überblick über ausgewählte Grundlagen der Agenten-Theorie.

Der Festlegung in [Cop04] folgend wird unter einem Agenten in der vorliegenden Arbeit eine Entität verstanden, welche in der Lage ist, eine bestimmte Aufgabe auszuführen, um einen Menschen zu unterstützen. In diesem Sinne kann es sich bei einem Agenten also um ein Lebewesen, um einen Roboter oder auch um Software handeln.

Ein Software-Agent ist dann gemäß [Cop04] wie folgt zu verstehen:

„A software agent is a computer program designed to carry out some task on behalf of a user.“

Es sei erwähnt, dass diese Begriffsfestlegung in der Fachliteratur alles andere als einheitlich ist. Folgende, spezifischere Definition wird z. B. in [Woo09] verwendet:

„An agent is a computer system that is situated in some environment and that is capable of autonomous action in this environment in order to meet its design objectives.“

Ein weiteres Beispiel stellt nachfolgende, [Das10] entnommene Definition dar:

„An agent is a computational entity with intentionality that performs user delegated tasks autonomously.“

Wie bereits aus den obigen Definitionen erkennbar, werden Software-Agenten oft anhand bestimmter Eigenschaften definiert bzw. dann auch noch feiner eingeordnet. Oftmals werden dabei Repräsentationsfähigkeit, Autonomie und Kommunikationsfähigkeit als grundlegende Eigenschaften gefordert (vgl. [Zar99]). Repräsentationsfähigkeit bedeutet, dass der Software-Agent in der Lage ist, anstelle seines Nutzers bestimmte Aktionen auszuführen. Der Begriff der Autonomie lässt sich in dem Sinne auffassen, dass der Agent selbstständig, d. h. vollständig ohne oder zumindest weitgehend ohne Eingriff eines anderen Agenten oder eines Menschen, handeln kann (vgl. [Wei05]).

Um seine Aufgabe zu erfüllen, muss der Software-Agent in der Regel auch in der Lage sein, mit seiner Umwelt, welche neben dem Nutzer u. a. andere Agenten, Computer-Programme und Informationsquellen beinhalten kann (vgl. [Wei99b]), interagieren können.

In der Fachliteratur wird Software-Agenten nicht selten insbesondere auch die Eigenschaft der Intelligenz zugewiesen. Der Begriff Intelligenz ist dabei erst einmal recht schwammig. Er lässt sich z. B. präzisieren, indem man fordert, dass ein intelligenter Software-Agent zusätzliches Wissen über seine Umwelt besitzt, und so auch in der Lage ist, seine Aufgabe zu erfüllen, wenn sich bestimmte Aspekte der Aufgabe ändern oder wenn nicht per se für ihn erwartbare Situationen in der Umwelt auftreten (vgl. [Cop04]). Aus der Forderung nach Intelligenz lassen sich dann wieder andere konkrete Eigenschaften, die der Software-Agent erfüllen muss, ableiten (vgl. [Wei99b]). Beispiele hierfür sind Reaktivität, Proaktivität und Lernfähigkeit.

Ein reaktiver Software-Agent kann auf Ereignisse in seiner Umgebung, welche er dann auch z. B. mittels Sensoren oder durch Kommunikation wahrnehmen können muss, reagieren. Ein proaktiver Software-Agent kann nicht nur reagieren, sondern auch, getrieben z. B. durch ein bestimmtes Ziel, die Initiative ergreifen. Ein lernfähiger Software-Agent ist in der Lage, neue Information z. B. aus gemachten Beobachtungen oder aus bei ihm eintreffenden Hinweisen aus seiner Umwelt zu nutzen. Dadurch kann er dann auch seine Fähigkeiten zur Lösung der Aufgabe verbessern.

Eine wichtige Eigenschaft ist oft auch die der Mobilität, welche besagt, dass der Software-Agent in der Lage ist, sich (üblicherweise über ein Netzwerk) von einem Ort an einen anderen zu transferieren. Mobile Software-Agenten könnten also insbesondere eingesetzt werden, um verteilt vorliegende Informationsbestände zur Lösung ihrer Aufgabe bandbreitenschonender auszuwerten und einzubeziehen, als es der Fall wäre, wenn sie die Information hierzu anfordern müssten (vgl. [Cop04]). Der Einsatz mobiler Software-Agenten bietet sich u. a. deshalb gerade auch in vernetzten Systeme an.

Die Kombination unterschiedlicher Software-Agenten in einem Multi-Agenten-System wird oft genutzt, um die Gesamtleistung eines Systems zu



verbessern und um auch komplexe Aufgaben von Software-Agenten ausführen lassen zu können. Dies ist dann möglich, ohne dass eine zentrale Kontrollinstanz vorgesehen werden muss. Gemäß [Cop04] verfügt jeder der Software-Agenten dabei in der Regel nur über unvollständige Information bzw. hat nicht die Fähigkeit, die gestellte Aufgabe alleine auszuführen. Durch den Einsatz eines solchen Multi-Agenten-Systems können dann auch eher einfach gestrickte Software-Agenten eingesetzt werden, um gemeinsam eine Aufgabe zu adressieren, welche ein eher hohes Maß an Intelligenz fordert. Auch können die Software-Agenten gemeinsam ein Gesamtproblem lösen, ohne dass die Anforderung besteht, dass der einzelne Software-Agent dieses komplett überschaut.

Bei einem Multi-Agenten-System lassen sich kooperative und konkurrierende Multi-Agenten-Umgebungen unterscheiden. Ein Multi-Agenten-System, in dem Software-Agenten zur Erreichung einer gemeinsamen Zielsetzung zusammenarbeiten, stellt in der Regel für die einzelnen Software-Agenten eine kooperative Multi-Agenten-Umgebung dar. Wie in [Cop04] noch angeführt, bietet ein Multi-Agenten-System dadurch, dass die einzelnen Software-Agenten ihre Arbeiten parallelisieren können, auch die Möglichkeit, Aufgaben schneller zu lösen, als es sonst der Fall wäre, und durch das Vorsehen zusätzlicher Software-Agenten lässt sich ein derartiges System prinzipiell auch mit zusätzlicher Robustheit gegenüber Ausfällen versehen. Legt man die eingangs angeführte, allgemeine Definition des Agenten-Begriffs zugrunde, so könnte ein Multi-Agenten-System auch so ausgelegt werden, dass z. B. menschliche Agenten und Agenten in Form von Software-Entitäten kooperieren und dabei ihre jeweiligen Stärken und Schwächen einbringen. In diesem Sinne könnte ein Multi-Agenten-System auch eine gute Basis zur Realisierung eines sog. Augmented Intelligence Systems (vgl. [Kir19]), in dem sich menschliche Intelligenz und künstliche Intelligenz bestmöglich zu sog. erweiterter Intelligenz ergänzen können, sein.

### 4.3.2 Agentenbasierte Fusionsarchitektur

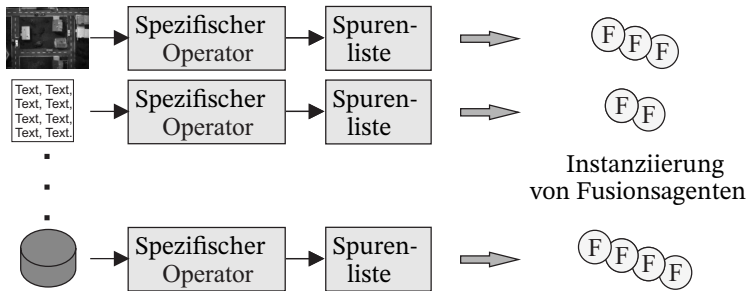
Das erstellte Konzept für die agentenbasierte Fusionsarchitektur, welches nachfolgend genauer beschrieben wird, sieht zur Durchführung der lokalen Bayes'schen Fusion zwei Phasen vor. In einer Initialisierungsphase werden in den zur Fusion vorliegenden Informationsbeiträgen Spuren im Hinblick auf den „wahren“ Wert der durch  $z$  verkörperten Größen des Interesses gewonnen. Mit diesen Spuren korrespondierende Hypothesen bezüglich des „wahren“ Werts von  $z$  werden dann zur weiteren Untersuchung an Fusionsagenten übergeben. In der darauf folgenden Ermittlungsphase beziehen die Fusionsagenten die bei Gewinnung der ihnen zugewiesenen Spuren noch nicht berücksichtigte Information ein, mit dem Ziel, die mit ihren Hypothesen korrespondierenden anfänglichen Degree-of-Beliefs weiterzuentwickeln. Zum Abschluss der Ermittlungsphase soll dann anhand der final resultierenden Degree-of-Beliefs über die Gültigkeit der Hypothesen entschieden werden. Wesentlich bei diesem Vorgehen ist, dass jede Spur nur konzentriert auf einen Ausschnitt des Wertebereichs  $Z$  der Größen des Interesses untersucht wird.

Die nachfolgende genauere Darstellung des Konzepts für die agentenbasierte Fusionsarchitektur zielt darauf ab, die grundsätzlichen und für den weiteren Verlauf der Arbeit wesentlichen Aspekte dazulegen. Im konkreten Anwendungsfall können, z. B. je nach Zusammenhang der Informationsbeiträge (vgl. Abschnitt 2.4.1), Detail-Anpassungen an der beschriebenen Vorgehensweise zielführend oder sogar notwendig sein.

#### 4.3.2.1 Initialisierungsphase

Wie in Abb. 4.1 illustriert werden die für die Fusion vorliegenden Informationsbeiträge mittels spezifischer Operatoren ausgewertet, um Spuren, d. h. Auffälligkeiten in  $Z$ , zu ermitteln. Der zur Auswertung eines bestimmten Informationsbeitrags eingesetzte Operator sollte dabei möglichst optimal auf die entsprechende Informationsquelle abgestimmt sein. Dies bedeutet, dass er

idealerweise das bestmögliche Ergebnis, das bei isolierter Betrachtung dieser Quelle bzw. der von ihr gelieferten Information möglich ist, liefern sollte.



**Abbildung 4.1:** Schematische Darstellung der Initialisierungsphase der agentenbasierten Fusionsarchitektur. Zum Abschluss dieser Phase werden für alle oder nur für die wichtigsten der Spuren in den Spurenlisten Fusionsagenten (F) instanziiert.

Die gewonnenen Spuren werden in Spurenlisten gesammelt. Werden sie dort entsprechend ihrer Signifikanz geordnet, so wird eine intrinsische Skalierbarkeit der lokalen Bayes'schen Fusion erreicht. Zum Abschluss der Initialisierungsphase kann dann nämlich für alle oder mindestens für die wichtigsten Spuren jeweils ein Fusionsagent instanziiert werden. Über die maximale Anzahl an eingesetzten Fusionsagenten kann dabei anhand der verfügbaren Ressourcen entschieden werden.

Besteht die Aufgabe z. B. darin, mittels lokaler Bayes'scher Fusion Objekte in einer Szene zu lokalisieren und zu klassifizieren, so macht es Sinn,  $Z$  als das kartesische Produkt  $Z = Z_1 \times Z_2 \times Z_3$  aller möglichen Objektpositionen  $Z_1 \times Z_2 \subset \mathbb{R}^2$  und aller möglichen Objekttypen  $Z_3$  festzulegen. Liefert die Informationsquelle Nummer  $s$  weiter ein Luftbild der Szene, so könnte der quellspezifische Operator  $o_s(\cdot)$  mittels eines Bildauswerteverfahrens zur Detektion und Klassifikation von Objekten der in  $Z_3$  festgelegten Typen umgesetzt werden. Der Einsatz des Bildauswerteverfahrens sollte darauf abzielen, ein Ergebnis  $\{o_s(z) | z \in Z\}$  über  $Z$  zu gewinnen, welches lokale Maxima für vor-

handene Objekte bestimmter Typen hat. Aus den durch die Auswertung resultierenden lokalen Maxima leiten sich die Spuren in der mit  $d_s$  korrespondierenden Spurenliste ab. Eine Spur hat dabei dann formal gesehen die Form

$$(c_{s,t}, p_{U(c_{s,t})}(z|d_s)) , \quad t \in \{1, \dots, T_s\} . \quad (4.13)$$

Dabei bezeichnen  $c_{s,t} \in Z$  das Argument des jeweiligen lokalen Maximums von  $o_s(\cdot)$  und  $T_s$  die Anzahl der durch die Auswertung des Informationsbeitrags  $d_s$  gewonnenen Spuren. Weiter stellt  $p_{U(c_{s,t})}(z|d_s)$  eine sog. lokale Degree-of-Belief-Verteilung dar. Diese wird geeignet, gegebenenfalls (d. h. bei entsprechender Eignung von  $o_s(\cdot)$  hierfür) direkt unter Einbeziehung des von  $o_s(\cdot)$  gelieferten Resultats, erzeugt. Lokal bedeutet, dass sie sich auf eine nicht allzu große Umgebung  $U(c_{s,t})$  um  $c_{s,t}$  bezieht, d. h. es gilt  $U(c_{s,t}) \subset Z$ ,  $|U(c_{s,t})| \ll |Z|$ .

Nach Auswertung aller vorliegenden Informationsbeiträge  $d_s$ ,  $s \in \{1, \dots, S\}$ , ist  $T := \sum_{s=1}^S T_s$  dann die Gesamtzahl aller in den Spurenlisten erfassten Spuren. Wie im Beispiel bereits dargestellt bestehen diese formal gesehen jeweils aus einem Paar der Form „Auffälligkeit“ und „korrespondierende lokale Degree-of-Belief-Verteilung“, vgl. Gleichung (4.13). Zum Abschluss der Initialisierungsphase werden dann  $J \leq T$  Fusionsagenten instanziiert.

Die in der Ermittlungsphase dann in kooperativer Weise von den Fusionsagenten umgesetzte lokale Bayes'sche Fusion erfolgt somit insgesamt gesehen konzentriert auf die folgende Teilmenge von  $Z$ :

$$U := \bigcup_{j=1}^J U^{(j)} , \quad U^{(j)} \subset Z . \quad (4.14)$$

Dabei bildet jeder Teilausschnitt  $U^{(j)}$  von  $Z$  eine Spur ab, welche einem Fusionsagenten zugewiesen wurde, d. h. er hat die Form  $U^{(j)} = U(c_{s,t})$  mit  $t \in \{1, \dots, T_s\}$ ,  $s \in \{1, \dots, S\}$ .

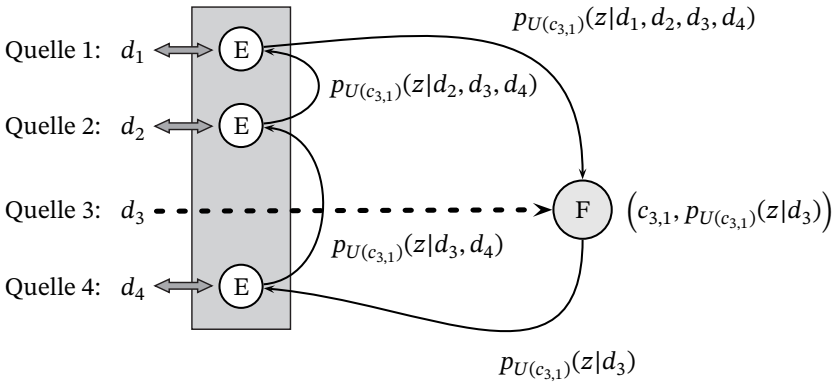
### 4.3.2.2 Ermittlungsphase

Die zum Ende der Initialisierungsphase instanziierten Fusionsagenten haben nun die Aufgabe, die mit ihren Spuren korrespondierenden lokalen Degree-of-Belief-Verteilungen weiterzuentwickeln. Dafür besucht jeder von ihnen die im Hinblick auf seine Spur noch nicht berücksichtigten Informationsquellen und betreibt, jeweils konzentriert auf den mit der Spur korrespondierenden Ausschnitt von  $Z$ , Bayes'sche Fusion.

In jedem der dabei durchzuführenden Fusions Schritte wird die lokale Degree-of-Belief-Verteilung des jeweiligen Fusionsagenten als (lokale) A-Priori-Verteilung aufgefasst, in welche die zusätzlich zu berücksichtigende Information (lokal) mittels des Satzes von Bayes (vgl. Gleichung (3.22)) integriert werden kann. Die so resultierende (lokale) A-Posteriori-Verteilung stellt dann wiederum die (lokale) A-Priori-Verteilung für den nächsten Fusions Schritt, den der entsprechende Fusionsagent durchzuführen hat, dar. Wie in Abschnitt 3.5.2 erläutert kann gerade beim Vorliegen heterogener Informationsquellen oftmals von einer bedingten Unabhängigkeit der Informationsbeiträge (gegeben  $z$ ) ausgegangen werden. In diesem Fall liefert dann auch die Anwendung des in Gleichung (3.25) formulierten sequentiellen Fusions schemas exakte Ergebnisse.

Zur Veranschaulichung des durch die Fusionsagenten umzusetzenden Vorgehens ist in nachfolgender Abb. 4.2 die sequentielle Weiterentwicklung einer lokalen Degree-of-Belief-Verteilung durch einen Fusionsagenten für den Fall von vier Informationsquellen illustriert. Der hier exemplarisch dargestellte Fusionsagent ist zuständig für die durch  $c_{3,1}$  induzierte Spur, welche mit dem ersten Eintrag in der durch Auswertung des Informationsbeitrags  $d_3$  resultierten Spurenliste korrespondiert. Dieser Fusionsagent hat in der Ermittlungsphase nun die Aufgabe, seine anfängliche lokale Degree-of-Belief-Verteilung  $p_{U(c_{3,1})}(z|d_3)$  durch Einbeziehung der in Form der weiteren Informationsbeiträge  $d_1$ ,  $d_2$  und  $d_4$  vorliegenden Information (lokal) weiterzuentwickeln. Hierzu besucht er die noch nicht berücksichtigten Quellen nach und nach und führt jeweils einen (lokalen) Fusions Schritt durch. Im letzten Schritt resultiert dann seine finale lokale A-Posteriori-Verteilung  $p_{U(c_{3,1})}(z|d_1, d_2, d_3, d_4)$ , wel-

che die von allen vier Informationsquellen gelieferte Information lokal, d. h. in Bezug auf den mit der ihm zugewiesenen Spur korrespondierenden Ausschnitt  $U(c_{3,1})$  von  $Z$ , berücksichtigt.



**Abbildung 4.2:** Schematische Darstellung der schrittweisen Weiterentwicklung einer lokalen Degree-of-Belief-Verteilung durch einen Fusionsagenten (F) unter Zuhilfenahme von Experten-Agenten (E) für die Fusion der Beiträge von vier Informationsquellen.

Falls ein Fusionsagent bei der Erledigung seiner Aufgabe nicht imstande ist, selbst auf bestimmte Information zuzugreifen oder diese selbst auszuwerten, d. h. die für die Fusion erforderliche (lokale) Likelihood-Funktion zu bilden, so kann er diese Aufgabe an einen sog. Experten-Agenten auslagern. Ein solcher Experten-Agent kann ein auf die Bereitstellung bzw. Auswertung der entsprechend benötigten Information spezialisierter Software-Agent oder auch ein menschlicher Agent sein. Der Einsatz menschlicher Experten-Agenten macht u. a. dann Sinn, wenn für eine bestimmte Art von Information keine ausreichend leistungsfähigen bzw. qualitativ ausreichend hochwertigen maschinellen Auswerte-Algorithmen verfügbar sind.

Nach Abschluss der Ermittlung, d. h. der (lokalen) Weiterentwicklung der lokalen Degree-of-Belief-Verteilungen durch die Fusionsagenten gilt es,

die resultierenden finalen lokalen Degree-of-Belief-Verteilungen in Zusammenschau zu bringen und ggf. weitere, auf dem insgesamt ermittelten Informationsstand basierende Schlüsse zu ziehen. Hierzu könnte in der agentenbasierten Fusionsarchitektur ggf. eine auf weiteren Software-Entitäten basierende Art finale Instanz (eine „Jury“) umgesetzt werden. Denkbar wäre darüber hinaus aber auch, dass die Fusionsagenten diese Aufgabe bzw. Teile davon selbst kollaborativ und/oder in Kooperation mit dem Menschen lösen.

#### **4.3.2.3 Diskussion des vorgestellten Konzepts**

Das Konzept für die agentenbasierte Fusionsarchitektur sieht ein hohes Maß an Skalierbarkeit vor. Das Ziel besteht letztlich darin, die konkrete Ausgestaltung der lokalen Bayes'schen Fusion bestmöglich an die verfügbaren Ressourcen anpassbar zu machen. Nachfolgend werden wesentliche Aspekte diesbezüglich dargestellt, auch unter Bezugnahme von Erkenntnissen aus Kapitel 2. Zusätzliche Aspekte werden sich im weiteren Verlauf der Arbeit, vor allem aus den angestellten mathematisch-formalen Untersuchungen ergeben. Diese Aspekte werden dann an der jeweiligen Stelle und unter Bezugnahme auf die jeweiligen Ergebnisse diskutiert werden.

Durch die in Abschnitt 4.3.2.1 darlegte Möglichkeit zur Ordnung der Spuren in den Spurenlisten entsprechend ihrer Signifikanz kann bei Bedarf erst einmal eine Priorisierung der lokalen Bayes'schen Fusion auf besonders „heiße“ Spuren vorgenommen werden. Falls sich dadurch kein Fusionsresultat von ausreichender Güte bestimmen lässt und/oder weiterer Informationsbedarf besteht, so könnten die von den Fusionsagenten durchgeführten Ermittlungen dann in einem zweiten Schritt auf weitere Spuren ausgeweitet werden.

Durch eine vom theoretischen Standpunkt aus unkompliziert realisierbare Modifikation des durch die Fusionsagenten umgesetzten Fusionsschemas könnte die Ermittlung in Bezug auf eine bestimmte Spur statt durch einen auch durch mehrere Fusionsagenten erfolgen, um so schneller ein Resultat zu erzielen. Wichtig ist dabei dann, dass sich die entsprechenden Fusionsagenten abstimmen, wer von ihnen den Beitrag einer bestimmten Informationsquelle in seiner lokalen Degree-of-Belief-Verteilung berücksichtigt. Weiter muss

dafür gesorgt werden, dass durch die parallele Ermittlung entlang derselben Spur das Vorwissen nicht mehrfach in die entsprechenden lokalen Degree-of-Belief-Verteilungen eingeht und so das zum Schluss in der Zusammenschau resultierende Ergebnis verfälscht wird. Prinzipiell könnte man diese Problematik auch lösen, indem man mehrfach berücksichtigtes Vorwissen nach Abschluss der Ermittlungen wieder entsprechend heraus dividiert.

Die agentenbasierte Fusionsarchitektur ist modular in dem Sinne, dass sie recht unkompliziert erweitert werden kann, wenn neue Informationsquellen, zu deren Auswertung ggf. auch zusätzliche Arten von Expertise erforderlich sind, hinzukommen. In dem Fall könnte man eine Erweiterung des Agenten-Teams um neue Experten-Agenten vorsehen, um so auch die Auswertung der neu verfügbaren Information zu ermöglichen. Auch könnten neue Experten-Agenten als Austausch für bereits vorhandene eingebunden werden. Dies bietet sich an, wenn neue („bessere“) Möglichkeiten zur Auswertung bestimmter Arten von Information verfügbar werden. Ein Anwendungsfall diesbezüglich könnte auch der sein, dass menschliche Experten-Agenten nach und nach durch Software-Agenten ersetzt werden, wenn maschinelle Auswerteverfahren, welche zuvor nicht in ausreichender Qualität und/oder Leistungsfähigkeit verfügbar waren, eingebunden werden können.

Im vorstellten Konzept für die agentenbasierte Fusionsarchitektur wurde davon ausgegangen, dass die zur Fusion heranzuziehenden Informationsquellen vorab bekannt sind. Eine naheliegende Erweiterung des Konzepts besteht darin, die Fusionsagenten oder andere Agenten auch nach a priori unbekanntem Informationsquellen recherchieren zu lassen und ihnen dabei ggf. auch Data-Mining-Aufgaben zu übertragen.

Es liegt auf der Hand, dass sich das Konzept für die agentenbasierte Fusionsarchitektur insbesondere auch für die Umsetzung in großen Netzwerken anbietet. Einen wichtigen Anwendungsfall stellen gerade auch die zuvor in der Arbeit schon mehrfach thematisierten system-, ebenen- und/oder organisationsübergreifenden Verbünde dar. Hier liegen oft umfangreiche Mengen an Daten und Information und ggf. auch hilfreiche (menschliche und/oder softwaretechnisch realisierte) Expertise weit verteilt vor und die Fusion muss i. A. möglichst bandbreitenschonend erfolgen.



Insbesondere (jedoch nicht ausschließlich) wenn die Informationsfusion in solchen Verbänden realisiert wird, kann die mit der Umsetzung der Bayes'schen Fusion verbundene Problematik nicht nur darin bestehen, dass der zur Berechnung der A-Posteriori-Verteilung anfallende Aufwand wie in Abschnitt 4.1 dargestellt rechnerisch nicht tragbar ist. Vielmehr kann auch bereits die Bereitstellung der zu ihrer Berechnung erforderlichen Größen (z. B. in Form der ursprünglichen Informationsbeiträge oder der Likelihood-Funktionen) problematisch bzw. nicht realisierbar sein, wenn man diese, wie bei der globalen (d. h. nicht lokalen) Bayes'schen Fusion der Fall, auf ganz  $Z$  bezogen benötigt und sie ggf. auch entsprechend übermittelt werden müssen. Dabei ist auch davon auszugehen, dass im Fall der lokalen Bayes'schen Fusion bestimmte, ggf. hinsichtlich ihrer Kapazitäten begrenzte Expertise zur Auswertung der Informationsbeiträge nur in geringerem Umfang in Anspruch genommen werden muss, als es im Fall der globalen Bayes'schen Fusion nötig wäre.

Die lokale Bayes'sche Fusion könnte im Kontext von Verbänden grundsätzlich weiter auch dazu eingesetzt werden, unterschiedliche Fragen/Aufgabenstellungen von unterschiedlichen Entscheidungsträgern im Verbund jeweils konzentriert auf die für diese relevanten Aspekte zu adressieren. In diesem Sinne würde es dann auch möglich, die resultierende Information gerade dort bereitzustellen, wo sie tatsächlich auch benötigt wird, was ebenfalls essentiell für ein optimales Ressourcenmanagement ist.



# 5 Mathematische Basis lokaler Bayes'scher Fusion

## 5.1 Einführung

Wie in Abschnitt 4.3 dargestellt wurde, besteht das Ziel lokaler Bayes'scher Fusionsansätze gerade darin, die Bayes'sche Fusion nicht global bezüglich des gesamten Wertebereichs  $Z$  der Größen des Interesses vorzunehmen, sondern nur lokal, d. h. möglichst konzentriert auf Bereiche von  $Z$ , in denen sich auch tatsächlich Aufgabenrelevantes abspielt. In anderen Worten soll die Bayes'sche Fusion auf eine Teilmenge  $U$  des Wertebereichs  $Z$  der Größen des Interesses konzentriert werden, welche so gewählt ist, dass sie mit hoher Wahrscheinlichkeit das Fusionsergebnis bzw. die zur Bestimmung desselben tatsächlich relevanten Teile von  $Z$  beinhaltet. Diese Teilmenge  $U \subset Z$  wird im weiteren Verlauf der Arbeit als lokaler Kontext bezeichnet.

Wird die lokale Bayes'sche Fusion auf Basis der agentenbasierten Fusionsarchitektur kollaborativ durch  $J \in \mathbb{N}$  Fusionsagenten umgesetzt, von welchen jeder in der Ermittlungsphase gerade für seine Spur zuständig ist, so betreibt (gemäß Gleichung (4.14)) jeder Fusionsagent lokale Bayes'sche Fusion in Bezug auf einen Anteil  $U^{(j)}$  des lokalen Kontexts  $U$ ,  $j \in \{1, \dots, J\}$ .

Werden die zur Berechnung der A-Posteriori-Verteilung auf ganz  $Z$  erforderlichen Operationen eingeschränkt auf den lokalen Kontext  $U = U_1 \times \dots \times U_N$ , so reduziert sich der in Gleichung (4.1) angegebene Aufwand von

$$O(|Z|) = O(\eta^N), \quad \eta := \sqrt[N]{\prod_{n=1}^N |Z_n|} \quad (5.1)$$

auf

$$O(|U|) = O(\theta^N), \quad \theta := \sqrt[N]{\prod_{n=1}^N |U_n|}. \quad (5.2)$$

Wie aus Gleichung (5.2) erkennbar ist, wächst der Aufwand zur Berechnung der (dann lokalen) A-Posteriori-Verteilung i. A. ebenfalls exponentiell, nun aber mit der mittleren Dimension des lokalen Kontexts  $U$ . Um tatsächlich einen deutlichen Vorteil bei der numerischen („punktweisen“) Berechnung der (dann lokalen) A-Posteriori-Verteilung zu erreichen, muss  $|U| \ll |Z|$  gelten.

Wie zuvor in Abschnitt 4.3.2.3 dargestellt wurde, wird die Herausforderung oftmals auch nicht nur darin bestehen, dass bei der Bayes'schen Fusion der Aufwand zur Berechnung (und ggf. auch zur weiteren Auswertung) der A-Posteriori-Verteilung nicht tragbar ist. Insbesondere können auch bereits die Bestimmung und Übermittlung der zu ihrer Berechnung erforderlichen Größen problematisch, möglicherweise sogar nicht realisierbar, sein. Hinzu kommt, dass gerade in Verbänden das Vorgehen, in Form der A-Posteriori-Verteilung jeweils stets zuerst ein maximal umfassendes Fusionsresultat zu berechnen, auch konzeptionell dem für ein optimales Ressourcenmanagement notwendigen Prinzip, dass Information letztlich nur dort und in der Form bereitgestellt werden sollte, wie sie auch tatsächlich benötigt wird, widerspricht. Das Konzept für die agentenbasierte Fusionsarchitektur hat verdeutlicht, dass Ansätze zur lokalen Bayes'schen Fusion gerade auch Potential bieten, diesen Anforderungen nachzukommen.

Aus theoretischer Hinsicht bedarf es zur genaueren Entwicklung lokaler Bayes'scher Fusionsansätze einer Anpassung der Bayes'schen Fusionsmethodik an lokale Betrachtungen. In Abschnitt 5.2 werden die mathematischen Grundlagen dazu geschaffen und die resultierenden lokalen Bayes'schen Modelle werden ausführlich analysiert. Ergänzend und in gewissem Sinne auch komplementär hierzu werden auf Basis einer in Teilen heuristischen Vorgehensweise ausgewählte Teilaspekte im Hinblick auf die lokale Bayes'sche Fusion in Abschnitt 5.3 anhand eines möglichst anwendungsnahen Beispiels umgesetzt und untersucht.

## 5.2 Lokale Bayes'sche Modelle

Im vorliegenden Abschnitt wird der Frage nachgegangen, ob und wie genau sich Lokalität konform zur Bayes'schen Theorie beschreiben lässt. Hierzu werden die in Abschnitt 3.3 und Abschnitt 3.4.2 aufbereiteten Grundlagen aus der Wahrscheinlichkeitstheorie und der Bayes'schen Inferenz aufgegriffen und erweitert, bevor auf dieser Basis dann lokale Bayes'sche Modelle mathematisch formalisiert und untersucht werden. Die nachfolgende Darstellung stellt eine überarbeitete und erweiterte Version der hierzu zuvor in [San06], [San07], [San08] und [San09b] veröffentlichten Ergebnisse zur lokalen Bayes'schen Fusion dar.

### 5.2.1 Bayes'sche Experimente

Gemäß Abschnitt 3.4.2 setzt sich ein Bayes'sches (statistisches) Modell aus der Annahme einer Verteilungsklasse  $\{p(d|z)|z \in Z\}$  und einer A-Priori-Verteilung  $p(z)$  auf  $Z$  zusammen. Betrachtet man, anders als in einer konkreten Aufgabenstellung, in der der Wert von  $d$  bereits fixiert ist, diese beide Größen als Funktionen von  $z$  und  $d$ , so wird durch diese eine Produktwahrscheinlichkeitsverteilung  $p(z,d) = p(z)p(d|z)$  auf  $Z \times D$  induziert.

In [Flo90] wird die einem Bayes'schen Modell zugrundeliegende mathematische Modellierung im abstrakten maßtheoretischen Sinne auf Basis des mit  $p(z,d)$  korrespondierenden Produktwahrscheinlichkeitsraums formuliert. Dieser wird in [Flo90] als Bayes'sches Experiment bezeichnet. Die Betrachtung eines solchen Bayes'schen Experiments ist für die vorliegende Arbeit hilfreich, da daraus deutlich wird, dass es verschiedene Ansatzpunkte gibt, den Gedanken, dass die Bayes'sche Fusion nicht global über ganz  $Z$ , sondern konzentriert auf einen lokalen Kontext  $U \subset Z$  erfolgen soll, umzusetzen. Weiter wird daraus gut erkennbar, wie das aus mathematischer Hinsicht jeweils geschehen kann.

Entsprechende Ansatzpunkte werden in den nachfolgenden Abschnitten genauer herausgearbeitet und untersucht. Hierfür muss im vorliegenden Ab-

schnitt jedoch zuerst die einem Bayes'schen Experiment zugrundeliegende mathematische Modellierung in den Grundzügen eingeführt werden.

Die Angabe der Verteilungsklasse  $\{p(d|z)|z \in Z\}$  korrespondiert mit der Angabe einer Struktur der Form  $(D, \mathcal{D}, \{P_z|z \in Z\})$ . Der Terminologie in [Flo90] folgend wird diese Struktur als statistisches Experiment bezeichnet. Es ist dabei  $(D, \mathcal{D})$  ein messbarer Raum mit einer  $\sigma$ -Algebra  $\mathcal{D}$  über  $D$  (vgl. Definition 3.1). Weiter ist  $\{P_z|z \in Z\}$  eine Menge von Wahrscheinlichkeitsmaßen über  $(D, \mathcal{D})$  (vgl. Definition 3.2).

Die Angabe der A-Priori-Verteilung  $p(z)$  korrespondiert mit der Erweiterung der in Form des statistischen Experiments vorliegenden mathematischen Modellierung durch Angabe eines geeignet gewählten Wahrscheinlichkeitsraums der Form  $(Z, \mathcal{Z}, P)$ . Hierbei sind  $\mathcal{Z}$  eine  $\sigma$ -Algebra über  $Z$  und  $P$  ein Wahrscheinlichkeitsmaß auf dem messbaren Raum  $(Z, \mathcal{Z})$ .

Hieraus lässt sich dann ein Bayes'sches Experiment in Form eines Produktwahrscheinlichkeitsraums festlegen<sup>1</sup>:

$$(Z \times D, \sigma(\mathcal{Z} \times \mathcal{D}), \Pi) . \quad (5.3)$$

Dabei bezeichnet  $\sigma(\mathcal{Z} \times \mathcal{D})$  die kleinste  $\sigma$ -Algebra über  $Z \times D$ , die alle Elemente der Form  $E \times F$  mit  $E \in \mathcal{Z}$  und  $F \in \mathcal{D}$  enthält.  $\Pi$  bezeichnet das durch  $P$  und  $P_z$  festgelegte Produktwahrscheinlichkeitsmaß auf  $(Z \times D, \sigma(\mathcal{Z} \times \mathcal{D}))$ .

Um die in den nachfolgenden Abschnitten eingeführten Ansätze, Lokalität im Bayes'schen Sinne zu formulieren, besser nachvollziehen zu können, ist es wichtig zu verstehen, wie genau in einem probabilistischen Modell festgelegt wird, welche Informationsstände in Bezug auf einen interessierenden Sachverhalt grundsätzlich überhaupt beschrieben werden können. Deshalb wird

<sup>1</sup> Der Vollständigkeit halber sei anmerkt, dass die umgekehrte Zerlegung eines derartigen Produktwahrscheinlichkeitsraums in die mit  $p(z|d)$  und  $p(d)$  korrespondierenden Komponenten nicht immer existiert (vgl. [Flo90]). Im Hinblick auf die in der vorliegenden Arbeit nachfolgend umgesetzte Vorgehensweise ist dieser Aspekt aber nicht relevant, da man davon ausgehen kann, dass diese Zerlegung existiert.

auf diesen Aspekt im nachfolgenden Abschnitt vorab kurz grundlegend eingegangen.

## 5.2.2 Beschreibbare Informationsstände

Es sei  $(\Omega, \mathcal{A}, P)$  ein allgemeiner, gemäß Definition 3.2 festgelegter Wahrscheinlichkeitsraum. Wie aus Definition 3.2 erkennbar ist, ist die Definitionsmenge des Wahrscheinlichkeitsmaßes  $P$  stets durch die zugehörige  $\sigma$ -Algebra  $\mathcal{A}$  vorgegeben. Hieraus ergibt sich direkt, dass diese gerade festlegt, welche Informationsstände im Hinblick auf einen interessierenden Sachverhalt auf Basis einer solchen probabilistischen Modellierung grundsätzlich überhaupt erreichbar sind und auch welche Information in Bezug auf den interessierenden Sachverhalt eingebracht werden kann.

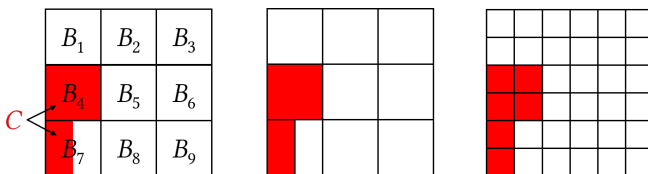
Die Wahrscheinlichkeit  $P(A)$  einer Menge  $A \subset \Omega$  existiert also nur, wenn  $A \in \mathcal{A}$  gilt. Ebenso existiert die bedingte Wahrscheinlichkeit  $P(A|B)$  für Mengen  $A, B \subset \Omega$  nur, wenn neben der Bedingung  $A \in \mathcal{A}$  auch die Bedingung  $B \in \mathcal{A}$  erfüllt ist und zusätzlich  $P(B) > 0$  gilt (vgl. Definition 3.3). Zur Illustration dieser Aussagen seien nachfolgend drei verschiedene Beispiele betrachtet. Es bezeichne hierfür  $\zeta$  in abstrakter Weise einen bestimmten interessierenden Sachverhalt.

Ist  $\mathcal{A}$  die größte  $\sigma$ -Algebra über  $\Omega$ , die überhaupt möglich ist, d. h.  $\mathcal{A} = \{\emptyset, \Omega\}$ , so bedeutet dies, dass man letztlich keinerlei (nichttriviale) Information über den interessierenden Sachverhalt erlangen bzw. einbringen kann. In diesem Fall können nämlich nur die Wahrscheinlichkeiten dafür, ob  $\zeta \in \emptyset$  oder ob  $\zeta \in \Omega$  gilt, angegeben werden. Es gilt notwendigerweise  $P(\emptyset) = 0$  und  $P(\Omega) = 1$ .

Ist  $\Omega$  abzählbar und ist  $\mathcal{A}$  die feinste  $\sigma$ -Algebra über  $\Omega$ , die überhaupt möglich ist, d. h.  $\mathcal{A} = \mathcal{P}(\Omega)$ , wobei  $\mathcal{P}(\Omega)$  die Potenzmenge von  $\Omega$  bezeichnet, so kann unabhängig davon, was der „wahre“ Wert von  $\zeta$  ist, prinzipiell (die gemäß der Festlegung von  $\Omega$ ) maximale Information diesbezüglich erlangt werden. In diesem Fall kann nämlich für jedes Elementarereignis der Form  $\{\omega\}$  mit  $\omega \in \Omega$  die Wahrscheinlichkeit dafür, dass  $\zeta = \omega$  gilt, angegeben werden

und damit auch für jede andere Teilmenge  $A$  von  $\Omega$  die Wahrscheinlichkeit dafür, dass  $\zeta \in A$  gilt. Ebenso kann auch jeder Informationsstand der Form  $\zeta \in A$  durch Berechnung der entsprechenden bedingten Wahrscheinlichkeit eingebracht werden.

Es sei  $\mathcal{B} = \{B_1, \dots, B_M\}$  eine Partition von  $\Omega$  in  $M$  Teilmengen für ein  $M \in \mathbb{N}$ . Betrachtet man für beliebige Teilmengen  $\tilde{\mathcal{B}}$  von  $\mathcal{B}$  die Vereinigung  $\bigcup_{B \in \tilde{\mathcal{B}}} B$ , so bildet die Gesamtheit der so resultierenden Mengen eine  $\sigma$ -Algebra über  $\Omega$  (vgl. [Beh13]). Diese wird als die von  $\mathcal{B}$  erzeugte  $\sigma$ -Algebra bezeichnet. Ergänzt man den dadurch resultierenden messbaren Raum  $(\Omega, \mathcal{A})$  um ein Wahrscheinlichkeitsmaß  $P$ , so stellen die Elemente der Partition  $\mathcal{B}$  die feingranularsten Ereignisse, über welche Wahrscheinlichkeitsaussagen gerade noch möglich sind, dar.



**Abbildung 5.1:** In der linken Teilabbildung ist eine Partition  $\mathcal{B} = \{B_1, \dots, B_9\}$  eines Grundraums  $\Omega$  illustriert. Die rot markierte Teilmenge  $C$  dieses Grundraums ist kein Element der von  $\mathcal{B}$  erzeugten  $\sigma$ -Algebra. Sie ist jedoch ein Element der von der Partition erzeugten  $\sigma$ -Algebra, wenn die Partition so verfeinert wird, dass  $C$  keine echten Teilmengen von Elementen der Partition mehr enthält. Die mittlere und rechte Teilabbildung illustrieren zwei Beispiele für entsprechende Verfeinerungen. Die kleinstmögliche Erweiterung  $\hat{C}$  von  $C$ , die ein Element der von der (ursprünglichen) Partition  $\mathcal{B}$  erzeugten  $\sigma$ -Algebra ist, ist  $B_4 \cup B_7$ .

Ist  $C$  eine Teilmenge von  $\Omega$  mit  $C \cap B_{m_0} \neq \emptyset$  und  $C \cap B_{m_0} \neq B_{m_0}$  für ein  $m_0 \in \{1, \dots, M\}$ , so ist  $P(C)$  nicht definiert. Ebenso ist auch die bedingte Wahrscheinlichkeit  $P(\cdot|C)$  nicht definiert. D.h. es ist nicht möglich, die Wahrscheinlichkeit dafür, dass  $\zeta \in C$  gilt, (exakt) anzugeben und es ist nicht möglich, Wissen dahingehend, dass  $\zeta \in C$  gilt, (in exakter Form) einzubringen. Hierfür wäre zuerst der Übergang zu einer feingranulareren  $\sigma$ -Algebra



$\tilde{\mathcal{A}}$ , bezüglich welcher  $C$  messbar ist, nötig. Dieser Sachverhalt ist in Abb. 5.1 exemplarisch für den Fall  $M = 9$  illustriert. Alternativ könnte man noch  $C$  so erweitern, dass die resultierende Teilmenge  $\tilde{C}$  in  $\mathcal{A}$  enthalten ist, und statt der Aussage  $\zeta \in C$  die (ungenauere) Aussage  $\zeta \in \tilde{C}$  betrachten.

### 5.2.3 Lokalität durch Einschränkung

Eine Möglichkeit, den Gedanken, dass die Bayes'sche Fusion nicht global über ganz  $Z$ , sondern konzentriert auf einen lokalen Kontext  $U \subset Z$  erfolgen soll, umzusetzen, besteht darin, die Annahme, dass sicher  $z \in U$  gilt, in die zugrundeliegende probabilistische Modellierung einzuspeisen. Diese Möglichkeit wird im vorliegenden Abschnitt genauer betrachtet.

#### 5.2.3.1 Lokales Bayes'sches Experiment

Die beim ursprünglichen globalen Bayes'schen (statistischen) Modell angenommene Verteilungsklasse reduziert sich durch die Annahme, dass  $z \in U$  sicher gilt, auf  $\{p(d|z)|z \in U\}$ . Das korrespondierende statistische Modell ist dementsprechend  $(D, \mathcal{D}, \{P_z|z \in U\})$ .

Im Hinblick auf den a priori verfügbaren Informationsstand in Bezug auf den „wahren“ Wert von  $z$  führt die Annahme, dass  $z \in U$  sicher gilt, dazu, dass das Wahrscheinlichkeitsmaß  $P$  auf  $(Z, \mathcal{Z})$  in das bedingte Wahrscheinlichkeitsmaß  $P(\cdot|U)$  überführt werden muss. Hierfür muss der lokale Kontext  $U$  messbar sein, d. h.  $U$  muss derart festgelegt worden sein, dass  $U \in \mathcal{Z}$  gilt, und es muss weiter  $P(U) > 0$  sein.

Aus Definition 3.3 ergibt sich sofort, dass für jedes Ereignis  $A \in \mathcal{Z}$  gilt

$$P(A|U) = \frac{P(A \cap U)}{P(U)}. \quad (5.4)$$

Es ergibt sich insbesondere

$$P(A|U) = \begin{cases} \frac{P(A)}{P(U)} & \text{falls } A \cap U = A \\ 0 & \text{falls } A \cap U = \emptyset . \end{cases} \quad (5.5)$$

Da das Wahrscheinlichkeitsmaß  $P(\cdot|U)$  außerhalb von  $U$  den Wert Null annimmt, lässt sich seine Definitionsmenge entsprechend einschränken. Hierzu wird das Konstrukt einer sog. Spur- $\sigma$ -Algebra benötigt. Diese ist in der Wahrscheinlichkeitstheorie wie folgt definiert:

**Definition 5.1.** *Es seien  $(\Omega, \mathcal{A})$  ein messbarer Raum (gemäß Definition 3.1) und  $\emptyset \neq B \subset \Omega$ . Dann heißt die  $\sigma$ -Algebra*

$$\mathcal{A}|_B := \{A \cap B | A \in \mathcal{A}\} \quad (5.6)$$

die Spur- $\sigma$ -Algebra von  $\mathcal{A}$  in  $B$ .

Bei der Spur- $\sigma$ -Algebra  $\mathcal{A}|_B$  gemäß Definition 5.1 handelt es sich um eine  $\sigma$ -Algebra über  $B$ . Ein Beweis dieser Aussage findet sich z. B. in [Mei05].

Da der lokale Kontext  $U$  als messbar angenommen wird, ergibt sich im konkret vorliegenden Fall

$$\mathcal{Z}|_U = \{A | A \subset U, A \in \mathcal{Z}\} . \quad (5.7)$$

$P(\cdot|U)$  stellt ein Wahrscheinlichkeitsmaß über dem messbaren Raum  $(U, \mathcal{Z}|_U)$  dar. Der a priori verfügbare Informationsstand in Bezug auf den „wahren“ Wert von  $z$  lässt sich beim Übergang zu einem lokalen Bayes'schen Modell, welches die Annahme, dass sicher  $z \in U$  gilt, macht, also modellieren, indem man den Wahrscheinlichkeitsraum  $(Z, \mathcal{Z}, P)$  durch den Wahrscheinlichkeitsraum  $(U, \mathcal{Z}|_U, P(\cdot|U))$  ersetzt.

Das lokale Bayes'sche Experiment ist damit dann wie folgt festgelegt:

$$(U \times D, \sigma(\mathcal{Z}|_U \times \mathcal{D}), \Pi) . \quad (5.8)$$

Analog zum ursprünglichen globalen Bayes'schen Experiment gemäß Gleichung (5.3) bezeichnet  $\sigma(\mathcal{Z}|_U \times \mathcal{D})$  die kleinste  $\sigma$ -Algebra über  $U \times D$ , welche alle Elemente der Form  $E \times F$  mit  $E \in \mathcal{Z}|_U$  und  $F \in \mathcal{D}$  enthält und  $\Pi$  bezeichnet das nun durch  $P(\cdot|U)$  und  $P_z$  festgelegte Produktwahrscheinlichkeitsmaß<sup>1</sup> auf  $(U \times D, \sigma(\mathcal{Z}|_U \times \mathcal{D}))$ .

### 5.2.3.2 Lokale Bayes'sche Fusion

Da sich im globalen Bayes'schen Modell der a priori vorliegende Informationsstand in Bezug auf  $z$  durch die A-Priori-Verteilung  $p(z)$  beschreiben lässt, ergibt sich in mathematischer Hinsicht sofort, dass dieser im lokalen Bayes'schen Modell durch die lokale A-Priori-Verteilung

$$p(z|U) := \frac{p(z)}{P(U)} \quad (5.9)$$

beschrieben werden kann.

Der nachfolgende Satz gibt Antwort auf die Frage, wie auf dieser Basis dann die eigentliche lokale Bayes'sche Fusion erfolgt:

**Satz 5.2.** *Es sei ein lokales Bayes'sches Modell durch die Annahme der Verteilungsklasse  $\{p(d|z)|z \in U\}$  und der lokalen A-Priori-Verteilung  $p(z|U)$  gegeben. Es bezeichne  $p(z|d,U)$  die zugehörige lokale A-Posteriori-Verteilung, welche für  $z \in U$  definiert ist. Diese berechnet sich gemäß der Regel*

$$p(z|d,U) = \frac{p(d|z)p(z|U)}{\int_U p(d|z)p(z|U) dz} \propto l(d|z)p(z|U). \quad (5.10)$$

und es gilt

$$p(z|d,U) = \frac{p(z|d)}{P(U|d)}. \quad (5.11)$$

<sup>1</sup> Da das Produktwahrscheinlichkeitsmaß im weiteren Verlauf der Arbeit nicht mehr benötigt wird, wird der Einfachheit halber keine separate Notation für dieses im lokalen Bayes'schen Experiment eingeführt.

*Beweis.* In Gleichung (5.10) ist gerade die Anwendung des Satzes von Bayes im lokalen Bayes'schen Modell zur Berechnung der (hier) lokalen A-Posteriori-Verteilung formuliert. Vgl. Gleichung (3.5), Gleichung (3.6) und Gleichung (3.7) im Hinblick das Analogon im globalen Bayes'schen Modell.

Hiervon ausgehend berechnet man unter Nutzung von Gleichung (5.9) leicht, dass gilt

$$p(z|d,U) = \frac{p(d|z)p(z|U)}{\int_U p(d|z)p(z|U) dz} \quad (5.12)$$

$$= \frac{p(d|z)p(z)}{\int_U p(d|z)p(z) dz} \quad (5.13)$$

$$= \frac{p(z|d)}{\int_U p(z|d) dz} \quad (5.14)$$

$$= \frac{p(z|d)}{P(U|d)}. \quad (5.15)$$

□

### 5.2.3.3 Bezug zur globalen Bayes'schen Fusion

Das lokale Bayes'sche Modell ergibt sich in eindeutiger Weise aus dem lokalen Kontext  $U \subset Z$ , auf den die Einschränkung erfolgt, und den Bestandteilen des globalen Bayes'schen Modells, welche sich tatsächlich auf  $U$  beziehen: Die Likelihood-Funktion im lokalen Bayes'schen Modell ist identisch mit der Likelihood-Funktion  $l(d|z)$  im globalen Bayes'schen Modell. Der einzige Unterschied besteht darin, dass sie nur für  $z \in U$  bestimmt und ausgewertet werden muss. Es sei angemerkt, dass diese Aussage offensichtlich auch speziell für die (bezüglich  $d$ ) normierte Version  $p(d|z)$  der Likelihood-Funktion gilt. Weiter ergibt sich beim Übergang vom globalen zum lokalen Bayes'schen Modell die lokale A-Priori-Verteilung  $p(z|U)$  durch das Einschränken der globalen A-Priori-Verteilung  $p(z)$  auf den lokalen Kontext  $U$ . Will man diesen Übergang mathematisch abbilden, so ist die globale A-Priori-Verteilung  $p(z)$  für  $z \in U$  zu bestimmen. Um sie gemäß Gleichung (5.9) auf  $U$  umzunormie-

ren, ist weiter die Kenntnis der globalen A-Priori-Wahrscheinlichkeit  $P(U)$  des lokalen Kontexts  $U$  nötig.

Tatsächlich muss die Bestimmung der lokalen A-Priori-Verteilung  $p(z|U)$  allerdings nicht notwendigerweise wie zuvor beschrieben ausgehend von der globalen A-Priori-Verteilung  $p(z)$  erfolgen. Sie kann auch durch die direkte Anwendung eines Mechanismus gemäß Abschnitt 3.5.1, um den bezüglich  $z$  a priori vorliegenden Informationsstand – nun nur im Hinblick auf  $U$  – in eine (objektive) Wahrscheinlichkeitsverteilung (im Sinne der Degree-of-Belief-Interpretation von Wahrscheinlichkeit) zu überführen, erfolgen. Dabei wird die globale A-Priori-Wahrscheinlichkeit  $P(U)$  des lokalen Kontexts  $U$  dann nicht explizit berechnet, was letztlich bedeutet, dass sie im lokalen Bayes'schen Modell per se nicht als bekannt angenommen werden kann.

Innerhalb des lokalen Bayes'schen Modells ist gemäß der Festlegung von  $\mathcal{Z}|_U$  der Informationsstand in Bezug auf eine Menge  $A \subset U$  probabilistisch beschreibbar, wenn dieser auch im globalen Bayes'schen Modell probabilistisch beschreibbar war, d. h. wenn  $A \in \mathcal{Z}$  gilt (vgl. Gleichung (5.7)). Allerdings erfolgt dies nun auf Basis der lokalen Wahrscheinlichkeiten. Global vorhandene Ereignisse  $A \subset \bar{U} = Z \setminus U$  sind kein Teil des lokalen Bayes'schen Modells, ihre Existenz wird (lokal) völlig ignoriert. Informationsstände im Hinblick auf entsprechende globale Ereignisse  $A \in \mathcal{Z}$  lassen sich im lokalen Bayes'schen Modell also weder beschreiben noch berücksichtigen. Dies steht im Einklang mit der Tatsache, dass die Bildung des lokalen Bayes'schen Modells auf der Annahme, dass sicher  $z \in U$  gilt, beruhte. Gilt für ein global vorhandenes Ereignis, dass es sowohl Anteile aus  $U$  als auch aus  $\bar{U}$  enthält, so muss im lokalen Bayes'schen Modell der in  $Z \setminus U$  enthaltene Anteil ignoriert werden. Gilt z. B.  $A = \{z^*, z^{**}\}$  mit  $z^* \in U, z^{**} \notin U$ , so kann im lokalen Bayes'schen Modell also statt  $A$  nur  $\bar{A} = \{z^*\}$  berücksichtigt werden.

Im Folgenden soll nun noch der Zusammenhang zwischen lokalen Wahrscheinlichkeitsaussagen und globalen Wahrscheinlichkeitsaussagen näher beleuchtet und hinsichtlich der resultierenden Konsequenzen untersucht werden.

Grundsätzlich ist der Grad des Dafürhaltens in ein Ereignis  $A$  umso größer anzusehen, je höher die Wahrscheinlichkeit dieses Ereignisses ist. Die globale A-Posteriori-Wahrscheinlichkeitsverteilung  $p(z|d)$  induziert auf dieser Basis eine Präferenzrelation  $\succsim$  auf den Ereignissen in  $\mathcal{Z}$ :

$$A \succsim B \Leftrightarrow \int_A p(z|d) dz \geq \int_B p(z|d) dz, \quad A, B \in \mathcal{Z}. \quad (5.16)$$

Ebenso induziert die lokale A-Posteriori-Wahrscheinlichkeitsverteilung  $p(z|d,U)$  eine Präferenzrelation  $\succsim_{|U}$  auf den Ereignissen in  $\mathcal{Z}_{|U}$ :

$$A \succsim_{|U} B \Leftrightarrow \int_A p(z|d,U) dz \geq \int_B p(z|d,U) dz, \quad A, B \in \mathcal{Z}_{|U}. \quad (5.17)$$

Da  $p(z|d)$  und  $p(z|d,U)$  bis auf den Faktor  $P(U|d)$  identisch sind, sind die (lokale) Präferenzordnung  $\succsim_{|U}$  und die (globale) Präferenzordnung  $\succsim$  konsistent. D. h. es gilt für Ereignisse  $A, B$  in  $\mathcal{Z}_{|U}$  (welche wegen der  $\mathcal{Z}$ -Messbarkeit von  $U$  auch in  $\mathcal{Z}$  enthalten sind):

$$A \succsim_{|U} B \Leftrightarrow A \succsim B. \quad (5.18)$$

Die analoge Aussage gilt auch a priori, da sich die globale A-Priori-Verteilung  $p(z)$  und die lokale A-Priori-Verteilung  $p(z|U)$  gerade um den Faktor  $P(U)$  unterscheiden. Ist der lokale Degree-of-Belief in ein Ereignis  $A \in \mathcal{Z}_{|U}$  also (a priori oder a posteriori) höher als der lokale Degree-of-Belief in ein Ereignis  $B \in \mathcal{Z}_{|U}$ , so weiß man, dass dies ebenso im globalen Bayes'schen Modell gilt.

Da der Faktor, um welchen sich der Wert von  $p(z)$  bzw.  $p(z|d)$  beim Übergang vom globalen zum lokalen Bayes'schen Modell ändert, gleich für alle  $z \in U$  ist, gilt weiter, dass die Prior-Odds sowie die Posterior-Odds von Ereignissen  $A, B \in \mathcal{Z}_{|U}$  invariant unter dem Übergang sind, d. h. es gelten:

$$\underbrace{\frac{\int_A p(z) dz}{\int_B p(z) dz}}_{\text{Globale Prior-Odds}} = \underbrace{\frac{\int_A p(z|U) dz}{\int_B p(z|U) dz}}_{\text{Lokale Prior-Odds}}, \quad (5.19)$$

und

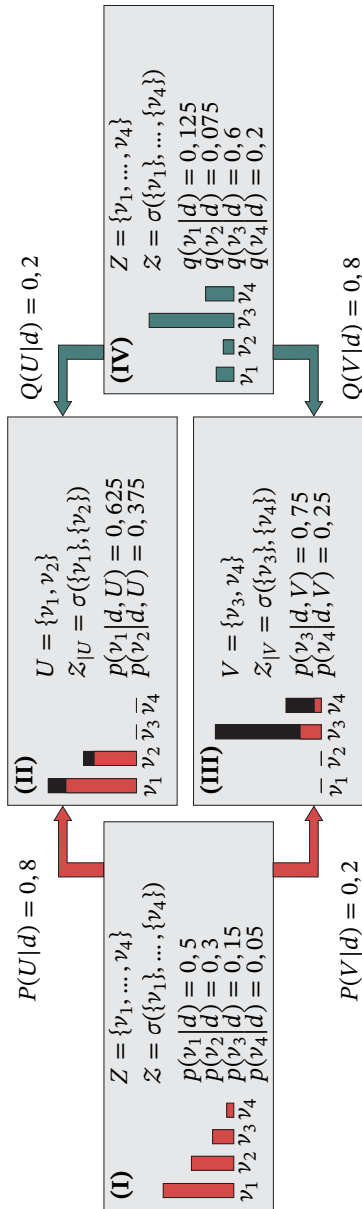
$$\underbrace{\frac{\int_A p(z|d) dz}{\int_B p(z|d) dz}}_{\text{Globale Posterior-Odds}} = \underbrace{\frac{\int_A p(z|d, U) dz}{\int_B p(z|d, U) dz}}_{\text{Lokale Posterior-Odds}}. \quad (5.20)$$

Ist also der lokale Degree-of-Belief in ein Ereignis  $A \in \mathcal{Z}_{|U}$  (a priori oder a posteriori)  $c$ -mal der lokale Degree-of-Belief in ein Ereignis  $B \in \mathcal{Z}_{|U}$ , so weiß man, dass dies ebenso im globalen Bayes'schen Modell gilt.

Die Schwierigkeit besteht allerdings darin, dass sich allein aus der Kenntnis des lokalen Bayes'schen Modells keine Aussage darüber treffen lässt, wie hoch der Degree-of-Belief eines lokal betrachteten Ereignisses im globalen Bayes'schen Modell absolut gesehen ist. Dieser Sachverhalt und die daraus resultierenden Konsequenzen sind in Abb. 5.2 für den Fall  $Z = \{\nu_1, \dots, \nu_4\}$  illustriert. Als zugehörige  $\sigma$ -Algebra ist dabei  $\mathcal{Z} = \mathcal{P}(Z)$  angenommen. Weiter sind im in Abb. 5.2 betrachteten Beispiel  $U = \{\nu_1, \nu_2\}$  und  $V = \{\nu_3, \nu_4\}$  zwei unterschiedliche, hier exemplarisch gewählte lokale Kontexte, auf die die lokale Bayes'sche Fusion jeweils eingeschränkt erfolgt(e).

In Abschnitt (I) von Abb. 5.2 ist die A-Posteriori-Verteilung  $p(z|d)$  im globalen Bayes'schen Modell dargestellt. Diese wird durch die Annahme, dass sicher  $z \in U$  gilt, letztlich in die in Abschnitt (II) dargestellte lokale A-Posteriori-Verteilung  $p(z|d, U)$  überführt. Aus der Annahme, dass sicher  $z \in V$  gilt, resultiert dagegen letztlich die in Abschnitt (III) dargestellte lokale A-Posteriori-Verteilung  $p(z|d, V)$ .

Die Annahme, dass  $z$  sicher im jeweiligen lokalen Kontext  $U$  bzw.  $V$  enthalten ist, führt dazu, dass die Wahrscheinlichkeit der Ereignisse, welche jeweils lokal betrachtet werden, lokal überschätzt wird. Das Ausmaß, in welchem das beim Übergang von  $p(z|d)$  zu  $p(z|d, U)$  bzw.  $p(z|d, V)$  passiert, ist in Abschnitt (II) und Abschnitt (III) von Abb. 5.2 für die im jeweiligen lokalen Kontext enthaltenen Elementarereignisse  $\{\nu_i\}$  anhand der schwarz markierten Abschnitte in den Balken, welche deren lokale A-Posteriori-Wahrscheinlichkeiten illustrieren, dargestellt.



**Abbildung 5.2:** Illustration des Übergangs (a posteriori) zum lokalen Bayes'schen Modell bei Umsetzung der Lokalität durch Einschränkung. In Abschnitt (I) und Abschnitt (IV) sind zwei mögliche globale A-Posteriori-Verteilungen angeführt. In Abschnitt (II) und Abschnitt (III) sind zwei lokale A-Posteriori-Verteilungen angeführt, welche bei unterschiedlicher Wahl des lokalen Kontexts daraus resultieren.



Da der Übergang von der globalen A-Posteriori-Verteilung zur lokalen A-Posteriori-Verteilung (für Ereignisse, welche im lokalen Kontext enthalten sind) anhand einer Division durch die A-Posteriori-Wahrscheinlichkeit des lokalen Kontexts erfolgt, ist diese Überschätzung absolut gesehen umso größer, je unwahrscheinlicher der lokale Kontext global gesehen ist. Da im in Abb. 5.2 dargestellten Beispiel die A-Posteriori-Wahrscheinlichkeit  $P(U|d) = 0,8$  des lokalen Kontexts  $U$  eher hoch ist, ist die Überschätzung bei der in Abschnitt (II) dargestellten, zugehörigen lokalen A-Posteriori-Verteilung  $p(z|d,U)$  eher gering. Anders ist die Situation jedoch bei der lokalen A-Posteriori-Verteilung  $p(z|d,V)$ , welche in Abschnitt (III) dargestellt ist. Da hier die A-Posteriori-Wahrscheinlichkeit des lokalen Kontexts  $V$  mit  $P(V|d) = 0,2$  eher gering ist, findet eine deutlich ausgeprägte Überschätzung statt.

Da aus einem lokalen Bayes'schen Modell nicht in eindeutiger Weise auf das zugehörige globale Bayes'sche Modell zurückgeschlossen werden kann, wäre in dem in Abb. 5.2 dargestellten Beispiel z. B. auch die in Abschnitt (IV) angegebene globale A-Posteriori-Verteilung  $q(z|d)$  kompatibel mit den beiden in Abschnitt (II) und Abschnitt (III) dargestellten lokalen A-Posteriori-Verteilungen, d. h. es gilt  $p(z|d,U) = q(z|d,U)$  und  $p(z|d,V) = q(z|d,V)$ . Anders als beim Übergang von Abschnitt (I) zu Abschnitt (II) und Abschnitt (III) ist in diesem Fall der lokale Kontext  $U$  mit  $Q(U|d) = 0,2$  global gesehen eher unwahrscheinlich, während der lokale Kontext  $V$  mit  $Q(V|d) = 0,8$  global gesehen eher wahrscheinlich ist. D. h. in diesem Fall ist die Überschätzung der lokalen A-Posteriori-Wahrscheinlichkeiten in Abschnitt (II) recht ausgeprägt, während sie in Abschnitt (III) eher gering ausfällt; dieser Sachverhalt ist in Abb. 5.2 nicht in den Balken in Abschnitt (II) und Abschnitt (III) illustriert.

Aus der Kenntnis der in Abschnitt (II) von Abb. 5.2 dargestellten lokalen A-Posteriori-Verteilung ergibt sich, dass auch global gesehen das Ereignis  $\{\nu_1\}$  a posteriori um den Faktor  $5/3$  mal wahrscheinlicher ist als das Ereignis  $\{\nu_2\}$ . Ebenso ergibt sich aus der Kenntnis der in Abschnitt (III) dargestellten lokalen A-Posteriori-Verteilung, dass auch global gesehen das Ereignis  $\{\nu_3\}$  a posteriori um den Faktor 3 mal wahrscheinlicher ist als das Ereignis  $\{\nu_4\}$ . Wie anhand der beiden in Abschnitt (I) und Abschnitt (IV) dargestellten möglichen globalen A-Posteriori-Verteilungen erkennbar, ermöglicht dieses Wissen jedoch

per se keine Aussage darüber, ob der jeweilige lokale Maximum-A-Posteriori-Schätzwert  $\nu_1$  bzw.  $\nu_3$  global gesehen tatsächlich ebenfalls besonders relevant ist. Eine entsprechende Aussage wäre nur möglich, wenn man einschätzen könnte ob die globale A-Posteriori-Wahrscheinlichkeit des lokalen Kontexts  $U$  bzw.  $V$  tatsächlich auch entsprechend hoch ist. Da im vorliegenden Beispiel  $Z = U \cup V$  ist, weiß man, dass der globale Maximum-A-Posteriori-Schätzwert entweder  $\nu_1$  oder  $\nu_3$  sein muss. Würde  $Z$  noch zusätzliche Elemente enthalten, welche weder in  $U$  noch in  $V$  enthalten ist, so könnte grundsätzlich nicht ausgeschlossen werden, dass der globale Maximum-A-Posteriori-Schätzwert im nicht betrachteten Anteil  $Z \setminus (U \cup V)$  enthalten ist.

Die zuvor anhand des in Abb. 5.2 dargestellten Beispiels beschriebene grundsätzliche Schwierigkeit im Hinblick auf die Einschätzung der (absoluten) Relevanz von Ergebnissen der lokalen Bayes'schen Fusion, welche eingeschränkt auf einen lokalen Kontext  $U \subset Z$  erfolgt, ist in Abb. 5.3 ergänzend noch einmal für den kontinuierlichen, zweidimensionalen Fall illustriert.

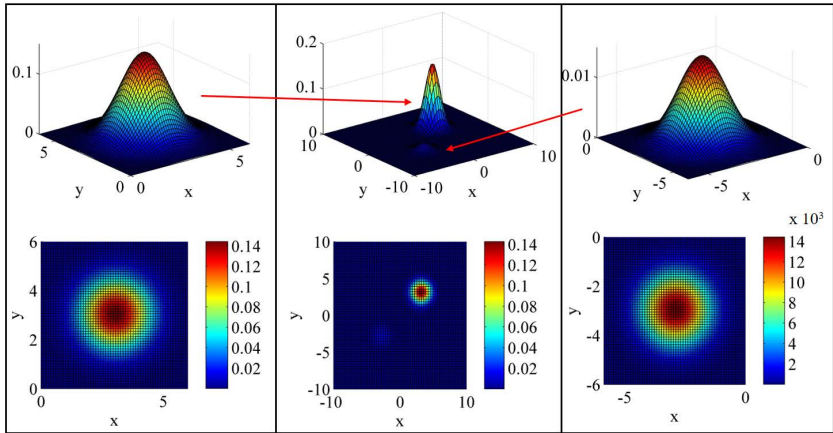
Die in der Mitte von Abb. 5.3 illustrierte globale A-Posteriori-Verteilung  $p(z_1, z_2 | d)$  stellt eine Mixtur zweier Normalverteilungen dar. Genauer gilt

$$p(z_1, z_2 | d) = \sum_{i=1}^2 \omega_i q_i(z_1, z_2). \quad (5.21)$$

Dabei ist  $q_1(z_1, z_2)$  gemäß  $\mathcal{N}((3,3)^T, I_2)$  und  $q_2(z_1, z_2)$  ist weiter gemäß  $\mathcal{N}((-3,-3)^T, I_2)$  verteilt, wobei  $I_2$  die zweidimensionale Einheitsmatrix bezeichnet. Die Gewichte in der Mixtur betragen  $\omega_1 = 10/11$  und  $\omega_2 = 1/11$ .

Auf der rechten und linken Seite von Abb. 5.3 sind zwei lokale A-Posteriori-Verteilungen illustriert, welche sich wiederum aus unterschiedlichen Festlegungen des lokalen Kontexts ergeben. Auf der linken Seite korrespondiert der lokale Kontext mit einem quadratischen, um das globale Maximum  $(3,3)^T$  von  $p(z_1, z_2 | d)$  zentrierten Bereich. Auf der rechten Seite korrespondiert der lokale Kontext mit einem analogen quadratischen Bereich, welcher hier jedoch um das sehr gering ausgeprägte lokale Maximum  $(-3,-3)^T$  gelegt wurde. Aus Abb. 5.3 ist gut erkennbar, dass sich rein auf Basis der (jeweils auf den lokalen Kontext eingeschränkten) lokalen A-Posteriori-Verteilung nicht erkennen

lässt, ob man in Form des lokalen Kontexts tatsächlich einen (global) relevanten Bereich von  $Z$  adressiert oder nicht.



**Abbildung 5.3:** Die in der Mitte dargestellte globale A-Posteriori-Verteilung ist auf der linken und auf der rechten Seite der Abbildung jeweils auf einen lokalen Kontext eingeschränkt. Während im links dargestellten Fall der lokale Kontext tatsächlich mit einem global gesehen äußerst relevanten Bereich korrespondiert, gilt dies im rechts dargestellten Fall nicht.

In Bezug auf eine Darstellung möglicher Fehlschlüsse, welche resultieren können, wenn ein Bayes'sches Modell bestimmte, a posteriori relevante Hypothesen nicht abbildet, sei ergänzend noch auf die Referenz [Fit08] verwiesen. Diese betrachtet das Thema im Hinblick auf die Anwendung Bayes'scher Inferenz für Aufgabenstellungen der Wissenschaftsgeschichte. Die hierzu aufgezeigte Problematik ist rein vom mathematischen Standpunkt aus gesehen verwandt mit der in der vorliegenden Arbeit aufgezeigten. Sie ist im Detail jedoch anders ausgeprägt, konkret wird untersucht, unter welchen Bedingungen a priori unplausible wissenschaftliche Theorien als mögliche Hypothesen vorab ausgeschlossen werden können. Dieser Zielsetzung entsprechend basieren die in [Fit08] dazu angestellten Untersuchungen auf anders gelagerten

mathematischen Ansatzpunkten, als es in der vorliegenden Arbeit, genauer in Kapitel 6 und Kapitel 7, der Fall sein wird.

#### 5.2.3.4 Fazit

Die Umsetzung des Gedankens, dass die Bayes'sche Fusion nicht global über ganz  $Z$ , sondern konzentriert auf einen lokalen Kontext  $U \subset Z$  erfolgen soll, auf Basis der Annahme, dass sicher  $z \in U$  gilt, lässt sich im Hinblick auf die Bestimmung der hierfür benötigten Größen, d. h. lokale A-Priori-Verteilung  $p(z|U)$  und Likelihood-Funktion  $l(d|z)$  für  $z \in U$ , methodisch gesehen recht einfach umsetzen. Ebenso lässt sich die eigentliche lokale Bayes'sche Fusion mittels des in Gleichung (5.10) angegebenen Berechnungsschemas unkompliziert durchführen.

Lokale berechnete Angaben hinsichtlich der Wahrscheinlichkeit von (global vorhandenen) Ereignissen, welche im lokalen Kontext enthalten sind, sind relativ zueinander genommen auch aus globaler Sicht vollständig korrekt. Die Absolut-Werte von lokal berechneten Wahrscheinlichkeiten entsprechen jedoch nicht denen, welche man auf Basis des globalen Bayes'schen Modells erhalten würde. Sie sind global gesehen nur dann näherungsweise korrekt, wenn sichergestellt werden kann, dass die Wahrscheinlichkeit des lokalen Kontexts  $U$  im globalen Bayes'schen Modell nahe Eins liegt. Diese Aussagen gelten gleichermaßen a priori und a posteriori.

Da in diesem Sinne rein auf Basis des lokalen Bayes'schen Modells nicht auf die (absolute) globale Relevanz der lokal betrachteten Ereignisse geschlossen werden kann, ist es wichtig, dass Mechanismen, um den lokalen Kontext  $U$  geeignet festzulegen, verfügbar sind. Wünschenswert wären auch Möglichkeiten, die Adäquatheit des lokalen Kontexts  $U$  bzw. des resultierenden lokalen Bayes'schen Modells tatsächlich (möglichst quantitativ) bewerten zu können.

#### 5.2.4 Lokalität durch Vergrößerung

Eine andere Möglichkeit, den Gedanken, dass die Bayes'sche Fusion nicht global über ganz  $Z$ , sondern konzentriert auf einen lokalen Kontext  $U \subset Z$  er-

folgen soll, umzusetzen, besteht darin, die probabilistische Modellierung so auszugestalten, dass im Hinblick auf  $U$  eine genaue, d. h. detaillierte, Fusion erfolgt, der Anteil  $\bar{U} := Z \setminus U$  dagegen nur vergrößert betrachtet wird.

Prinzipiell wäre es denkbar,  $\bar{U}$  hierzu in  $m \in \mathbb{N}$  unterschiedliche Teilmengen mit  $m \ll |\bar{U}|$  aufzuteilen. Nachfolgend wird der naheliegendste Fall  $m = 1$  betrachtet, eine Adaption der Ergebnisse auf den Fall  $m > 1$  ist grundsätzlich möglich.

In der Fachliteratur finden sich speziell für das Anwendungsgebiet Bayes'sche Netze (vgl. z. B. [Jen07]) recht umfangreiche Forschungsergebnisse im Hinblick auf die Vergrößerung und auch Verfeinerung der möglichen Zustände von Zufallsvariablen. Konkret angeführt seien die Referenzen [Cha90, Liu02, Liu98]. Bis auf die grundlegenden mathematischen Formeln im Hinblick auf die Konstitution der Vergrößerungen auf Basis der ursprünglichen Wahrscheinlichkeitsverteilungen sind diese Arbeiten allerdings nicht im engeren Bezug zur vorliegenden Arbeit zu sehen. Zum einen ergeben sich durch das anders gelagerte Anwendungsgebiet andere (Bayes'sche) Modelle und anders gelagerte Anforderungen an die benötigten Verfahren. Zu anderem ist der in der vorliegenden Arbeit gewählte formal wahrscheinlichkeitstheoretische Ausgangspunkt der Betrachtungen ein anderer.

#### 5.2.4.1 Lokales Bayes'sches Experiment

Die anvisierte vergrößerte Sichtweise im Hinblick auf  $\bar{U}$  lässt sich umsetzen, indem man das lokale Bayes'sche Experiment wie folgt festlegt:

$$(Z \times D, \sigma(\sigma(\mathcal{Z}_{|U}, \bar{U}) \times \mathcal{D}), \Pi). \quad (5.22)$$

Dabei bezeichnet  $\sigma(\mathcal{Z}_{|U}, \bar{U})$  die kleinste  $\sigma$ -Algebra, welche sowohl die Bestandteile der bereits in Abschnitt 5.2.3 verwendeten Spur- $\sigma$ -Algebra  $\mathcal{Z}_{|U}$  von  $\mathcal{Z}$  in  $U$  und zusätzlich das Komplement  $\bar{U}$  des lokalen Kontexts  $U$  als atomares Element enthält. Wieder analog zum ursprünglichen (globalen) Bayes'schen Experiment gemäß Gleichung (5.3) bezeichnet  $\sigma(\sigma(\mathcal{Z}_{|U}, \bar{U}) \times \mathcal{D})$  dann die kleinste  $\sigma$ -Algebra über  $Z \times D$ , welche alle Elemente der Form  $E \times F$  mit

$E \in \sigma(\mathcal{Z}|_U, \bar{U})$  und  $F \in \mathcal{D}$  enthält. Außerdem bezeichnet  $\Pi$  das resultierende Produktwahrscheinlichkeitsmaß, welches nun über der Menge der Ereignisse aus  $\sigma(\sigma(\mathcal{Z}|_U, \bar{U}) \times \mathcal{D})$  definiert ist. Wie bereits in Abschnitt 5.2.3.1 wird angenommen, dass der lokale Kontext  $U$  messbar bezüglich  $\mathcal{Z}$  ist, d. h. dass  $U \in \mathcal{Z}$  gilt, und dass  $P(U) > 0$  ist.

Das mit der A-Priori-Verteilung  $p(z)$  korrespondierende Wahrscheinlichkeitsmaß  $P$  auf  $(Z, \mathcal{Z})$  wird dann beim Übergang zum lokalen Bayes'schen Experiment nur insoweit verändert, als dass seine Definitionsmenge eingeschränkt wird auf  $\sigma(\mathcal{Z}|_U, \bar{U})$ . Die Wahrscheinlichkeiten aller dabei dann weiterhin betrachteten Ereignisse  $A \in \sigma(\mathcal{Z}|_U, \bar{U})$  werden dabei nicht verändert. Unter Zuhilfenahme von Gleichung (5.7) erkennt man, dass für ein in  $\sigma(\mathcal{Z}|_U, \bar{U})$  enthaltenes Ereignis  $A$  entweder  $A \subset U$  mit  $A \in \mathcal{Z}$  oder  $A = B \cup \bar{U}$  mit  $B \subset U$  und  $B \in \mathcal{Z}$  gilt. Dies bedeutet, dass Wahrscheinlichkeitsaussagen bezüglich aller in  $U$  enthaltenen Ereignisse also weiterhin in exakter Form möglich sind, der Anteil  $\bar{U}$  von  $Z$  dagegen nur vergrößert (im Sinne von: kumuliert) betrachtet wird.

Die beim ursprünglichen (globalen) Bayes'schen (statistischen) Modell angenommene Verteilungsklasse  $\{p(d|z)|z \in Z\}$  muss hierfür dann nur für  $z \in U$  detailliert ausgewertet werden. Für  $z \in \bar{U}$  wird sie im Rahmen der lokalen Bayes'schen Fusion nur vergrößert, d. h. in Bezug auf die Annahme  $z \in \bar{U}$  insgesamt, ausgewertet. In diesem Sinne kann das korrespondierende statistische Modell technisch gesehen modifiziert werden zu  $(D, \mathcal{D}, \{P_z|z \in U\} \cup \{P_{\bar{U}}\})$ . Dabei muss das Wahrscheinlichkeitsmaß  $P_{\bar{U}}$  so beschaffen sein, dass es die Wahrscheinlichkeit der Ereignisse  $A \in \mathcal{D}$  unter der Hypothese, dass  $z \in \bar{U}$  gilt, widerspiegelt.

#### 5.2.4.2 Lokale Bayes'sche Fusion

Da im globalen Bayes'schen Modell der a priori verfügbare Informationsstand in Bezug auf  $z$  durch die A-Priori-Verteilung  $p(z)$  beschrieben wird, ergibt sich, dass sich dieser im lokalen Bayes'schen Modell, welches eine kumulierte Betrachtung von  $\bar{U}$  umsetzt, durch folgende Größen beschreiben lässt:

$$p(z) \quad (\text{nur im Detail betrachtet}) \text{ f\"ur } z \in U, \quad (5.23)$$

$$P(\bar{U}) = \int_{\bar{U}} p(z) dz. \quad (5.24)$$

Im Rahmen der lokalen Bayes'schen Fusion m\"ussen diese, den a priori vorliegenden Informationsstand beschreibenden Gr\"o\ss en dann in die entsprechenden, den a posteriori vorliegenden Informationsstand beschreibenden Gr\"o\ss en \u00fcberf\"uhrt werden. Berechnet werden m\"ussen also

$$p(z|d) \quad (\text{nur im Detail betrachtet}) \text{ f\"ur } z \in U, \quad (5.25)$$

$$P(\bar{U}|d) = \int_{\bar{U}} p(z|d) dz. \quad (5.26)$$

Der nachfolgende Satz gibt analog zu Satz 5.2 Antwort auf die Frage, wie die Durchf\"uhrung der eigentlichen lokalen Bayes'sche Fusion dann erfolgt:

**Satz 5.3.** *Ein globales Bayes'sches Modell, welches durch die Annahme der Verteilungsklasse  $\{p(d|z)|z \in Z\}$  und der (globalen) A-Priori-Verteilung  $p(z)$  gegeben ist, werde lokal beschrieben, indem der Informationsstand im Hinblick auf  $\bar{U}$  nur noch kumuliert betrachtet wird. Dann gilt:*

$$p(z|d) = \frac{p(d|z)p(z)}{p(d)} \propto l(d|z)p(z) \quad \text{f\"ur } z \in U, \quad (5.27)$$

$$P(\bar{U}|d) = \frac{p(d|\bar{U})p(\bar{U})}{p(d)} \propto l(d|\bar{U})p(\bar{U}). \quad (5.28)$$

mit

$$p(d|\bar{U}) := \frac{\int_{\bar{U}} p(d|z)p(z) dz}{P(\bar{U})}, \quad (5.29)$$

$$l(d|\bar{U}) := \frac{\int_{\bar{U}} l(d|z)p(z) dz}{P(\bar{U})}. \quad (5.30)$$

Die Proportionalit\"atsbeziehungen in Gleichung (5.27) und Gleichung (5.28) sind

dabei so aufzufassen, dass insgesamt gelten muss

$$\int_U p(z|d) dz + P(\bar{U}|d) = 1. \quad (5.31)$$

Es sei angemerkt, dass im Hinblick auf das Anwendungsgebiet Bayes'sche Netze das Analogon zu Gleichung (5.29) bzw. Gleichung (5.30) z. B. in [Liu98] formuliert und als Basis für weitere Arbeiten genommen wird.

*Beweis.* Gemäß des Satzes von Bayes gilt für alle  $z \in Z$ :

$$p(z|d) = \frac{p(d|z)p(z)}{p(d)} \propto l(d|z)p(z). \quad (5.32)$$

Vgl. Gleichung (3.5), Gleichung (3.6) und Gleichung (3.7). Gleichung (5.27) ist gerade diese Beziehung formuliert für  $z \in U$ .

Mit Gleichung (5.29) und Gleichung (5.30) gilt weiter:

$$P(\bar{U}|d) = \int_{\bar{U}} p(z|d) dz \quad (5.33)$$

$$= \int_{\bar{U}} \frac{p(d|z)p(z)}{p(d)} dz \quad (5.34)$$

$$= \frac{p(d|\bar{U})P(\bar{U})}{p(d)} \quad (5.35)$$

$$\propto l(d|\bar{U})P(\bar{U}). \quad (5.36)$$

□

### 5.2.4.3 Bezug zur globalen Bayes'schen Fusion

Anhand der in Gleichung (5.23) und Gleichung (5.24) bzw. in Gleichung (5.25) und Gleichung (5.26) angegebenen Größen wird im lokalen Bayes'schen Modell der Informationsstand im Hinblick auf den „wahren“ Wert von  $z$  a priori bzw. a posteriori probabilistisch exakt so abgebildet, wie es im globalen



Bayes'schen Modell der Fall ist. Alle A-Priori-Wahrscheinlichkeiten und A-Posteriori-Wahrscheinlichkeiten, welche im lokalen Bayes'schen Modell definiert sind, sind also identisch zu denen im globalen Bayes'schen Modell. Jedoch wird der Informationsstand im Hinblick auf  $\bar{U}$  nur noch kumuliert probabilistisch adressiert.

Das lokale Bayes'sche Modell ergibt sich in eindeutiger Weise aus dem lokalen Kontext  $U$  und dem globalen Bayes'schen Modell. Anders als in Abschnitt 5.2.3, wo die Lokalität durch eine vollständige Einschränkung auf den lokalen Kontext  $U$  modelliert wurde, sind hier nun aber nicht mehr nur die Anteile des globalen Bayes'schen Modells relevant, welche sich auf den lokalen Kontext  $U$  beziehen. Genauer gesagt müssen zur Durchführung der lokalen Bayes'schen Fusion Likelihood-Funktion  $l(d|z)$  und der entsprechende Ausschnitt der (globalen) A-Priori-Verteilung  $p(z)$  für  $z \in U$  bekannt sein. Weiter müssen im Hinblick auf  $\bar{U}$  die globale A-Priori-Wahrscheinlichkeit  $P(\bar{U})$  des Komplements  $\bar{U}$  des lokalen Kontexts  $U$  und der Wert von  $l(d|\bar{U})$  bekannt sein bzw. bestimmt werden.

Die Bestimmung der globalen A-Priori-Verteilung  $p(z)$  für  $z \in U$  und der A-Priori-Wahrscheinlichkeit  $P(\bar{U})$  des Komplements  $\bar{U}$  des lokalen Kontexts  $U$  können prinzipiell durch die Anwendung eines Mechanismus gemäß Abschnitt 3.5.1, um den a priori im Hinblick auf  $z$  vorliegenden Informationsstand in eine (objektive) probabilistische Repräsentation (im Sinne der Degree-of-Belief-Interpretation von Wahrscheinlichkeit) zu überführen, erfolgen. Gemäß der Struktur des lokalen Bayes'schen Modells kann dabei letztlich zum einen Information  $\bar{I}$  berücksichtigt werden, welche sich darauf bezieht, wie die Wahrscheinlichkeitsmasse zwischen  $U$  und  $\bar{U}$  aufgeteilt ist. Zum anderem kann bezüglich  $U$ , d. h. für  $z \in U$ , auch detailliertere Information  $I$  im Hinblick auf den „wahren“ Wert der durch  $z$  verkörperten Größen des Interesses eingebracht werden. Um das in Abschnitt 3.5.1 detaillierter behandelte Prinzip der Maximalen Entropie möglichst effizient anzuwenden, kann man dazu wie folgt in zwei Schritten vorgehen: Mittels Anwendung des Prinzips der Maximalen Entropie und unter Berücksichtigung des in Form von  $\bar{I}$  vorliegenden Informationsstands wird die Maximum-Entropie-Verteilung über der Partition  $\{U, \bar{U}\}$  von  $Z$  bestimmt. Hieraus resultieren

dann gerade die Werte von  $P(U)$  und  $P(\bar{U}) = 1 - P(U)$ . Weiter wird das Prinzip der Maximalen Entropie angewandt, um den in Form von  $I$  vorliegenden Informationsstand in eine (wie in Abschnitt 5.2.3) auf  $U$  eingeschränkte Maximum-Entropie-Verteilung zu überführen. Anschließend wird anhand der gemäß Gleichung (5.9) geltenden Beziehung  $p(z) = P(U)p(z|U)$  für  $z \in U$  der benötigte Ausschnitt der (globalen) A-Priori-Verteilung  $p(z)$  für  $z \in U$  berechnet.

Zur Durchführung der lokalen Bayes'schen Fusion gemäß Satz 5.3 (auf Basis der in Gleichung (5.27) und Gleichung (5.28) angegebenen Proportionalitätsbeziehungen) könnte man dabei auch alternativ für  $z \in U$  statt  $p(z)$  direkt  $p(z|U)$  und (gleichzeitig) statt  $P(\bar{U})$  die Größe  $P(\bar{U})/P(U)$  in die Fusion einbringen.

Die Sinnhaftigkeit der zuvor beschriebenen Vorgehensweise hinsichtlich der Anwendung des Prinzips der Maximalen Entropie ergibt sich aus der nachfolgenden Proposition:

**Proposition 5.4.** *Es bezeichne  $p_C(x)$  die Wahrscheinlichkeitsverteilung über der Partition  $X := \{U, \bar{U}\}$  von  $Z$ , welche  $U$  gerade die Wahrscheinlichkeit  $P(U)$  zuweist. Dann gilt*

$$\begin{aligned} & H[p_C(x)] + P(U) H[p(z|U)] \\ &= -P(\bar{U}) \log P(\bar{U}) - \int_U p(z) \log p(z) dz . \end{aligned} \quad (5.37)$$

*Beweis.* Unter Berücksichtigung von Gleichung (5.9) ergibt sich

$$\begin{aligned} & H[p_C(x)] + P(U) H[p(z|U)] \\ &= -P(U) \log P(U) - P(\bar{U}) \log P(\bar{U}) \end{aligned} \quad (5.38)$$

$$\begin{aligned} & -P(U) \int_U p(z|U) \log p(z|U) dz \\ &= -P(U) \left( \log P(U) + \int_U p(z|U) \log p(z|U) dz \right) \end{aligned} \quad (5.39)$$

$$- P(\bar{U}) \log P(\bar{U}) \quad (5.40)$$

$$= - P(U) \left( \int_U p(z|U) (\log p(z|U) + \log P(U)) dz \right) \\ - P(\bar{U}) \log P(\bar{U}) \quad (5.41)$$

$$= - \int_U p(z|U) P(U) \log (p(z|U) P(U)) dz \\ - P(\bar{U}) \log P(\bar{U}) \quad (5.42)$$

$$= - \int_U p(z) \log p(z) dz - P(\bar{U}) \log P(\bar{U}) . \quad (5.43)$$

□

Es sei angemerkt, dass im Fall, dass die globale A-Priori-Verteilung  $p(z)$  auf  $Z$  eine diskrete Wahrscheinlichkeitsverteilung darstellt, die rechte Seite von Gleichung (5.43) gerade die Entropie einer wiederum diskreten Wahrscheinlichkeitsverteilung notiert, bei welcher (anders als bei der globalen A-Priori-Verteilung  $p(z)$ ) der Anteil  $\bar{U}$  von  $Z$  als atomares Ereignis betrachtet wird.

Die Bestimmung der Likelihood-Funktion  $l(d|z)$  bezüglich  $z \in U$  kann für das lokale Bayes'sche Modell analog wie im globalen Bayes'schen Modell erfolgen. Wie aus Gleichung (5.29) bzw. (5.30) erkennbar ist, ergibt sich die außerdem für die lokale Bayes'sche Fusion benötigte Größe  $l(d|\bar{U})$  als das über  $\bar{U}$  mit der globalen A-Priori-Verteilung  $p(z)$  gewichtete Mittel der Likelihood-Funktion  $l(d|z)$ . Will man diese Größe anhand von Gleichung (5.29) bzw. Gleichung (5.30) berechnen und nicht anderweitig herleiten, so müssen insgesamt gesehen Likelihood-Funktion  $l(d|z)$  und A-Priori-Verteilung  $p(z)$  wie im globalen Bayes'schen Modell also doch für alle möglichen Werte  $z \in Z$  bestimmt werden. Da sich die in Gleichung (5.29) bzw. Gleichung (5.30) angegebene Berechnung wiederum auf ganz  $\bar{U}$  (im Detail) bezieht, widerspricht diese Vorgehensweise nicht zuletzt dem Wunsch, die Berechnungen auf  $U$  zu konzentrieren, um so deutlich Aufwand zu sparen.

Im Hinblick auf eine alternative direkte Bestimmung der Größe  $l(d|\bar{U})$  (gemeinsam mit  $l(d|z)$  für  $z \in U$ ) lässt sich feststellen, dass sich dies i. A. schwierig

rig gestalten wird, da diese Größe letztlich nicht den intuitiven Erwartungen an ein (relatives) Maß für die Plausibilität des Werts von  $d$  unter der mit der Annahme  $z \in \bar{U}$  verbundenen Hypothese (vis-à-vis Hypothesen bezüglich anderer möglicher Werte für  $z$ ) entspricht: Erweitert man die mit  $\bar{U}$  korrespondierende Hypothese im Hinblick auf den „wahren“ Wert von  $z$ , indem man ein Ereignis  $A \subset U$  zu  $\bar{U}$  hinzufügt, so kann, abhängig von den Werten der globalen A-Priori-Verteilung  $p(z)$  für  $z \in A$ , die Beziehung  $l(d|\bar{U} \cup A) \leq l(d|\bar{U})$  gelten. Dieser aus mathematischer Hinsicht geltende Sachverhalt dürfte es aufgrund seiner Nichtintuitivität schwierig machen,  $l(d|\bar{U})$  gemeinsam mit  $l(d|z)$  für  $z \in U$  direkt anzugeben. Es sei angemerkt, dass diese Schwierigkeit allerdings nicht besteht, falls die globale A-Priori-Verteilung  $p(z)$  mindestens bezüglich  $\bar{U} \cup A$  nichtinformativ ist, d. h. dass sie so beschaffen ist, dass sie keinen Wert von  $z \in \bar{U} \cup A$  favorisiert. In diesem Fall ergeben sich immerhin auch gewisse rechnerische Vorteile hinsichtlich der Berechnung von  $l(d|\bar{U})$  (vgl. auch [Liu98]). Die oben angeführte Tatsache, dass bei den Berechnungen  $\bar{U}$  im Detail einbezogen werden muss, gilt jedoch weiterhin.

Wie in Abschnitt 3.5.2 dargestellt wurde, lässt sich im Fall, dass die  $d$  in zusammengefassten Informationsbeiträge  $d_s$ ,  $s \in \{1, \dots, S\}$ , als bedingt unabhängig gegeben  $z$  angesehen werden können, ein sequentielles Bayes'sches Fusionsschema realisieren (vgl. Gleichung (3.24) und Gleichung (3.25)). Wie nachfolgend anhand eines Zahlenbeispiels aufgezeigt werden wird, impliziert die Annahme der bedingten Unabhängigkeit der Informationsbeiträge gegeben  $z$ , d. h.

$$p(d|z) = \prod_{s=1}^S p(d_s|z), \quad z \in Z, \quad (5.44)$$

nicht, dass auch ein entsprechend sequentielles Fusionsschema im lokalen Bayes'schen Modell, in welchem  $\bar{U}$  kumuliert betrachtet wird, realisiert werden kann. Zwar impliziert Gleichung (5.44) direkt, dass insbesondere

$$p(d|z) = \prod_{s=1}^S p(d_s|z), \quad z \in U \quad (5.45)$$

gilt. Allerdings folgt daraus nicht die Gültigkeit der im lokalen Bayes'schen Modell weiterhin benötigten Beziehung

$$p(d|\bar{U}) = \prod_{s=1}^S p(d_s|\bar{U}) \quad (5.46)$$

mit  $p(d|\bar{U})$  gemäß Gleichung (5.29) und (in analoger Weise)

$$p(d_s|\bar{U}) = \frac{\int_{\bar{U}} p(d_s|z)p(z) dz}{P(\bar{U})}, \quad s \in \{1, \dots, S\}. \quad (5.47)$$

Dies bedeutet, dass im lokalen Bayes'schen Modell, in dem  $\bar{U}$  kumuliert betrachtet wird, die in  $d$  zusammengefassten Informationsbeiträge nur noch kontext-spezifisch bedingt unabhängig sind. Es sei angemerkt, dass die Thematik der kontextspezifischen Unabhängigkeit wiederum für das (anders gelagerte) Anwendungsgebiet Bayes'sche Netze auch ausführlich in [Bou96] adressiert wird.

Zur Verdeutlichung der beschriebenen Problematik im Hinblick auf den teilweisen Verlust der bedingten Unabhängigkeit im lokalen Bayes'schen Modell sei nun wie bereits angekündigt ein konkretes Zahlenbeispiel betrachtet: Es gelte  $n = 1$  und  $Z = \{\nu_1, \nu_2, \nu_3\}$ . Die A-Priori-Verteilung sei gegeben durch  $p(\nu_i) = 1/3$  für alle  $i \in \{1, 2, 3\}$ . Weiter gelte  $s = 2$ , d. h. es seien zwei Informationsbeiträge  $d_1$  und  $d_2$  gegeben, so dass  $d = (d_1, d_2)$  gilt. Setzt man

$$\begin{aligned} p(d_1, d_2 | \nu_1) &= 0,2025, & p(d_1, d_2 | \nu_2) &= 0,045, & p(d_1, d_2 | \nu_3) &= 0,045, \\ p(d_1 | \nu_1) &= 0,45, & p(d_1 | \nu_2) &= 0,45, & p(d_1 | \nu_3) &= 0,1, \\ p(d_2 | \nu_1) &= 0,45, & p(d_2 | \nu_2) &= 0,1, & p(d_2 | \nu_3) &= 0,45, \end{aligned} \quad (5.48)$$

so ist die Annahme der bedingten Unabhängigkeit gemäß Gleichung (5.44) erfüllt. Es macht im globalen Bayes'schen Modell also keinen Unterschied, ob die Informationsbeiträge in einem Schritt gemäß der Regel

$$p(\nu_i | d_1, d_2) \propto p(d_1, d_2 | \nu_i) p(\nu_i), \quad i \in \{1, 2, 3\}, \quad (5.49)$$

oder sequentiell in zwei Schritten gemäß der Regel

$$p(\nu_i|d_1, d_2) \propto p(d_1|\nu_i)p(d_2|\nu_i)p(\nu_i), \quad i \in \{1, 2, 3\}, \quad (5.50)$$

fusioniert werden. In beiden Fällen ergibt sich für die A-Posteriori-Verteilung

$$\begin{aligned} p(\nu_1|d_1, d_2) &\approx 0,6923, & p(\nu_2|d_1, d_2) &\approx 0,1538, \\ p(\nu_3|d_1, d_2) &\approx 0,1538. \end{aligned} \quad (5.51)$$

Weiter sei im vorliegenden Beispiel exemplarisch  $U = \{\nu_1\}$  gewählt. Man verifiziert leicht, dass in diesem Fall dann zwar die in Gleichung (5.45) formulierte Bedingung für  $z \in U$  trivialerweise erfüllt ist, nicht jedoch die in Gleichung (5.46) formulierte Bedingung im Hinblick auf  $\bar{U}$ . Die bedingte Unabhängigkeit der Informationsbeiträge im globalen Bayes'schen Modell hat sich also zu einer kontextspezifischen bedingten Unabhängigkeit reduziert. Die Durchführung der Fusion in einem Fusionsschritt liefert die korrekte lokale A-Posteriori-Verteilung, welche wie folgt gegeben ist:

$$p(\nu_1|d_1, d_2) \approx 0,6923, \quad P(\bar{U}|d_1, d_2) \approx 0,3077 \quad (5.52)$$

mit  $\bar{U} = \{\nu_2, \nu_3\}$ . Die Anwendung eines sequentiellen Fusionschemas, innerhalb dessen zuerst  $d_1$  und dann  $d_2$  berücksichtigt wird, liefert in diesem Beispiel dagegen ein deutlich anderes (falsches) Ergebnis für die lokale A-Posteriori-Verteilung. In diesem Fall ergibt sich nämlich (fälschlicherweise)

$$p(\nu_1|d_1, d_2) \approx 0,5724, \quad P(\bar{U}|d_1, d_2) \approx 0,4276 \quad (5.53)$$

mit  $\bar{U} = \{\nu_2, \nu_3\}$ .

#### 5.2.4.4 Fazit

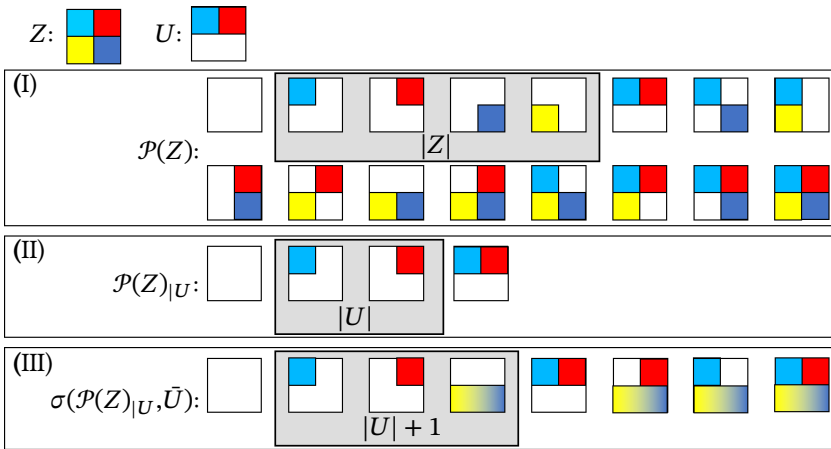
Bei Umsetzung des Gedankens, dass die Bayes'sche Fusion nicht global über ganz  $Z$ , sondern konzentriert auf einen lokalen Kontext  $U \subset Z$  erfolgen soll, durch Ausgestaltung des lokalen Bayes'schen Modells dahingehend, dass im Hinblick auf  $U$  eine genaue Fusion erfolgt, der Anteil  $\bar{U}$  dagegen nur

kumuliert betrachtet wird, sind die lokal definierten Wahrscheinlichkeiten (a priori und a posteriori) identisch zu denen im globalen Bayes'schen Modell. Die Bestimmung der zur Beschreibung des a priori verfügbaren Informationsstands benötigten Größen  $p(z)$  für  $z \in U$  und  $P(U)$  lässt sich recht praktikabel durchführen. Dasselbe gilt für die Bestimmung der Likelihood-Funktion  $l(d|z)$  für  $z \in U$ . Schwierig gestaltet sich dagegen die Bestimmung von  $l(d|\bar{U})$ . Eine exakte Berechnung anhand der in diese Größe eingehenden, im globalen Bayes'schen Modell vorhandenen Größen ist aufwändig, da das Komplement  $\bar{U}$  des lokalen Kontexts  $U$  hierzu doch im Detail berücksichtigt werden muss. Eine direkte Bestimmung von  $l(d|\bar{U})$  (zusammen mit  $l(d|z)$  für  $z \in U$ ) gestaltet sich i. A. ebenso schwierig, hier vor allem, weil sich  $l(d|\bar{U})$  nicht so verhält, wie man es intuitiv erwarten würde.

Beim Übergang vom globalen Bayes'schen Modell zum lokalen Bayes'schen Modell reduziert sich eine bedingte Unabhängigkeit der Informationsbeiträge i. A. zu einer kontextspezifischen bedingten Unabhängigkeit. Dies bedeutet, dass mindestens im Hinblick auf  $\bar{U}$  die Anwendung eines sequentiellen Fusionschemas i. A. nicht möglich ist. Die Vorteile, welche sich beim Vorliegen heterogener Informationsbeiträge in diesem Sinne oftmals durch die Rechtfertigkeit einer bedingten Unabhängigkeitsnahme ergeben, gehen also i. A. verloren.

### 5.2.5 Zusammenfassende Betrachtung

In Abb. 5.4 sind die beiden in Abschnitt 5.2.3 und Abschnitt 5.2.4 eingeführten Ansätze, Lokalität im Hinblick auf einen lokalen Kontext  $U \subset Z$  probabilistisch zu modellieren, in illustrativer Weise einander gegenübergestellt. Das Ziel der Darstellung ist es, noch einmal anhand eines Beispiels aufzuzeigen, welche Informationsstände im Hinblick auf den „wahren“ Wert der Größen des Interesses im resultierenden lokalen Bayes'schen Modell jeweils (a priori und a posteriori) noch erlangt werden können und wodurch die entsprechenden Wahrscheinlichkeiten festgelegt sind.



**Abbildung 5.4:** Beispiel zwecks Vergleich der beiden eingeführten Ansätze, Lokalität im Hinblick auf einen lokalen Kontext  $U \subset Z$  probabilistisch zu modellieren.

Wie durch die farbigen Quadrate illustriert wird, wird im in Abb. 5.4 dargestellten Beispiel angenommen, dass  $Z$  aus vier (diskreten) Elementen besteht und dass der lokale Kontext  $U$  zwei dieser Elemente umfasst. Weiter wird angenommen, dass  $\mathcal{Z} = \mathcal{P}(Z)$  gilt. Im globalen Bayes'schen Modell sind dann die Wahrscheinlichkeiten sämtlicher, in Abschnitt (I) von Abb. 5.4 dargestellten Mengen  $A \in \mathcal{P}(Z)$  definiert und berechenbar. Sie sind a priori und a posteriori eindeutig festgelegt durch die entsprechende (jeweilige) Wahrscheinlichkeitsverteilung über den  $|Z|$  Elementarereignissen, welche in Abschnitt (I) von Abb. 5.4 grau hinterlegt sind.

Bildet man das lokale Bayes'sche Modell wie in Abschnitt 5.2.3 auf Basis der Annahme dass  $z \in U$  sicher gilt, so sind nur noch die Wahrscheinlichkeiten der in Abschnitt (II) von Abb. 5.4 illustrierten Elemente von  $\mathcal{P}(Z)|_U$  definiert. D. h. es kann nur Information im Hinblick auf diese lokal erlangt werden. Die global vorhandenen Anteile von  $Z$ , welche sich auf  $\bar{U}$  beziehen, werden im lokalen Bayes'schen Modell vollständig ignoriert, wodurch es dann zu der in Abschnitt 5.2.3.3 beschriebenen Überbewertung der Degree-of-Beliefs der lo-



kal vorhandenen Ereignisse kommt. Die lokalen Wahrscheinlichkeiten dieser Ereignisse sind a priori und a posteriori eindeutig festgelegt durch die entsprechende (jeweilige) lokale Wahrscheinlichkeitsverteilung über den lokal nun nur noch  $|U|$  Elementarereignissen, welche in Abschnitt (II) von Abb. 5.4 wiederum grau hinterlegt sind.

Bildet man das lokale Bayes'sche Modell dagegen wie in Abschnitt 5.2.4, indem man den lokalen Kontext  $U$  weiterhin detailliert adressiert, sein Komplement  $\bar{U}$  aber nur kumuliert behandelt, so sind lokal die Wahrscheinlichkeiten von mehr Ereignissen definiert. Dieser Sachverhalt ist in Abschnitt (III) von Abb. 5.4 illustriert. Jedes dieser in  $\sigma(\mathcal{P}(Z)|_U, \bar{U})$  enthaltenen Ereignisse zeichnet sich gerade dadurch aus, dass es  $\bar{U}$  entweder nicht oder vollständig enthält. Da dadurch die Anteile von  $Z$ , welche sich auf  $\bar{U}$  beziehen, weiterhin als existent angenommen werden, kommt es hier zu keiner Überbewertung der Degree-of-Beliefs der lokal vorhandenen Ereignisse. Die lokalen Wahrscheinlichkeiten dieser Ereignisse sind a priori und a posteriori eindeutig festgelegt durch die entsprechende (jeweilige) lokale Wahrscheinlichkeitsverteilung über den nun  $|U| + 1$  Elementarereignissen, welche in Abschnitt (III) von Abb. 5.4 wiederum grau hinterlegt sind.

Betrachtet man wie in Abb. 5.4 die beiden unterschiedlichen Möglichkeiten, Lokalität zu definieren, rein vom Standpunkt aus, welche Informationsstände im Hinblick auf den „wahren“ Wert von  $z$  probabilistisch repräsentiert werden können (und damit verbunden auch die Frage, inwieweit die lokal berechneten Wahrscheinlichkeiten den globalen entsprechen) und welcher Aufwand im Hinblick auf die Berechnung der lokalen A-Posteriori-Verteilung nötig ist, so ergibt sich, dass der in Abschnitt 5.2.4 eingeführte Ansatz, Lokalität durch Vergrößerung des globalen Bayes'schen Modells zu beschreiben, klar dem in Abschnitt 5.2.3 eingeführten Ansatz, bei welchem das globale Bayes'sche Modell komplett auf den lokalen Kontext  $U$  eingeschränkt wird, vorzuziehen ist. Man gewinnt deutlich mehr Aussagekraft im Hinblick auf die globale Situation, indem man Wahrscheinlichkeitsverteilungen berechnet, welche im entsprechenden lokalen Bayes'schen Modell durch  $|U| + 1$  statt  $|U|$  Elementarereignisse festgelegt sind.

Im Hinblick auf die Zielsetzung, einen möglichst breit anwendbaren Ansatz zur Modellierung von Lokalität insbesondere auch für die Fusion heterogener Informationsquellen umzusetzen, sind allerdings die in Abschnitt 5.2.4.3 ermittelten Schwierigkeiten im Hinblick auf die Konstitution der entsprechenden lokalen Bayes'schen Modelle (Schwierigkeiten bezüglich der Bildung von  $l(d|\bar{U})$ , teilweiser Verlust bedingter Unabhängigkeit) gravierender anzusehen als die in Abschnitt 5.2.3.3 ermittelten (Verzerrung lokal berechneter Wahrscheinlichkeitsaussagen). Voraussetzung für die Gültigkeit dieser Aussage ist dabei, dass Prinzipien und Methoden bereitstehen müssen, auf deren Basis man den lokalen Kontext  $U$  tatsächlich geeignet festlegen und möglichst auch ohne vollständige Kenntnis des globalen Bayes'schen Modells wenigstens in einem Mindestmaß hinsichtlich seiner Adäquatheit bewerten kann. Auch wäre es wünschenswert, die lokal berechneten Ergebnisse explizit in den umfassenderen globalen Bezug setzen zu können. Diese beiden Aspekte werden in Kapitel 6 und Kapitel 7 ausführlich adressiert.

Eine rudimentärere, aber (von der zugrundeliegenden Idee her) manchmal praxistaugliche mögliche Vorgehensweise diesbezüglich wird außerdem auch bereits zuvor in dem in Abschnitt 5.3 vorgestellten Beispiel umgesetzt werden. In dem in Abschnitt 5.3 vorgestellten Beispiel wird dabei auch eine rein qualitative, in Teilen heuristische Analyse auf Basis des Ansatzes, Lokalität durch Einschränkung zu modellieren, vorgenommen werden. Dazu werden statt normierten nur unnormierte Wahrscheinlichkeitsverteilungen geeignet modelliert und weiterentwickelt. Wie anhand des Beispiels erkennbar werden wird, lassen sich bereits durch diese vereinfachte, mathematisch nicht vollständig exakte Herangehensweise an die lokale Bayes'sche Fusion Ergebnisse ableiten, welche (mit methodischen Einschränkungen) in der Praxis im Hinblick auf die Beantwortung spezieller Teilfragestellungen hilfreich sein können.

Wenngleich sich die vorliegende Arbeit nachfolgend vor allem auf den in Abschnitt 5.2.3 eingeführten Ansatz der Lokalität durch Einschränkung beziehen wird, so sei dennoch betont, dass zumindest für spezielle Aufgabenstellungen auch der in Abschnitt 5.2.4 eingeführte Ansatz der Lokalität durch Vergrößerung durchaus als vielversprechend zu erachten ist. Es wäre ein Thema für

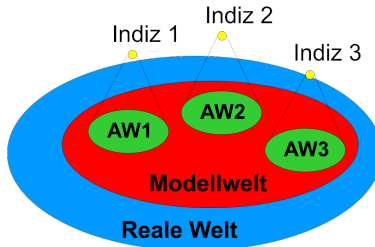
Folgearbeiten, diesen gerade in Bezug auf spezielle Aufgabenstellungen der Informationsfusion hinsichtlich der Möglichkeiten, die mit ihm einhergehenden Schwierigkeiten zu adressieren, zu untersuchen und weiterzuentwickeln. Im Hinblick auf an die vorliegende Arbeit anschließende Folgearbeiten wäre es weiter auch als zielführend anzusehen, eine Weiterentwicklung der in der vorliegenden Arbeit letztlich erarbeiteten Ansätze zur lokalen Bayes'schen Fusion vorzunehmen, indem man diese (hinsichtlich der Idee einer Konzentration auf einen lokalen Kontext  $U$ ) mit dem in Abschnitt 4.2.1 behandelten Prinzip konjugierter Verteilungsfamilien bzw. mit den in Abschnitt 4.2.2 und Abschnitt 4.2.3 vorgestellten Verfahren zur Approximation der A-Posteriori-Verteilung bzw. von Kenngrößen derselben kombiniert.

## 5.2.6 Verteilte Umsetzung

Im vorliegenden Abschnitt soll noch auf die Frage eingegangen werden, wie sich die in Abschnitt 5.2.3 und Abschnitt 5.2.4 eingeführten Ansätze, Lokalität im Hinblick auf einen lokalen Kontext  $U \subset Z$  probabilistisch zu modellieren, ausgestalten, wenn die lokale Bayes'sche Fusion kollaborativ durch  $J \in \mathbb{N}$  Agenten, von denen jeder für einen Anteil  $U^{(j)} \subset Z$  mit  $U = \bigcup_{j=1}^J U^{(j)}$  zuständig ist, durchgeführt wird. Wie nachfolgend aufgezeigt werden wird, besteht das Vorgehen dabei dann gerade darin, dass man  $J$  lokale Bayes'sche Modelle ansetzt, welche die jeweiligen Zuständigkeitsbereiche  $U^{(j)}$ ,  $j \in \{1, \dots, J\}$ , der  $J$  Agenten widerspiegeln. Diese Modelle werden so ausgestaltet, dass sie insgesamt das (lokale) Fusionsergebnis liefern, welches auf Basis eines auf den lokalen Kontext  $U$  insgesamt bezogenen lokalen Bayes'schen Modells erreicht worden wäre.

In Abb. 5.5 ist das dieser Vorgehensweise zugrundeliegende Prinzip unter Bezugnahme auf den in [Sav54] eingeführten und in [Las93, Las94] erweiterten sog. Small World Formalismus illustriert. Ein globales Bayes'sches Modell konstituiert gemäß Abb. 5.5 eine Modellwelt, d. h. eine Small World, welche grob gesagt einer geeigneten Vereinfachung der realen Welt entspricht. Reale Welt und Modellwelt können sich hinsichtlich der vorhandenen Größen

(deren Menge, Wertebereiche und Granularität) sowie hinsichtlich deren Beziehungen untereinander (z. B. Abhängigkeitsstrukturen) unterscheiden.



**Abbildung 5.5:** Illustration des Prinzips verteilter lokaler Bayes'scher Fusion unter Bezugnahme auf den Small World Formalismus. Siehe hierzu auch [San09c].

Lokale Bayes'sche Fusionsansätze zeichnen sich dadurch aus, dass die Lösung der Fusionsaufgabe wiederum geeignet auf einen Teil der Modellwelt beschränkt wird. Das Ziel ist es, diesen Teil so zu wählen, dass das Fusionsergebnis für die lokale Welt auch aussagekräftig für die Modellwelt (und damit für die reale Welt) ist. Bei der verteilten Umsetzung lokaler Bayes'scher Fusion wird die lokale Welt in einzelne Agentenwelten (in der Abbildung 3 Stück, welche durch AW1, AW2, AW3 notiert sind) aufgespalten. Das Zusammenfügen der lokalen Fusionsergebnisse der Agentenwelten entspricht der Bestimmung eines Fusionsresultats für die Modellwelt. Eine etwas ausführlichere Darstellung dieser Sachverhalte wurde zuvor in [San09c] veröffentlicht.

### 5.2.6.1 Bei Lokalität durch Vergrößerung

Im Falle der Lokalität durch Vergrößerung korrespondiert mit der probabilistischen Modellierung eines jeden der  $J$  Agenten ein Produktwahrscheinlichkeitsraum der Form von Gleichung (5.22), wobei beim Agenten Nummer  $j$  der

lokale Kontext  $U$  durch  $U^{(j)}$  ersetzt wird:

$$(Z \times D, \sigma(\sigma(\mathcal{Z}_{|U^{(j)}}), \bar{U}^{(j)}) \times \mathcal{D}), \Pi). \quad (5.54)$$

Der Agent Nummer  $j$  besitzt die detaillierte globale A-Priori-Verteilung  $p(z)$  im Hinblick auf den ihm zugewiesenen Anteil  $U^{(j)}$  von  $Z$ , während er im Hinblick auf dessen Komplement  $\bar{U}^{(j)} = Z \setminus U^{(j)}$  nur die kumulierte globale A-Priori-Wahrscheinlichkeit  $P(\bar{U}^{(j)})$  besitzt. Diese beiden Größen können dann wie in Satz 5.3 beschrieben in das finale Fusionsergebnis des Agenten überführt werden. Dieses ist gegeben in Form des Ausschnitts aus der globalen A-Posteriori-Verteilung  $p(z|d)$  für  $z \in U^{(j)}$  zusammen mit der A-Posteriori-Wahrscheinlichkeit  $P(\bar{U}^{(j)}|d)$  des Anteils von  $Z$ , für den der Agent Nummer  $j$  nicht zuständig ist. Man beachte, dass aufgrund der in Abschnitt 5.2.4.3 diesbezüglich dargestellten Schwierigkeit die Durchführung einer sequentiellen Fusion durch den Agenten i. A. nicht möglich ist.

Die resultierenden Fusionsergebnisse unterschiedlicher Agenten sind direkt kompatibel. D. h. sie können direkt zu einem gesamthafteren Fusionsresultat zusammengefügt werden. Dieses ist dann so beschaffen, dass der gesamte lokale Kontext  $U = \bigcup_{j=1}^J U^{(j)}$  detailliert durch entsprechende, von den Agenten gelieferte Anteile der globalen A-Posteriori-Verteilung  $p(z|d)$  adressiert wird, während das Komplement  $\bar{U}$  des lokalen Kontexts  $U$  wiederum nur kumuliert durch die Angabe von

$$P(\bar{U}|d) = 1 - \sum_{j=1}^J \int_{U^{(j)}} p(z|d) dz \quad (5.55)$$

betrachtet wird. Gleichung (5.55) gibt gerade an, wie sich die A-Posteriori-Wahrscheinlichkeit  $P(\bar{U}|d)$  des Komplements  $\bar{U}$  des lokalen Kontexts  $U$  anhand der von den  $J$  Agenten gelieferten (hier als disjunkt angenommenen) Ausschnitte der globalen A-Posteriori-Verteilung  $p(z|d)$  berechnet.

### 5.2.6.2 Bei Lokalität durch Einschränkung

Im Falle der Lokalität durch Einschränkung korrespondiert mit der probabilistischen Modellierung eines jeden der  $J$  Agenten ein Produktwahrscheinlichkeitsraum der Form von Gleichung (5.8), wobei beim Agenten Nummer  $j$  wiederum der lokale Kontext  $U$  durch  $U^{(j)}$  ersetzt wird:

$$(U^{(j)} \times D, \sigma(\mathcal{Z}_{|U^{(j)}} \times \mathcal{D}), \Pi) . \quad (5.56)$$

Der Agent Nummer  $j$  besitzt eine auf den in seine Zuständigkeit fallenden Anteil  $U^{(j)} \subset Z$  eingeschränkte lokale A-Priori-Verteilung  $p(z|U^{(j)})$ . Diese kann er wie in Satz 5.2 beschrieben in eine lokale, wiederum auf den Anteil  $U^{(j)} \subset Z$  eingeschränkte A-Posteriori-Verteilung  $p(z|d, U^{(j)})$  überführen. Lässt sich im globalen Bayes'schen Modell die Annahme der bedingten Unabhängigkeit der in  $d$  zusammengefassten Informationsbeiträge gegeben  $z$  rechtfertigen, so gilt das hier auch lokal, d. h. der Agent kann die lokale Bayes'sche Fusion auch sequentiell umsetzen. Dies entspricht dann im Wesentlichen der im Konzept für die agentenbasierte Fusionsarchitektur beschriebenen Vorgehensweise der Fusionsagenten (vgl. Abschnitt 4.3.2.2).

Die so insgesamt resultierenden Fusionsergebnisse unterschiedlicher Agenten sind per se allerdings nicht kompatibel. D. h. sie können nicht direkt zu einem gesamthafteren Fusionsergebnis zusammengefügt werden. Diesen Sachverhalt kann man sich gut veranschaulichen, indem man die Beispiele in Abb. 5.2 und Abb. 5.3 heranzieht und sich vorstellt, dass die beiden unterschiedlichen Teilbereiche von  $Z$ , die jeweils betrachtet werden, in die Zuständigkeit zweier unterschiedlicher Agenten fallen. Ein Ausweg aus dieser Problematik ergibt sich jedoch, wenn man vorsieht, dass den Agenten (oder je zwei von ihnen) jeweils ein gemeinsamer (ggf. auch sehr geringer) Anteil von  $Z$  zugewiesen wird, denn dann können, wie im Folgenden beschrieben wird, die Ergebnisse der Agenten im Hinblick auf diesen Anteil genutzt werden, um die von ihnen insgesamt gelieferten Ergebnisse abzugleichen.

Es gelte  $j_1, j_2 \in \{1, \dots, J\}$ ,  $B \in \mathcal{Z}_{|U^{(j_i)}}$  für  $i \in \{1, 2\}$ . Dann ergibt sich unter Ausnutzung von Gleichung (5.11) die Beziehung

$$\frac{P(B|d, U^{(j_1)})}{P(B|d, U^{(j_2)})} = \frac{\int_B p(z|d, U^{(j_1)}) dz}{\int_B p(z|d, U^{(j_2)}) dz} = \frac{P(U^{(j_2)}|d)}{P(U^{(j_1)}|d)} := \rho_{(j_1, j_2)}. \quad (5.57)$$

Agent Nummer  $j_1$  kann anhand von  $\rho_{(j_1, j_2)}$  die von Agent Nummer  $j_2$  berechnete lokale A-Posteriori-Verteilung  $p(z|d, U^{(j_2)})$  in Bezug zu seiner lokalen A-Posteriori-Verteilung  $p(z|d, U^{(j_1)})$  setzen, indem er für  $z \in U^{(j_2)} \setminus U^{(j_1)}$

$$\tilde{p}(z|d, U^{(j_1)}) := \rho_{(j_1, j_2)} \cdot p(z|d, U^{(j_2)}) \quad (5.58)$$

berechnet. Das Zusammenfassen und neu normieren von  $\tilde{p}(z|d, U^{(j_1)})$  und  $p(z|d, U^{(j_1)})$  liefert dann eine weiterhin lokale, nun aber nur noch auf den Anteil  $U^{(j_1)} \cup U^{(j_2)}$  von  $Z$  eingeschränkte A-Posteriori-Verteilung  $p(z|d, U^{(j_1)} \cup U^{(j_2)})$ . Das beschriebene Vorgehen lässt sich schrittweise fortführen oder ggf. auch (wenn alle Agenten das gleiche Ereignis  $B$  betrachten) in einem Schritt umsetzen, um letztlich die auf  $U$  eingeschränkte lokale A-Posteriori-Verteilung  $p(z|d, U)$  zu bestimmen.

Unter der Annahme, dass der Wert der globalen A-Posteriori-Verteilung  $p(z|d)$  auf den mit  $\bar{U}$  korrespondierenden (als nicht relevant betrachteten) Anteilen von  $Z$  in etwa gleich (da gering) ist, lässt sich der Faktor  $\rho_{(j_1, j_2)}$  weiter auch wie folgt unter Bezugnahme auf unterschiedliche Anteile von  $\bar{U}$ , welche den gleiche Umfang besitzen, approximieren: Angenommen Agent  $j_1$  nimmt  $B \in \mathcal{Z}$  in seinen Zuständigkeitsbereich mit auf und Agent  $j_2$  nimmt  $C \in \mathcal{Z}$  in seinen Zuständigkeitsbereich mit auf. Dann gilt

$$\rho_{(j_1, j_2)} \approx \frac{P(B|d, U^{(j_1)})}{P(C|d, U^{(j_2)})}. \quad (5.59)$$

Die in Gleichung (5.59) angegebene Approximation lässt sich (unter der zugrundeliegenden Annahme) insbesondere auch berechnen, wenn die den Agenten zugewiesenen Anteile  $U^{(j_1)}$ ,  $U^{(j_2)}$  gerade nicht nur die eigentlichen jeweiligen Auffälligkeiten in  $Z$ , sondern auch gewisse, nicht allzu große Umgebungen um diese umfassen, wie es beim Konzept für die agentenbasierte Fusionsarchitektur bereits vorgesehen war (vgl. Abschnitt 4.3.2.1). Verallgemeinert man die angestellte Betrachtung wiederum auf alle  $J$  Agenten, so ergibt sich, dass sich (unter der zugrundeliegenden Annahme) so

letztlich eine zumindest näherungsweise korrekte Version der lokalen A-Posteriori-Verteilung  $p(z|d, U)$  berechnen lässt und zwar ohne dass die Agenten tatsächlich gleiche Anteile von  $Z$  betrachten müssen.

## 5.3 Beispiel für eine qualitative Analyse

### 5.3.1 Einführung

Im vorliegenden Abschnitt wird ein möglichst anwendungsnahe Beispiel im Hinblick auf eine Anwendung lokaler Bayes'scher Fusion umgesetzt und untersucht. Dabei gilt es, mittels lokaler Bayes'scher Fusion unterschiedliche Informationsbeiträge, wie sie in ähnlicher Weise im Bereich der Aufklärung als Basis für entsprechende Intelligence Produkte vorliegen können, zu fusionieren und darauf basierend aufklärungsrelevante Erkenntnisse abzuleiten. Das genaue Ziel besteht darin, das Vorhandensein sowie den Typ und die Position von Fahrzeugen, welche sich innerhalb eines bestimmten geographischen Interessengebiets (im Englischen: Area of Interest (AOR); vgl. z. B. [Wun09]) befinden, festzustellen.

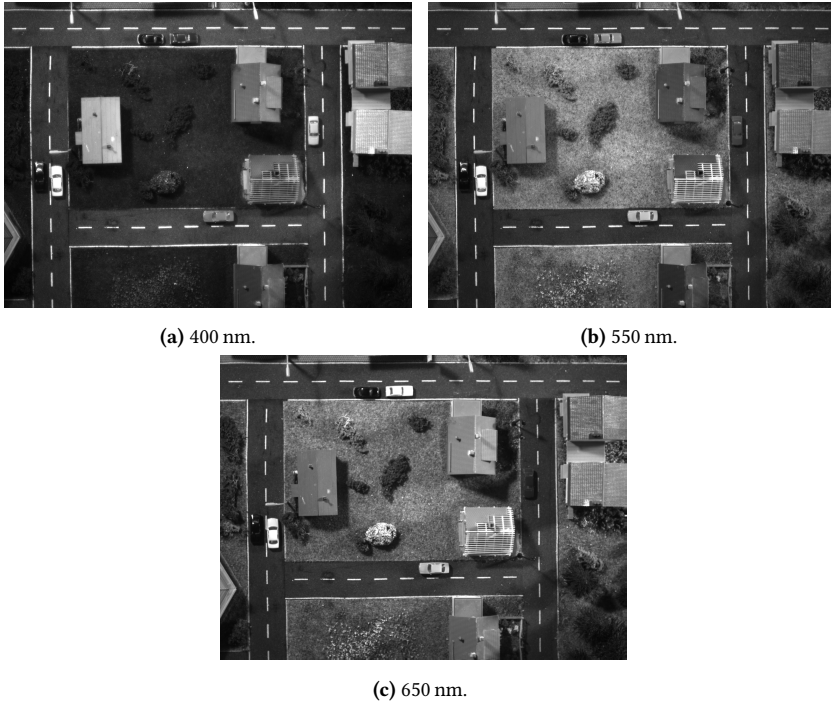
Variationen der in [San10b] veröffentlichten und nachfolgend im Wesentlichen wiedergegebenen Version dieses Beispiels und entsprechende Vorarbeiten wurden zuvor in [Bey06b], [Bey07] und [San09c] veröffentlicht. Weitere Aspekte, vor allem im Hinblick auf die Anwendung unterschiedlicher Bildverarbeitungsverfahren zur Auswertung und letztlich Fusion von Bildinformation, d. h. in Form von Bilddaten vorliegenden Informationsbeiträgen, auf welche nachfolgend nur in begrenztem Umfang eingegangen werden kann, wurden auch darüber hinaus im Rahmen betreuter studentischer Arbeiten untersucht (siehe [Gor09, Gou08, Gou10]). Um die flexible Generierung der notwendigen Beispieldaten zu ermöglichen und um die Verifizierbarkeit der jeweils erzielten Ergebnisse sicherzustellen, wurde das geographische Interessengebiet (und die konkret in ihm vorliegende Situation) auf Basis einer Modelllandschaft (Maßstab 1:160) und unter Verwendung maßstabsgerechter Miniaturmodelle real existierender Fahrzeugtypen exemplarisch umgesetzt.



Bedingt durch die hohe Komplexität und auch Variabilität der im Bereich der Aufklärung eingesetzten Verfahren und der zugrundeliegenden, sich teilweise auch noch in Präzisierung befindenden operationellen Prozesse, auf welche in Abschnitt 2.4.3.5 ausschnittsweise eingegangen wurde, kann es die vorliegende Arbeit nicht leisten, ein vollständig realitätsgetreues Beispiel hierzu zu liefern. Die konkret vorgenommene Ausgestaltung des Beispiels zielt statt dessen darauf ab, ergänzend und teilweise komplementär zu den in Abschnitt 5.2 vorgenommenen Arbeiten ausgewählte Teilaspekte im Hinblick auf die lokale Bayes'sche Fusion umzusetzen und zu untersuchen. Hinsichtlich der Auswertung der Informationsbeiträge wird dabei rein exemplarisch vor allem auf etablierte Verfahren und Methoden aus der Fachliteratur zurückgegriffen.

Nachfolgend wird wie in [San10b] der Fall adressiert, dass für die Fusion ein Informationsbeitrag  $d_1$  in Form von durch menschliche Beobachter gemeldeten Aufklärungsergebnissen sowie drei Informationsbeiträge  $d_2$ ,  $d_3$ ,  $d_4$  in Form von Luftbildern des Interessengebiets zur Verfügung stehen. Weiter wird davon ausgegangen, dass a priori vorliegendes Kartenmaterial, welches Information im Hinblick auf das im Interessengebiet vorhandene Straßennetz liefert, das in der Fusionsaufgabe verfügbare Vorwissen darstellt.

Die drei Luftbilder korrespondieren konkret mit den in Abb. 5.6 dargestellten Grauwertbildern des Interessengebiets, welche von einer festen Kameraposition als Draufsicht aufgenommen wurden. Um exemplarisch unterschiedliche Arten von Bildinformation zu generieren, wurde die Kamera für die Aufnahmen mit unterschiedlichen Spektralfiltern (vgl. z. B. [Ghe11]) ausgestattet. Die mittlere Wellenlänge des Durchlassbereichs des jeweils verwendeten Spektralfilters beträgt 400 nm bei  $d_2$ , 550 nm bei  $d_3$  und 650 nm bei  $d_4$ . Die Bandbreite der Spektralfilter beträgt jeweils 50 nm. Die von den menschlichen Beobachtern gemeldeten Aufklärungsergebnisse, welche nachfolgend noch genauer beschrieben werden, sind grob gesagt in Form von Angaben hinsichtlich der konkret gemachten Beobachtungen im Hinblick auf im Interessengebiet vorhandene Fahrzeuge und von mit den Angaben korrespondierenden Unsicherheiten gegeben.



**Abbildung 5.6:** Grauwertbilder des Interessengebiets.

Zur Lösung der Aufgabenstellung wird (wie bereits in Abschnitt 5.2.5 skizziert) eine qualitative Analyse auf Basis lokaler Bayes'scher Fusion vorgenommen. Hierfür wird ein Ausschnitt aus einer nicht normierten Version der A-Posteriori-Verteilung berechnet. Das zugrundeliegende globale Bayes'sche Modell wird so angesetzt, dass durch die Bestimmung der a posteriori signifikantesten Werte für die im Hinblick auf das konkrete Ziel der Aufklärung relevanten Größen des Interesses ein Rückschluss auf Position und Typ der im Interessengebiet vorhandenen Fahrzeuge ermöglicht wird. Die Bestimmung dieser Werte (im Hinblick auf den lokal betrachteten Anteil des Wertebereichs der Größen des Interesses) erfolgt durch eine Auswertung der Verhältnisse

von A-Posteriori-Wahrscheinlichkeiten, bei welcher lokale Maxima der (nicht normierten) A-Posteriori-Verteilung untereinander und im Hinblick auf sie umgebende lokale Ausschnitte betrachtet werden.

### 5.3.2 (Globale) Bayes'sche Modellierung

Entsprechend der konkret vorliegenden Informationsbeiträge und der zu ihrer Auswertung eingesetzten Verfahren wird für die Durchführung der Fusion neben Typ und Position auch die Fahrtrichtung eines Fahrzeugs als Größe des Interesses im Bayes'schen Modell mit modelliert. Genauer gilt

$$z = (z_1, z_2, z_3, z_4) \in Z = Z_1 \times Z_2 \times Z_3 \times Z_4 . \quad (5.60)$$

Dabei gibt  $z_1 \in Z_1$  die mögliche horizontale und  $z_2 \in Z_2$  die mögliche vertikale Position eines Fahrzeugs im Interessengebiet an<sup>1</sup>. Weiter bezeichnet  $z_3 \in Z_3$  den möglichen Fahrzeugtyp und  $z_4 \in Z_4$  bezeichnet die mögliche Fahrtrichtung. Um das Beispiel überschaubar zu halten, wird angenommen, dass hinsichtlich des Fahrzeugtyps

$$Z_3 = \{A \text{ (Audi)}, D \text{ (DKW)}, F \text{ (Ford)}, M \text{ (Mercedes)}, O \text{ (Opel)}\} \quad (5.61)$$

und hinsichtlich der Fahrtrichtung<sup>2</sup>

$$Z_4 = \{n \text{ (nördlich)}, o \text{ (östlich)}, s \text{ (südlich)}, w \text{ (westlich)}\} \quad (5.62)$$

gelten. Im Hinblick auf das genaue Ziel der Aufklärung korrespondiert die Fahrtrichtung  $z_4$  im Bayes'schen Sinne mit einem Nuisance-Parameter, welcher a posteriori aus dem Fusionsergebnis eliminiert werden kann.

<sup>1</sup> Die horizontale Position korrespondiert mit der entsprechenden Spaltenkoordinate und die vertikale Position korrespondiert mit der entsprechenden Zeilenkoordinate eines vom Interessengebiet aufgenommenen Bildes.

<sup>2</sup> Eine Variation des Beispiels für den Fall, dass die Fahrtrichtung einen Parameter mit Werten von 0 bis 359 Grad darstellt, ist grundsätzlich möglich (vgl. [Gor09, Gou10]), allerdings müssen dafür die eingesetzten Auswerteverfahren und die Modellbildung angepasst werden.

Die Informationsbeiträge  $d_2, d_3, d_4$  werden als bedingt unabhängig gegeben  $z$  angenommen, da sich die Durchlassbereiche der zur Generierung der Bilder verwendeten Spektralfilter nicht überlappen<sup>1</sup>. Aufgrund der heterogenen Natur der zugrundeliegenden Quellen kann insbesondere auch  $d_1$  gegenüber  $d_2, d_3, d_4$  als bedingt unabhängig gegeben  $z$  angenommen werden.

Die A-Priori-Verteilung  $p(z)$ , welche das in Form der Straßenkarte vorliegende Vorwissen unter Berücksichtigung einer Messunsicherheit  $\sigma_{\text{MU}}$  für jeden Punkt der Straßenkarte repräsentiert, kann unter Bezugnahme auf das Prinzip der Maximalen Entropie (vgl. Abschnitt 3.5.1) für  $(z_1, z_2) \in Z_1 \times Z_2$  näherungsweise als proportional zur Faltung der Gleichverteilung  $\mathcal{U}(Y)$  auf

$$Y := \{y = (y_1, y_2) | y \in Z_1 \times Z_2 \text{ und } y \text{ ist laut Karte Teil einer Straße}\} \quad (5.63)$$

mit der (zweidimensionalen) Normalverteilung  $\mathcal{N}((0,0)^\top, \sigma_{\text{MU}}^2 I_2)$  gesetzt werden (vgl. [Bey07]). Dabei bezeichnet  $I_2$  die zweidimensionale Einheitsmatrix. D. h. es gilt

$$p(z_1, z_2) \propto q_{\text{MU}}(z_1, z_2) * * q_{\text{GV}}(z_1, z_2) \quad (5.64)$$

wobei  $q_{\text{MU}}(z_1, z_2)$  gemäß  $\mathcal{N}((0,0)^\top, \sigma_{\text{MU}}^2 I_2)$  und  $q_{\text{GV}}(z_1, z_2)$  gemäß  $\mathcal{U}(Y)$  verteilt sind. Da kein Vorwissen in Bezug auf den Fahrzeugtyp und die Fahrtrichtung vorliegt, muss die A-Priori-Verteilung  $p(z)$  weiter nichtinformativ im Hinblick auf  $z_3$  und  $z_4$  sein. Dementsprechend wird festgelegt:

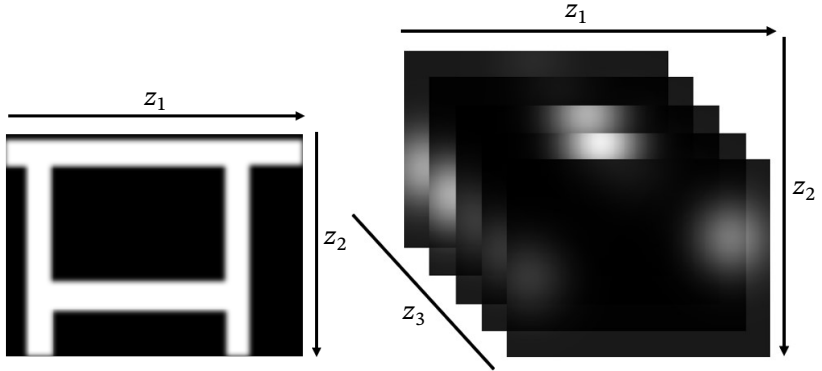
$$p(z) = p(z_1, z_2, z_3, z_4) \propto p(z_1, z_2) . \quad (5.65)$$

Die so resultierende A-Priori-Verteilung  $p(z)$  ist in Abb. 5.7a in nicht normierter Version visualisiert. Die Darstellung erfolgt dabei anhand eines Grauwertbilds mit Abmessungen gemäß  $Z_1 \times Z_2$ . Je höher der Intensitätswert des Grauwertbilds an einer bestimmten Position  $(z_1, z_2)$  ausfällt, desto höher ist

---

<sup>1</sup> Diese Annahme wird faktisch nur näherungsweise gelten, allein schon da dieselbe Kamera zur Aufnahme der Bilder verwendet wurde, was zu systematischen Einflüssen führen kann. Weiter ist davon auszugehen, dass das Sensorrauschen nicht unabhängig ist. Vgl. hierzu auch [Ghe08a].

(der für alle möglichen Typen  $z_3$  gemäß  $Z_3$  und für alle möglichen Fahrtrichtungen  $z_4$  gemäß  $Z_4$  gleiche) Wert der A-Priori-Verteilung  $p(z)$ .



(a) Nicht normierte Version der A-Priori-Verteilung.

(b) Likelihood-Funktion für den Informationsbeitrag  $d_1$ . Die Grauwertbilder sind so angeordnet, dass sie von vorne nach hinten mit den möglichen Fahrzeugtypen A, D, F, M und O korrespondieren.

Abbildung 5.7: A-Priori-Verteilung  $p(z) \propto p(z_1, z_2)$  und Likelihood-Funktion  $l(d_1|z)$ .

Der Informationsbeitrag  $d_1$ , welcher die von den menschlichen Beobachtern gelieferte Information verkörpert, ist wie folgt als Zusammenstellung von Fakten und mit diesen korrespondierenden Unsicherheiten beschrieben:

$$d_1 = \left\{ \left( (\mu_1(k), \mu_2(k))^T, \sigma_\mu(k), r_H(k), t(k) \right) \mid k \in \{1, \dots, K\} \right\}. \quad (5.66)$$

Dabei bezeichnet  $K \in \mathbb{N}$  die Gesamtzahl der gemeldeten Fahrzeuge. Es ist  $(\mu_1(k), \mu_2(k))$  die gemeldete Position des  $k$ -ten gemeldeten Fahrzeugs und  $\sigma_\mu(k)$  ist die mit dieser Positionsangabe korrespondierende räumliche Unsicherheit. Weiter notiert  $r_H(k)$  die für das  $k$ -te gemeldete Fahrzeug zusätzlich gemeldete Fahrtrichtung. Dabei wird angenommen, dass ausschließlich gemeldet wird, ob sich das jeweilige Fahrzeug in horizontaler (d. h. östlicher

oder westlicher) oder vertikaler (d. h. nördlicher oder südlicher) Fahrtrichtung fortbewegt. Es gilt also

$$r_H(k) \in R_C := \{o \vee w, n \vee s\}. \quad (5.67)$$

Weiter bezeichnet  $t(k)$  den Grad des Dafürhaltens des Beobachters dafür, dass das  $k$ -te gemeldete Fahrzeug von einem bestimmten Typ ist. Genauer gilt

$$t(k) = \{q_T(z_3; k) | z_3 \in Z_3\} \quad (5.68)$$

wobei  $q_T(z_3; k)$  eine Degree-of-Belief-Verteilung über der Menge der möglichen Fahrzeugtypen  $Z_3$  darstellt. Beschränkt sich die Angabe des Beobachters darauf, dass er nur für bestimmte Typen aus  $Z_3$  einen konkreten Degree-of-Belief-Wert angibt, so wird die restliche Wahrscheinlichkeitsmasse (konsistent zum Prinzip der Maximalen Entropie) als gleichverteilt auf die bei der Angabe nicht genannten Fahrzeugtypen angenommen.

Die in Form von  $d_1$  vorliegende Information wird wie folgt auf Basis von Mixturen von Normalverteilungen als Likelihood-Funktion  $l(d_1|z)$  modelliert:

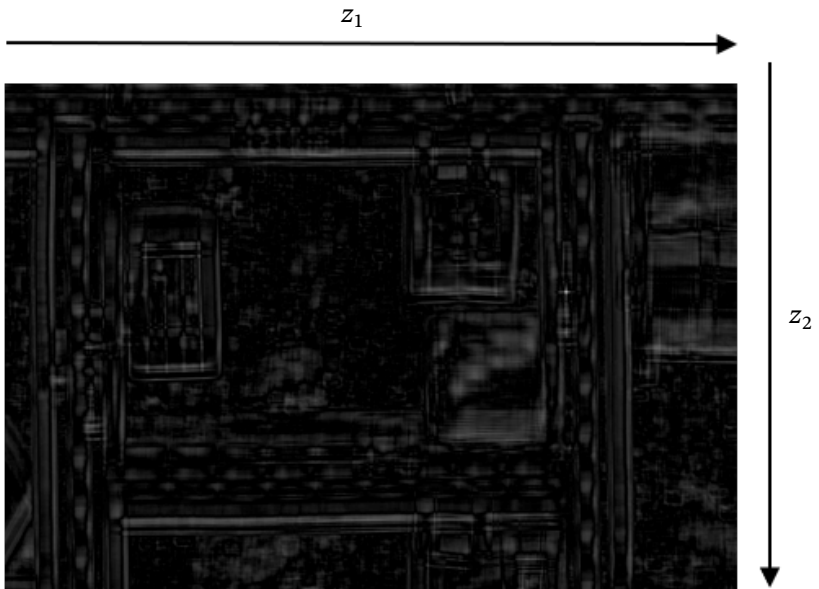
$$l(d_1|z) = l(d_1|z_1, z_2, z_3, z_4) = \sum_{k=1}^K q_T(z_3; k) q_P(z_1, z_2). \quad (5.69)$$

Dabei wird für  $q_P(z_1, z_2)$  die  $\mathcal{N}((\mu_1(k), \mu_2(k))^T, \sigma_\mu(k)^2 I_2)$  Verteilung über der Menge der möglichen Fahrzeugpositionen angesetzt. Diese wird dementsprechend für  $(z_1, z_2) \in Z_1 \times Z_2$  ausgewertet. Da der menschliche Beobachter keine explizite Angabe macht, wie sicher er sich jeweils hinsichtlich der gemeldeten Fahrtrichtung  $r_H(k)$  ist, wird die Likelihood-Funktion  $l(d_1|z)$  vereinfachend als nichtinformativ im Hinblick auf die Fahrtrichtung  $z_4$  modelliert.

Die resultierende Likelihood-Funktion  $l(d_1|z)$  ist in Abb. 5.7b visualisiert. Sie stellt für jeden in  $Z_3$  enthaltenen möglichen Fahrzeugtyp  $z_3$  eine Funktion über der Menge  $Z_1 \times Z_2$  der möglichen Fahrzeugpositionen dar. Die Darstellung der entsprechenden (hier) fünf Funktionen erfolgt in Abb. 5.7b anhand von fünf Grauwertbildern mit Abmessungen gemäß  $Z_1 \times Z_2$ . Je höher der Intensitätswert des mit einem möglichen Typ  $z_3$  korrespondierenden Grau-

wertbilds an einer bestimmten Position  $(z_1, z_2)$  ausfällt, desto höher ist der entsprechende (für jede mögliche Fahrtrichtung  $z_4$  aus  $Z_4$  gleiche) Wert der Likelihood-Funktion  $l(d_1|z)$ .

Zur Gewinnung der Likelihood-Funktionen  $l(d_s|z)$ ,  $s \in \{2,3,4\}$ , wird im Beispiel eine an die möglichen Typen gemäß  $Z_3$  angepasste Matched-Filter Bank, welche Grauwertbilder der möglichen Typen als Templates enthält, verwendet. Die Templates korrespondieren im Beispiel mit Grauwertbildern, welche ohne den Einsatz der Spektralfilter gewonnen wurden. Das Template Matching wird auf Gradientenbildern, welche aus  $d_2$ ,  $d_3$ ,  $d_4$  und den Templates durch Anwendung des Sobel-Operators (vgl. z. B. [Bey12]) gewonnen wurden, angewandt und basiert auf der Auswertung der normalisierten Kreuzkorrelation (vgl. z. B. [Asc93], [Bur08] oder [Mar95]).



**Abbildung 5.8:** Likelihood-Funktion  $l(d_2|z)$  für festen Fahrzeugtyp  $z_3 = A$  und feste Fahrtrichtung  $z_4 = s$  als Funktion von  $(z_1, z_2)$ .

Die Likelihood-Funktionen  $l(d_s|z)$ ,  $s \in \{2,3,4\}$ , ergeben sich gemäß dieser Umsetzung des Beispiels also durch eine Auswertung der einzelnen möglichen Werte für  $z \in Z$  und sind nicht durch eine überschaubare Anzahl an Parametern in geschlossener Form beschreibbar. In Abb. 5.8 ist die Likelihood-Funktion  $l(d_2|z)$  in Bezug auf den Fahrzeugtyp  $z_3 = A$  und die Fahrtrichtung  $z_4 = s$  als Funktion über der Menge  $Z_1 \times Z_2$  der möglichen Positionen als Grauwertbild visualisiert. Je höher der Intensitätswert des Grauwertbilds an einer bestimmten Position  $(z_1, z_2)$  ausfällt, desto höher ist der Wert von  $l(d_s|z_1, z_2, A, s)$ .

Wird die Bayes'sche Fusion global bezüglich ganz  $Z$  durchgeführt, so müssen die Ergebnisse der Auswertung von Gleichung (5.64) für alle möglichen Werte  $(z_1, z_2) \in Z_1 \times Z_2$  und die Ergebnisse der Auswertung von Gleichung (5.69) für alle möglichen Werte  $(z_1, z_2, z_3) \in Z_1 \times Z_2 \times Z_3$  bereitgestellt werden. Weiter ist das Ergebnis des Template Matching in Bezug auf jeden der Informationsbeiträge  $d_2$ ,  $d_3$  und  $d_4$  jeweils für jeden möglichen Wert  $(z_1, z_2, z_3, z_4) \in Z_1 \times Z_2 \times Z_3 \times Z_4$  zu berechnen und bereitzustellen. Erfolgt dies, z. B. bei Umsetzung der Bayes'schen Fusion innerhalb eines großen Netzwerks mit spezialisierten Auswerteeinheiten, nicht an der Stelle, an der auch die Berechnung und weitere Analyse des eigentlichen Fusionsergebnisses stattfindet, so müssen diese Größen dann auch entsprechend komplett übermittelt werden, was gerade bei weit verteilten Systemen und/oder vorliegenden Bandbreitenbeschränkungen kritisch sein kann. Auch werden hinsichtlich ihrer Kapazitäten begrenzte Auswerteeinheiten möglicherweise zu umfangreich in Anspruch genommen. Vgl. hierzu insbesondere auch Abschnitt 4.3.2.3.

Anhand der vorgenannten Größen muss dann eine nicht normierte Version der A-Posteriori-Verteilung für jeden möglichen Wert  $(z_1, z_2, z_3, z_4) \in Z$  (in einem Schritt oder wie in Gleichung (3.25) dargestellt sequentiell) berechnet werden:

$$p(z_1, z_2, z_3, z_4 | d_1, d_2, d_3, d_4) \propto p(z_1, z_2, z_3, z_4) \prod_{s=1}^4 l(d_s | z_1, z_2, z_3, z_4). \quad (5.70)$$



Anschließend müssen die signifikantesten Werte für  $(z_1, z_2, z_3)$  auf Basis der so resultierenden (nicht normierten) Version der A-Posteriori-Verteilung ermittelt werden.

Im konkret vorliegenden Beispiel wird entsprechend den in Form der Informationsbeiträge  $d_2$ ,  $d_3$  und  $d_4$  zur Auswertung vorliegenden Bilddaten, welche (gemäß Annahme) gerade das Interessengebiet abdecken und jeweils die Abmessungen  $1280 \text{ Pixel} \times 960 \text{ Pixel}$  haben, exemplarisch davon ausgegangen, dass ungefähr<sup>1</sup>  $|Z_1| = 1280$  und  $|Z_2| = 960$  gelten. Bereits für den hier weiter exemplarisch angenommenen Fall  $|Z_3| = 5$  und  $|Z_4| = 4$  ergibt sich somit, dass rein zur Berechnung der nicht normierten Version der A-Posteriori-Verteilung gemäß Gleichung (5.70)  $|Z| = \prod_{n=1}^4 |Z_n| > 2 \cdot 10^7$  mögliche Werte für  $z$  ausgewertet werden müssen.

### 5.3.3 Lokale Bayes'sche Fusion

Der erste Schritt im Hinblick auf die Umsetzung der lokalen Bayes'schen Fusion besteht in der Bestimmung des lokalen Kontexts  $U$ . Hierfür müssen die für die Fusion vorliegenden Informationsbeiträge bestmöglich, aber noch mit in der Praxis vertretbarem Aufwand hinsichtlich möglicher Auffälligkeiten in  $Z$  ausgewertet werden (vgl. Abschnitt 4.3.2.1), um den lokalen Kontext  $U$  auf dieser Basis konstituieren zu können.

Eine im Hinblick auf die (globale) Bayes'sche Modellierung gemäß Abschnitt 5.3.2 theoretisch gesehen optimale Vorauswertung der in Form von  $d_2$ ,  $d_3$ ,  $d_4$  vorliegenden Bilddaten lässt sich durch die Anwendung des zur Berechnung der Likelihood-Funktionen  $l(d_s|z)$ ,  $s \in \{2,3,4\}$ , verwendeten Template Matching Verfahrens erzielen. Die prinzipielle Machbarkeit dieses Vorgehens wurde u. a. in [Bey07] anhand der dort konkret adressierten Version des Beispiels aufgezeigt.

---

<sup>1</sup> Da beim Template Matching Verfahren die Bildränder jeweils bis zur halben Template-Größe nicht exakt ausgewertet werden können, sollten  $|Z_1|$  und  $|Z_2|$  genau genommen minimal geringer angesetzt werden.

Um zu vermeiden, dass die exakten, auf dem Template Matching Verfahren basierenden Likelihood-Funktionen  $l(d_s|z)$ ,  $s \in \{2,3,4\}$ , jeweils für alle möglichen Werte aus  $Z$  berechnet werden müssen, was nicht tragbar sein könnte, kann wie in [San09c] und [San10b] aufgezeigt im vorliegenden exemplarischen Anwendungsfall alternativ wie folgt vorgegangen werden: Die Gradientenbilder, welche aus  $d_2$ ,  $d_3$  und  $d_4$  durch Anwendung des Sobel-Operators gewonnen werden, werden nach dem Vorhandensein horizontal und vertikal ausgerichteter Kantenpaare, deren Abmessung und Anordnung in etwa den seitlichen Außenmaßen eines Fahrzeugs im Bild entspricht, abgesucht. Dabei wird von der durchschnittlichen Länge und Breite eines Fahrzeugs ausgegangen und es werden weiter gewisse Variationen hinsichtlich Fahrzeug-Länge, Fahrzeug-Breite und Kanten-Offset berücksichtigt (vgl. [San09c] zu Details). Das Ergebnis stellen dann  $J \in \mathbb{N}$  mögliche Kantenpaare dar.

Die Mittelpunkte  $m_j$ ,  $j \in \{1, \dots, J\}$ , der rechteckigen Strukturen, welche mit den aufgefundenen Kantenpaaren korrespondieren, werden dann als mögliche Fahrzeugpositionen erfasst. Um real zu erwartenden Abweichungen zwischen den so ermittelten Mittelpunkten und den tatsächlichen Mittelpunkten der Fahrzeuge Rechnung zu tragen, werden weiter auch kleine, um die Punkte  $m_j$  gelagerte räumliche Umgebungen  $U_1^{(j)} \times U_2^{(j)} \subset Z_1 \times Z_2$  als mögliche Fahrzeugpositionen erfasst. Diese Umgebungen konstituieren dann Ausschnitte aus  $Z$ , welche von folgender Form sind:

$$U^{(j)} = U_1^{(j)} \times U_2^{(j)} \times Z_3 \times U_4^{(j)} \subset Z, \quad j \in \{1, \dots, J\}, \quad (5.71)$$

d. h. sie schränken jeweils die Menge  $Z_1 \times Z_2$  der möglichen Fahrzeugpositionen und die Menge der möglichen Fahrtrichtungen  $Z_4$  ein. Dabei gilt  $U_4^{(j)} = \{n, s\}$ , falls das zugehörige  $j$ -te Kantenpaar vertikal ausgerichtet ist, und  $U_4^{(j)} = \{o, w\}$ , falls es horizontal ausgerichtet ist.

Der beschriebene Ansatz zur Vorauswertung der in Form von  $d_2$ ,  $d_3$  und  $d_4$  als Bilddaten vorliegenden Informationsbeiträge lässt sich anhand von in [Zha01] veröffentlichten Ergebnissen zu psychologischen Tests, in denen wichtige Merkmale, anhand deren Menschen Autos in Bildern erkennen, identifiziert wurden, motivieren. Die Ursache dafür, dass er sich trotz seiner

Einfachheit im konkret vorliegenden Beispiel recht gut zur Vorauswertung der Bilddaten eignet, gründet sich dabei natürlich vor allem auch auf die Tatsache, dass es im Hinblick auf die Korrektheit von mittels lokaler Bayes'scher Fusion erzielten Ergebnissen genügt, dass die Vorauswertung so gestaltet ist, dass der lokale Kontext  $U$  insgesamt gesehen mindestens die tatsächlich relevanten Auffälligkeiten enthält. Vgl. hierzu insbesondere auch die Evaluation und Diskussion der Ergebnisse im nachfolgenden Abschnitt 5.3.4.

Im Hinblick auf den Informationsbeitrag  $d_1$ , welcher die von den menschlichen Beobachtern gelieferte Information verkörpert, können mögliche Auffälligkeiten in  $Z$  direkt auf Basis von Gleichung (5.66) festgelegt werden. Hinsichtlich der tatsächlichen Positionen der  $K$  gemeldeten Fahrzeuge kann davon ausgegangen werden, dass diese (sofern die Fahrzeuge tatsächlich vorhanden sind) mit hoher Wahrscheinlichkeit innerhalb von räumlichen Regionen um die gemeldeten Positionen  $(\mu_1(k), \mu_2(k))$ ,  $k \in \{1, \dots, K\}$ , liegen. Die Größe der räumlichen Regionen lässt sich dabei anhand der für die Positionsangaben jeweils bekannten räumlichen Unsicherheiten  $\sigma_\mu(k)$  festlegen. Die Annahme, dass der menschliche Beobachter recht zuverlässig meldet, ob sich ein Fahrzeug in horizontaler oder vertikaler Fahrtrichtung bewegt, führt damit dann letztlich auf (hier exemplarisch bezüglich der Menge der möglichen Positionen als quadratisch angesetzte) Ausschnitte von  $Z$  der folgenden Form:

$$U^{(j)} = U_1^{(j)} \times U_2^{(j)} \times Z_3 \times U_4^{(j)} \subset Z, \quad j \in \{J+1, \dots, J+K\}, \quad (5.72)$$

mit

$$U_1^{(j)} \times U_2^{(j)} = \{(z_1, z_2) \mid \mu_i(j-J) - \sigma(j-J) \leq z_i \leq \mu_i(j-J) + \sigma(j-J), i \in \{1, 2\}\} \quad (5.73)$$

und  $U_4^{(j)} = \{n, s\}$ , falls  $r_H(j-J) = n \vee s$ , und  $U_4^{(j)} = \{o, w\}$ , falls  $r_H(j-J) = o \vee w$  gilt. Konkret wird für die nachfolgenden Berechnungen angenommen, dass für alle  $j \in \{J+1, \dots, J+K\}$  gerade  $\sigma(j-J) = \sigma_\mu(j-J) = 0,1 \cdot |Z_1|$  gilt. Sofern die gemäß Gleichung (5.68) gemeldete Einschätzung  $t(k)$  hinsichtlich des Typs des  $k$ -ten gemeldeten Fahrzeugs besonders signifikante Werte aufweist, lässt

sich die in Gleichung (5.72) getroffene Festlegung für den entsprechenden  $J + k$ -ten Ausschnitt aus  $Z$  noch weiter verfeinern, indem man  $Z_3$  durch die Teilmenge der entsprechend signifikanten Fahrzeugtypen ersetzt. Diese Maßnahme wurde im vorliegenden Beispiel der Übersichtlichkeit halber nicht vorgesehen. Es sei jedoch betont, dass sich gerade beim Vorliegen einer umfangreicheren Menge an möglichen Fahrzeugtypen  $Z_3$  als im vorliegenden Beispiel der Fall hierdurch wesentliches Potential zur weiteren, auch deutlichen Verkleinerung der entsprechenden Ausschnitte  $U^{(j)}$ ,  $j \in \{J + 1, \dots, J + K\}$ , von  $Z$  ergibt.

Insgesamt ergibt sich der lokale Kontext  $U$  zu

$$U = \bigcup_{j=1}^{J+K} U^{(j)}. \quad (5.74)$$

Zur Durchführung der lokalen Bayes'schen Fusion auf Basis einer Einschränkung auf  $U$  genügt es nun grundsätzlich, die (nicht normierte) A-Priori-Verteilung  $p(z)$  gemäß Gleichung (5.64) und die Likelihood-Funktion  $l(d_1|z)$  gemäß Gleichung (5.69) nur ausschnittsweise für  $z \in U$  bereitzustellen. Ebenso muss bezüglich der Likelihood-Funktionen  $l(d_s|z)$ ,  $s \in \{2,3,4\}$ , das Ergebnis des Template Matching nur für  $z \in U$  bereitgestellt werden. Weiter sind die Berechnung der nicht normierten A-Posteriori-Verteilung gemäß Gleichung (5.70) und die Bestimmung der signifikantesten Werte für  $(z_1, z_2, z_3)$  nur im Hinblick auf  $z \in U$  durchzuführen.

Die Berechnung der normalisierten Kreuzkorrelation im Ortsbereich kann unnötig hohen Aufwand verursachen. In diesem Fall lässt sich eine Beschleunigung durch den Einsatz der schnellen Fourier-Transformation (im Englischen: fast Fourier transform (FFT)) und den Einsatz von Integralbildern erzielen (siehe hierzu z. B. [Bey12], [Gon02], [Lew95] und [Vio01]). Hierfür müssen die auszuwertenden Anteile von  $Z_1 \times Z_2$  jedoch in (jeweils geschlossen) rechteckiger Form vorliegen. Da diese Situation einen guten Beispielfall dafür darstellt, dass es nötig sein kann, einen lokalen Kontext aus praktischen Gründen wieder auszuweiten und hinsichtlich seiner Zerlegung in einzelne Anteile

umzustrukturieren, soll auf das mögliche Vorgehen hierbei anhand des vorliegenden Beispiels noch etwas genauer eingegangen werden.

Als Vorbereitung für das Template Matching werden dazu für jedes mögliche Richtungspaar  $r_i \in R_i = \{\{n,s\}, \{o,w\}\}$  die zusammenhängenden Anteile in den durch den lokalen Kontext  $U$  als auszuwertend notierten räumlichen Gebieten

$$G_1(r_i) \times G_2(r_i) := \bigcup_{\substack{j \in \{1, \dots, J+K\}: \\ U_4^{(j)} = r_i}} U_1^{(j)} \times U_2^{(j)} \subset Z_1 \times Z_2 \quad (5.75)$$

auf eine rechteckige Form von minimaler Größe ausgeweitet. Hierfür kann ein Region Growing Verfahren (siehe z. B. [Aza07] oder [Gon02]) angewandt werden. Ein Region Growing Verfahren ist ein Segmentierungsverfahren, welches einzelne Punkte oder Regionen eines Bildes nach einem vordefinierten Kriterium in größere Regionen gruppiert. Es bezeichne  $J_1$  die Anzahl der für  $r_i = \{n,s\}$  und  $J_2$  die Anzahl der für  $r_i = \{o,w\}$  resultierenden Gebiete. Zur Durchführung des Template Matchings wird der lokale Kontext  $U$  dann entsprechend der so konstituierten rechteckigen Gebiete  $V_1^{(j)} \times V_2^{(j)} \subset Z_1 \times Z_2$ ,  $j \in \{1, \dots, J_1 + J_2\}$ , auf  $V$  mit  $U \subset V$  ausgeweitet:

$$V = \bigcup_{j=1}^{J_1+J_2} V_1^{(j)} \times V_2^{(j)} \times Z_3 \times V_4^{(j)}. \quad (5.76)$$

Dabei gilt  $V_4^{(j)} = \{n,s\}$  für  $j \in \{1, \dots, J_1\}$  und es gilt weiter  $V_4^{(j)} = \{o,w\}$  für  $j \in \{J_1 + 1, \dots, J_1 + J_2\}$ .

Um dafür zu sorgen, dass die Differenz  $|V| - |U|$  möglichst moderat ausfällt, kann dabei noch das folgende Schema rekursiv für die Identifikation und Verkleinerung von zu umfangreich ausfallenden Anteilen  $V_1^{(j)} \times V_2^{(j)}$ ,  $j \in \{1, \dots, J_1 + J_2\}$ , angewandt werden: Es wird geprüft, ob  $V_1^{(j)} \times V_2^{(j)}$  einen recht großen nicht relevanten, rechteckigen Anteil  $B$  aus  $Z_1 \times Z_2$  enthält. Ein solcher Anteil  $B \subset V_1^{(j)} \times V_2^{(j)}$  liegt vor, wenn er mehr als  $\lambda$  Prozent von  $|V_1^{(j)} \times V_2^{(j)}|$  ausmacht und (für das entsprechende Richtungspaar) kein Teil

von den in Form des lokalen Kontexts  $U$  identifizierten Bereichen von  $Z_1 \times Z_2$  ist. Enthält  $V_1^{(j)} \times V_2^{(j)}$  einen derartigen Anteil  $B$ , so wird er aus  $V_1^{(j)} \times V_2^{(j)}$  entfernt, und der Rest von  $V_1^{(j)} \times V_2^{(j)}$  wird (in Abhängigkeit von der Lage von  $B$  innerhalb von  $V_1^{(j)} \times V_2^{(j)}$ ) in zwei, drei oder vier neue rechteckige Gebiete aufgespalten. Es sei angemerkt, dass sich dieses Schema zur Aufspaltung der Anteile  $V_1^{(j)} \times V_2^{(j)}$ ,  $j \in \{1, \dots, J_1 + J_2\}$ , grundsätzlich gesehen auch noch adaptieren lässt, um eine bessere verteilte Umsetzung der lokalen Bayes'schen Fusion (falls eine solche vorgesehen ist) zu ermöglichen. So könnte man z. B. auch allgemein recht umfangreiche resultierende Anteile aufsplitten.

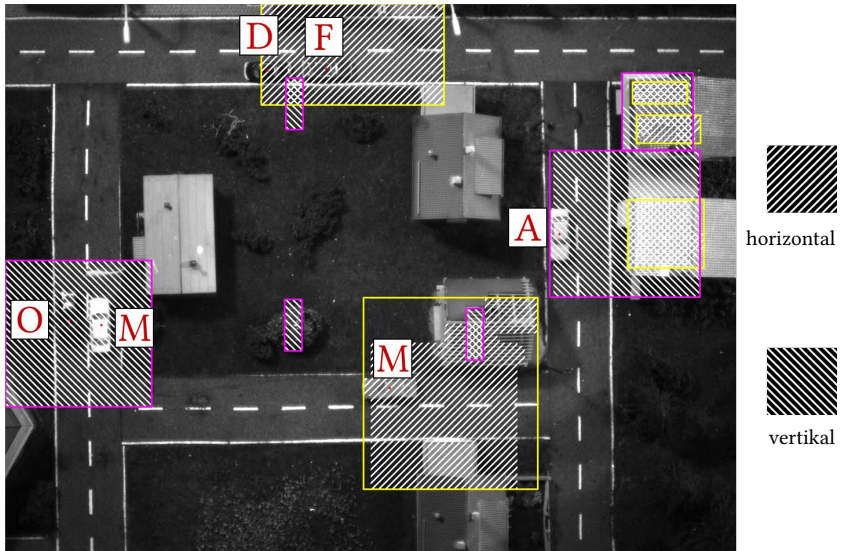
### 5.3.4 Evaluation und Diskussion

In Abb. 5.9 ist das Endergebnis der zuvor beschriebenen, in Teilen heuristischen Umsetzung der lokalen Bayes'schen Fusion im Hinblick auf die gemäß dem Ziel der Aufklärung vorzunehmende Feststellung des Vorhandenseins sowie von Typ und Position der Fahrzeuge im Interessengebiet illustriert. Alle Fahrzeuge wurden korrekt detektiert und klassifiziert, das Endergebnis enthält auch keine Fehldetektionen.

Die schraffierten Gebiete im in Abb. 5.9 dargestellten Ergebnisbild markieren die räumlichen Bereiche, für welche (nicht normierte) Werte der A-Posteriori-Verteilung berechnet wurden und welche dann weiter (lokal) hinsichtlich der signifikantesten Werte für  $(z_1, z_2, z_3)$  ausgewertet wurden. Die genaue Ausprägung der Schraffur kennzeichnet, ob dabei die Hypothese  $z_4 \in \{o, w\}$  (horizontale Fahrtrichtung) oder die Hypothese  $z_4 \in \{n, s\}$  (vertikale Fahrtrichtung) betrachtet wurde.

Die farbig markierten Rechtecke in Abb. 5.9 kennzeichnen ergänzend in räumlicher Hinsicht die Bereiche, welche sich wie in Abschnitt 5.3.3 beschrieben bei der Ausweitung von  $U$  zu  $V$  ergeben. Die dabei beim Template Matching jeweils auszuwertenden Fahrtrichtungen sind in der Abbildung anhand der Schraffur der durch die farbigen Rechtecke jeweils umschlossenen Bereiche erkennbar. Zur finalen Konstitution von  $V$  wurde  $\lambda = 15$  (Prozent) gewählt. In diesem Fall ergibt sich, dass 89 Prozent der beim Template Matching berechneten Werte der Likelihood-Funktionen  $l(d_s|z)$ ,  $s \in \{2, 3, 4\}$ , auch tatsächlich

zur Berechnung der (nicht normierten) A-Posteriori-Verteilung benötigt werden.



**Abbildung 5.9:** Endergebnis der auf Basis lokaler Bayes'scher Fusion durchgeführten Detektion und Klassifikation der im Interessengebiet vorhandenen Fahrzeuge sowie Darstellung von  $U$  und  $V$ .

Bei der vorgenommenen Umsetzung (mit den zuvor jeweils im Text angegebenen Parametrisierungen der Verfahren) gilt  $|U| \approx 3 \cdot 10^6$ , während hinsichtlich des gesamten, bei der Modellierung der Fusionsaufgabe festgelegten Wertebereichs der Größen des Interesses wie in Abschnitt 5.3.2 dargestellt  $|Z| > 2 \cdot 10^7$  gilt. Dabei ist anzumerken, dass sich in einer realen Anwendung eine deutlichere Einschränkung des Wertebereichs der Größen des Interesses ergeben kann, als es das hier konkret dargestellte, diesbezüglich durchaus mit Einschränkungen versehene Beispiel zu verdeutlichen vermag. Entsprechende Faktoren hierfür stellen, wiederum bezogen auf das konkrete

Beispiel, das Vorliegen eines umfangreicheren Interessengebiets mit räumlich verteilter angeordneter Fahrzeugen sowie das Vorliegen einer umfangreicheren Menge  $Z_3$  von möglichen Fahrzeugtypen verbunden mit dem Vornehmen einer Einschränkung auch hinsichtlich dieser dar. Auf weitere Faktoren in methodischer Hinsicht wird nachfolgend noch eingegangen.

Bei der konkret vorgenommenen Umsetzung der lokalen Bayes'schen Fusion für das im Rahmen des vorliegenden Abschnitts betrachtete Beispiel wurde bewusst der Fall adressiert, dass nicht das (aus Sicht der (globalen) Bayes'schen Modellierung) optimale Verfahren zur Auswertung der in Form von  $d_2$ ,  $d_3$  und  $d_4$  vorliegenden Bilddaten verwendet wird. Hierdurch wird deutlich, dass auch der Einsatz eines nicht optimalen Verfahrens zur Vorauswertung der Informationsbeiträge im Hinblick auf die Bestimmung von Auffälligkeiten in  $Z$  zu korrekten Ergebnissen bei der lokalen Bayes'schen Fusion führen kann. Wesentlich hierfür ist, dass der lokale Kontext  $U$  tatsächlich die relevanten Auffälligkeiten enthält. Die Tatsache, dass der lokale Kontext  $U$  (wie hier der Fall) dadurch zusätzliche mögliche Werte aus  $Z$  enthält, hat dabei keine negativen Auswirkungen im Hinblick auf die Korrektheit der lokalen Bayes'schen Fusion. Sie wirkt sich durch den resultierenden größeren Umfang des lokalen Kontexts  $U$  aber natürlich auf den Aufwand zur Berechnung der A-Posteriori-Verteilung aus, welcher sich entsprechend erhöht. Die Entscheidung, ob es im Realfall vorteilhafter ist, mehr Aufwand in eine optimale Vorauswertung zu investieren, oder alternativ (wie hier beispielhaft umgesetzt) bedingt durch einen größeren Umfang des lokalen Kontexts  $U$  mehr Aufwand zur Durchführung der eigentlichen Fusion in Kauf zu nehmen, wird i. A. an den konkret vorliegenden Ressourcen (auch: Auswertefähigkeiten) und ggf. auch deren Verteilung und Konnektivität z. B. innerhalb eines größeren Netzwerks festzumachen sein.

Durch die Verwendung der Spektralfilter bei der Aufnahme der in Form von  $d_2$ ,  $d_3$  und  $d_4$  vorliegenden Bilddaten ergibt sich ein gutes Beispiel dafür, wie sich die unterschiedlichen, bei der Fusion vorliegenden Informationsbeiträge gegenseitig ergänzen können. Dies ist bei der konkret durchgeführten Umsetzung bereits bei der nicht optimalen Vorauswertung der Bilddaten zur Konstitution der entsprechenden Anteile von  $U$  der Fall. Auch trägt im vorliegenden



Beispiel die Berücksichtigung der von den menschlichen Beobachtern gelieferten Information wesentlich dazu bei, dass der lokale Kontext  $U$  tatsächlich alle relevanten Werte für  $(z_1, z_2, z_3)$  enthält. Da die Fahrzeuge vom Typ D und O sich in allen drei Spektralbildern nur sehr schlecht vom Untergrund abheben, werden die entsprechenden Anteile von  $Z$  bei der vorgenommenen nicht optimalen Vorauswertung der Bilddaten nämlich nicht als auffällig identifiziert. In einer realen Anwendung könnte die von menschlichen Beobachtern gelieferte Information auch entsprechend komplettierend wirken, wenn die zu detektierenden Objekte (hier: Fahrzeuge) in den Bilddaten aufgrund von Verdeckungen nicht oder nur teilweise sichtbar sind.

Da dem Menschen in der Regel die Feststellung der genauen Positionen nicht besonders exakt gelingt, sind (wie in Abb. 5.9 gut erkennbar) aufgrund der resultierenden ausgeprägten Unsicherheiten diesbezüglich die auf Basis des Informationsbeitrags  $d_1$  resultierenden Anteile von  $U$  in räumlicher Hinsicht deutlich umfassender ausgeprägt als die Anteile, welche sich durch Vorauswertung der in Form von  $d_2$ ,  $d_3$  und  $d_4$  vorliegenden Bilddaten ergeben. Das im vorliegenden Abschnitt adressierte Beispiel verdeutlicht dadurch auch sehr gut, dass ein Abgleich der sich durch die Vorauswertung unterschiedlicher Informationsbeiträge ergebenden Ausschnitte von  $Z$  hilfreich sein kann. Liefert die Vorauswertung der in Form von  $d_2$ ,  $d_3$  und  $d_4$  vorliegenden Bilddaten Auffälligkeiten in  $Z$ , von welchen auszugehen ist, dass sie mit Auffälligkeiten, welche auf Basis von  $d_1$  ermittelt wurden, korrespondieren, so könnte man die lokale Bayes'sche Fusion in Bezug auf diese z. B. in einem ersten Schritt nur bezogen auf die in räumlicher Hinsicht weniger umfangreichen Anteile, welche sich aus der Vorauswertung von  $d_2$ ,  $d_3$  und  $d_4$  ergeben, durchführen. Gegebenenfalls könnte man diese unter Berücksichtigung der aus der Betrachtung von  $d_1$  resultierenden Ausschnitte von  $Z$  dann auch hinsichtlich der im ersten Schritt berücksichtigten möglichen Fahrzeugtypen noch weiter einschränken. Man könnte speziell für den Fall der hier durchgeführten qualitativen Analyse dann prüfen, ob sich bereits durch diesen ersten Schritt signifikante Werte  $(z_1, z_2, z_3)$  insgesamt ausreichend im Hinblick auf die Ableitung erster konkreter (möglicherweise zeitkritischer) Aktionen identifizieren lassen. Die in diesem ersten Schritt nicht betrachteten Anteile von  $U$  könnten dann im Rahmen einer Zweitanalyse, bei der man die lokale Bayes'sche Fusion im

Hinblick auf die noch nicht analysierten Anteile von  $U$  komplettiert, miteinbezogen werden. Dadurch könnte man dann die erzielten Ergebnisse weiter verifizieren und/oder erforderliche Aktionen weiter präzisieren. Ein Beispiel für eine entsprechende Aktion könnte allgemein gesehen die Planung weiterer Aufklärungsaktivitäten, z. B. in Form eines Cross Cueings (siehe z. B. [Koc14]), bezogen auf einen als besonders beobachtungswert identifizierten Teilbereich des Interessengebiets sein.

Insgesamt lässt sich festhalten, dass das im vorliegenden Abschnitt 5.3 adressierte Beispiel trotz seiner (notwendigerweise) nicht vollständigen Realitäts-treue und der vereinfachten und in Teilen auch heuristischen Umsetzung der lokalen Bayes'schen Fusion relevante Erkenntnisse für die vorliegende Arbeit liefert, welche die aus den theoretischen Untersuchungen resultierenden Erkenntnisse komplettieren. Besonderes erwähnt werden diesbezüglich sollen insbesondere auch noch einmal die Vorteile, die sich im Hinblick auf die weitere Auswertung von Informationsbeiträgen und letztlich auch die Handhabung der lokal zu betrachtenden Anteile von  $Z$  ergeben, wenn man einen lokalen Kontext auf eine besser verarbeitbare Struktur ausweitet. In der Praxis müssen dann auch für solche Aufgaben geeignete und ausreichend effiziente Verfahren zur Verfügung stehen.

## 6 Mathematische Kriterien zur Wahl des lokalen Kontexts

### 6.1 Einführung

Für die lokale Bayes'sche Fusion müssen in einem ersten Schritt die zur Fusion vorliegenden Informationsbeiträge im Hinblick auf Auffälligkeiten in  $Z$  untersucht werden. Auf Basis dieser Auffälligkeiten wird dann der lokale Kontext  $U \subset Z$ , auf den die Fusion konzentriert wird, festgelegt. Im in Abschnitt 5.3 bereits adressierten Beispiel zur lokalen Bayes'schen Fusion wurde der lokale Kontext  $U$  recht intuitiv anhand der konkreten Aufgabenstellung, der vorliegenden Informationsbeiträge und der zu ihrer Auswertung verwendeten Verfahren festgelegt. Wesentlich im Hinblick auf die Praktikabilität eines derartigen Vorgehens ist dabei vor allem die Tatsache, dass es für die Korrektheit der lokalen Bayes'schen Fusion genügt, wenn der lokale Kontext  $U$  insgesamt gesehen mindestens die tatsächlich relevanten Auffälligkeiten aus  $Z$  enthält.

Der Begriff Korrektheit ist dabei in dem Sinne zu verstehen, dass der lokale Kontext  $U$  die tatsächlich relevanten Auffälligkeiten aus  $Z$  enthalten muss, da ansonsten die mittels lokaler Bayes'scher Fusion erzielten Ergebnisse kaum nützlich sind und sich gerade bei einer nicht nur qualitativen Analyse zudem die Gefahr, dass sie zu falschen Schlüssen verleiten, deutlich erhöht. Die in Abschnitt 5.2 formulierten Regeln zur Bildung eines lokalen Bayes'schen Modells und zur Durchführung der (lokalen) Berechnungen sind dementsprechend mathematisch gesehen immer korrekt – unabhängig davon, ob der lokale Kontext  $U$  sinnvoll gewählt wurde oder nicht. Das in Abschnitt 5.3 behandelte Beispiel sowie die in Form von Gleichung (5.59) angegebene Möglichkeit, anhand welcher Agenten ihre lokalen Ergebnisse näherungsweise in Bezug

setzen können, haben auch deutlich gemacht, dass es im Hinblick auf die Einordnung der mittels lokaler Bayes'scher Fusion erzielten Ergebnisse grundsätzlich auch hilfreich sein kann, wenn der lokale Kontext  $U$  zusätzliche Elemente aus  $Z$  enthält, welche als Referenz im Hinblick auf die Einschätzung der Signifikanz lokal berechneter (im Beispiel nicht normierter) Wahrscheinlichkeitsaussagen dienen können.

In diesem Kapitel wird der Frage nachgegangen, was aus mathematischer Hinsicht sinnvolle Vorgehensweisen zur Festlegung eines lokalen Kontexts  $U \subset Z$  (auch von möglichst minimalem Umfang) sind und wie sich die Güte lokaler Bayes'scher Modelle generell in Abhängigkeit von der Wahl des lokalen Kontexts  $U$  bewerten lässt. Hierzu werden zum einen in Abschnitt 6.2 die in Abschnitt 3.4.3 aufbereiteten Grundlagen der Likelihood-Inferenz aufgegriffen und erweitert, bevor auf dieser Basis mögliche Vorgehensweisen zur Festlegung eines lokalen Kontexts  $U$  und probabilistische Fehlerschranken im Hinblick auf die voraussichtliche Stimmigkeit eines lokalen Bayes'schen Modells abgeleitet werden. Zum anderen werden in Abschnitt 6.3 die in Abschnitt 3.5.1 eingeführten Prinzipien aus dem Kontext der Informationstheorie aufgegriffen und erweitert, um die Frage nach der Korrektheit der lokalen Bayes'schen Fusion aus informationstheoretischer Sicht zu beleuchten.

## 6.2 Prinzipien statistischer Evidenz

### 6.2.1 Irreführende statistische Evidenz

Gilt für  $z^*, z^{**} \in Z$  dass  $l(d|z^*) > l(d|z^{**})$  ist, so liefern die in  $d$  zusammengefassten Informationsbeiträge gemäß dem bereits in Abschnitt 3.4.3 eingeführten Law of Likelihood statistische Evidenz für  $z^*$  vis-à-vis  $z^{**}$ . Weiter stellt gemäß dem Law of Likelihood die Likelihood-Ratio  $l(d|z^*)/l(d|z^{**})$  ein quantitatives (relatives) Maß für die Stärke der von  $d$  gelieferten statistischen Evidenz dar.

Zwecks verbaler Interpretation der Stärke statistischer Evidenz ist es in der Fachliteratur üblich, die kontinuierlichen Werte der Likelihood-Ratio in

unterschiedliche Kategorien einzuteilen (siehe z. B. [Ait04], [Blu02], [Edw92], [Jef61], [Luc05], [Kas95], [Roy97], [Roy00]). Gängig ist besonders die in [Roy00] vorgeschlagene Einteilung, welche vom Autor bereits zuvor in [Roy97] unter Bezugnahme auf ein kanonisches, auf dem Ziehen aus unterschiedlichen Urnen basierendes Zufallsexperiment (vgl. z. B. auch [Hen10]) begründet wurde. Dieser Einteilung zufolge stellt (für  $z^*$  vis-à-vis  $z^{**}$ ) eine Likelihood-Ratio  $l(d|z^*)/l(d|z^{**}) = k$  schwache (im Englischen: weak) statistische Evidenz dar, falls  $1 < k < 8$  gilt. Sie stellt recht starke (im Englischen: fairly strong) statistische Evidenz dar, falls  $8 \leq k < 32$  gilt, und starke (im Englischen: strong) statistische Evidenz, falls  $k \geq 32$  gilt.

Statistische Evidenz kann grundsätzlich auch irreführend sein. Die in  $d$  zusammengefassten Informationsbeiträge liefern (insgesamt gesehen) irreführende statistische Evidenz in Bezug auf die in Form von  $z^*$  und  $z^{**}$  vorliegenden Hypothesen für den „wahren“ Wert der (durch  $z$  verkörperten) Größen des Interesses, wenn gilt: Es ist  $l(d|z^*)/l(d|z^{**}) > 1$ , tatsächlich ist aber  $z^{**}$  der „wahre“ Wert von  $z$ .

Dass das Law of Likelihood trotz der Tatsache, dass statistische Evidenz grundsätzlich irreführend sein kann, stimmig ist, liegt u. a. daran, dass die Wahrscheinlichkeit dafür, dass irreführende statistische Evidenz von einer bestimmten Stärke auftritt, grundsätzlich in Abhängigkeit von deren Stärke begrenzt ist. Genauer gilt allgemein (mit der bisherigen Notation) für

$$A := \left\{ d \in D \mid \frac{l(d|z^*)}{l(d|z^{**})} \geq k \right\} \quad (6.1)$$

die Abschätzung

$$P_{z^{**}}(A) = \int_{d \in A} p(d|z^{**}) dd \leq \frac{1}{k}. \quad (6.2)$$

Ist also  $z^{**}$  der „wahre“ Wert der (durch  $z$  verkörperten) Größen des Interesses, so beträgt die Wahrscheinlichkeit dafür, dass  $d$  dementsprechend irreführende statistische Evidenz von mindestens der Stärke  $k$  für  $z^*$  vis-à-vis  $z^{**}$  liefert, höchstens  $1/k$ . Die in Form von Gleichung (6.1) und Gleichung (6.2)

formulierte allgemeingültige Schranke für die Wahrscheinlichkeit irreführender statistischer Evidenz wird von Royall in [Roy97] und [Roy00] bewiesen und intensiv diskutiert. Sie wurde bereits davor von anderen Wissenschaftlern in unterschiedlichen Kontexten hergeleitet. Zu nennen sind diesbezüglich [Dem97] und [Las91] sowie weitere Autoren, welche Royall in [Roy00] anführt. Da der Beweis recht einfach ist, soll er an dieser Stelle in Anlehnung an [Roy97] kurz wiedergegeben werden: Für  $d \in A$  gilt wegen Gleichung (6.1), dass  $p(d|z^{**}) \leq 1/k p(d|z^*)$  ist. Hieraus und aus der Tatsache, dass die Wahrscheinlichkeit von Ereignissen immer kleiner gleich Eins ist, ergibt sich sofort

$$P_{z^{**}}(A) = \int_{d \in A} p(d|z^{**}) dd \leq \frac{1}{k} \int_{d \in A} p(d|z^*) dd \leq \frac{1}{k}. \quad (6.3)$$

Es sei angemerkt, dass sich beim Vorliegen spezieller Verteilungsannahmen, insbesondere Normalverteilungsannahmen, auch genauere Abschätzungen ergeben, siehe hierzu z. B. [Roy00].

Die Wahrscheinlichkeit für irreführende statistische Evidenz ist relevant für die Planung eines statistischen Experiments. Sie spielt jedoch keine Rolle im Hinblick auf die letztliche Feststellung, ob  $z^{**}$  tatsächlich der „wahre“ Wert der (durch  $z$  verkörperten) Größen des Interesses ist. Basis dieser Feststellung muss allein die aus der Anwendung des Satz von Bayes (vgl. Gleichung (3.22)) resultierende A-Posteriori-Wahrscheinlichkeitsverteilung  $p(z|d)$  sein (vgl. [Roy97] und insbesondere auch [Blu11]). Wie in Abschnitt 3.4.3 bereits erläutert wurde, ist für diese Feststellung gerade der bezüglich des „wahren“ Werts von  $z$  insgesamt vorliegende Informationsstand, welcher auch das vorhandene Vorwissen umfasst, relevant.

Aufgrund ihrer grundsätzlichen Bedeutung im Hinblick auf die Stimmigkeit des Law of Likelihood soll an dieser Stelle noch kurz auf einige weitere, u. a. in [Roy97] und [Roy00] angeführte und diskutierte Ergebnisse im Hinblick auf die Wahrscheinlichkeit irreführender statistischer Evidenz eingegangen werden. Hierzu sei (ausnahmsweise) angenommen, dass  $d_1, d_2, \dots$  eine beliebig fortsetzbare Folge unabhängiger und identisch gemäß  $p(d_s|z^{**})$  verteilter

Zufallsvariablen darstellt ( $s \in \mathbb{N}$ ). Gemäß [Rob70] gilt

$$P_{z^{**}} \left( \prod_1^s \frac{l(d_s|z^*)}{l(d_s|z^{**})} \geq k \text{ für (mindestens) ein } s \in 1, 2, \dots \right) \leq \frac{1}{k}. \quad (6.4)$$

Gleichung (6.4) sagt gerade aus, dass man die in Form von Gleichung (6.1) und Gleichung (6.2) formulierte Schranke für die Wahrscheinlichkeit irreführender statistischer Evidenz letztlich auch nicht umgehen kann, indem man das mit der Generierung einer gemäß  $p(d_s|z^{**})$  verteilten Beobachtung verbundene Zufallsexperiment gezielt wiederholt, um irreführende statistische Evidenz der Stärke mindestens  $k$  für  $z^*$  vis-à-vis  $z^{**}$  zu generieren. Dieses Vorhaben wird mit Wahrscheinlichkeit  $1 - 1/k$  fehlschlagen (vgl. hierzu auch [Blu02]). Da die in Gleichung (6.4) angegebene Likelihood-Ratio  $\prod_1^s l(d_s|z^*)/l(d_s|z^{**})$  für  $s \rightarrow \infty$  fast sicher (d. h. mit Wahrscheinlichkeit Eins) gegen Null konvergiert (vgl. [Roy97] und [Roy00]), kann man weiter davon ausgehen, dass im Fall, dass  $z^{**}$  der „wahre“ Wert ist, bei ausreichend großer Wahl von  $S$  durch  $d = (d_1, \dots, d_S)$  statistische Evidenz von mindestens der Stärke  $k$  für  $z^{**}$  vis-à-vis  $z^*$  geliefert wird.

## 6.2.2 Anwendung für die lokale Bayes'sche Fusion

Im vorliegenden Abschnitt werden Ergebnisse formuliert, welche die in Abschnitt 6.2.1 in erweiterter und aufbereiteter Form dargestellten Prinzipien statistischer Evidenz als Basis nutzen, um eine aus statistischer Sicht sinnvolle und quantitativ bewertbare Vorgehensweise zu Festlegung des lokalen Kontexts  $U$  zu formulieren. Diese Ergebnisse und die erforderlichen Grundlagen zu den Prinzipien statistischer Evidenz wurden zuvor in [San10c] und anteilig in [San08] und [San09c] veröffentlicht.

Die konkrete Vorgehensweise zur Festlegung des lokalen Kontexts  $U$  leitet sich letztlich aus dem Law of Likelihood und der in Form von Gleichung (6.1) und Gleichung (6.2) formulierten Schranke für die Wahrscheinlichkeit irreführender statistischer Evidenz ab. Wie bereits in Abschnitt 3.4.3 dargestellt wurde, eignet sich die standardisierte Likelihood-Funktion  $l_{\text{st}}(d|z)$ , welche resultiert, wenn man die Likelihood-Funktion  $l(d|z)$  auf einen Maximalwert von

Eins skaliert, recht gut, um einen Überblick über die von  $d$  gelieferte statistische Evidenz zu erhalten. Aus diesem Grund wird im Folgenden stets gleich konkret auf diese Bezug genommen.

Wählt man den lokalen Kontext  $U$  wie folgt durch Anlegen einer geeigneten Schranke  $1/k$  an die standardisierte Likelihood-Funktion,

$$U := \left\{ z \in Z \mid l_{\text{st}}(d|z) > \frac{1}{k} \right\}, \quad (6.5)$$

so enthält der lokale Kontext  $U$  gerade die möglichen Werte für  $z$ , welche am konsistentesten mit der in Form von  $d$  vorliegenden Information sind: Der lokale Kontext  $U$  enthält dann gerade die möglichen Werte aus  $Z$ , gegenüber welchen der von  $d$  am besten gestützte mögliche Wert aus  $Z$ , d. h. der Maximum-Likelihood-Schätzwert, um einen Faktor, welcher geringer als  $k$  ist, besser gestützt wird. Für jeden nicht im lokalen Kontext  $U$  enthaltenen möglichen Wert aus  $Z$  gibt es (mindestens) einen Wert in  $U$ , der von  $d$  um einen Faktor, welcher mindestens  $k$  ist, besser gestützt wird.

Legt man den lokalen Kontext  $U$  wie in Gleichung (6.5) fest, so fungiert die in Gleichung (6.1) und Gleichung (6.2) formulierte Schranke für die Wahrscheinlichkeit irreführender statistischer Evidenz in bestimmtem Sinne als probabilistische Fehlerschranke für die Korrektheit (im Sinne von Abschnitt 6.1) lokaler Bayes'scher Fusion. Die Wahrscheinlichkeit (hier zu verstehen im Sinne der Degree-of-Belief-Interpretation) dafür, dass man auf Basis dieser Festlegung ein lokales Bayes'sches Modell konstruiert, bei dem der „wahre“ Wert der (durch  $z$  verkörperten) Größen des Interesses nicht im lokalen Kontext  $U$  enthalten ist, lässt sich durch  $1/k$  abschätzen. Wie aus Gleichung (6.2) erkennbar, wird diese Wahrscheinlichkeit letztlich von einem pre-experimentellen Standpunkt aus formuliert. Konkret quantifiziert wird der Degree-of-Belief dafür, dass die in  $d$  zusammengefassten Informationsbeiträge sich so ausgestalten, dass das (dann auf Basis der konkret vorliegenden Informationsbeiträge umgesetzte) Vorgehen zur Festlegung des lokalen Kontexts  $U$  nicht stimmig im Hinblick auf die konkret in Form von  $d$  vorliegende Information ist.

Die statistische Evidenz, welche von einem in  $d$  enthaltenen Informationsbeitrag  $d_s$ ,  $s \in \{1, \dots, S\}$ , alleine geliefert wird, lässt sich in Form der (individu-



ellen) standardisierten Likelihood-Funktion  $l_{st}(d_s|z)$  fassen. Auf dieser Basis ist eine individuelle Auswertung der von den einzelnen Informationsquellen erfassten Informationsbeiträge im Hinblick auf die von ihnen jeweils gelieferte statistische Evidenz möglich. Dieses Vorgehen ist besonders dann interessant, wenn die Informationsbeiträge als statistisch unabhängig gegeben  $z$  angenommen werden können und wie in Abschnitt 3.5.2 beschrieben individuell in die Fusion eingebracht werden. Wie ebenfalls in Abschnitt 3.5.2 begründet lässt sich diese Annahme oftmals beim Vorliegenden heterogener Informationsquellen rechtfertigen. Es liegt auf der Hand, dass in diesem Fall die Möglichkeit zur individuellen Vorauswertung der Informationsbeiträge im Hinblick auf die Festlegung des lokalen Kontexts  $U$  besonders günstig sein kann, da hierfür (genauer: zur Formulierung der entsprechenden Likelihood-Funktionen) in der Regel unterschiedliche Arten von Expertise nötig sind.

An dieser Stelle wird auch die in Abschnitt 4.3 skizzierte Analogie zwischen lokaler Bayes'scher Fusion und kriminalistischen Ermittlungen zur Aufklärung eines Verbrechens wieder besonders deutlich. Ein forensischer Experte, der bestimmte Daten in Bezug auf eine mögliche Straftat auswerten soll, muss, wie auch in Abschnitt 3.4.3 dargestellt, ein allein auf diesen Daten beruhendes Ergebnis liefern. Das von ihm gelieferte Ergebnis muss dabei jedoch so beschaffen sein, dass es durch die ermittelten Beamten bzw. später final durch ein Gericht mit Ergebnissen, welche aus der Auswertung weiterer Daten resultieren, und dem verfügbaren Vorwissen kombiniert werden kann.

Für die lokale Bayes'sche Fusion macht es in dieser Situation Sinn, den lokalen Kontext  $U$  so festzulegen, dass er die möglichen Werte aus  $Z$  enthält, welche am konsistentesten mit den (nun individuell betrachteten) Informationsbeiträgen sind. Eine entsprechende Festlegung kann durch Anlegen einer geeigneten Schranke  $1/k$  an die (individuellen) standardisierten Likelihood-Funktionen gemacht werden, d. h.

$$U := \left\{ z \in Z \mid l_{st}(d_s|z) > \frac{1}{k} \text{ für mindestens ein } s \in \{1, \dots, S\} \right\}. \quad (6.6)$$

Auch in diesem Fall ergibt sich aus Gleichung (6.1) und Gleichung (6.2), dass die Wahrscheinlichkeit dafür, dass man auf Basis dieser Festlegung ein loka-

les Bayes'sches Modell konstruiert, bei dem der „wahre“ Wert der (durch  $z$  verkörperten) Größen des Interesses nicht im lokalen Kontext  $U$  enthalten ist, durch  $\epsilon$  beschränkt ist. Im Falle der bedingten Unabhängigkeit der Informationsbeiträge gegeben  $z$  verschärft sich diese Schranke zu  $1/k^S$ . Man beachte dabei, dass in der Situation von Gleichung (6.6) das Komplement des lokalen Kontexts  $U$  gerade wie folgt festgelegt ist:

$$\bar{U} = \left\{ z \in Z \mid l_{st}(d_s|z) \leq \frac{1}{k} \text{ für alle } s \in \{1, \dots, S\} \right\}. \quad (6.7)$$

Hinsichtlich des in Abschnitt 6.2.1 angegebenen Schemas zur verbalen Interpretation der Stärke statistischer Evidenz scheint es auf Basis der im Rahmen der vorliegenden Arbeit hierzu angestellten Überlegungen sinnvoll, dieses im Fall von Gleichung (6.6) und der bedingten Unabhängigkeit der Informationsbeiträge auf  $1/k^S$  anzuwenden. Dabei ist anzumerken, dass notwendige Anpassungen an dieser Festlegung, insbesondere auch eventuelle (wohl geringfügige) Anpassungen am zugrundeliegenden Schema selbst, für reale Anwendungen der Informationsfusion nicht ausgeschlossen werden können.

Soll die Bestimmung des lokalen Kontexts  $U$  möglichst genau an den verfügbaren Ressourcen festgemacht werden, so ist es vorstellbar, dass man in der Praxis die gemäß Gleichung (6.5) bzw. Gleichung (6.6) anzulegenden Schranken in Abhängigkeit vom resultierenden Umfang des lokalen Kontexts  $U$  festlegt. Die verbale Interpretation der Stärke statistischer Evidenz und die probabilistische Fehlerschranke, auf Basis deren man die Planung ansonsten machen würde, können dabei dann herangezogen werden, um die sich so ergebende Festlegung des lokalen Kontexts  $U$  im Hinblick auf die Durchführung der eigentlichen lokalen Bayes'schen Fusion grundsätzlich zu beurteilen.

Falls der lokale Kontext  $U$  einen bestimmten Umfang nicht überschreiten soll und/oder der Aufwand zur Bestimmung der zur Auswertung von Gleichung (6.5) bzw. Gleichung (6.6) benötigten standardisierten Likelihood-Funktionen sowie der Aufwand zur Auswertung derselben an sich mitberücksichtigt werden sollen, so bietet es sich an, diese Vorauswertung auf geeignete Approximationen der standardisierten Likelihood-Funktionen anzuwenden. Es könnte sich weiter auch anbieten, die entsprechende Vorauswertung der

in  $d$  zusammengefassten Informationsbeiträge auf eine Teilmenge derselben zu konzentrieren und dafür aber eine niedrigere Schranke an die betrachteten standardisierten Likelihood-Funktionen anzulegen.

Gemäß Gleichung (6.6) wird bei individueller Betrachtung der Informationsbeiträge ein möglicher Wert aus  $Z$  gerade dann ignoriert, wenn bezüglich aller herangezogenen Informationsbeiträge der entsprechende Wert der standardisierten Likelihood-Funktionen kleiner gleich  $1/k$  ist (vgl. Gleichung (6.7)). Es ist zu erwarten, dass für einen festen Wert von  $1/k$  die Auswertung von nur  $T < S$  Informationsbeiträgen zu einem geringeren Umfang des resultierenden lokalen Kontexts  $U$  führen wird. In diesem Fall besteht also die Möglichkeit, die an die standardisierten Likelihood-Funktionen angelegte Schranke von  $1/k$  geeignet auf  $1/m$  mit  $m > k$  zu reduzieren, ohne dass der zur Durchführung der Fusion anfallende Aufwand (bezogen auf den Umfang des lokalen Kontexts  $U$ ) zunimmt. Die niedrigere angelegte Schranke  $1/m$  korrespondiert wiederum mit der Planung auf Basis einer geringeren Fehlerschranke der Größe  $1/m$  bzw.  $1/m^T$  im Falle der bedingten Unabhängigkeit der Informationsbeiträge.

Durch dieses Vorgehen lassen sich die Bestimmung und Auswertung der nicht betrachteten  $S-T$  standardisierten Likelihood-Funktionen für die Vorauswertung vermeiden. Erfolgt die lokale Bayes'sche Fusion wie in Abschnitt 5.2.3 beschrieben eingeschränkt auf den lokalen Kontext  $U$ , so müssen die für die Vorauswertung nicht genutzten, aber zur Durchführung der Fusion benötigten Likelihood-Funktionen dann nur für die in  $U$  enthaltenen möglichen Werte aus  $Z$  bestimmt und für die Fusion bereitgestellt werden<sup>1</sup>.

Wird angestrebt, einen lokalen Kontext  $U$  eines bestimmten Umfangs zu erhalten, so lässt sich die Frage, ob dafür besser nur  $T < S$  Informationsbeiträge

<sup>1</sup> Es sei angemerkt, dass anhand von dem in Abschnitt 5.3 dargestellten Beispiel deutlich wurde, dass es in der Praxis technisch, d. h. zur besseren Handhabung bzw. Auswertung der resultierenden Anteile von  $Z$  sinnvoll sein kann, den lokalen Kontext  $U$  auszuweiten. In diesem Fall müssen die nur zur Durchführung der Fusion benötigten Likelihood-Funktionen dann (sofern die Ausweitung auch die Verarbeitung der mit ihnen korrespondierenden Informationsbeiträge betrifft) für die mit der Ausweitung korrespondierenden Anteile von  $Z$  bestimmt und bereitgestellt werden.

oder alle  $S$  Informationsbeiträge vorausgewertet werden sollen, nicht pauschal beantworten. Resultiert z. B. im Falle der bedingten Unabhängigkeit der Informationsbeiträge als anzulegende Schranke  $1/m$ , wenn nur  $T < S$  Informationsbeiträge vorausgewertet werden, und  $1/k$ , wenn alle  $S$  Informationsbeiträge vorausgewertet werden, so hängt es von den sich bei einer konkreten Fusionsaufgabe für  $m, k, T$  und  $S$  ergebenden Werten ab, ob  $1/m^T < 1/k^S$  gilt oder nicht. Es sei angemerkt, dass neben dieser rein formal-mathematischen Betrachtung hier natürlich auch (inhaltliche) Aspekte im Hinblick auf die konkret vorliegende Aufgabenstellung berücksichtigt werden sollten. So sollte man in der Regel auch berücksichtigen, in welchem Zusammenhang die Informationsquellen bzw. die von ihnen gelieferten Beiträge zueinander stehen (vgl. Abschnitt 2.4.1), bzw. auch darüber hinaus, was hinsichtlich der jeweiligen Stärken und Schwächen der Informationsquellen bekannt ist.

Es liegt auf der Hand, dass zur Festlegung des lokalen Kontexts  $U$  neben einer Auswertung der standardisierten Likelihood-Funktion(en) auch die Auswertung des durch  $p(z)$  verkörperten Vorwissens in Betracht gezogen werden sollte, um dafür zu sorgen, dass  $P(U|d)$  möglichst groß ausfällt. Wie in [San08] aufgezeigt lässt sich dies technisch bewerkstelligen, indem man das Vorwissen als zusätzlich vorliegenden Informationsbeitrag  $d_0$  ummodelliert. Mit dieser expliziten Notation für das Vorwissen ergibt sich gemäß dem Satz von Bayes (vgl. Gleichung (3.22)) die Proportionalitätsbeziehung

$$p(z|d, d_0) \propto l(d, d_0|z)p(z|d_0) . \quad (6.8)$$

Da man normalerweise davon ausgeht, dass das Vorwissen  $d_0$  und die in  $d$  zusammengefassten Informationsbeiträge gegeben  $z$  bedingt unabhängig voneinander sind, lässt sich Gleichung (6.8) umformen zu

$$p(z|d, d_0) \propto l(d|z)p(z|d_0) \quad (6.9)$$

$$\propto l(d|z)l(d_0|z)p_{\text{NI}}(z) . \quad (6.10)$$

Dabei bezeichnet  $l(d_0|z)$  die Likelihood-Funktion, welche mit dem als Informationsbeitrag  $d_0$  ummodellierten Vorwissen korrespondiert, und  $p_{\text{NI}}(z)$  bezeichnet eine geeignete nichtinformativ A-Priori-Verteilung auf  $Z$ .

### 6.2.3 Anwendungsbeispiel

Das Ziel des vorliegenden Abschnitts besteht darin, die in Abschnitt 6.2.2 vorgestellten Ergebnisse im Hinblick auf die Festlegung des lokalen Kontexts  $U$  auf Basis der Prinzipien statistischer Evidenz anhand eines überschaubaren und möglichst gut nachvollziehbaren Beispiels zu verdeutlichen. Adressiert wird dabei der Fall, dass die vorliegenden Informationsbeiträge zu Bestimmung des lokalen Kontexts  $U$  individuell ausgewertet und auch individuell in die Fusion eingebracht werden, da dieses Vorgehen in der Praxis gerade beim Vorliegen heterogener Informationsquellen oftmals besonders interessant sein wird. Als konkrete Anwendung wurde die naive Bayes'sche Klassifikation beim Ansetzen einer 0-1-Verlustfunktion gewählt. Die Informationsbeiträge  $d_s \in D_s$  korrespondieren dabei mit den zur Klassifikation vorliegenden Attributen und  $z \in Z$  korrespondiert mit der zugehörigen Klasse. Die nachfolgend dargestellte Untersuchung und die aus ihr resultierenden Ergebnisse wurden zuvor in ähnlicher Form in [San10c] veröffentlicht.

Als Datengrundlage wurden zwei Datensätze aus dem UCI (University of California at Irvine) Machine Learning Repository [Asu07, Dua19] ausgewählt und zwar konkret der Pendigits Datensatz ( $S = 16$  Attribute,  $|Z| = 10$  Klassen) und der Letter Recognition Datensatz ( $S = 16$  Attribute,  $|Z| = 26$  Klassen). In Anbetracht der letztlich doch recht überschaubaren Anzahl an Klassen in den beiden Datensätzen ist die Anwendung der lokalen Bayes'schen Fusion auf diese vor dem Hintergrund der Zielsetzung der vorliegenden Arbeit natürlich vor allem exemplarisch zu sehen.

Die A-Posteriori-Wahrscheinlichkeit  $p(z|d)$  dafür, dass ein bestimmtes, in Form der in  $d = (d_1, \dots, d_S)$  zusammengefassten Attributwerte vorliegendes Datum zu einer bestimmten Klasse  $z \in Z$  gehört, wird bei der naiven Bayes'schen Klassifikation unter Annahme der bedingten Unabhängigkeit von  $d_1, \dots, d_S$  gegeben  $z$  berechnet (vgl. Abschnitt 3.5.2). Das Ansetzen einer 0-1-Verlustfunktion  $v(e, z)$  als Bewertungskriterium im Hinblick auf die Bestimmung eines optimalen Schätzwerts  $\hat{z}$  für  $z$ , d. h.

$$v(e,z) = \begin{cases} 1 & \text{falls } e \neq z \\ 0 & \text{falls } e = z \end{cases}, \quad (6.11)$$

liefert für  $\hat{z}$  gerade den Maximum-A-Posteriori-Schätzwert. Es gilt also, dass  $\hat{z} = \arg \max_{z \in Z} p(z|d)$  ist. Vgl. Abschnitt 3.5.3 sowie [Bey08].

Konkret umgesetzt für die Untersuchung wurde die nachfolgende, in [Dom97] zu seiner Evaluation herangezogene, recht einfache Variante des naiven Bayes'schen Klassifikators: Die numerischen Attributwerte der Datensätze wurden in 10 Intervalle von jeweils gleicher Länge aufgeteilt. Nullen in den probabilistischen Repräsentationen, welche auftreten können, wenn in den Trainingsdaten zu einer Klasse ein bestimmter Attributwert niemals auftritt, wurden mittels Laplace-Korrektur vermieden. Die resultierenden Ergebnisse hinsichtlich der Korrektheit des naiven Bayes'schen Klassifikators wurden jeweils über 20 Testdurchläufe gemittelt berechnet. Für jeden Testdurchlauf wurden die Daten zufällig in Trainings- und Testdaten aufgeteilt, wobei die Trainingsdaten jeweils 2/3 der insgesamt vorliegenden Daten ausmachen.

Die Vorab-Evaluation der beiden verwendeten Datensätze im Hinblick auf die Frage, ob sich eine Anwendung des naiven Bayes'schen Klassifikators grundsätzlich rechtfertigen lässt, ergab, dass dieser (über die 20 Testdurchläufe) eine mittlere Korrektheit von circa 87,81 Prozent mit einer Standardabweichung von circa 0,58 Prozent auf dem Pendigits Datensatz und eine mittlere Korrektheit von circa 70,77 Prozent mit einer Standardabweichung von circa 0,57 Prozent auf dem Letter Recognition Datensatz erreicht, so dass seine exemplarische Anwendung im Rahmen des vorliegenden Beispiels nicht völlig abwegig ist.

Im Rahmen der angestellten Untersuchung wurde auf Basis der aus dem Training resultierenden Likelihood-Funktionen evaluiert, wie genau die Korrektheit des naiven Bayes'schen Klassifikators abnimmt, wenn man vor Durchführung der Klassifikation mögliche Klassen aus  $Z$  gemäß bestimmter Regeln schrittweise entfernt.

Bei Regel 1 wird eine Klasse  $z \in Z$  entfernt, wenn für alle  $s \in \{1, \dots, S\}$  gilt  $l_{st}(d_s|z) \leq \epsilon$ . Komplementär gesprochen wird eine Klasse  $z \in Z$  also ge-

rade dann nicht entfernt, wenn für mindestens ein  $s \in \{1, \dots, S\}$  gilt, dass  $l_{\text{st}}(d_s|z) > \epsilon$  ist. Diese Regel entspricht gerade der in Gleichung (6.6) angegebenen Vorgehensweise zur Bildung eines lokalen Kontexts (wenn man  $\epsilon = 1/k$  setzt).

Bei Regel 2 wird eine Klasse  $z \in Z$  entfernt, wenn für alle  $s \in \{1, \dots, S\}$  gilt  $p(d_s|z) \leq \epsilon$ . Komplementär gesprochen wird eine Klasse  $z \in Z$  also gerade dann nicht entfernt, wenn für mindestens ein  $s \in \{1, \dots, S\}$  die Beziehung  $p(d_s|z) > \epsilon$  gilt.

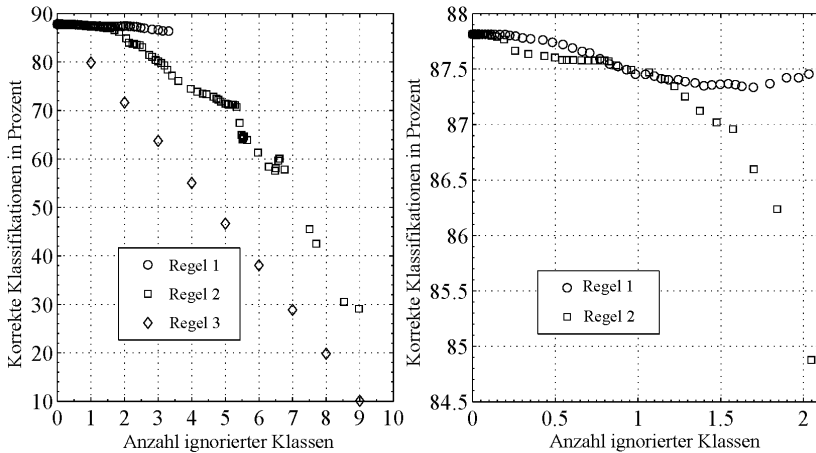
Regel 1 und Regel 2 unterscheiden sich dadurch, dass bei Regel 2 die absoluten Werte von  $p(d_s|z)$  für die Entscheidung herangezogen werden, während bei Regel 1 hierfür die Werte von  $p(d_s|z)$  jeweils relativ zum von  $d$  am besten gestützten Wert betrachtet werden. Es sei angemerkt, dass auf die mit der Betrachtung der absoluten Werte von  $p(d_s|z)$  verbundenen Problematiken, welche sich letztlich daraus ergeben, dass der Wert von  $p(d_s|z)$  recht gering für alle  $z \in Z$  ausfallen kann, bereits in Abschnitt 3.4.3 hingewiesen wurde.

Bei beiden Regeln wurde die angewandte Schranke  $\epsilon$  in kleinen, äquidistanten Schritten erhöht, konkret wurde für jedes  $\epsilon \in \{i/100 \mid i \in \{0,1,2, \dots, 99\}\}$  eine Auswertung vorgenommen.

Bei Regel 3 werden die zu entfernenden Klassen  $z \in Z$  zufällig, d. h. auf Basis einer Gleichverteilung über der Menge der im jeweiligen Schritt verbliebenen Klassen, ausgewählt.

Die Ergebnisse der Untersuchung für den Pendigits Datensatz sind in Abb. 6.1 dargestellt. Bezüglich Regel 1 und Regel 2 setzt jeder der (runden bzw. quadratischen) Marker in der Grafik für einen bestimmten Wert der Schranke  $\epsilon \in \{i/100 \mid i \in \{0,1,2, \dots, 99\}\}$  die Anzahl der durchschnittlich entfernten Klassen und den durchschnittlichen Prozentsatz korrekter Klassifikationen zueinander in Bezug. Wie aus Abb. 6.1 erkennbar ist, führt das Entfernen von möglichen Klassen aus  $Z$  gemäß Regel 2 immerhin noch zu einer deutlich höheren verbleibenden durchschnittlichen Korrektheit, als wenn die Klassen gemäß Regel 3 völlig zufällig aus  $Z$  entfernt werden. Jedoch liegt die verbleibende Korrektheit des naiven Bayes'schen Klassifikators auch bei Regel 2 deutlich unter der verbleibenden Korrektheit, welche resultiert, wenn man die Klassen

gemäß Regel 1, welche konsistent mit den Prinzipien statistischer Evidenz und den aus diesen im Rahmen der vorliegenden Arbeit abgeleiteten Vorgehensweisen zur Festlegung eines lokalen Kontexts ist, entfernt.



**Abbildung 6.1:** Veränderung des Anteils korrekter Klassifikationen beim Pendigits Datensatz, wenn mögliche Klassen gemäß Regel 1, Regel 2 und Regel 3 entfernt werden. Die angegebenen Werte sind wie im Text beschrieben gemittelt über 20 Testdurchläufe. Das rechte Diagramm stellt einen vergrößerten Ausschnitt von Teilen des linken Diagramms dar.

Auch in den Fällen, in denen  $\epsilon$  schon bereits recht nahe bei Eins liegt, lässt Regel 1 nicht zu, dass ein extrem umfangreicher Anteil möglicher Klassen aus  $Z$  entfernt wird. Konkret werden für  $\epsilon = 0,99$  im Schnitt nur circa 33,13 Prozent der möglichen Klassen aus  $Z$  entfernt, wodurch sich die durchschnittliche Korrektheit (von ursprünglich circa 87,81 Prozent) auf circa 86,40 Prozent (mit einer Standardabweichung von circa 0,64 Prozent) reduziert. Dieses Verhalten lässt sich damit erklären, dass mit  $S = 16$  eine recht hohe Anzahl an (standardisierten) Likelihood-Funktionen vorausgewertet wurde.



Für  $\epsilon = 0,81 \approx \sqrt[16]{1/32}$  werden im Schnitt circa 14,96 Prozent der möglichen Klassen aus  $Z$  entfernt, wodurch sich die durchschnittliche Korrektheit auf circa 87,36 Prozent (mit einer Standardabweichung von circa 0,54 Prozent) reduziert. Für  $\epsilon = 0,88 \approx \sqrt[16]{1/8}$  werden im Schnitt circa 19,70 Prozent der möglichen Klassen aus  $Z$  entfernt, wodurch sich die durchschnittliche Korrektheit auf circa 87,42 Prozent (mit einer Standardabweichung von circa 0,59 Prozent) reduziert. Diese konkrete Auswahl der betrachteten Werte für  $\epsilon$  korrespondiert mit dem in Abschnitt 6.2.2, genauer nach Gleichung (6.7), vorgeschlagenen (auf Basis der im Rahmen der Arbeit angestellten Überlegungen sinnvoll erscheinenden) Vorgehen. Durch die Betrachtung gerade dieser Werte im Rahmen des vorliegenden Beispiels soll aber nicht der Anspruch erhoben werden, die Sinnhaftigkeit exakt dieser Festlegung anhand des Beispiels zu verifizieren. Diese beiden Aussage gelten gleichermaßen für weitere, nachfolgend betrachtete Werte der Form  $\sqrt[16]{1/k}$ , in der  $T$  die Anzahl der ausgewerteten Quellen notiert.

Ergebnisse der Untersuchung für den Letter Recognition Datensatz sind im Hinblick auf die für die Untersuchung wesentlichen Aspekte vergleichbar mit denen für den Pendigits Datensatz. Konkret werden für  $\epsilon = 0,99$  im Schnitt hier circa 55,34 Prozent der möglichen Klassen aus  $Z$  entfernt, wodurch sich die durchschnittliche Korrektheit (von ursprünglich circa 70,77 Prozent) auf circa 64,77 Prozent (mit einer Standardabweichung von circa 0,99 Prozent) reduziert. Für  $\epsilon = 0,81 \approx \sqrt[16]{1/32}$  werden im Schnitt circa 5,16 Prozent der möglichen Klassen aus  $Z$  entfernt, wodurch sich die durchschnittliche Korrektheit auf circa 70,71 Prozent (mit einer Standardabweichung von circa 0,59 Prozent) reduziert. Für  $\epsilon = 0,88 \approx \sqrt[16]{1/8}$  werden im Schnitt circa 19,35 Prozent der möglichen Klassen aus  $Z$  entfernt, wodurch sich die durchschnittliche Korrektheit auf circa 69,89 Prozent (mit einer Standardabweichung von circa 0,66 Prozent) reduziert.

Um zu demonstrieren, dass sich die Prinzipien statistischer Evidenz auch sinnvoll anwenden lassen, wenn zur Bestimmung des lokalen Kontexts eine andere (sinnvolle) probabilistische Repräsentation der vorliegenden Information verwendet wird als zur anschließenden Durchführung der eigentlichen lokalen Bayes'schen Fusion, und dass die Vorauswertung nur eines Teils ver-

fügbarer Information ein gangbarer Weg sein kann, wurde für den Pendigits Datensatz exemplarisch noch eine zweite Untersuchung angestellt. Bei diesem Datensatz korrespondieren die Attribute mit  $(x,y)$ -Koordinaten handgeschriebenen Zahlen (genauer: den Ziffern 0 bis 9). Genauer korrespondiert  $d_s$  mit einer  $x$ -Koordinate, wenn  $s$  ungerade ist, und mit einer  $y$ -Koordinate, wenn  $s$  gerade ist ( $s \in \{1, \dots, 16\}$ ).

Für die Untersuchung wurde auf Basis der Attribute  $d_1$ ,  $d_5$  und  $d_7$  exemplarisch ein neues Attribut  $d_x$ , welches die Veränderungen der entsprechenden  $x$ -Koordinaten im Datensatz angibt, wie folgt bestimmt:

$$d_x := \begin{cases} 0, & d_1 < d_5 < d_7 \\ 1, & d_1 < d_5 \text{ und } d_7 \leq d_5 \\ 2, & d_5 \leq d_1 \text{ und } d_5 < d_7 \\ 3, & d_7 \leq d_5 \leq d_1 \end{cases} . \quad (6.12)$$

Dieses Attribut enthält also einen Teil der in Form von  $d_1$ ,  $d_5$  und  $d_7$  vorliegenden Information, genauer hinsichtlich ihres Zusammenhangs. In analoger Weise wurde auf Basis der Attribute  $d_2$ ,  $d_6$  und  $d_8$  ein weiteres neues Attribut  $d_y$  bestimmt. Weiter wurden die nachfolgenden beiden zusätzlichen Regeln, welche konzeptionell gesehen konsistent mit den Prinzipien statistischer Evidenz und dem im Rahmen der vorliegenden Arbeit daraus abgeleiteten Vorgehen für die lokale Bayes'sche Fusion sind, formuliert:

Bei Regel 4 wird eine Klasse  $z \in Z$  ignoriert, wenn gilt  $l_{\text{st}}(d_x|z) \leq \epsilon$ . Komplementär gesprochen wird eine Klasse  $z \in Z$  also gerade dann nicht ignoriert, wenn gilt, dass  $l_{\text{st}}(d_x|z) > \epsilon$  ist.

Bei Regel 5 wird eine Klasse  $z \in Z$  ignoriert, wenn  $l_{\text{st}}(d_x|z) \leq \epsilon$  und  $l_{\text{st}}(d_y|z) \leq \epsilon$  gelten. Komplementär gesprochen wird eine Klasse  $z \in Z$  also gerade dann nicht ignoriert, wenn  $l_{\text{st}}(d_x|z) > \epsilon$  oder  $l_{\text{st}}(d_y|z) > \epsilon$  ist.

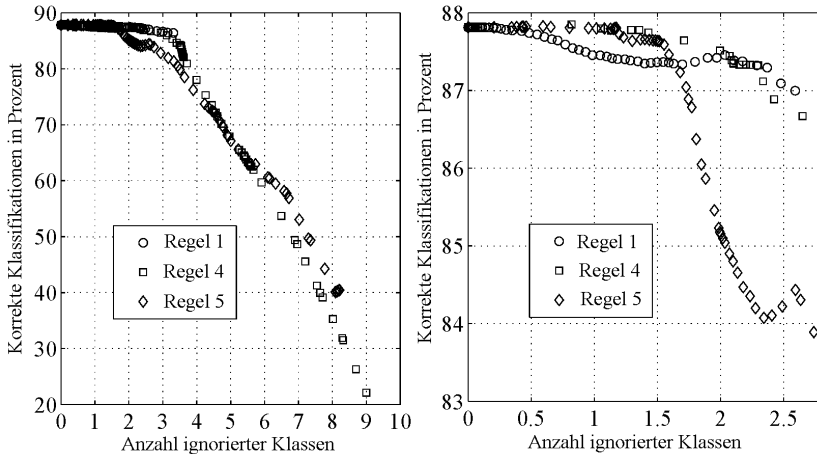
Wie bei Regel 1 und Regel 2 wurde auch bei Regel 4 und Regel 5 die jeweils angewandte Schranke  $\epsilon$  in kleinen, äquidistanten Schritten erhöht, konkret wurde wieder für jedes  $\epsilon \in \{i/100 \mid i \in \{0,1,2, \dots, 99\}\}$  eine Auswertung vorgenommen. Zur Bildung der für die Vorauswertung gemäß Regel 4 und Regel 5 benötigten standardisierten Likelihood-Funktionen wurden  $p(d_x|z)$  und

$p(d_y|z)$  in den Trainingsdurchläufen jeweils mitbestimmt und dann entsprechend auf einen Maximalwert von Eins skaliert. Die eigentliche (lokale) naive Bayes'sche Klassifikation wurde nach der Anwendung von Regel 4 bzw. Regel 5 jeweils unverändert auf Basis der ursprünglichen probabilistischen Repräsentationen der vorliegenden Information, d. h. auf Basis von  $p(d_s|z)$ ,  $s \in \{1, \dots, S\}$ , durchgeführt.

In Abb. 6.2 sind die Ergebnisse, welche sich durch Anwendung von Regel 1, Regel 4 und Regel 5 ergeben, einander gegenübergestellt. Man erkennt, dass die Anwendung von Regel 4 und Regel 5 grundsätzlich Sinn ergibt. Sie verhindern zwar nicht, dass bei größeren, hier auch kaum realistischen Werten von  $\epsilon$  im Schnitt eine recht hohe Anzahl an möglichen Klassen aus  $Z$  ignoriert wird, was notwendigerweise dazu führt, dass die durchschnittliche Korrektheit des naiven Bayes'schen Klassifikators auch abnimmt. Solange dies noch nicht der Fall ist (d. h. für kleinere Werte von  $\epsilon$ ), liefern Regel 4 und Regel 5 jedoch passable Ergebnisse.

Bei Regel 4 werden für  $\epsilon = 0,04 \approx 1/32$  im Schnitt circa 12,91 Prozent der möglichen Klassen aus  $Z$  entfernt, wodurch sich die durchschnittliche Korrektheit (von ursprünglich circa 87,81 Prozent) auf circa 87,78 Prozent (mit einer Standardabweichung von circa 0,58 Prozent) reduziert. Weiter werden für  $\epsilon = 0,13 \approx 1/8$  im Schnitt circa 20,72 Prozent der möglichen Klassen aus  $Z$  entfernt, wodurch sich die durchschnittliche Korrektheit auf circa 87,45 Prozent (mit einer Standardabweichung von circa 0,62 Prozent) reduziert.

Bei Regel 5 werden für  $\epsilon = 0,18 \approx \sqrt[3]{1/32}$  im Schnitt circa 11,61 Prozent der möglichen Klassen aus  $Z$  entfernt, wodurch sich die durchschnittliche Korrektheit (von ursprünglich circa 87,81 Prozent) auf circa 87,80 Prozent (mit einer Standardabweichung von circa 0,58 Prozent) reduziert. Weiter werden für  $\epsilon = 0,36 \approx \sqrt[3]{1/8}$  im Schnitt circa 17,71 Prozent der möglichen Klassen aus  $Z$  entfernt, wodurch sich die durchschnittliche Korrektheit auf circa 86,78 Prozent (mit einer Standardabweichung von circa 0,79 Prozent) reduziert.



**Abbildung 6.2:** Veränderung des Anteils korrekter Klassifikationen beim Pendigits Datensatz, wenn mögliche Klassen gemäß Regel 1, Regel 4 und Regel 5 entfernt werden. Die angegebenen Werte sind wie im Text beschrieben gemittelt über 20 Testdurchläufe. Die rechte Grafik stellt einen vergrößerten Ausschnitt eines Teils der linken Grafik dar.

In Tabelle 6.1 sind die genannten Werte im Hinblick auf Regel 1, Regel 4 und Regel 5 noch einmal einander gegenübergestellt. Es bezeichnet dabei  $T$  die Anzahl der ausgewerteten (zuvor im Text jeweils genauer genannten) Likelihood-Funktionen. Weiter bezeichnen  $\nu$  den durchschnittlichen Anteil der für den angegebenen Wert der angelegten Schranke  $\epsilon$  entfernten Klassen,  $\mu$  die durchschnittliche verbleibende Korrektheit des naiven Bayes'schen Klassifikators und  $\sigma$  die korrespondierende Standardabweichung; diese Angaben sind jeweils in Prozent und auf zwei Nachkommastellen gerundet. Die näherungsweise Angabe von  $\epsilon$  ist so zu verstehen, dass jeweils die Ergebnisse für den entsprechend exakten oder (falls dieser nicht ausgewertet wurde) nächst größeren Wert aus  $\{i/100 \mid i \in \{0,1,2, \dots, 99\}\}$  herangezogen wurden. Bei Regel 4 und Regel 5 wurde auf die Angabe der für  $\epsilon = 0,99$  erzielten Werte bewusst verzichtet, da eine entsprechende Wahl von  $\epsilon$ , welche bereits schon

für den Fall  $T = 16$  sehr hoch und damit diskutabel ist, in diesen beiden Fällen sicher nicht als sinnvoll zu erachten ist.

**Tabelle 6.1:** Zusammenstellung der zuvor im Text bezüglich Regel 1, Regel 4 und Regel 5 genannten Werte für den Pendigits Datensatz ( $|Z| = 10$ ) und den Letter Recognition Datensatz ( $|Z| = 16$ ).

	Pendigits Regel 4 (T=1)	Pendigits Regel 5 (T=2)	Pendigits Regel 1 (T=16)	Letter Recognition Regel 1 (T=16)
$\epsilon = 0$	$v = 0$ $\mu = 87,81$ $\sigma = 0,58$	$v = 0$ $\mu = 87,81$ $\sigma = 0,58$	$v = 0$ $\mu = 87,81$ $\sigma = 0,58$	$v = 0$ $\mu = 70,77$ $\sigma = 0,57$
$\epsilon \approx \sqrt[3]{1/32}$	$v = 12,91$ $\mu = 87,78$ $\sigma = 0,58$	$v = 11,61$ $\mu = 87,80$ $\sigma = 0,58$	$v = 14,96$ $\mu = 87,36$ $\sigma = 0,54$	$v = 5,16$ $\mu = 70,71$ $\sigma = 0,56$
$\epsilon \approx \sqrt[3]{1/8}$	$v = 20,72$ $\mu = 87,45$ $\sigma = 0,62$	$v = 17,71$ $\mu = 86,78$ $\sigma = 0,79$	$v = 19,70$ $\mu = 87,42$ $\sigma = 0,59$	$v = 19,35$ $\mu = 69,89$ $\sigma = 0,66$
$\epsilon = 0,99$	N/A	N/A	$v = 33,31$ $\mu = 86,40$ $\sigma = 0,64$	$v = 55,34$ $\mu = 64,77$ $\sigma = 0,99$

Die in der Tabelle angegebenen Ergebnisse sollten u. a. aufgrund der doch überschaubaren Anzahl Klassen in den Datensätzen, der vereinfachenden Annahme der bedingten Unabhängigkeit der Attribute gegeben die Klasse und der recht einfachen Umsetzung des naiven Bayes'schen Klassifikators sicher nicht überinterpretiert werden. Sie lassen jedoch gerade auch unter Berücksichtigung der insgesamt auch für Regel 2 und Regel 3 erzielten Ergebnisse (vgl. insbesondere auch Abb. 6.1 und Abb. 6.2) die grundsätzliche Sinnhaftigkeit des Vorgehens auf Basis der Prinzipien statistischer Evidenz erkennen.

### 6.3 Informationstheoretische Betrachtung

Informationstheoretische Prinzipien wurden in Abschnitt 3.5.1 bereits im Hinblick auf die Aufgabe, einen vorliegenden Informationsstand in eine objektive probabilistische Repräsentation im Sinne der Degree-of-Belief-Interpretation von Wahrscheinlichkeit zu überführen, aufgegriffen. Bezüge zwischen der Bayes'schen Theorie und der Informationstheorie werden in der Fachliteratur auch darüber hinausgehend hergestellt und genutzt. Nachfolgend werden weitere, für die vorliegende Arbeit relevante Ansätze und Ergebnisse aus der Fachliteratur dargestellt und in Bezug gesetzt, um letztlich die Frage nach der Korrektheit der lokalen Bayes'schen Fusion aus informationstheoretischer Sicht zu beleuchten.

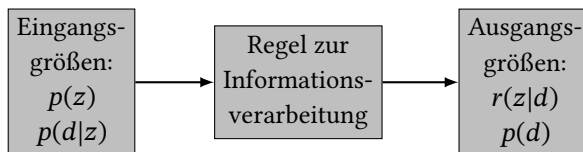
Der Kern der resultierenden Ergebnisse für die lokale Bayes'sche Fusion wurde samt komprimierter Zusammenstellung der erforderlichen Grundlagen zuvor in [San09a] und [San09c] veröffentlicht. Im vorliegenden Abschnitt werden diese Ergebnisse demgegenüber deutlich umfangreicher motiviert und auch deutlich umfangreicher in Bezug auf die sich ergebenden Schlussfolgerungen hinsichtlich der Wahl des lokalen Kontexts  $U \subset Z$  betrachtet. Während in den vorgenannten Veröffentlichungen die nachfolgend in Abschnitt 6.3.2 eingeführte, von Walker formulierte Minimierungsregel und zusätzlich (vollständig) alternativ das in Abschnitt 3.5.1 eingeführte Prinzip der Minimalen Information als Ausgangspunkt für die Untersuchungen gewählt wurden, fungiert in der vorliegenden Arbeit zwecks besserer Anschaulichkeit vor allem das im nachfolgenden Abschnitt 6.3.1 eingeführte Informationsverarbeitungsschema nach Zellner als Basis. Die von Walker formulierte Minimierungsregel verdeutlicht dabei vor allem auch in unabhängiger Weise zusätzlich die Stimmigkeit der Resultate. Das Prinzip der Minimalen Information wird (angewandt über  $Z$ ) in die Argumentation miteinbezogen. Auf eine alternative, vollständige Herleitung der Ergebnisse auf Basis des Prinzips der Minimalen Information (welches dazu über  $Z \times D$  angewandt werden muss) wird im Folgenden verzichtet, vor allem auch um den Umfang der Darstellung nicht ausufern zu lassen.

### 6.3.1 Informationsverarbeitungsschema nach Zellner

Zellner zeigt in [Zel88] die Optimalität Bayes'scher Inferenz von einem informationstheoretischen Standpunkt aus auf. Basis ist dabei das in Abb. 6.3 dargestellte Informationsverarbeitungsschema. Auf der linken Seite in Abb. 6.3 stehen die in Abschnitt 3.4.2 eingeführten Komponenten eines Bayes'schen Modells, d. h. die A-Priori-Verteilung  $p(z)$  und die Likelihood-Funktion  $p(d|z)$ . Zwecks Einheitlichkeit wird letztere hier und in den nachfolgenden Abschnitten (genauer: bis Gleichung (6.75) in Abschnitt 6.3.4.5) als normiert bezüglich  $d$  angenommen und durch  $l(d|z)$  notiert.

Beim Informationsverarbeitungsschema wird angenommen, dass die Eingangsgrößen durch eine Regel zur Informationsverarbeitung prozessiert werden. Dadurch resultieren die auf der rechten Seite von Abb. 6.3 dargestellten Größen. Dabei ergibt sich  $p(d)$  in eindeutiger Weise aus der Verteilungsannahme  $\{p(d|z)|z \in Z\}$  und der A-Priori-Verteilung  $p(z)$  zu  $p(d) = \int_Z p(d|z)p(z) dz$ . Die Ausgangsgröße  $r(z|d)$  stellt eine Wahrscheinlichkeitsverteilung über  $Z$  dar, welche sich abhängig von der angewandten Regel zur Informationsverarbeitung ergibt<sup>1</sup>.

Es sei angemerkt, dass in [Zel88] explizit die Abhängigkeit der A-Priori-Verteilung  $p(z)$  und damit auch der Größen  $r(z|d)$  und  $p(d)$  vom Vorwissen notiert wird.



**Abbildung 6.3:** Informationsverarbeitungsschema nach Zellner. Abbildung angelehnt an [Zel88].

<sup>1</sup> In Abschnitt 6.3.1 und Abschnitt 6.3.2 sollen die Notationen  $r(z|d)$  und  $q(z|d)$  verdeutlichen, dass es sich um aus den Berechnungen resultierende Ausgangsgrößen handelt.

Unter Einnahme eines nach Durchführung der Informationsverarbeitung angesiedelten Standpunkts wird die in den Eingangs- und Ausgangsgrößen (erwartungsgemäß) enthaltene Information in [Zel88] wie folgt quantifiziert:

(Maß für die) Information in  $p(z)$ :

$$E_{r(z|d)}[\log p(z)] = \int_{\mathcal{Z}} r(z|d) \log p(z) dz, \quad (6.13)$$

(Maß für die) Information in  $p(d|z)$ :

$$E_{r(z|d)}[\log p(d|z)] = \int_{\mathcal{Z}} r(z|d) \log p(d|z) dz, \quad (6.14)$$

(Maß für die) Information in  $r(z|d)$ :

$$E_{r(z|d)}[\log r(z|d)] = \int_{\mathcal{Z}} r(z|d) \log r(z|d) dz, \quad (6.15)$$

(Maß für die) Information in  $p(d)$ :

$$E_{r(z|d)}[\log p(d)] = \int_{\mathcal{Z}} r(z|d) \log p(d) dz = \log p(d). \quad (6.16)$$

Die Regel zur Informationsverarbeitung in Abb. 6.3 ist gerade dann optimal, wenn die Wahrscheinlichkeitsverteilung  $r(z|d)$  so beschaffen ist, dass die aus der Anwendung der Regel zur Informationsverarbeitung resultierende Ausgangsinformation so weit als möglich der in die Regel zur Informationsverarbeitung eingehenden Information entspricht. Das Verhältnis zwischen ausgehender und eingehender Information sollte also nahezu Eins sein.

Sofern möglich, d. h. falls eine entsprechende Regel zur Informationsverarbeitung existiert, sollte die eingehende Information vollständig der ausgehenden Information entsprechen. In diesem Fall erfüllt die Regel zur Informationsverarbeitung das in [Zel88] formulierte Prinzip der Informationserhaltung (im Englischen: Information Conservation Principle (ICP)).

Existiert keine derartige optimale Regel zur Informationsverarbeitung, welche das Prinzip der Informationserhaltung erfüllt, so sollte man die bestmögliche Regel wählen. Diese ist dadurch charakterisiert, dass das Verhältnis von ausgehender und eingehender Information möglichst Eins ist. Dabei wird in



[Zel88] verständlicherweise noch gefordert, dass die Regel zur Informationsverarbeitung im Ausgang nicht mehr Information liefern darf als im Eingang vorhanden ist.

Gemäß [Zel88] muss die bestmögliche Regel zur Informationsverarbeitung letztlich so beschaffen sein, dass sie die Wahrscheinlichkeitsverteilung  $r(z|d)$  auf  $Z$  liefert, welche das Funktional

$$\begin{aligned}
 F_1[q(z|d)] &:= \int_Z q(z|d) \log q(z|d) dz + \log p(d) \\
 &\quad - \int_Z q(z|d) \log p(z) dz - \int_Z q(z|d) \log p(d|z) dz \quad (6.17) \\
 &= \underbrace{\int_Z q(z|d) \log q(z|d) dz + \log p(d)}_{\text{(Maß für die) Information in den Ausgangsgrößen}} \\
 &\quad - \underbrace{\int_Z q(z|d) [\log p(z) + \log p(d|z)] dz}_{\text{(Maß für die) Information in den Eingangsgrößen}} \\
 &\hspace{15em} (6.18)
 \end{aligned}$$

minimiert, d. h.

$$r(z|d) = \arg \min_{q(z|d) \in \mathfrak{P}_z} F_1[q(z|d)] \quad (6.19)$$

wobei  $\mathfrak{P}_z$  die Menge aller (wohldefinierten) Wahrscheinlichkeitsverteilungen auf  $Z$  bezeichnet.

In [Zel88] wird durch Minimierung dieses Funktionals gezeigt, dass die bestmögliche Regel zur Informationsverarbeitung im Ausgang (neben  $p(d)$ ) gerade die Wahrscheinlichkeitsverteilung  $r(z|d)$ , welche

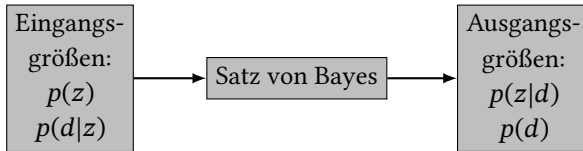
$$r(z|d) \propto p(d|z)p(z) \quad (6.20)$$

erfüllt, liefern muss. Die Wahrscheinlichkeitsverteilung  $r(z|d)$  muss also der

A-Posteriori-Verteilung  $p(z|d)$ , welche bei der Bayes'schen Inferenz resultiert, entsprechen. Die bestmögliche Regel zur Informationsverarbeitung ergibt sich somit gemäß dem Satz von Bayes (vgl. Gleichung (3.6)). Aufgrund der Tatsache, dass

$$F_1[p(z|d)] = 0 \quad (6.21)$$

gilt, ist dabei auch das Prinzip der Informationserhaltung erfüllt, d. h. die Bayes'sche Inferenz gewährleistet in dieser (informationstheoretischen) Hinsicht also eine optimale Informationsverarbeitung. In Abb. 6.4 ist das Informationsverarbeitungsschema aus Abb. 6.3 für diesen Fall dargestellt.



**Abbildung 6.4:** Informationsverarbeitung gemäß dem Satz von Bayes.

Bemerkenswert an der im vorliegenden Abschnitt wiedergegebenen Formulierung Bayes'scher Inferenz als optimale Informationsverarbeitung ist vor allem auch, dass dieser Ansatz über die in [Zel88] aufgezeigte Optimalitätsaussage hinausgehend auch weitere Schlüsse im Hinblick auf die Funktionsweise Bayes'scher Inferenz ermöglicht. Vgl. hierzu insbesondere auch [Zel02]. Dadurch lassen sich dann insbesondere auch Möglichkeiten zur gezielten Abwandlung der Bayes'schen Inferenz ableiten.

Im Kontext der Bayes'schen Fusion könnte vor allem auch ein entsprechend abgeleitetes Vorgehen zur gezielten Steuerung des Einflusses der Eingangsgrößen interessant sein. Will man z. B. die Tatsache berücksichtigen, dass die A-Priori-Verteilung  $p(z)$  nicht in optimaler Qualität bestimmt werden konnte und sie dementsprechend mit vermindertem Einfluss in die Bayes'sche Fusion einbringen, so lässt sich dies gemäß [Zel02] aus informationstheoretisch

optimaler Sicht umsetzen, indem man statt  $p(z)$  gerade  $p^\alpha(z)$  mit  $\alpha \in [0,1]$  als A-Priori-Verteilung in die Bayes'sche Fusion einbringt. Der Exponent  $\alpha$  ist umso kleiner zu wählen, je geringer die Qualität von  $p(z)$  ist. Im Grenzfall  $\alpha = 0$  wird  $p(z)$  überhaupt nicht berücksichtigt, im anderen Grenzfall  $\alpha = 1$  geht  $p(z)$  wie üblich in die Bayes'sche Fusion ein. Eine analoge Aussage gilt im Hinblick auf die Likelihood-Funktion. Ihr Einfluss könnte also gleichermaßen mittels eines (anderen) Exponenten  $\gamma \in [0,1]$  gewichtet werden. Es liegt auf der Hand, dass dieses Vorgehen beim Vorliegen der bedingten Unabhängigkeit der (in  $d$  zusammengefassten) Informationsbeiträge  $d_s$  gegeben  $z$  ( $s \in \{1, \dots, S\}$ ) grundsätzlich auch auf den Fall eines sequentiellen Fusionschemas gemäß Gleichung (3.25) erweitert werden könnte, indem man die individuellen Likelihood-Funktionen mit entsprechend geeignet gewählten Exponenten  $\gamma_s \in [0,1]$ ,  $s \in \{1, \dots, S\}$ , versieht.

### 6.3.2 Minimierungsregel nach Walker

Walker stellt in [Wal06] dar, dass die bei der Bayes'schen Inferenz resultierende A-Posteriori-Verteilung  $p(z|d)$  sich gerade aus der folgenden Beziehung, welche von ihm als Minimierungsregel bezeichnet wird, ergibt:

$$p(z|d) = \arg \min_{q(z|d) \in \mathfrak{P}_z} F_2[q(z|d)] \quad \text{mit} \quad (6.22)$$

$$F_2[q(z|d)] := - \int_Z q(z|d) \log p(d|z) dz + \text{KD}[q(z|d), p(z)]. \quad (6.23)$$

Hierbei bezeichnet  $\text{KD}[q(z|d), p(z)]$  die bereits in Definition 3.6 eingeführte Kullback-Leibler-Divergenz, hier konkret von  $q(z|d)$  relativ zu  $p(z)$ . Weiter bezeichnet  $\mathfrak{P}_z$  in Gleichung (6.22) die Menge aller Wahrscheinlichkeitsverteilungen auf  $Z$  mit  $\text{KD}[q(z|d), p(z)] < \infty$ .

Walker weist in [Wal06] darauf hin, dass sich die Gültigkeit der von ihm aufgestellten Minimierungsregel aus dem in Zellner formulierten Prinzip der Informationserhaltung ergibt. Vor allem aber wird in [Wal06] die Minimierungsregel auch direkt, d. h. ohne Ausnutzung des Bezugs zu den Ergebnissen von Zellner, bewiesen und weitergehend interpretiert. Das Funktional  $F_2[q(z|d)]$

in Gleichung (6.23) besteht demnach aus zwei Termen, von denen der erste die Nähe von  $q(z|d)$  zur (hier bezüglich  $d$  normierten) Likelihood-Funktion forciert, während der zweite Term verhindert, dass  $q(z|d)$  zu stark von der A-Priori-Verteilung  $p(z)$  abweicht.

Vergleicht man Gleichung (6.17) und Gleichung (6.23) genauer, so stellt man sofort fest, dass die von Zellner und Walker formulierten Funktionale in folgendem Zusammenhang stehen:

$$F_2[q(z|d)] = F_1[q(z|d)] - \log p(d) . \quad (6.24)$$

Aus Gleichung (6.21) ergibt sich für den Minimalwert des Funktional  $F_2[q(z|d)]$  gerade

$$F_2[p(z|d)] = -\log p(d) . \quad (6.25)$$

Der Wert von  $F_2[p(z|d)]$  lässt sich damit aus informationstheoretischer Sicht auffassen als Informationszuwachs, der insgesamt resultiert, wenn  $d$  verfügbar wird.

### 6.3.3 Von den Informationsbeiträgen gelieferte Information

Ist  $q(z|d)$  die A-Posteriori-Verteilung  $p(z|d)$ , so wird der zweite Term im von Walker gemäß Gleichung (6.23) festlegten Funktional  $F_2[q(z|d)]$  zu

$$I_{p(z)}[d] := \text{KD}[p(z|d), p(z)] \quad (6.26)$$

$$= \int_{\mathcal{Z}} p(z|d) \log \frac{p(z|d)}{p(z)} dz \quad (6.27)$$

$$= E_{p(z|d)} [\log p(z|d)] - E_{p(z|d)} [\log p(z)] . \quad (6.28)$$

Gemäß dem von Zellner formulierten Informationsverarbeitungsschema liefert  $I_{p(z)}[d]$  für den Fall, dass  $r(z|d)$  identisch zu  $p(z|d)$  ist, gerade die Differenz zwischen dem in Gleichung (6.15) angegebenen Maß für die Information, welche (erwartungsgemäß) in der Ausgangsgröße  $p(z|d)$  enthalten ist, und

dem in Gleichung (6.13) angegebenen Maß für die Information, welche (erwartungsgemäß) in der Eingangsgröße  $p(z)$  enthalten ist. In anderen Worten stellt die Größe  $I_{p(z)}[d]$  ein Maß für die Veränderung des bezüglich  $z$  vorliegenden Informationsstands, welche sich durch die in Form der Anwendung des Satzes von Bayes erfolgende Informationsverarbeitung ergibt, dar.

Die in der vorliegenden Arbeit als  $I_{p(z)}[d]$  notierte Größe wird in der Fachliteratur zur Bayes'schen Theorie auch allgemein, d. h. unabhängig von dem in [Zel88] entwickelten Ansatz, aufgefasst als Maß für den Informationszuwachs, welchen die konkret in Form von  $d$  vorliegenden Informationsbeiträge in Bezug auf  $z \in Z$  liefern, wenn der zuvor im Hinblick auf  $z$  verfügbare Informationsstand (in objektiver Weise) durch die A-Priori-Verteilung  $p(z)$  probabilistisch repräsentiert wird. Vgl. hierzu insbesondere [Ber04]. Diese Interpretation ist auch konsistent zur in Abschnitt 3.5.1 angeführten Interpretation der Kullback-Leibler-Divergenz.

Der Formulierung von  $I_{p(z)}[d]$  liegt ein vollständig post-experimenteller Standpunkt zugrunde. Die Menge an Information, welche  $d$  in Bezug auf  $z \in Z$  liefert, wird im Hinblick auf den konkret vorliegenden Wert von  $d$  bestimmt und die in Gleichung (6.28) vorgenommene Erwartungswertbildung erfolgt jeweils bezüglich der A-Posteriori-Verteilung  $p(z|d)$ .

Geht man einen Schritt zurück und untersucht, welche (Menge an) Information ein Bayes'sches Experiment mit A-Priori-Verteilung  $p(z)$ , in dem  $d$  gemäß  $p(d)$  verteilt ist, liefert, so ergibt sich durch zusätzliche Erwartungsbildung im Hinblick auf  $d$  folgendes Informationsmaß (vgl. [Ber04]):

$$J[z,d] := E_{p(d)}[I_{p(z)}[d]] \quad (6.29)$$

$$= \int_D p(d) \int_Z p(z|d) \log \frac{p(z|d)}{p(z)} dz dd \quad (6.30)$$

$$= \int_D \int_Z p(z,d) \log \frac{p(z,d)}{p(z)p(d)} dz dd \quad (6.31)$$

$$= \text{KD}[p(z,d), p(z)p(d)] . \quad (6.32)$$

Die in der vorliegenden Arbeit als  $J[z,d]$  notierte Größe wird in der Informationstheorie üblicherweise als Transinformation bezeichnet. Wie erkennbar aus Gleichung (6.32) stellt sie ein Maß für den statistischen Zusammenhang zwischen  $z$  und  $d$  dar. Sie gibt in diesem Sinne die (Menge an) Information, welche  $d$  erwartungsgemäß bezüglich  $z$  liefert, als auch die (Menge an) Information, welche  $z$  wiederum erwartungsgemäß über  $d$  enthält, an (vgl. [Bey99]). Diese Interpretation kann man sich auch sehr gut verdeutlichen, wenn man Gleichung (6.29) aufgespalten nach den in  $I_{p(z)}[d]$  enthaltenen Komponenten (gemäß Gleichung (6.28)) wiederum im Lichte des von Zellner formulierten Informationsverarbeitungsschemas betrachtet.

Der Vollständigkeit halber soll kurz darauf eingegangen werden, dass man in der Fachliteratur auch eine alternative, auf [Lin56] zurückgehende Definition für die (Menge an) Information, welche konkret in Form von  $d$  vorliegende Informationsbeiträge in Bezug auf  $z \in Z$  liefern, findet. Diese ist (statt durch  $I_{p(z)}[d]$ ) gegeben durch

$$\tilde{I}_{p(z)}[d] = E_{p(z|d)}[\log p(z|d)] - E_{p(z)}[\log p(z)] . \quad (6.33)$$

Anders als bei der Berechnung von  $I_{p(z)}[d]$  gemäß Gleichung (6.28) wird hier die in  $p(z)$  enthaltene Information also (trotz konkret angenommenem Wert für  $d$ ) nicht a posteriori als  $E_{p(z|d)}[\log p(z)]$  sondern a priori als  $E_{p(z)}[\log p(z)]$  bewertet. Man beachte, dass  $\tilde{I}_{p(z)}[d]$  damit gerade die Differenz zwischen der Entropie der A-Priori-Verteilung  $p(z)$  und der Entropie der A-Posteriori-Verteilung  $p(z|d)$  angibt.

Die in Gleichung (6.33) angegebene Festlegung ist für sich genommen konzeptionell gesehen mit Schwierigkeiten verbunden. Wie in [Ber04] im Hinblick auf die Frage nach ihrer Eignung als Maß für die (Menge an) Information, welche konkret in Form von  $d$  vorliegende Informationsbeiträge in Bezug auf  $z \in Z$  liefern, angemerkt und wie auch bereits in [Lin56] festgestellt ist sie auch nicht invariant unter Eins-zu-Eins-Transformationen von  $z$ .

Der Unterschied zwischen  $I_{p(z)}[d]$  und  $\tilde{I}_{p(z)}[d]$  erübrigt sich allerdings, wenn man (wie auch bereits in [Lin56] durchgeführt) die Erwartungsbildung gemäß

$p(d)$  vornimmt, d. h. es gilt neben Gleichung (6.29) ebenfalls

$$J[z,d] = E_{p(d)}[\tilde{I}_{p(z)}[d]]. \quad (6.34)$$

Siehe hierzu [Lin56] und insbesondere auch [Cov91].

### 6.3.4 Anwendung für die lokale Bayes'schen Fusion

Ein globales Bayes'sches Modell sei in Form der Verteilungsklasse  $\{p(d|z)|z \in Z\}$  und der A-Priori-Verteilung  $p(z)$  über  $Z$  gegeben. Weiter sei angenommen, dass die Bayes'sche Fusion wie in Abschnitt 5.2.3 beschrieben eingeschränkt auf einen lokalen Kontext  $U \subset Z$  durchgeführt werden soll. Diese Einschränkung entspricht der Annahme, dass sicher  $z \in U$  gilt.

#### 6.3.4.1 A priori erfolgende Einschränkung auf den lokalen Kontext

Beim Übergang vom globalen zum lokalen Bayes'schen Modell reduziert sich die angenommene Verteilungsklasse auf  $\{p(d|z)|z \in U\}$  und die globale A-Priori-Verteilung  $p(z)$  wird gemäß Gleichung (5.9) durch die lokale A-Priori-Verteilung  $p(z|U) = p(z)/P(U)$  ersetzt.

Nachfolgend wird, ähnlich zu einer in [Wil80] vorgenommenen Anwendung des Prinzips der Minimalen Information, skizziert, dass das Vorgehen, die globale A-Priori-Verteilung  $p(z)$  durch die lokale A-Priori-Verteilung  $p(z|U)$  zu ersetzen, konsistent zum Prinzip der Minimalen Information ist. Hieraus ergibt sich dann insbesondere auch eine Möglichkeit, den Informationszuwachs bezüglich  $z$ , welcher aus der Annahme  $z \in U$  resultiert, aus informationstheoretischer Sicht stimmig zu quantifizieren.

Ausgehend vom globalen Bayes'schen Modell, in dem der bezüglich  $z$  a priori verfügbare Informationsstand durch die globale A-Priori-Verteilung  $p(z)$  repräsentiert wird, stellt die Annahme, dass sicher  $z \in U$  gilt, neu im Hinblick auf  $z$  verfügbar gewordene Information dar. Folgt man dem Prinzip der

Minimalen Information, so ist (gemäß<sup>1</sup> Gleichung (3.14)) die globale A-Priori-Verteilung  $p(z)$  so durch eine neue Wahrscheinlichkeitsverteilung  $p_{\text{MI}}(z)$  zu ersetzen, dass gilt

$$p_{\text{MI}}(z) = \arg \min_{q(z) \in \mathfrak{P}_z} \text{KD}[q(z), p(z)] \quad (6.35)$$

wobei  $\mathfrak{P}_z$  die Menge aller Wahrscheinlichkeitsverteilungen auf  $Z$  bezeichnet, welche konsistent mit der neu hinzugekommenen Information, dass (gemäß der Annahme) sicher  $z \in U$  gilt, sind. Es entspricht  $\mathfrak{P}_z$  dann gerade der Menge aller Wahrscheinlichkeitsverteilungen  $q(z)$ , welche die Bedingung  $\int_U q(z) dz = 1$  erfüllen. Damit ergibt sich

$$\text{KD}[q(z), p(z)] = \int_Z q(z) \log \frac{q(z)}{p(z)} dz \quad (6.36)$$

$$= \int_U q(z) \log \frac{q(z)}{p(z)} dz \quad (6.37)$$

$$= \int_U q(z) \log \frac{q(z) P(U)}{p(z) P(U)} dz \quad (6.38)$$

$$= \int_U q(z) \log \frac{q(z)}{p(z|U)} dz - \log P(U) . \quad (6.39)$$

$$= \text{KD}[q(z), p(z|U)] - \log P(U) . \quad (6.40)$$

Es gilt  $\text{KD}[q(z), p(z|U)] \geq 0$  mit Gleichheit genau dann, wenn  $q(z)$  und  $p(z|U)$  übereinstimmen (vgl. z. B. [Cov91] hinsichtlich den Eigenschaften der Kullback-Leibler-Divergenz). D. h. für  $p_{\text{MI}}(z)$  ergibt sich gerade die lokale A-Priori-Verteilung  $p(z|U)$ .

Im Einklang mit dem Prinzip der Minimalen Information lässt sich der in Gleichung (6.40) dann noch verbleibende Term

---

<sup>1</sup> In Gleichung (6.35) sind gegenüber Gleichung (3.14)  $q(z)$  und  $p(z)$  bewusst vertauscht, da die globale A-Priori-Verteilung  $p(z)$  bei der im vorliegenden Abschnitt vorgenommenen Anwendung des Prinzips der Minimalen Information anders als in Abschnitt 3.5.1 die vor Bekanntwerden der neu verfügbaren Information vorhandene Wahrscheinlichkeitsverteilung darstellt.



$$I_{p(z)}[U] := -\log P(U) \quad (6.41)$$

auffassen als (Maß für den) Informationszuwachs bezüglich  $z \in Z$ , welcher aus der Annahme, dass  $z \in U$  gilt, resultiert (vgl. hierzu auch [Wil80]). Die Berechnung von  $I_{p(z)}[U]$  setzt voraus, dass die A-Priori-Wahrscheinlichkeit  $P(U)$  des lokalen Kontexts  $U$  bekannt ist.

### 6.3.4.2 Durchführung der lokalen Bayes'schen Fusion

Liegt ein lokales Bayes'sches Modell bereits in Form der angenommenen Verteilungsklasse  $\{p(d|z)|z \in U\}$  und der lokalen A-Priori-Verteilung  $p(z|U)$  vor, so wird die lokale A-Priori-Verteilung  $p(z|U)$  durch Anwendung des Satzes von Bayes wie folgt in optimaler Weise in die lokale A-Posteriori-Verteilung  $p(z|d,U)$  überführt (vgl. Gleichung (5.10)):

$$p(z|d,U) = \frac{p(d|z)p(z|U)}{\int_U p(d|z)p(z|U) dz} \quad (6.42)$$

In Abb. 6.5 ist dieser Prozess anhand des in Abschnitt 6.3.1 behandelten und in Abb. 6.4 für den Fall globaler Bayes'scher Inferenz bereits illustrierten Informationsverarbeitungsschemas dargestellt. Die lokale Anwendung des Satzes von Bayes entspricht dabei gerade Gleichung (6.42).

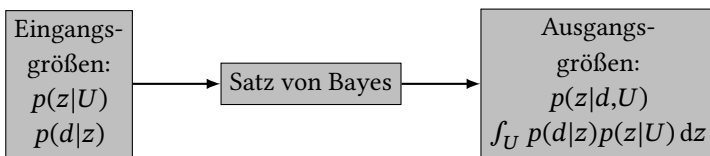


Abbildung 6.5: Lokale Informationsverarbeitung gemäß dem Satz von Bayes.

Der Informationszuwachs, welchen die konkret in Form von  $d$  vorliegenden Informationsbeiträge in Bezug auf  $z \in U$  liefern, ist bei der Anwendung des

Satzes von Bayes im lokalen Bayes'schen Modell gegeben durch (vgl. Gleichung (6.26))

$$I_{p(z|U)}[d] = \text{KD}[p(z|d,U), p(z|U)] . \quad (6.43)$$

Die Größe  $I_{p(z|U)}[d]$  kann im lokalen Bayes'schen Modell anhand der lokalen A-Priori-Verteilung  $p(z|U)$  und der lokalen A-Posteriori-Verteilung  $p(z|d,U)$  berechnet werden.

Fasst man den Informationszuwachs im Hinblick auf  $z \in Z$ , welcher sich gemäß Gleichung (6.41) a priori durch die Annahme, dass  $z \in U$  sicher gilt, ergibt, und den durch die Anwendung des Satzes von Bayes resultierenden Informationszuwachs bezüglich  $z \in U$  gemäß Gleichung (6.43), welcher lokal gesehen durch das Bekanntwerden von  $d$  resultiert, zusammen, so ergibt sich

$$I_{p(z)}[d,U] := I_{p(z)}[U] + I_{p(z|U)}[d] \quad (6.44)$$

$$= -\log P(U) + \int_U p(z|d,U) \log \frac{p(z|d,U)}{p(z|U)} dz \quad (6.45)$$

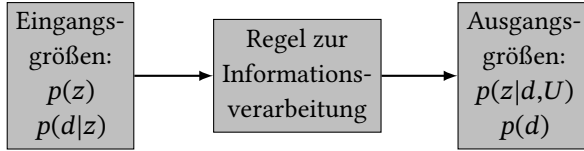
$$= \int_U p(z|d,U) \log \frac{p(z|d,U)}{p(z|U)P(U)} dz \quad (6.46)$$

$$= \int_U p(z|d,U) \log \frac{p(z|d,U)}{p(z)} dz \quad (6.47)$$

$$= \text{KD}[p(z|d,U), p(z)] . \quad (6.48)$$

### 6.3.4.3 A posteriori erfolgende Einschränkung auf den lokalen Kontext

Nimmt man wie in nachfolgender Abb. 6.6 dargestellt im in Abb. 6.3 dargestellten Informationsverarbeitungsschema an, dass die Ausgangsgröße  $r(z|d)$  gerade die lokale A-Posteriori-Verteilung  $p(z|d,U)$  ist, so ist die Regel zur Informationsverarbeitung nicht optimal.



**Abbildung 6.6:** Globale Sicht auf die Informationsverarbeitung im lokalen Bayes'schen Modell.

Wertet man das von Zellner gemäß Gleichung (6.18) formulierte Funktional für diesen Fall aus, so ergibt sich

$$\begin{aligned}
 F_1[p(z|d,U)] &= \int_Z p(z|d,U) \log p(z|d,U) dz + \log p(d) \\
 &\quad - \int_Z p(z|d,U) [\log p(z) + \log p(d|z)] dz \quad (6.49)
 \end{aligned}$$

$$= \int_Z p(z|d,U) \log \frac{p(z|d,U)p(d)}{p(z)p(d|z)} dz \quad (6.50)$$

$$= \text{KD}[p(z|d,U), p(z|d)] \quad (6.51)$$

$$=: I_{p(z|d)}[U] . \quad (6.52)$$

Konsistent zum in Abschnitt 6.3.4.1 bereits auf die a priori vorliegende Situation angewandten Prinzip der Minimalen Information lässt sich  $I_{p(z|d)}[U]$  auffassen als Zuwachs an Information bezüglich  $z \in Z$ , der sich a posteriori gesehen durch die Annahme, dass sicher  $z \in U$  gilt, ergibt.

Gemäß Gleichung (5.11) ist  $p(z|d,U) = p(z|d)/P(U|d)$ , d. h. es ergibt sich

$$I_{p(z|d)}[U] = \text{KD} \left[ \frac{p(z|d)}{P(U|d)}, p(z|d) \right] = -\log P(U|d) . \quad (6.53)$$

Die Größe  $I_{p(z|d)}[U]$  ist im lokalen Bayes'schen Modell offensichtlich nicht berechenbar.

### 6.3.4.4 Alternative Betrachtung

An dieser Stelle ist es interessant, noch einmal einen Blick auf das von Walker gemäß Gleichung (6.23) formulierte Funktional zu werfen, um zu sehen, dass sich die Ergebnisse aus Abschnitt 6.3.4.2 und Abschnitt 6.3.4.3 auch unabhängig von Zellner's Ansatz direkt hieraus ableiten lassen.

Wertet man Gleichung (6.23) von einem globalen Gesichtspunkt für die lokale A-Posteriori-Verteilung  $p(z|d,U)$  aus, so ergibt sich

$$F_2[p(z|d,U)] \quad (6.54)$$

$$= - \int_Z p(z|d,U) \log p(d|z) dz + \text{KD}[p(z|d,U), p(z)] \quad (6.55)$$

$$= - \int_U p(z|d,U) \log p(d|z) dz + \int_U p(z|d,U) \log \frac{p(z|d,U)}{p(z)} dz \quad (6.56)$$

$$= - \int_U p(z|d,U) \log \frac{p(d|z) p(z)}{p(z|d,U)} dz \quad (6.57)$$

$$= - \int_U p(z|d,U) \log \frac{p(z|d) p(d)}{p(z|d,U)} dz \quad (6.58)$$

$$= - \int_U p(z|d,U) \log \frac{p(z|d) p(d) P(U|d)}{p(z|d)} dz \quad (6.59)$$

$$= - \log P(U|d) - \log p(d) . \quad (6.60)$$

Da gemäß Gleichung (6.25) gilt, dass  $F_2[p(z|d)] = - \log p(d)$  ist, ergibt sich

$$F_2[p(z|d,U)] - F_2[p(z|d)] = - \log P(U|d) . \quad (6.61)$$

Man erhält also

$$F_2[p(z|d,U)] - F_2[p(z|d)] = \text{KD}[p(z|d,U), p(z|d)] = I_{p(z|d)}[U] . \quad (6.62)$$

Wählt man statt der globalen A-Posteriori-Verteilung  $p(z|d)$  die lokale A-Posteriori-Verteilung  $p(z|d,U)$ , so erhöht sich also der Wert des von Walker formulierten Funktionals gerade um den Informationsgewinn, der sich a posteriori aus dem Bekanntwerden der Tatsache, dass  $z \in U$  gilt, ergibt.

Wertet man Gleichung (6.23) für die lokale A-Posteriori-Verteilung  $p(z|U,d)$  dagegen unter den im lokalen Bayes'schen Modell gelten Annahmen aus, so ergibt sich mit Gleichung (6.42) ein anderer Wert für  $F_2[p(z|d,U)]$ :

$$F_2[p(z|d,U)] = -\log \left( \int_U p(d|z)p(z|U) dz \right). \quad (6.63)$$

Der Wert des Funktionals verändert sich gegenüber der für die globale Bayes'sche Fusion vorliegenden Situation, d. h. gegenüber  $F_2[p(z|d)]$ , hier also wie folgt:

$$F_2[p(z|d,U)] - F_2[p(z|d)] \quad (6.64)$$

$$= -\log \left( \int_U p(d|z)p(z|U) dz \right) + \log p(d) \quad (6.65)$$

$$= -\log \int_U \frac{p(d|z)p(z|U)}{p(d)} dz \quad (6.66)$$

$$= -\log \int_U \frac{p(d|z)p(z)}{p(d)P(U)} dz \quad (6.67)$$

$$= -\log \int_U \frac{p(z|d)}{P(U)} dz \quad (6.68)$$

$$= -\log P(U|d) + \log P(U) \quad (6.69)$$

$$= \text{KD}[p(z|d,U), p(z|d)] - \text{KD}[p(z|U), p(z)] \quad (6.70)$$

$$= I_{p(z|d)}[U] - I_{p(z)}[U]. \quad (6.71)$$

Gegenüber Gleichung (6.62) verringert sich diese Differenz also um den Informationszuwachs bezüglich  $z \in Z$ , welcher a priori aus der Annahme, dass  $z \in U$  gilt, resultiert (vgl. Gleichung (6.41)).

Weiterhin gelten die in Gleichung (6.44) bis Gleichung (6.48) durchgeführten Berechnungen, so dass sich auch die Gleichheit

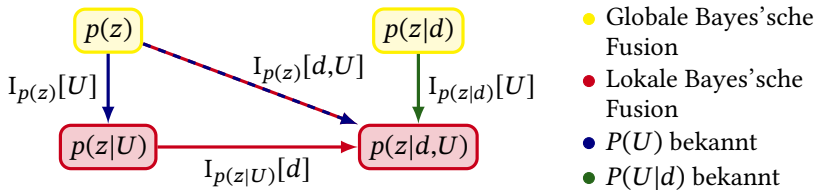
$$I_{p(z)}[d,U] = I_{p(z)}[U] + I_{p(z|U)}[d] = \text{KD}[p(z|d,U), p(z)] \quad (6.72)$$

ergibt.

Dadurch, dass  $F_2[p(z|d)]$  aus informationstheoretischer Sicht den Informationsszuwachs, der insgesamt resultiert, wenn  $d$  verfügbar wird, angibt, lassen sich Gleichung (6.62) und Gleichung (6.71) auch so auffassen, dass sich dieser Wert durch die Tatsache, dass man die globale A-Posteriori-Verteilung  $p(z|d)$  durch die lokale A-Posteriori-Verteilung  $p(z|d,U)$  ersetzt, um  $I_{p(z|d)}[U]$  bzw.  $I_{p(z|d)}[U] - I_{p(z)}[U]$  verringert, je nachdem ob man die Annahme  $z \in U$  als gegeben nimmt (Fall von Gleichung (6.71)) oder nicht (Fall von Gleichung (6.62)).

### 6.3.4.5 Zusammenfassende Betrachtung

Abb. 6.7 fasst die in Abschnitt 6.3.4.1 bis Abschnitt 6.3.4.3 hergeleiteten Bezüge, welche aus informationstheoretischer Sicht zwischen einem globalen Bayes'schen Modell und einem lokalen Bayes'schen Modell, welches aus ersterem durch Einschränkung auf einen lokalen Kontext  $U \subset Z$  resultiert, zusammen. Dabei ist jeweils farblich markiert, inwieweit die Kenntnis der A-Priori-Wahrscheinlichkeit  $P(U)$  und der A-Posteriori-Wahrscheinlichkeit  $P(U|d)$  des lokalen Kontexts  $U$  zur Berechnung der angegebenen informationstheoretischen Größen benötigt wird.



**Abbildung 6.7:** Schematische Darstellung der Bezüge zwischen einem globalen Bayes'schen Modell und einem lokalen Bayes'schen Modell, welches aus ersterem durch Einschränkung auf einen lokalen Kontext  $U \subset Z$  resultiert.

Ist die globale A-Priori-Wahrscheinlichkeit  $P(U)$  des lokalen Kontexts  $U$  bekannt, so kann man zusätzlich den Informationszuwachs bezüglich  $z \in Z$ , welcher a priori aus der Annahme, dass  $z \in U$  gilt, resultiert, berechnen. Ebenso kann man in diesem Fall zusätzlich den gemeinsamen Informationszuwachs bezüglich  $z \in Z$ , welcher gleichzeitig durch diese Annahme und das Bekanntwerden von  $d$  resultiert, berechnen.

Zur Berechnung des Zuwachses an Information bezüglich  $z \in Z$ , der a posteriori gesehen durch die Annahme, dass sicher  $z \in U$  gilt, resultiert, wird jedoch die Kenntnis der globalen A-Posteriori-Wahrscheinlichkeit  $P(U|d)$  des lokalen Kontexts  $U$  benötigt. Die vorangegangene Darstellung, insbesondere Abschnitt 6.3.4.3, hat verdeutlicht, dass man letztlich gerade diese Größe benötigen würde, um die Korrektheit (im Sinne von Abschnitt 6.1) lokaler Bayes'scher Modelle im Hinblick auf die Wahl des lokalen Kontexts  $U$  aus informationstheoretischer Sicht tatsächlich exakt quantifizieren zu können.

Die globale A-Posteriori-Wahrscheinlichkeit  $P(U|d)$  des lokalen Kontexts  $U$  gibt gerade an, wie wahrscheinlich es global gesehen ist, dass der „wahre“ Wert von  $z$  im lokalen Kontext  $U$  enthalten ist. Es ist letztlich nicht verwunderlich, dass diese Größe nicht bekannt ist, wenn die lokale Bayes'sche Fusion wie in Abschnitt 5.2.3 beschrieben eingeschränkt auf den lokalen Kontext  $U$  durchgeführt wird.

Genauer gilt

$$P(U|d) = \frac{P(U|d)}{P(U|d) + P(\bar{U}|d)} \quad (6.73)$$

$$= \left( 1 + \frac{P(\bar{U}|d)}{P(U|d)} \right)^{-1} \quad (6.74)$$

$$= \left( 1 + \frac{\int_{\bar{U}} p(d|z) p(z) dz}{\int_U p(d|z) p(z) dz} \right)^{-1} \quad (6.75)$$

$$= \left( 1 + \frac{\int_{\bar{U}} l(d|z) p(z) dz}{\int_U l(d|z) p(z) dz} \right)^{-1} \quad (6.76)$$

$$= \left( 1 + \frac{\int_{\bar{U}} l_{\text{st}}(d|z) p(z) dz}{\int_U l_{\text{st}}(d|z) p(z) dz} \right)^{-1}. \quad (6.77)$$

Das Analogon zu Gleichung (6.74) und Gleichung (6.76) für den Fall eines statistischen Modells mit zwei exklusiven, sich ausschließenden Hypothesen wurde in [Blu11] formuliert.

Sind die Informationsbeiträge bedingt unabhängig gegeben  $z$ , so gilt außerdem

$$P(U|d) = \left( 1 + \frac{\int_{\bar{U}} \prod_{s=1}^S l_{\text{st}}(d_s|z) p(z) dz}{\int_U \prod_{s=1}^S l_{\text{st}}(d_s|z) p(z) dz} \right)^{-1}. \quad (6.78)$$

Aus Gleichung (6.77) und Gleichung (6.78) lässt sich erkennen, dass die in Abschnitt 6.2.2 aus den Prinzipien statistischer Evidenz abgeleitete Vorgehensweise zur Festlegung des lokalen Kontexts  $U$ , welche grob gesagt darin besteht, diesen durch Anlegen geeigneter Schranken an die standardisierte(n) Likelihood-Funktion(en)  $l_{\text{st}}(d|z)$  bzw.  $l_{\text{st}}(d_s|z)$  zu bestimmen, grundsätzlich Sinn macht. Dieses Vorgehen kann wesentlich dazu beitragen, dass bei dem in Gleichung (6.77) bzw. Gleichung (6.78) jeweils angegebenen Bruch der Zähler möglichst klein und der Nenner möglichst groß ausfällt, d. h. dass Gleichung (6.77) bzw. Gleichung (6.78) insgesamt möglichst nahe bei Eins liegen. Man erkennt ebenfalls die Sinnhaftigkeit der am Schluss von Abschnitt 6.2.2 vorgeschlagenen Vorgehensweise, bei der Vorauswertung der Informationsbeiträge ggf. auch (durch eine entsprechende technische Ummodellierung) das in Form der A-Priori-Verteilung  $p(z)$  vorliegende Vorwissen mitzuberücksichtigen.

Ebenfalls erkennbar aus Gleichung (6.77) und Gleichung (6.78) wird eine weitere grundsätzliche Tatsache in Bezug auf die lokale Bayes'sche Fusion, welche eingeschränkt auf einen lokalen Kontext  $U \subset Z$  erfolgt: Ist der Umfang des Komplements  $\bar{U}$  des lokalen Kontexts  $U$  recht hoch im Vergleich zum Umfang des lokalen Kontexts selbst, was man ja anstreben mag, um den Aufwand der eigentlichen lokalen Bayes'schen Fusion möglichst gering zu halten, so kann auch bei noch so guter Wahl des lokalen Kontexts  $U$  dessen A-



Posteriori-Wahrscheinlichkeit  $P(U|d)$  möglicherweise deutlich ungleich Eins ausfallen. Dies bedeutet insbesondere auch, dass sich in diesem Fall die in Abschnitt 5.2.3.3 dargestellte Schwierigkeit, dass Wahrscheinlichkeitsaussagen im lokalen Bayes'schen Modell Überschätzungen der tatsächlichen (globalen) Wahrscheinlichkeitsaussagen darstellen, nicht beliebig gut umgehen lässt. Gerade auch deshalb ist es sinnvoll, die Güte lokaler Bayes'scher Modelle (auch) von einem pre-experimentellen Standpunkt aus zu betrachten.

Dieser pre-experimentelle Standpunkt lässt sich wie in Abschnitt 6.2 dargestellt durch die Ausnutzung der Prinzipien statistischer Evidenz, genauer auf Basis der Betrachtung der Wahrscheinlichkeit für irreführende statistische Evidenz, erreichen. Diese Wahrscheinlichkeit bildet sich unabhängig von ihrem von einem post-experimentellen Standpunkt und unter Berücksichtigung von evtl. zusätzlich vorhandenem Vorwissen formulierten Pendant, welches hier gerade der A-Posteriori-Wahrscheinlichkeit  $P(U|d)$  des lokalen Kontexts  $U$  entspricht. Vgl. hierzu insbesondere auch [Blu11]. Es sei angemerkt, dass das in Gleichung (6.29) in Form der Transinformation formulierte Maß für die Menge an Information, welche ein Bayes'sches Experiment mit A-Priori-Verteilung  $p(z)$ , in dem  $d$  gemäß  $p(d)$  verteilt ist, erwartungsgemäß liefert, keine solche Trennung abbildet. Hier wird gewissermaßen die (letztlich andere) Zielsetzung verfolgt, von einem pre-experimentellen Standpunkt aus vollständig auf die a posteriori vorliegende Situation zu schließen.

Man beachte weiterhin, dass der in Abschnitt 5.2.5 angeführte Vorschlag, in Folgearbeiten eine Weiterentwicklung der in der vorliegenden Arbeit erarbeiteten Ansätze zur lokalen Bayes'schen Fusion vorzunehmen, indem man diese mit dem in Abschnitt 4.2.1 behandelten Prinzip konjugierter Verteilungsfamilien bzw. mit den in Abschnitt 4.2.2 und Abschnitt 4.2.3 vorgestellten Verfahren zur Approximation der A-Posteriori-Verteilung  $p(z|d)$  bzw. von Kenngrößen derselben kombiniert, gerade auch vor dem Hintergrund der Tatsache, dass man möglicherweise den Umfang des Komplements  $\bar{U}$  des lokalen Kontexts  $U$  nicht beliebig gering ausfallen lassen möchte, als zielführend zu erachten ist. Dies gilt gerade auch bei einer verteilten Umsetzung der lokalen Bayes'schen Fusion. So ist es insbesondere zu erwarten, dass sich die im lokalen Bayes'schen Modell benötigten probabilistischen Größen deutlich bes-

ser durch überschaubare konjugierte Verteilungsfamilien approximieren lassen als ihre Pendants im globalen Bayes'schen Modell. Im Erfolgsfall reduziert sich dann die lokale Bayes'sche Fusion auf die analytische Verknüpfung einer moderaten Anzahl an Verteilungsparametern. Bei der Umsetzung entsprechender Ansätze könnte man auch versuchen, gezielt eine möglichst gute Approximation der globalen A-Posteriori-Wahrscheinlichkeit  $P(U|d)$  des lokalen Kontexts  $U$  anzustreben, z. B. unter Zuhilfenahme von Ansätzen zur Bayes'schen Robustheitsanalyse (siehe hierzu z. B. [Ins00]).

Abschließend sei noch angemerkt, dass in Abschnitt 7.2 eine Abschätzung für die A-Posteriori-Wahrscheinlichkeit  $P(U|d)$  des lokalen Kontexts  $U$  hergeleitet wird, welche unter gewissen Zusatzannahmen auf Basis der auf  $U$  eingeschränkten lokalen Bayes'schen Fusion berechnet werden kann.

## 7 Einbettung in den globalen Kontext

### 7.1 Einführung

Die aus der lokalen Bayes'schen Fusion resultierende lokale A-Posteriori-Verteilung stellt partielle Information im Hinblick auf die globale A-Posteriori-Verteilung  $p(z|d)$  dar. Dies lässt sich so auffassen, dass sie letztlich eine Menge von möglichen globalen A-Posteriori-Verteilungen über  $Z$  festlegt. Im vorliegenden Kapitel wird im Hinblick auf die lokale Bayes'sche Fusion, welche wie in Abschnitt 5.2.3 beschrieben durch Einschränkung auf den lokalen Kontext  $U$  erfolgt, untersucht, wie sich das in Form der auf  $U$  eingeschränkten lokalen A-Posteriori-Verteilung  $p(z|d,U)$  ergebende Fusionsresultat auf Basis dieses Gedankens und unter zusätzlicher Verwendung von Information darüber, wie der lokale Kontext  $U$  gebildet wurde, expliziter als in Abschnitt 5.2.3 in Bezug zum in Form der globalen A-Posteriori-Verteilung  $p(z|d)$  resultierenden Resultat der globalen Bayes'schen Fusion setzen und auf dieser Basis auch zur Ableitung weiterführender Entscheidungen nutzen lässt.

Als erster Schritt hierzu wird, vor allem auch unter Berücksichtigung der aus Kapitel 6 resultierenden Erkenntnisse hinsichtlich mathematischer Kriterien zur Wahl des lokalen Kontexts  $U$ , in Abschnitt 7.2 ein Wahrscheinlichkeitsintervall-Schema für die globalen A-Posteriori-Wahrscheinlichkeiten von Ereignissen hergeleitet. Die globale A-Posteriori-Verteilung  $p(z|d)$  und anhand von ihr berechnete globale A-Posteriori-Wahrscheinlichkeiten von Ereignissen verkörpern gerade die Tatsache, dass in der Regel aufgrund der auch nach der Durchführung der Informationsfusion noch vorhandenen Restunsicherheit nicht zweifelsfrei auf den „wahren“ Wert

der durch  $z$  verkörperten Größen des Interesses zurückgeschlossen werden kann (vgl. auch Abschnitt 3.2). Das Wahrscheinlichkeitsintervall-Schema repräsentiert in expliziter und nicht probabilistischer Weise die aufgrund der Einschränkung auf den lokalen Kontext  $U$  a posteriori zusätzlich noch verbleibende Unsicherheit in Bezug auf  $z$ . Dies geschieht unter Berücksichtigung vorhandener Zusatzinformation im Hinblick auf die Festlegung des lokalen Kontexts  $U$ . Wie vor allem auch anhand eines Beispiels illustriert werden wird, kann das Wahrscheinlichkeitsintervall-Schema dadurch dazu beitragen, in der Praxis einen möglichst intuitiven Überblick über die auf Basis der lokalen Bayes'schen Fusion erzielten Erkenntnisse hinsichtlich der globalen A-Posteriori-Wahrscheinlichkeiten von Ereignissen zu erhalten und so besser feststellen zu können, ob die im Rahmen der lokalen Bayes'schen Fusion erzielten Resultate einen ausreichend guten Informationsstand im Hinblick auf  $z$  darstellen. Ist dies nicht der Fall, so könnte man sich entschließen, die lokale Bayes'sche Fusion z. B. durch Einbeziehung zusätzlicher Informationsquellen oder durch Ausweitung des lokal betrachteten Anteils von  $Z$  fortzuführen.

Ist in einer vorliegenden Fusionsaufgabe letztlich nur der final resultierende Informationsstand im Hinblick auf  $z$  von Interesse, d. h. soll die beschriebene Trennung der Unsicherheiten unterschiedlicher Herkunft nicht explizit weiter genutzt werden, so macht es gerade im Hinblick auf die Allgemeingültigkeit der Degree-of-Belief-Interpretation von Wahrscheinlichkeit (vgl. Abschnitt 3.3.2.2) Sinn, diese Trennung aufzuheben und eine eindeutige A-Posteriori-Wahrscheinlichkeitsverteilung zu bestimmen, welche den insgesamt vorliegenden Informationsstand (auch unter Berücksichtigung der Information darüber, wie der lokale Kontext gebildet wurde) in objektiver Weise ausschließlich probabilistisch repräsentiert. Dies ist das Thema von Abschnitt 7.3, in dem zu diesem Zweck das in Abschnitt 3.5.1 eingeführte Prinzip der Maximalen Entropie angewandt und hinsichtlich seiner Arbeitsweise speziell in diesem Fall genauer untersucht wird. Insbesondere wird dabei durch analytische Lösung des mit der Anwendung des Prinzips der Maximalen Entropie korrespondierenden Optimierungsproblems eine geschlossene Lösung für die resultierende eindeutige (Maximum-Entropie) A-Posteriori-Verteilung hergeleitet werden.

In der statistischen Entscheidungstheorie ist neben der Weiterentwicklung des Informationsstands bezüglich der Größen des Interesses auch noch relevant, welche Konsequenzen darauf basierende Entscheidungen nach sich ziehen. Dadurch lässt sich dann bestmöglich eine bestimmte Aktion aus einer Menge von möglichen Aktionen auswählen. Die Bayes'sche Theorie zeichnet sich dadurch aus, dass sie konsistent mit der statistischen Entscheidungstheorie ist (vgl. z. B. [Ber85] oder [Ber04]). Die in Abschnitt 7.3 adressierte Festlegung einer eindeutigen (Maximum-Entropie) A-Posteriori-Verteilung entspricht letztlich einer bestimmten, konkreten Entscheidung. Es ist i. A. nicht sinnvoll, eine derartige Entscheidung vorzusehen, wenn das Ergebnis der Bayes'schen Fusion nachgelagert zur Lösung eines Entscheidungsproblems im Sinne der statistischen Entscheidungstheorie verwendet werden soll, denn in diesem Fall würden zwei harte Entscheidungen unabhängig voneinander getroffen werden. Dementsprechend wird in Abschnitt 7.4 untersucht, wie genau man das Wahrscheinlichkeitsintervall-Schema methodisch gesehen direkt in die Lösung eines Entscheidungsproblems einbringen kann und welche Schlüsse auf dieser Basis möglich sind. Wie aufgezeigt werden wird, ist es dabei zielführend, das korrespondierende Entscheidungsproblem unter partieller Information auf ein spezielleres Entscheidungsproblem unter sog. linearer partieller Information (LPI) zu überführen, und es werden auf dieser Basis konkrete Entscheidungskriterien abgeleitet.

## 7.2 Wahrscheinlichkeitsintervall-Schema

Der vorliegende Abschnitt fasst Ergebnisse aus dem bisherigen Verlauf der Arbeit zusammen und entwickelt diese in Hinblick auf die Zielsetzung des vorliegenden Kapitels gezielt weiter. Die entsprechenden Anteile wurden zuvor in ähnlicher Form als Teil der Veröffentlichungen [San09a], [San10b] und [San14] publiziert.

Der in Abschnitt 5.2.3 beschriebene Übergang von einem globalen zu einem lokalen Bayes'schen Modell durch die Annahme, dass  $z \in U$  sicher gilt, bedeutet aus globaler Sicht, dass allen Ereignissen  $A \subset \bar{U}$  a priori und a posteriori die Wahrscheinlichkeit Null zugeordnet wird. Durch das Bedingen auf

den lokalen Kontext  $U$  wird die Wahrscheinlichkeitsmasse, welche im globalen Bayes'schen Modell auf  $\bar{U}$  liegt, beim Übergang zum lokalen Bayes'schen Modell auf  $U$  umverteilt. Die Umsetzung dieses Vorgehens ist notwendig, um wieder ein aus mathematischer Hinsicht valides Bayes'sches Modell zu erhalten und so letztlich auch das gesamte Spektrum der durch die Bayes'sche Theorie bereitgestellten Konzepte und Methoden verfügbar zu haben.

Wie bereits in Abschnitt 5.2.3.3 dargestellt wurde, erhöht sich dadurch die Wahrscheinlichkeit von Ereignissen  $A \subset U$  a priori um den Faktor  $1/P(U)$  und a posteriori um den Faktor  $1/P(U|d)$ . Insgesamt gesehen gilt a posteriori für ein Ereignis  $A \subset Z$ :

$$P(A|d,U) = \int_A p(z|d,U) dz \quad (7.1)$$

mit

$$p(z|d,U) = \begin{cases} \frac{p(z|d)}{P(U|d)}, & \text{falls } z \in U, \\ 0, & \text{falls } z \notin U. \end{cases} \quad (7.2)$$

Gleichung (7.2) ergibt sich aus Satz 5.2. Anders als dort wird  $p(z|d,U)$  hier jedoch als Wahrscheinlichkeitsverteilung auf  $Z$  statt nur auf  $U$  aufgefasst. Dies ist möglich, da man wie in Abschnitt 5.2.3.1 beschrieben sowohl  $(U, \mathcal{Z}|_U)$  als auch  $(Z, \mathcal{Z})$  als den dem zugehörigen Wahrscheinlichkeitsmaß  $P(\cdot|U)$  zugrundeliegenden messbaren Raum auffassen kann.

Insbesondere gilt also für Ereignisse  $A \subset U$

$$P(A|d) = P(A|d,U)P(U|d) \quad (7.3)$$

$$\leq P(A|d,U) \quad (7.4)$$

und für Ereignisse  $A \subset \bar{U}$

$$P(A|d) \leq 1 - P(U|d). \quad (7.5)$$

Weiter gilt wie in Gleichung (5.20) dargestellt für Ereignisse  $A, B \subset U$ :

$$\frac{P(A|d,U)}{P(B|d,U)} = \frac{P(A|d)}{P(B|d)} =: o(A,B). \quad (7.6)$$

Es sei das Ereignis  $B \subset U$  beliebig aber fest gewählt. Nimmt man an, dass die (im lokalen Bayes'schen Modell unbekannt) globale A-Posteriori-Wahrscheinlichkeit  $P(B|d)$  von  $B$  einen bestimmten Wert hat, so ergibt sich aus Gleichung (7.6) gerade der Wert der globalen A-Posteriori-Wahrscheinlichkeit  $P(A|d)$  für jedes andere Ereignis  $A \subset U$  in eindeutiger Weise als

$$P(A|d) = o(A,B)P(B|d). \quad (7.7)$$

Der Faktor  $o(A,B)$  lässt sich dabei im lokalen Bayes'schen Modell berechnen.

Wurde der lokale Kontext  $U$  so festgelegt, dass er alle möglichen Werte  $z \in Z$  mit  $l(d|z) > \delta$  enthält, was z. B. der Fall ist, wenn  $U$  wie in Abschnitt 6.2.2 beschrieben durch Ansetzen einer unteren Schranke an die standardisierte Likelihood-Funktion festgelegt wird, so lässt sich die globale A-Posteriori-Wahrscheinlichkeit  $P(U|d)$  des lokalen Kontexts  $U$  wie folgt abschätzen:

**Satz 7.1.** *Aus der Bedingung  $l(d|z) \leq \delta$  für alle  $z \in \bar{U}$  ergibt sich*

$$1 \geq P(U|d) \geq \frac{\int_U l(d|z)p(z|U) dz}{\int_U l(d|z)p(z|U) dz + \delta \left( \frac{1}{P(U)} - 1 \right)} =: \beta. \quad (7.8)$$

*Beweis.* Da  $P(U|d)$  eine Wahrscheinlichkeit darstellt, ist gilt  $1 \geq P(U|d)$  trivialerweise.

Die untere Schranke für  $P(U|d)$  lässt sich im Wesentlichen durch Anwendung des Satzes von Bayes (vgl. Gleichung (3.5)) und unter Ausnutzung der zusätzlich für  $z \in U$  geltenden Beziehung  $p(z) = p(z|U)P(U)$  (vgl. Gleichung (5.9)) herleiten:

$$P(U|d) = \int_U p(z|d) dz \quad (7.9)$$

$$= \frac{\int_U p(d|z)p(z) dz}{p(d)} \quad (7.10)$$

$$= \frac{\int_U p(d|z)p(z) dz}{\int_U p(d|z)p(z) dz + \int_{\bar{U}} p(d|z)p(z) dz} \quad (7.11)$$

$$= \frac{\int_U l(d|z)p(z) dz}{\int_U l(d|z)p(z) dz + \int_{\bar{U}} l(d|z)p(z) dz} \quad (7.12)$$

$$\geq \frac{\int_U l(d|z)p(z) dz}{\int_U l(d|z)p(z) dz + \delta \int_{\bar{U}} p(z) dz} \quad (7.13)$$

$$= \frac{\int_U l(d|z)p(z) dz}{\int_U l(d|z)p(z) dz + \delta(1 - P(U))} \quad (7.14)$$

$$= \frac{\int_U l(d|z)p(z|U)P(U) dz}{\int_U l(d|z)p(z|U)P(U) dz + \delta(1 - P(U))} \quad (7.15)$$

$$= \frac{\int_U l(d|z)p(z|U) dz}{\int_U l(d|z)p(z|U) dz + \delta\left(\frac{1}{P(U)} - 1\right)}. \quad (7.16)$$

□

Die in Gleichung (7.8) angegebene untere Schranke  $\beta$  für die A-Posteriori-Wahrscheinlichkeit  $P(U|d)$  des lokalen Kontexts  $U$  kann auf Basis des lokalen Bayes'schen Modells bestimmt werden, wenn neben der an die Likelihood-Funktion angelegten Schranke  $\delta$  die A-Priori-Wahrscheinlichkeit  $P(U)$  des lokalen Kontext bekannt ist. Ist Letzteres nicht der Fall, es aber möglich, für ein geeignetes  $\alpha \in (0,1)$  Abschätzung der Form  $P(U) \geq \alpha$  zu formulieren, so kann durch Fortführung der in Gleichung (7.8) vorgenommenen Abschätzung immerhin eine schwächere Schranke für  $P(U|d)$  bestimmt werden.

Die in Satz 7.1 angegebene Möglichkeit zur Abschätzung der globalen A-Posteriori-Wahrscheinlichkeit  $P(U|d)$  des lokalen Kontexts  $U$  kann auch auf den Fall, dass die in  $d$  zusammengefassten Informationsbeiträge  $d_s$ ,  $s \in \{1, \dots, S\}$ , auf Basis individueller Likelihood-Funktionen ausgewertet werden, übertragen werden. Dieser Sachverhalt ist besonders dann relevant,



wenn die Informationsbeiträge  $d_s$  als bedingt unabhängig gegeben  $z$  angenommen werden können und sich die Likelihood-Funktion  $l(d|z)$  somit als das Produkt der quellenspezifischen Likelihood-Funktionen bei der Fusion berücksichtigen lässt (vgl. Gleichung (3.24) und Gleichung (3.25)). Das nachfolgende Korollar stellt eine mögliche Adaption von Satz 7.1 für diesen Fall dar:

**Korollar 7.2.** *Es sei angenommen, dass die in  $d$  zusammengefassten Informationsbeiträge  $d_s$  bedingt unabhängig gegeben  $z$  sind. Dann ergibt sich aus der Bedingung  $l(d_s|z) \leq \delta_s$  für alle  $z \in \bar{U}$ ,  $s \in \{1, \dots, S\}$ , die Abschätzung*

$$1 \geq P(U|d) \geq \frac{\int_U l(d|z)p(z|U) dz}{\int_U l(d|z)p(z|U) dz + \left(\prod_{s=1}^S \delta_s\right) \left(\frac{1}{P(U)} - 1\right)} =: \tilde{\beta}. \quad (7.17)$$

*Beweis.* Die Bedingung  $1 \geq P(U|d)$  ist offensichtlich wieder trivialerweise erfüllt.

Analog wie im Beweis von Satz 7.1 ergibt sich

$$P(U|d) = \frac{\int_U l(d|z)p(z) dz}{\int_U l(d|z)p(z) dz + \int_{\bar{U}} l(d|z)p(z) dz}. \quad (7.18)$$

Gemäß der Annahme der bedingten Unabhängigkeit der Informationsbeiträge  $d_s$  gegeben  $z$  lässt sich in Gleichung (7.18) die Likelihood-Funktion  $l(d|z)$  durch das Produkt  $\prod_{s=1}^S l(d_s|z)$  ersetzen und die Abschätzung

$$\prod_{s=1}^S l(d_s|z) \leq \prod_{s=1}^S \delta_s \quad (7.19)$$

für  $z \in \bar{U}$  liefert damit weiter analog zum Beweis von Satz 7.1 die in Gleichung (7.17) formulierte untere Schranke  $\tilde{\beta}$  für  $P(U|d)$ .  $\square$

Im weiteren Verlauf der Arbeit wird der Einfachheit halber stets Bezug auf die in Satz 7.1 formulierte Abschätzung der A-Posteriori-Wahrscheinlichkeit

$P(U|d)$  des lokalen Kontexts  $U$  Bezug genommen. Die Herleitungen sind dabei jedoch auf die Situation von Korollar 7.2 übertragbar, wenn man  $\beta$  gemäß Gleichung (7.8) durch  $\tilde{\beta}$  gemäß Gleichung (7.17) ersetzt.

Das Einsetzen von Gleichung (7.8) in Gleichung (7.3) liefert eine untere Schranke für die globale A-Posteriori-Wahrscheinlichkeit von Ereignissen  $A \subset U$ . Kombiniert man dieses Ergebnis mit den oberen Schranken für die globalen A-Posteriori-Wahrscheinlichkeiten von Ereignissen gemäß Gleichung (7.4) und Gleichung (7.5), so erhält man insgesamt das folgende Intervall-Schema:

$$P(A|d) \in [l(A), r(A)] \quad \text{mit}$$

$$[l(A), r(A)] = \begin{cases} [\beta \cdot P(A|d, U), P(A|d, U)] , & \text{falls } A \subset U , \\ [0, 1 - \beta] , & \text{falls } A \subset \bar{U} . \end{cases} \quad (7.20)$$

Mit Gleichung (7.3) ergibt sich für die Länge

$$d(A) := r(A) - l(A) \quad (7.21)$$

der in Gleichung (7.20) angegebenen Wahrscheinlichkeitsintervalle:

$$d(A) = \begin{cases} \frac{1-\beta}{P(U|d)} P(A|d) , & \text{falls } A \subset U , \\ 1 - \beta , & \text{falls } A \subset \bar{U} . \end{cases} \quad (7.22)$$

$d(A)$  wird für  $A \subset U$  also erst einmal unabhängig von  $A$  zum einen von der A-Posteriori-Wahrscheinlichkeit  $P(U|d)$  des lokalen Kontexts  $U$  und zum anderen von der für diese berechneten unteren Schranke  $\beta$  beeinflusst. Für  $A \subset \bar{U}$  ist nur die Schranke  $\beta$  relevant.  $d(A)$  kann letztlich in beiden Fällen nur dann eher gering ausfallen, wenn der lokale Kontext  $U$  so konstituiert ist, dass  $P(U|d)$  auch dementsprechend hoch ist. Ebenso muss weiter auch die untere Schranke  $\beta$  für  $P(U|d)$  entsprechend aussagekräftig sein. Im Fall  $A \subset U$  hängt  $d(A)$  jedoch auch von der globalen A-Posteriori-Wahrscheinlichkeit des betrachteten Ereignisses  $A$  ab. Ein global gesehen eher wahrscheinliches Ereignis korrespondiert grundsätzlich mit einem längeren Wahrscheinlichkeitsintervall als ein global gesehen unwahrscheinlicheres Ereignis. Genauer gilt

wegen Gleichung (7.7) im Fall  $A, B \subset U$  für das Verhältnis der Längen der entsprechenden Wahrscheinlichkeitsintervalle gerade  $d(A)/d(B) = o(A, B)$ .

Im Hinblick auf die Auswertung der Wahrscheinlichkeitsintervalle ist dementsprechend auch insgesamt zu berücksichtigen, dass letztlich nicht allein Gleichung (7.20) das Wissen zusammenfasst, welches aus der lokalen Bayes'schen Fusion im Hinblick auf die globalen A-Posteriori-Wahrscheinlichkeiten resultiert. Zu berücksichtigen ist weiter Gleichung (7.7). Hieraus ergibt sich nämlich auch, dass die Werte der globalen A-Posteriori-Wahrscheinlichkeiten von Ereignissen, welche Teil des lokalen Kontexts  $U$  sind, nicht voneinander unabhängig in den Wahrscheinlichkeitsintervallen variieren können. Nimmt man an, dass die A-Posteriori-Wahrscheinlichkeit des Ereignisses  $B \subset U$  einen bestimmten Wert im zugehörigen Wahrscheinlichkeitsintervall annimmt, so ist wegen Gleichung (7.7) der mögliche Wert der A-Posteriori-Wahrscheinlichkeit für alle anderen Ereignisse  $A \subset U$  hieraus ja gerade eindeutig festgelegt. Es liegt auf der Hand, dass durch diesen Sachverhalt die Aussagekraft des Wahrscheinlichkeitsintervall-Schemas deutlich erhöht wird.

In Abschnitt 7.3.2.1 werden die Struktur des Wahrscheinlichkeitsintervall-Schemas und Möglichkeiten zur seiner Auswertung noch genauer anhand eines konkreten Beispiels verdeutlicht werden. Zuvor wird jedoch in Abschnitt 7.3.1 aufgezeigt, wie man durch Anwendung des Prinzips der Maximalen Entropie aus der Menge der möglichen A-Posteriori-Verteilungen, welche konsistent mit dem Wahrscheinlichkeitsintervall-Schema (verbunden mit Gleichung (7.7)) sind, eine dementsprechend aus informationstheoretischer Sicht optimale Auswahl einer bestimmten A-Posteriori-Verteilung treffen kann.

## 7.3 Anwendung des Prinzips der Maximalen Entropie

### 7.3.1 Analytische Bestimmung der (Maximum-Entropie) A-Posteriori-Verteilung

Gemäß dem Prinzip der Maximalen Entropie sollte die globale A-Posteriori-Verteilung als die Lösung des folgenden Optimierungsproblems festgelegt werden (vgl. Gleichung (3.11)):

$$p_{\text{ME}}(z|d) = \arg \max_{p(z|d) \in \mathfrak{P}_z} H[p(z|d)], \quad (7.23)$$

wobei  $\mathfrak{P}_z$  die Menge aller Wahrscheinlichkeitsverteilungen auf  $Z$  bezeichnet, welche konsistent mit Gleichung (7.7) und Gleichung (7.20) (oder äquivalent dazu mit Gleichung (7.2) und Gleichung (7.8)) sind.

Im vorliegenden Abschnitt wird eine explizite Formel für  $p_{\text{ME}}(z|d)$  angegeben und bewiesen. Anschließend wird die Arbeitsweise des Prinzips der Maximalen Entropie für diesen Fall genauer betrachtet. Diese Ergebnisse wurden zuvor als Teil der Publikationen [San10a] und [San14] veröffentlicht.

Für den Beweis der Formel für  $p_{\text{ME}}(z|d)$  wird das nachfolgende Lemma benötigt werden:

**Lemma 7.3.** *Es bezeichne  $p_C(x|d)$  die Wahrscheinlichkeitsverteilung über  $X := \{U, \bar{U}\}$ , welche  $U$  gerade die Wahrscheinlichkeit  $P(U|d)$  zuweist. Dann gilt die folgende Beziehung:*

$$H[p(z|d)] = H[p_C(x|d)] + P(U|d) H[p(z|d, U)] + P(\bar{U}|d) H[p(z|d, \bar{U})]. \quad (7.24)$$

*Beweis.* Sowohl  $U$  als auch  $\bar{U}$  lassen sich vom mathematischen Standpunkt als lokale Kontexte auffassen. Gemäß Gleichung (5.11) gilt somit

$$p(z|d) = P(V|d)p(z|d, V), \quad V \in \{U, \bar{U}\}. \quad (7.25)$$

Unter Ausnutzung dieser Beziehung ergibt sich:

$$H[p(z|d)] = - \int_{\mathcal{Z}} p(z|d) \log p(z|d) dz \quad (7.26)$$

$$= - \int_U p(z|d) \log p(z|d) dz$$

$$- \int_{\bar{U}} p(z|d) \log p(z|d) dz \quad (7.27)$$

$$= - \int_U P(U|d) p(z|d, U) \log (P(U|d) p(z|d, U)) dz$$

$$- \int_{\bar{U}} P(\bar{U}|d) p(z|d, \bar{U}) \log (P(\bar{U}|d) p(z|d, \bar{U})) dz \quad (7.28)$$

$$= -P(U|d) \log P(U|d) - P(\bar{U}|d) \log P(\bar{U}|d)$$

$$- P(U|d) \int_U p(z|d, U) \log p(z|d, U) dz$$

$$- P(\bar{U}|d) \int_{\bar{U}} p(z|d, \bar{U}) \log p(z|d, \bar{U}) dz \quad (7.29)$$

$$= H[p_C(x|d)] + P(U|d) H[p(z|d, U)]$$

$$+ P(\bar{U}|d) H[p(z|d, \bar{U})] . \quad (7.30)$$

□

Mit Lemma 7.3 lässt sich nun folgender Satz beweisen, welcher eine explizite Formel für die (Maximum-Entropie) A-Posteriori-Verteilung  $p_{\text{ME}}(z|d)$  gemäß Gleichung (7.23) angibt:

**Satz 7.4.** *Das Optimierungsproblem gemäß Gleichung (7.23) besitzt bezogen auf den Anteil  $U \subset \mathcal{Z}$ , d. h. für  $z \in U$ , folgende Lösung:*

$$p_{\text{ME}}(z|d) = \begin{cases} \pi p(z|d, U) , & \text{falls } \beta \leq \pi , \\ \beta p(z|d, U) , & \text{falls } \pi < \beta . \end{cases} \quad (7.31)$$

Hierbei bezeichnet  $\beta$  die in Satz 7.1 hergeleitete Schranke für die globale A-Posteriori-Wahrscheinlichkeit  $P(U|d)$  des lokalen Kontexts  $U$ . Weiter ist  $\pi$  gegeben durch

$$\pi = \frac{2^{\text{H}[p(z|d,U)]}}{2^{\text{H}[p(z|d,U)]} + 2^{\text{H}[p_{\text{ME}}(z|d,\bar{U})]}} . \quad (7.32)$$

Bei der Angabe in Gleichung (7.32) ist angenommen, dass die Logarithmen in den Entropie-Termen zur Basis 2 gebildet wurden. Der Ausdruck  $\text{H}[p_{\text{ME}}(z|d,\bar{U})]$  in Gleichung (7.32) notiert die Entropie einer nichtinformativen Wahrscheinlichkeitsverteilung über  $\bar{U}$ .

Hinsichtlich der Fortsetzung des in Gleichung (7.31) für  $z \in U$  angegebenen Anteils von  $p_{\text{ME}}(z|d)$  auf  $\bar{U}$  gilt: Da keine Information über die Struktur der globalen A-Posteriori-Verteilung innerhalb von  $\bar{U}$  vorliegt, muss  $p_{\text{ME}}(z|d)$  notwendigerweise nichtinformativ in Bezug auf  $\bar{U}$  sein.

*Beweis.* Die Maximierung der Entropie  $\text{H}[p(z|d)]$  von  $p(z|d)$  ist offensichtlich äquivalent zur Minimierung des negativen Gegenstücks  $-\text{H}[p(z|d)]$ .

Unter Verwendung der Abkürzung  $u_p := P(U|d)$  liefert Lemma 7.3

$$\begin{aligned} -\text{H}[p(z|d)] &= u_p \log u_p + (1 - u_p) \log(1 - u_p) \\ &\quad - u_p \text{H}[p(z|d,U)] - (1 - u_p) \text{H}[p(z|d,\bar{U})] . \end{aligned} \quad (7.33)$$

Schränkt man sich bei der lokalen Bayes'schen Fusion auf den lokalen Kontext  $U$  ein, so bedeutet dies, dass das Komplement  $\bar{U}$  des lokalen Kontexts  $U$  vollständig ignoriert wird. Es liegt im lokalen Bayes'schen Modell also absolut keine Information im Hinblick auf  $\bar{U}$  vor. Dies spiegelt sich auch gerade in Gleichung (7.7) und Gleichung (7.20) bzw. Gleichung (7.2) und Gleichung (7.8) wieder.

Gleichung (7.33) ergibt deshalb sofort, dass der Ausdruck  $-\text{H}[p(z|d)]$  minimal wird, wenn man  $p(z|d,\bar{U})$  so wählt, dass  $\text{H}[p(z|d,\bar{U})]$  maximal wird. Man muss also die Festlegung  $p(z|d,\bar{U}) = p_{\text{ME}}(z|d,\bar{U})$  treffen, wobei  $p_{\text{ME}}(z|d,\bar{U})$  eine nichtinformativ Wahrscheinlichkeitsverteilung über  $\bar{U}$  notiert.

Es gilt damit also, dass die Funktion

$$f(u_p) := u_p \log u_p + (1 - u_p) \log(1 - u_p) - u_p H[p(z|d,U)] - (1 - u_p) H[p_{\text{ME}}(z|d,\bar{U})] \quad (7.34)$$

unter der Bedingung  $\beta \leq u_p$ , welche mit Gleichung (7.8) korrespondiert, zu minimieren ist.

Durch Bildung der zweiten Ableitung von  $f(u_p)$  erkennt man, dass  $f(u_p)$  konvex ist für  $u_p \in (0,1)$ . Es liegt also ein konvexes Optimierungsproblem vor.

Die Lösung eines derartigen konvexen Optimierungsproblems basiert auf der Betrachtung der entsprechenden Lagrange-Funktion [Boy04], welche im vorliegenden Fall die folgende Form besitzt:

$$L(u_p, \lambda) = u_p \log u_p + (1 - u_p) \log(1 - u_p) - u_p H[p(z|d,U)] - (1 - u_p) H[p_{\text{ME}}(z|d,\bar{U})] + \lambda(\beta - u_p). \quad (7.35)$$

Die Karush-Kuhn-Tucker (KKT) Bedingungen [Boy04], welche erfüllt sein müssen, sind gegeben durch:

$$\begin{aligned} & \text{(a) } \beta \leq u_p, \quad \text{(b) } \lambda \geq 0, \quad \text{(c) } \lambda = 0 \vee \beta = u_p, \\ & \text{(d) } \log\left(\frac{u_p}{1 - u_p}\right) - H[p(z|d,U)] + H[p_{\text{ME}}(z|d,\bar{U})] - \lambda = 0. \end{aligned}$$

An dieser Stelle ist nun eine Fallunterscheidung hinsichtlich der Bedingung (c) notwendig:

Im Fall, dass  $\lambda = 0$  gilt, liefert Bedingung (d)

$$\frac{u_p}{1 - u_p} = 2^{H[p(z|d,U)] - H[p_{\text{ME}}(z|d,\bar{U})]}, \quad (7.36)$$

was äquivalent ist zu

$$u_p = \frac{2^{H[p(z|d,U)]}}{2^{H[p(z|d,U)]} + 2^{H[p_{ME}(z|d,\bar{U})]}} \cdot \quad (7.37)$$

Gemäß Bedingung (a) muss dabei  $\beta \leq u_p$  gelten.

Im anderen Fall, dass  $\beta = u_p$  gilt, liefert Bedingung (d)

$$\lambda = \log\left(\frac{\beta}{1-\beta}\right) - H[p(z|d,U)] + H[p_{ME}(z|d,\bar{U})] \cdot \quad (7.38)$$

Gemäß Bedingung (b) muss dabei gelten

$$\frac{\beta}{1-\beta} \geq 2^{H[p(z|d,U)] - H[p_{ME}(z|d,\bar{U})]} \cdot \quad (7.39)$$

Dies ist äquivalent zu

$$\beta \geq \frac{2^{H[p(z|d,U)]}}{2^{H[p(z|d,U)]} + 2^{H[p_{ME}(z|d,\bar{U})]}} \cdot \quad (7.40)$$

□

Wie in Abschnitt 3.5.1 dargestellt wurde, liefert das Prinzip der Maximalen Entropie aus der Menge aller Wahrscheinlichkeitsverteilungen, welche konsistent mit der vorliegenden Information sind, diejenige Wahrscheinlichkeitsverteilung, durch welche die Unsicherheit bezüglich des „wahren“ Werts von  $z$  nur in dem Umfang verringert wird, in dem dies notwendig ist, um die Konsistenz mit der vorliegenden Information sicherzustellen. Wirft man einen genaueren Blick auf die in Gleichung (7.31) angegebene Formel für  $p_{ME}(z|d)$ , so wird dieser Sachverhalt für den im vorliegenden Abschnitt betrachteten spezifischen Anwendungsfall auch offensichtlich.

Gilt  $\beta \leq \pi$ , so weist das Prinzip der Maximalen Entropie dem lokalen Kontext  $U$  einen umso größeren Anteil an globaler A-Posteriori-Wahrscheinlichkeit zu, je unsicherer die lokale A-Posteriori-Verteilung  $p(z|d,U)$  ist. Umgekehrt gilt, dass das Prinzip der Maximalen Entropie den lokalen Kontext  $U$  (a



posteriori) als global umso weniger wahrscheinlich annimmt, je informativer  $p(z|d,U)$  ist.

Dieser Sachverhalt entspricht der generellen Beobachtung, dass das Prinzip der Maximalen Entropie gewissermaßen den Worst Case annimmt (vgl. hierzu auch [Wal91]). Die Tatsache, dass die lokale A-Posteriori-Verteilung  $p(z|d,U)$  recht konzentriert ist, muss wie bereits dargestellt global gesehen nicht der Realität entsprechen (vgl. Abschnitt 5.2.3.3), denn durch den Übergang von der globalen A-Posteriori-Verteilung  $p(z|d)$  zur lokalen A-Posteriori-Verteilung  $p(z|d,U)$  werden die Wahrscheinlichkeiten von Ereignissen  $A \subset U$  verzerrt (genauer: als zu hoch eingeschätzt). Wie ebenfalls dargestellt ist der Grad dieser Verzerrung dabei umso höher, je geringer die globale A-Posteriori-Wahrscheinlichkeit  $P(U|d)$  des lokalen Kontexts  $U$  tatsächlich ist (vgl. auch Gleichung (7.3)). Das Prinzip der Maximalen Entropie geht davon aus, dass das Vorliegen einer recht konzentrierten lokalen A-Posteriori-Verteilung  $p(z|d,U)$  mit geringer Entropie auf eine entsprechend hohe Verzerrung zurückzuführen sein könnte und versucht diese dann auszugleichen, indem es die lokale A-Posteriori-Verteilung  $p(z|d,U)$  bei der Überführung in die (Maximum-Entropie) A-Posteriori-Verteilung  $p_{\text{ME}}(z|d)$  entsprechend umfangreicher abflacht. Letzteres korrespondiert gerade damit, dass die mittels  $p_{\text{ME}}(z|d)$  berechnete A-Posteriori-Wahrscheinlichkeit des lokalen Kontexts  $U$  umso geringer ausfällt, also umso mehr Wahrscheinlichkeitsmasse auf den bei der lokalen Bayes'schen Fusion nicht betrachteten Anteil  $\bar{U}$  von  $Z$  gelegt wird. Der beschriebene Sachverhalt wird nachfolgend auch in den in Abschnitt 7.3.2 dargelegten Berechnungsbeispielen noch genauer veranschaulicht.

Ist man sich der beschriebenen Arbeitsweise des Prinzips der Maximalen Entropie bewusst und interpretiert die gelieferten Ergebnisse auch entsprechend, so kann seine Anwendung durchaus eine vielversprechende Möglichkeit sein, um die lokale A-Posteriori-Verteilung  $p(z|d,U)$  nachgelagert an die eigentliche lokale Bayes'sche Fusion in eine globale A-Posteriori-Verteilung auf  $Z$  zu überführen. Seine Arbeitsweise kann dazu beitragen, dass die Ergebnisse der lokalen Bayes'schen Fusion von einem vorsichtigen Standpunkt

aus interpretiert werden und so letztlich Schlussfolgerungen, welche global gesehen nicht der Realität entsprechen, vermieden werden.

Voraussetzung dabei ist natürlich, dass die untere Schranke  $\beta$  für die globale A-Posteriori-Wahrscheinlichkeit  $P(U|d)$  des lokalen Kontexts  $U$  auch entsprechend aussagekräftig bestimmt werden muss. Eine wenig aussagekräftige Schranke kann letztlich dazu führen, dass bei Anwendung des Prinzips der Maximalen Entropie ein extrem hoher Anteil an Wahrscheinlichkeitsmasse auf  $\bar{U}$  gelegt wird, mit der Konsequenz, dass die (Maximum-Entropie) A-Posteriori-Verteilung  $p_{\text{ME}}(z|d)$  kaum noch zur Ableitung nützlicher Schlussfolgerungen verwendet werden kann.

Es sei angemerkt, dass der Sachverhalt, dass das Prinzip der Maximalen Entropie umso mehr Wahrscheinlichkeitsmasse auf  $\bar{U}$  umverteilt, je konzentrierter  $p(z|d, U)$  ist, speziell für den Fall, dass  $|U| = |\bar{U}| = 2$  gilt, in anderem Zusammenhang bereits 1981 von Van Fraassen in [Van81] beschrieben und untersucht wurde. Konkret geschah dies in Bezug auf das sog. Judy Benjamin Problem, welches auch in neueren Veröffentlichungen noch aufgegriffen wird (siehe z. B. [Bov09, Dou11]). Van Fraassen bezieht sich nicht direkt auf das Prinzip der Maximalen Entropie, sondern auf das Prinzip der Minimalen Information, welches wie in Abschnitt 3.5.1 dargestellt als Verallgemeinerung des Prinzips der Maximalen Entropie aufgefasst werden kann.

## 7.3.2 Berechnungsbeispiele

### 7.3.2.1 Wahrscheinlichkeitsintervalle, (Maximum-Entropie) A-Posteriori-Verteilung

Zur Verdeutlichung der Ergebnisse im Hinblick auf das in Abschnitt 7.2 adressierte Wahrscheinlichkeitsintervall-Schema für die globale A-Posteriori-Wahrscheinlichkeit von Ereignissen und damit zusammenhängend auch der Ergebnisse im Hinblick auf die durch Anwendung des Prinzips der Maximalen Entropie gemäß Abschnitt 7.3.1 resultierende eindeutige globale (Maximum-Entropie) A-Posteriori-Verteilung wird im vorliegenden Abschnitt der Übersichtlichkeit halber wieder ein einfaches und illustratives

Beispiel betrachtet. Dieses Beispiel wurde in der nachfolgenden Form zuvor bereits in [San10b] veröffentlicht.

Die Aufgabe bestehe darin, den Typ eines Fahrzeugs festzustellen, d. h. es gilt  $Z = Z_1$  wobei  $Z_1$  die Menge der möglichen Fahrzeugtypen notiert. Es liege eine Informationsquelle vor, welche als Informationsbeitrag  $d_1$  gerade ein Grauwertbild des Fahrzeugs liefert. Die Menge der möglichen Fahrzeugtypen wird exemplarisch als  $Z = \{A \text{ (Audi)}, D \text{ (DKW)}, F \text{ (Ford)}, M \text{ (Mercedes)}, O \text{ (Opel)}\}$  festgelegt. A priori sei das Vorhandensein dieser Typen gleich plausibel, so dass als A-Priori-Verteilung  $p(z)$  die Gleichverteilung auf  $Z$  zu setzen ist. Zur Gewinnung der Likelihood-Funktion  $l(d_1|z)$  wird (rein exemplarisch) eine an die möglichen Typen gemäß  $Z$  angepasste Matched-Filter Bank, welche Grauwertbilder von gleicher Größe wie  $d_1$  als Templates enthält, verwendet und es wird (wie im Beispiel in Abschnitt 5.3) die normalisierte Kreuzkorrelation berechnet. Zur Bestimmung des lokalen Kontexts  $U$  wird eine Schranke an die Likelihood-Funktion  $l(d_1|z)$  angelegt. Ein möglicher Wert  $z$  aus  $Z$  wird lokal ignoriert, d. h. nicht in den lokalen Kontext  $U$  aufgenommen, wenn gilt

$$l(d_1|z) \leq \delta \quad \text{mit } \delta = \epsilon \cdot \max_{z \in Z} l(d_1|z). \quad (7.41)$$

In Abb. 7.1 sind in Form der dargestellten Balken die resultierenden Wahrscheinlichkeitsintervalle für die globalen A-Posteriori-Wahrscheinlichkeiten der verschiedenen Fahrzeugtypen in Abhängigkeit von unterschiedlichen Festlegungen für  $\epsilon$  aufgezeigt. Der tatsächliche Wert der (im lokalen Bayes'schen Modell unbekannt) A-Posteriori-Wahrscheinlichkeit befindet sich jeweils an der Stelle, an der die Balken ihre Farbe wechseln.

Für  $\epsilon = 0,0227$  und  $\epsilon = 0,1$  ergibt sich als lokaler Kontext  $U = \{A, F, O\}$ . Für  $\epsilon = 0,1129$  resultiert  $U = \{A, O\}$ , d. h. der mögliche Typ F wird zusätzlich ignoriert im lokalen Bayes'schen Modell. Für  $\epsilon = 0,7$  resultiert  $U = \{O\}$ , d. h. der Typ O wird im lokalen Bayes'schen Modell als sicher vorliegend erachtet.

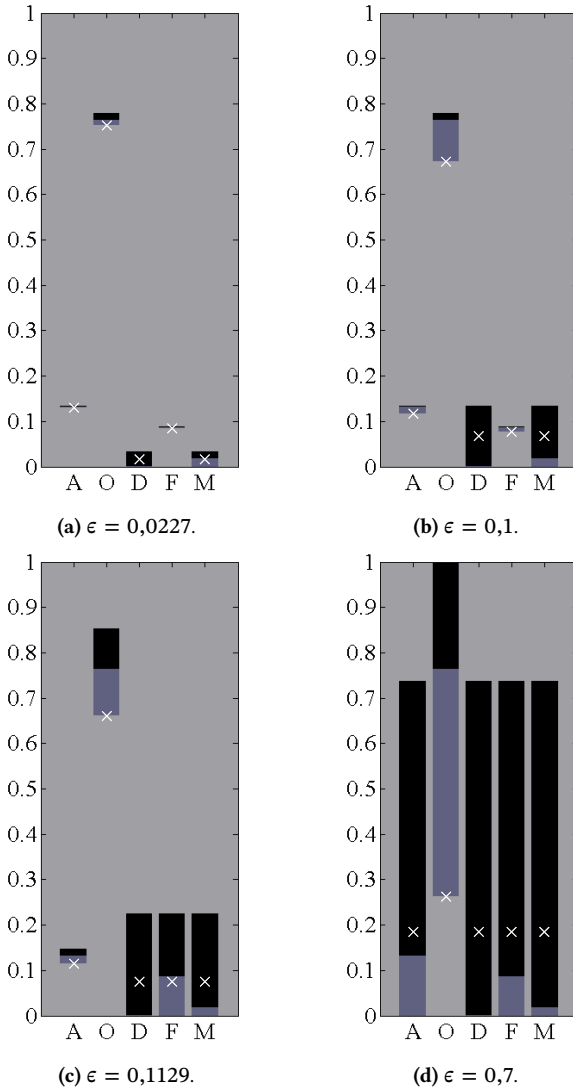


Abbildung 7.1: Wahrscheinlichkeitsintervalle und Maximum-Entropie Verteilung.

In allen vier hier exemplarisch dargestellten Fällen ist es möglich, den tatsächlich vorliegenden Typ O als den (eindeutigen) globalen Maximum-A-Posteriori-Schätzwert zu identifizieren. Es gilt nämlich, dass der Typ O in allen Fällen im lokalen Kontext  $U$  enthalten ist und dass im Hinblick auf die oberen Grenzen der Wahrscheinlichkeitsintervalle die Beziehung  $r(O) \geq r(z)$  für alle  $z \in U \setminus \{O\}$  besteht. Man beachte, dass die Frage, ob sich die entsprechenden Wahrscheinlichkeitsintervalle überlappen, nicht entscheidend für diese Schlussfolgerung ist.

Die Länge  $d(z)$  der resultierenden Wahrscheinlichkeitsintervalle wird nicht nur von der A-Posteriori-Wahrscheinlichkeit  $P(U|d)$  des lokalen Kontexts  $U$  und den Werten der A-Posteriori-Verteilung  $p(z|d)$  für  $z \in U$  beeinflusst, sondern auch von der an die Likelihood-Funktion angelegten Schranke. Diesen Sachverhalt erkennt man recht gut, wenn man die für  $\epsilon = 0,0227$  und  $\epsilon = 0,1$  resultierenden Wahrscheinlichkeitsintervalle, für die die vorgenannten beiden Größen identisch sind, vergleicht. Der Grund hierfür ist, dass die an die Likelihood-Funktion angelegte Schranke die Abschätzung für  $\beta$  und damit letztlich auch die Wahrscheinlichkeitsintervalle an sich beeinflusst (vgl. Gleichung (7.8) und Gleichung (7.22)).

Der Wert  $\epsilon = 0,0227$  wurde experimentell so bestimmt, dass er eine recht scharfe Schranke dafür, dass gerade die Typen D und M im lokalen Bayes'schen Modell ignoriert werden, liefert. Man sieht, dass auch in diesem Fall die tatsächlichen Werte der A-Posteriori-Verteilung entfernt von den entsprechenden Intervallschranken  $l(z)$  und  $r(z)$  liegen. Der Grund hierfür ist, dass in diesem Fall zwar  $l(d_1|M) \approx \epsilon \cdot \max_{z \in Z} l(d_1|z)$  gilt, aber weiter auch  $l(d_1|D) \ll \epsilon \cdot \max_{z \in Z} l(d_1|z)$ .

Der Wert  $\epsilon = 0,1129$  wurde in analoger Weise experimentell so bestimmt, dass er eine recht scharfe Schranke dafür, dass der Typ F zusätzlich im lokalen Bayes'schen Modell ignoriert wird, liefert. Da die resultierende Schranke aber wiederum (noch) weniger scharf im Hinblick auf  $l(d_1|D)$  und auch nicht scharf im Hinblick auf  $l(d_1|M)$  ist, nimmt die Länge der entsprechenden Wahrscheinlichkeitsintervalle weiter zu. Dieser Effekt verstärkt sich für die Wahl  $\epsilon = 0.7$  weiter.

Für das vorliegende Beispiel wurde weiter die durch Anwendung des Prinzips der Maximalen Entropie resultierende (Maximum-Entropie) A-Posteriori-Verteilung  $p_{\text{ME}}(z|d)$  bestimmt. Diese ist in Abb. 7.1 jeweils durch die Kreuze in den Wahrscheinlichkeitsintervallen dargestellt. Die (Maximum-Entropie) A-Posteriori-Verteilung korrespondiert für  $z \in U$  hier in allen vier Fällen gerade mit der unteren Schranke  $l(z)$  der Wahrscheinlichkeitsintervalle. Das Prinzip der Maximalen Entropie schätzt die A-Posteriori-Wahrscheinlichkeit  $P(U|d)$  des lokalen Kontexts  $U$  also in allen vier Fällen so gering, wie es in Anbetracht der zu berücksichtigenden Schranke  $\beta$  gemäß Gleichung (7.8) nur möglich ist, ein.

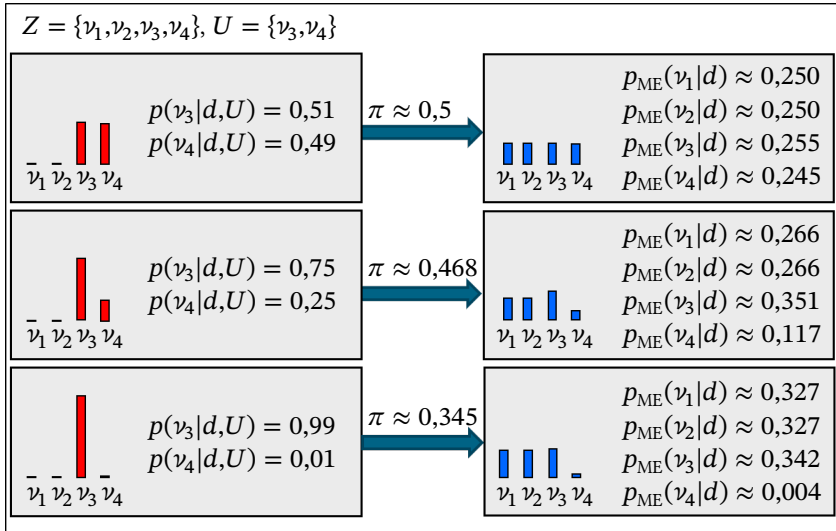
### 7.3.2.2 Arbeitsweise des Prinzips der Maximalen Entropie

Im nachfolgenden Beispiel wird die Arbeitsweise des Prinzips der Maximalen Entropie noch genauer verdeutlicht. Dies geschieht anhand des Falls, dass anders als in Gleichung (7.31) keine untere Schranke  $\beta$  für die globale A-Posteriori-Wahrscheinlichkeit  $P(U|d)$  des lokalen Kontexts  $U$  zu berücksichtigen ist, die A-Posteriori-Wahrscheinlichkeit des lokalen Kontexts  $U$  vom Prinzip der Maximalen Entropie also beliebig gering eingestuft werden kann.

In Abb. 7.2 ist anhand des Falls  $n = 1$  und  $Z = \{\nu_1, \nu_2, \nu_3, \nu_4\}$  veranschaulicht, wie die Struktur der lokalen A-Posteriori-Verteilung  $p(z|d, U)$  die resultierende (Maximum-Entropie) A-Posteriori-Verteilung  $p_{\text{ME}}(z|d)$  grundsätzlich gesehen beeinflussen kann. Es wird dabei angenommen, dass der lokale Kontext durch  $U = \{\nu_3, \nu_4\}$  gegeben ist.

Unter diesen Annahmen werden drei mögliche, unterschiedlich stark konzentrierte Ausprägungen der lokalen A-Posteriori-Verteilung  $p(\nu_i|d, U)$ ,  $i \in \{3, 4\}$ , betrachtet, welche auf der linken Seite von Abb. 7.2 angeführt sind. Auf der rechten Seite von Abb. 7.2 ist die jeweils durch Anwendung des Prinzips der Maximalen Entropie gemäß Satz 7.4 resultierende globale (Maximum-Entropie) A-Posteriori-Verteilung  $p_{\text{ME}}(z|d)$  angeführt, jeweils unter der Annahme, dass die Schranke  $\beta$  nicht greift. D. h. es gilt in allen Fällen  $p_{\text{ME}}(z|d) = \pi p(z|d, U)$  mit  $\pi$  gemäß Gleichung (7.32). Wie aus Abb. 7.2 gut erkennbar ist,

fällt  $p_{\text{ME}}(z|d)$  im Hinblick auf den lokalen Kontext  $U$  umso niedriger gegenüber der lokalen A-Posteriori-Verteilung  $p(z|d,U)$  aus, je geringer die Entropie von  $p(z|d,U)$  ist.



**Abbildung 7.2:** Illustratives Beispiel zur Verdeutlichung der Arbeitsweise des Prinzips der Maximalen Entropie im Hinblick auf die Bestimmung der globalen (Maximum-Entropie) A-Posteriori-Verteilung, wenn die Schranke  $\beta$  nicht greift.

Ist  $Z$  diskret, so wird die (Shannon) Entropie  $H[q(z)]$  einer Wahrscheinlichkeitsverteilung  $q(z)$  über  $Z$  genau dann maximal, wenn  $q(z)$  die Gleichverteilung auf  $Z$  ist, und es gilt  $H[q(z)] = \log |Z|$ . Bezeichnet  $q(z)$  die Gleichverteilung einer kontinuierlichen Zufallsgröße  $z$  über einem entsprechend gewählten Träger  $Z$ , so gilt für ihre (Boltzmann) Entropie  $H[q(z)] = \log \left( \int_Z dz \right)$ . Der Wert der Entropie der Maximum-Entropie-Verteilung (insbesondere) hängt also grob gesagt vom Umfang von  $Z$  ab. Vgl. hierzu auch [Bey99].

Betrachtet man Gleichung (7.31) und insbesondere Gleichung (7.32), so erkennt man sofort, dass die aus Satz 7.4 resultierende (Maximum-Entropie)

A-Posteriori-Verteilung  $p_{ME}(z|d)$  also letztlich auch von der Struktur von  $Z$  und insbesondere auch  $\bar{U}$  abhängt. Die nachfolgende Abbildung illustriert diesen Sachverhalt exemplarisch. Das in Abb. 7.3 angegebene Beispiel stellt das Analogon zum in Abb. 7.2 angebenen Beispiel für den Fall, dass  $Z$  zusätzliche Elemente enthält, welche  $\bar{U}$  zugeordnet sind, dar. Es wird wieder exemplarisch angenommen, dass die Schranke  $\beta$  nicht greift, d. h. dass also gilt  $p_{ME}(z|d) = \pi p(z|d, U)$  mit  $\pi$  gemäß Gleichung (7.32). Man erkennt gut, dass die globale (Maximum-Entropie) A-Posteriori-Verteilung  $p_{ME}(z|d)$  weniger Wahrscheinlichkeitsmasse über dem lokalen Kontext  $U$  besitzt, als es in der Situation von Abb. 7.2 der Fall ist. Die Umverteilung von Wahrscheinlichkeitsmasse auf  $\bar{U}$  fällt also noch stärker aus.

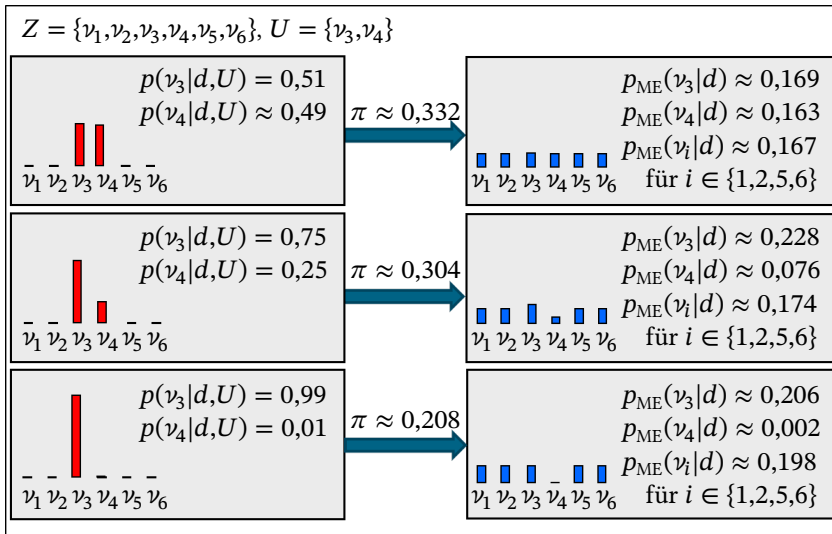


Abbildung 7.3: Analogon zu Abb. 7.2 für den Fall, dass  $Z$  zusätzliche Elemente enthält.



## 7.4 Anwendung der Entscheidungstheorie

Die im vorliegenden Abschnitt dargelegten Herleitungen und Ergebnisse wurden samt den hierzu notwendigen Grundlagen zuvor in [San11] und als Anteil von [San10a] veröffentlicht.

### 7.4.1 Entscheidungen unter Risiko, partieller Information und linearer partieller Information

Es sei  $A$  eine Menge an möglichen Aktionen und es sei  $u(a,z)$  eine Nutzenfunktion über  $A \times Z$ , welche jeweils den Nutzen einer Aktion  $a \in A$  in Abhängigkeit vom Wert von  $z \in Z$  angibt. Die Aufgabe bestehe darin, auf Basis des a posteriori vorliegenden Informationsstands bezüglich  $z$  bestmöglich eine Aktion aus  $A$  auszuwählen.

Wäre die globale A-Posteriori-Verteilung  $p(z|d)$  vollständig bekannt, so läge ein Entscheidungsproblem unter Risiko vor (vgl. z. B. [Rüg99]). Gemäß dem sog. Prinzip der Maximierung des erwarteten Nutzens (vgl. z. B. [Ber04]) sollte ein rationaler Entscheider in dieser Situation eine Aktion  $a_{\text{opt}} \in A$  auswählen, welche den a posteriori erwarteten Nutzen maximiert, d. h.

$$a_{\text{opt}} = \arg \max_{a \in A} E_{p(z|d)}[u(a,z)] \quad (7.42)$$

mit

$$E_{p(z|d)}[u(a,z)] = \int_Z u(a,z)p(z|d) dz . \quad (7.43)$$

Im Folgenden wird (sinnvollerweise) angenommen, dass die verwendeten Nutzenfunktionen beschränkt sind und dass jeweils eine Aktion, welche optimal im Sinne von Gleichung (7.42) ist, existiert.

Die (globale) Entscheidungsfindung auf Basis eines lokalen Bayes'schen Modells stellt ein Entscheidungsproblem unter partieller Information (vgl. z. B. [Pre02]) dar. In diesem Fall ist nur bekannt, dass  $p(z|d)$  in der Menge

aller Wahrscheinlichkeitsverteilungen über  $Z$ , welche konsistent mit Gleichung (7.7) und Gleichung (7.20) (oder äquivalent dazu mit Gleichung (7.2) und Gleichung (7.8)) sind, enthalten ist. Als Konsequenz hieraus ergibt sich, dass für den erwarteten Nutzen einer Aktion  $a \in A$  in der Regel jeweils eine Menge an möglichen Werten resultiert, welche bei der (globalen) Entscheidungsfindung berücksichtigt werden muss.

In der Fachliteratur finden sich unterschiedliche Ansätze im Hinblick die Erarbeitung von Entscheidungskriterien unter partieller Information. Konkret angeführt hierzu seien die Referenzen [Kof76, Pre02, Utk05, Wal91]. Manche der vorhandenen Ansätze sind, ggf. nach geringfügigen Anpassungen, in der Lage, auch mit kontinuierlichen Wahrscheinlichkeitsverteilungen umzugehen. Siehe hierzu z. B. [Aug98, Wal91, Wei01]. Effiziente Modelle und Algorithmen sind letztlich aber vor allem für den Fall von diskreten Wahrscheinlichkeitsverteilungen über einem endlichen Grundraum vorhanden. Die Referenzen [Obe07] und [Utk09] stellen Beispiele für Arbeiten dar, welche die Adaption entsprechender Ansätze auf den allgemeinen Fall adressieren.

Nachfolgend wird das bei der lokalen Bayes'schen Fusion vorliegende Entscheidungsproblem unter partieller Information in eine spezielle Ausprägung eines derartigen Entscheidungsproblems, genauer auf ein Entscheidungsproblem unter linearer partieller Information (LPI) überführt. Die Grundlagen zu der mit derartigen Entscheidungsproblemen korrespondierenden Theorie linearer partieller Information wurden in [Kof76] umfassend eingeführt. Diese Theorie besitzt eine recht hohe Ausdruckstärke und ermöglicht dabei gleichzeitig eine vom theoretischen Standpunkt aus recht unkomplizierte und auch methodisch exakte Handhabung der von ihr abgedeckten Klasse von Entscheidungsproblemen. Eine wichtige Basis für die vorliegende Arbeit stellt dabei weiter auch der in [Pre02] veröffentlichte Ansatz des sog. Lazy Decision Making dar. Der diesem Ansatz zugrundeliegende Gedanke besteht darin, dem Entscheider die Möglichkeit zu bieten, die zur Formulierung des Entscheidungsproblems notwendige Information zuerst einmal nur grob zu spezifizieren und diese dann im Rahmen der Entscheidungsfindung zielgerichtet weiter zu präzisieren, falls sie nicht genügt, um eine ausreichend gute Aktion zu identifizieren. Es ist gut erkennbar, dass dieser Gedanke, speziell für den Kon-

text der Entscheidungsfindung, konzeptionell gesehen Parallelen zu dem der lokalen Bayes'schen Fusion zugrundeliegenden Gedanken aufweist.

Der Festlegung in [Kof76] folgend wird lineare partielle Information in der vorliegenden Arbeit wie folgt definiert:

**Definition 7.5.** *Partielle Information bezüglich einer Wahrscheinlichkeitsverteilung über einer endlichen Menge der Kardinalität  $k \in \mathbb{N}$  stellt lineare partielle Information dar, wenn das entsprechende Teilgebiet  $W$  des  $k$ -dimensionalen Verteilungssimplexes  $[0,1]^k$  wie folgt durch ein System von Ungleichungen festgelegt werden kann: Es existieren  $B \in \mathbb{R}^{l \times k}$  und  $c \in \mathbb{R}^l$  mit  $l \in \mathbb{N}$  so dass gilt*

$$W = \left\{ p = (p^1, \dots, p^k)^T \in \mathbb{R}^k \mid \sum_{i=1}^k p^i = 1, 0 \leq p^i \text{ für } 1 \leq i \leq k, Bp \geq c \right\}. \quad (7.44)$$

Wie gut erkennbar ist, legt das Ungleichungssystem  $W$  in Definition 7.5 gerade eine Menge von Wahrscheinlichkeitsverteilungen über der zugrundeliegenden  $k$ -dimensionalen Menge von atomaren Ereignissen fest. Diese Wahrscheinlichkeitsverteilungen sind dabei in Form des Vektors  $p$  beschrieben, wobei die  $i$ -te Komponente  $p^i$  von  $p$  gerade die Wahrscheinlichkeit des  $i$ -ten atomaren Elements darstellt,  $i \in \{1, \dots, k\}$ .

Die Rückführung des bei der lokalen Bayes'schen Fusion vorliegenden Entscheidungsproblems unter partieller Information auf ein Entscheidungsproblem unter linearer partieller Information basiert auf der (manchmal, aber nicht immer ohnehin schon geltenden) Annahme, dass das Komplement  $\bar{U}$  des lokalen Kontexts  $U$  eine (ggf. auch umfangreiche) endliche Menge darstellt oder in eine solche überführt werden kann. Es wird also angenommen, dass  $\bar{U} = \{\nu_1, \dots, \nu_M\}$  für ein  $M \in \mathbb{N}$  gilt.

Das folgende Lemma zeigt auf, dass und wie genau sich die (globale) Entscheidungsfindung auf Basis eines lokalen Bayes'schen Modells dann auf ein Entscheidungsproblem unter linearer partieller Information zurückführen lässt:

**Lemma 7.6.** *Hinsichtlich der (globalen) A-Posteriori-Verteilung  $p(z|d)$  sei gerade bekannt, dass diese in der Menge  $\mathfrak{P}_z$  aller Wahrscheinlichkeitsverteilungen auf  $Z$ , welche konsistent mit Gleichung (7.2) und Gleichung (7.8) sind, enthalten ist. Dann ist in einem (globalen) Entscheidungsproblem mit Nutzenfunktion  $u(a,z)$  für jede Aktion  $a \in A$  die Menge der möglichen Werte ihres (globalen) a posteriori erwarteten Nutzens*

$$\left\{ E_{p(z|d)}[u(a,z)] \mid p(z|d) \in \mathfrak{P}_z \right\} \quad (7.45)$$

*identisch zur Menge der möglichen Werte für den erwarteten Nutzen der Aktion  $a$  in einem Entscheidungsproblem unter linearer partieller Information. Das korrespondierende Entscheidungsproblem unter linearer partieller Information ist ein Entscheidungsproblem über*

$$Z_{\text{LF}} := \{\nu_1, \dots, \nu_M, \nu_{M+1}\} \quad \text{mit } \bar{U} = \{\nu_1, \dots, \nu_M\} \text{ und } \nu_{M+1} := U \quad (7.46)$$

*und das zugehörige Teilgebiet von  $[0,1]^{M+1}$  ist gegeben durch*

$$W_{\text{LF}} := \left\{ p = (p^1, \dots, p^{M+1})^\top \mid \sum_{i=1}^{M+1} p^i = 1, 0 \leq p^i \text{ für } 1 \leq i \leq M, \beta \leq p^{M+1} \right\}. \quad (7.47)$$

*Weiter ist die Nutzenfunktion in diesem Entscheidungsproblem unter linearer partieller Information gegeben durch*

$$u_{\text{LF}}(a, \nu_i) := \begin{cases} u(a, \nu_i), & \text{für } i \in \{1, \dots, M\}, \\ E_{p(z|d, U)}[u(a, z)], & \text{für } i = M + 1. \end{cases} \quad (7.48)$$

*Beweis.* Es gelte

$$p^i := p(\nu_i|d), i \in \{1, \dots, M\}, \quad \text{und} \quad p^{M+1} := P(U|d). \quad (7.49)$$

$W_{\text{LF}}$  enthält damit dann gerade alle möglichen (globalen) A-Posteriori-Verteilungen über  $Z$ , die die Bedingung  $P(U|d) \geq \beta$  gemäß Gleichung (7.8) erfüllen. Man erkennt recht leicht, dass diese Bedingung im Hinblick auf

$p$  lineare partielle Information darstellt: Gleichung (7.47) ist identisch zu Gleichung (7.44), wenn man die redundante Bedingung  $0 \leq p^{M+1}$  zufügt und wenn man  $B = (0, \dots, 0, 1)$  sowie  $c = \beta$  setzt.

Aufgrund des Zusammenhangs zwischen der globalen A-Posteriori-Verteilung  $p(z|d)$  und der lokalen A-Posteriori-Verteilung  $p(z|d, U)$  gemäß Gleichung (7.2) gilt allgemein für jede Aktion  $a \in A$ :

$$E_{p(z|d)}[u(a, z)] = \int_{\bar{U}} u(a, z) p(z|d) dz + P(U|d) E_{p(z|d, U)}[u(a, z)]. \quad (7.50)$$

Mit der in Gleichung (7.46) gemachten Festlegung ist  $E_{p(z|d)}[u(a, z)]$  hier gerade identisch ist zu

$$E_p[u_{\text{LF}}(a, \nu_i)] = \sum_{i=1}^M p^i u(a, \nu_i) + p^{M+1} E_{p(z|d, U)}[u(a, z)]. \quad (7.51)$$

Dabei bezeichnet  $E_p[u_{\text{LF}}(a, \nu_i)]$  den erwarteten Nutzen der Aktion  $a \in A$  bezüglich der Nutzenfunktion  $u_{\text{LF}}(a, \nu_i)$ , der beim Zugrundelegen der  $p(z|d)$  gemäß Gleichung (7.50) entsprechenden Wahrscheinlichkeitsverteilung  $p$  aus der Menge  $W_{\text{LF}}$  resultiert. Es ergibt sich damit, dass die in Gleichung (7.48) eingeführte Nutzenfunktion auf  $A \times Z_{\text{LF}}$  wie in Lemma 7.6 formuliert konsistent zur ursprünglichen Nutzenfunktion  $u(a, z)$  auf  $A \times Z$  ist.  $\square$

Partielle Information bezüglich einer Wahrscheinlichkeitsverteilung stellt bekanntermaßen lineare partielle Information gemäß Definition 7.5 dar, wenn  $W$  geometrisch gesehen ein konvexes Polyeder [Fis01] darstellt. Siehe hierzu [Kof76]. Auf Basis des nachfolgenden Lemmas aus der Grundlagenliteratur kann man diese geometrische Betrachtung im Hinblick auf die lineare Optimierung nutzen.

Es bezeichne dabei im Folgenden  $V(W)$  die Menge der Ecken [Fis01] eines konvexen Polyeders  $W \subset \mathbb{R}^k$ .

**Lemma 7.7.** *Es seien  $f : \mathbb{R}^k \rightarrow \mathbb{R}$  eine lineare Funktion und  $W \subset \mathbb{R}^k$  ein konvexes Polyeder für ein  $k \in \mathbb{N}$ . Dann existieren Punkte  $w_{\min} \in V(W)$  und  $w_{\max} \in V(W)$  so, dass gilt  $f(w_{\min}) = \min_{w \in W} f(w)$  und  $f(w_{\max}) = \max_{w \in W} f(w)$ .*

Hinsichtlich eines Beweises von Lemma 7.7 sei auf die entsprechende Grundlagenliteratur, konkret z. B. auf [Fis01] oder [Neu04], verwiesen.

Das nächste Lemma zeigt, dass in der Situation von Lemma 7.6 die Menge der Ecken  $V(W_{\text{LF}})$  des konvexen Polyeders  $W_{\text{LF}}$  eine rechnerisch sehr günstige Struktur aufweist:

**Lemma 7.8.** *In der Situation von Lemma 7.6 ist  $p^\top = (p^1, \dots, p^{M+1})^\top \in \mathbb{R}^{M+1}$  genau dann eine Ecke von  $W_{\text{LF}}$  wenn gilt*

$$p^i = \begin{cases} 0, & \text{für } 1 \leq i \leq M, \\ 1, & \text{für } i = M + 1, \end{cases} \quad (7.52)$$

oder

$$p^i = \begin{cases} 1 - \beta, & \text{für ein } i_0 \in \{1, \dots, M\}, \\ 0, & \text{für } 1 \leq i \leq M \text{ und } i \neq i_0, \\ \beta, & \text{für } i = M + 1. \end{cases} \quad (7.53)$$

*Beweis.* Aus den in der Grundlagenliteratur zu linearer partieller Information dokumentierten Ergebnissen (vgl. [Kof76] oder [Pre02]) ergibt sich sofort, dass  $p^\top = (p^1, \dots, p^{M+1})^\top \in \mathbb{R}^{M+1}$  genau dann eine Ecke von  $W_{\text{LF}}$  ist, wenn  $M + 1$  der  $M + 2$  Bedingungen in Gleichung (7.47) mit Gleichheit erfüllt sind. Im vorliegenden Fall bedeutet dies, dass die Werte von  $M$  Komponenten von  $p$  den in Gleichung (7.47) angegebenen unteren Schranken entsprechen müssen.  $\square$

## 7.4.2 Entscheidungskriterien unter linearer partieller Information

Die Ergebnisse des vorangegangenen Abschnitts werden nun genutzt, um Kriterien im Hinblick auf eine (global) optimale Entscheidungsfindung auf Basis eines lokalen Bayes'schen Modells zu formulieren. Zum einen wird der Frage nachgegangen, unter welchen Umständen man sicher sein kann, dass eine auf Basis eines lokalen Bayes'schen Modells getroffene Entscheidung tatsächlich (global) optimal ist. Zum anderen wird der Frage nachgegangen, um wie viel schlechter (im Form des (globalen) a posteriori erwarteten Nutzens) eine auf Basis eines lokalen Bayes'schen Modells getroffene Entscheidung maximal sein kann im Vergleich zur (global) optimalen Entscheidung. Dieses Vorgehen ist letztlich inspiriert von in [Pre02] dokumentierten Ansätzen.

Als Grundlage für die Ableitung der nachfolgenden Ergebnisse dient die in Lemma 7.6 formulierte Rückführung der (globalen) Entscheidungsfindung auf Basis eines lokalen Bayes'schen Modells auf ein Entscheidungsproblem unter linearer partieller Information.

**Satz 7.9.** *Hinsichtlich der globalen A-Posteriori-Verteilung  $p(z|d)$  sei gerade bekannt, dass diese in der Menge  $\mathfrak{P}_z$  aller Wahrscheinlichkeitsverteilungen auf  $Z$ , welche konsistent mit Gleichung (7.2) und Gleichung (7.8) sind, enthalten ist. Dann stellt mit der Notation aus Lemma 7.6 für jede Aktion  $a \in A$  die Menge der möglichen Werte ihres (globalen) a posteriori erwarteten Nutzens*

$$\left\{ E_{p(z|d)}[u(a,z)] \mid p(z|d) \in \mathfrak{P}_z \right\} \quad (7.54)$$

ein Intervall der folgenden Form dar:

$$I_{W_{LF}}[u_{LF}(a, \nu_i)] := \left[ \underline{E}_{W_{LF}}[u_{LF}(a, \nu_i)], \bar{E}_{W_{LF}}[u_{LF}(a, \nu_i)] \right] \quad (7.55)$$

wobei

$$\underline{E}_{W_{\text{LF}}} [u_{\text{LF}}(a, \nu_i)] = \min \left\{ (1 - \beta) \min_{1 \leq i \leq M} u(a, \nu_i) + \beta E_{p(z|d, U)} [u(a, z)], \right. \\ \left. E_{p(z|d, U)} [u(a, z)] \right\}, \quad (7.56)$$

$$\bar{E}_{W_{\text{LF}}} [u_{\text{LF}}(a, \nu_i)] = \max \left\{ (1 - \beta) \max_{1 \leq i \leq M} u(a, \nu_i) + \beta E_{p(z|d, U)} [u(a, z)], \right. \\ \left. E_{p(z|d, U)} [u(a, z)] \right\}. \quad (7.57)$$

*Beweis.*  $E_p [u_{\text{LF}}(a, \nu_i)]$  stellt gemäß Gleichung (7.51) eine lineare Funktion bezüglich  $p$  dar. Da  $W_{\text{LF}}$  ein konvexes Polyeder darstellt, kann somit Lemma 7.7 angewandt werden. Die Menge der möglichen Werte des (globalen) a posteriori erwarteten Nutzens einer Aktion  $a \in A$ , welche resultiert, wenn  $p$  in  $W_{\text{LF}}$  variiert, stellt somit ein Intervall der folgenden Form dar:

$$\mathbb{I}_{W_{\text{LF}}} [u_{\text{LF}}(a, \nu_i)] = \left[ \min_{p \in V(W_{\text{LF}})} E_p [u_{\text{LF}}(a, \nu_i)], \max_{p \in V(W_{\text{LF}})} E_p [u_{\text{LF}}(a, \nu_i)] \right]. \quad (7.58)$$

Die Menge  $V(W_{\text{LF}})$  der Ecken von  $W_{\text{LF}}$  wurde bereits in Lemma 7.8 bestimmt. Aus Gleichung (7.51) ergibt sich damit weiter, dass gilt

$$\left\{ E_p [u_{\text{LF}}(a, \nu_i)] \mid p \in V(W_{\text{LF}}) \right\} = \left\{ E_{p(z|d, U)} [u(a, z)] \right\} \\ \bigcup \left\{ (1 - \beta) u(a, \nu_i) + \beta E_{p(z|d, U)} [u(a, z)] \mid i = 1, \dots, M \right\}. \quad (7.59)$$

$\nu_{i_0} \in \{\nu_1, \dots, \nu_M\}$  minimiert den Ausdruck in der zweiten Menge auf der rechten Seite von Gleichung (7.59), wenn  $u(a, \nu_{i_0}) = \min_{1 \leq i \leq M} u(a, \nu_i)$  gilt. Hieraus ergibt sich Gleichung (7.56). Eine analoge Betrachtung im Hinblick auf die Maximierung dieses Terms liefert Gleichung (7.57).  $\square$

In der vorliegenden Arbeit wird (anders als in [Pre02]) die Festlegung verwendet, dass eine Aktion  $a^* \in A$  eine andere Aktion  $a^{**} \in A \setminus \{a^*\}$  dominiert, wenn  $a^*$  im Hinblick auf den a posteriori erwarteten Nutzen definitiv min-



destens so gut ist wie  $a^{**}$ . Offensichtlich ist sichergestellt, dass die Aktion  $a^*$  die Aktion  $a^{**}$  (global) dominiert, wenn gilt

$$\bar{E}_{W_{LF}}[u_{LF}(a^{**}, \nu_i)] \leq \underline{E}_{W_{LF}}[u_{LF}(a^*, \nu_i)]. \quad (7.60)$$

Der nächste Satz liefert ein zusätzliches Kriterium, auf Basis dessen auch in (bestimmten) Fällen, in denen das Dominanzkriterium gemäß Gleichung (7.60) nicht erfüllt ist, festgestellt werden kann, dass eine Aktion eine andere Aktion dominiert.

**Satz 7.10.** *Hinsichtlich der globalen A-Posteriori-Verteilung  $p(z|d)$  sei gerade bekannt, dass diese in der Menge  $\mathfrak{P}_z$  aller Wahrscheinlichkeitsverteilungen auf  $Z$ , welche konsistent mit Gleichung (7.2) und Gleichung (7.8) sind, enthalten ist. Ist für zwei (unterschiedliche) Aktionen  $a^*, a^{**} \in A$  die Beziehung*

$$\begin{aligned} & \max \left\{ (1 - \beta) \max_{1 \leq i \leq M} \{u(a^{**}, \nu_i) - u(a^*, \nu_i)\} \right. \\ & \left. + \beta E_{p(z|d, U)}[u(a^{**}, z) - u(a^*, z)], E_{p(z|d, U)}[u(a^{**}, z) - u(a^*, z)] \right\} \leq 0 \end{aligned} \quad (7.61)$$

*erfüllt, so dominiert die Aktion  $a^*$  die Aktion  $a^{**}$ . Die in Gleichung (7.61) formulierte Bedingung kann dabei grundsätzlich nur dann erfüllt sein, wenn die Aktion  $a^*$  die Aktion  $a^{**}$  im lokalen Bayes'schen Modell dominiert, d. h. falls*

$$E_{p(z|d, U)}[u(a^{**}, z)] \leq E_{p(z|d, U)}[u(a^*, z)] \quad (7.62)$$

*erfüllt ist.*

*Beweis.* Die Aktion  $a^*$  dominiert die Aktion  $a^{**}$ , wenn für alle  $p \in W_{LF}$  gemäß Gleichung (7.47) die folgende Beziehung gilt:

$$E_p[u_{LF}(a^{**}, \nu_i)] - E_p[u_{LF}(a^*, \nu_i)] = E_p[u_{LF}(a^{**}, \nu_i) - u_{LF}(a^*, \nu_i)] \leq 0. \quad (7.63)$$

Diese Bedingung ist dann erfüllt, wenn gilt

$$\max_{p \in W_{LF}} E_p[u_{LF}(a^{**}, \nu_i) - u_{LF}(a^*, \nu_i)] \leq 0. \quad (7.64)$$

Da  $E_p[u_{LF}(a^{**}, \nu_i) - u_{LF}(a^*, \nu_i)]$  wiederum eine lineare Funktion bezüglich  $p$  darstellt, ist Lemma 7.7 anwendbar. Die Funktion  $E_p[u_{LF}(a^{**}, \nu_i) - u_{LF}(a^*, \nu_i)]$  nimmt also ihr Maximum in einer der Ecken von  $W_{LF}$  an. Mit der abkürzenden Notation

$$u(a^*, a^{**}, z) := u(a^{**}, z) - u(a^*, z) \quad (7.65)$$

ergibt sich

$$\begin{aligned} \{E_p[u_{LF}(a^{**}, \nu_i) - u_{LF}(a^*, \nu_i)] \mid p \in V(W_{LF})\} &= \{E_{p(z|d,U)}[u(a^*, a^{**}, z)]\} \\ \bigcup \{(1 - \beta)u(a^*, a^{**}, \nu_i) + \beta E_{p(z|d,U)}[u(a^*, a^{**}, z)] \mid i = 1, \dots, M\}. \end{aligned} \quad (7.66)$$

Die Gültigkeit des Dominanzkriteriums gemäß Gleichung (7.61) folgt nun, indem man das Maximum von Gleichung (7.66) nimmt und aus der zweiten Menge auf der rechten Seite die Elemente entfernt, für die der Ausdruck sicher nicht maximal wird.

Weiter gilt trivialerweise, dass die Gültigkeit der in Gleichung (7.61) formulierten Bedingung nur dann gegeben sein kann, wenn gilt

$$E_{p(z|d,U)}[u(a^{**}, z)] \leq E_{p(z|d,U)}[u(a^*, z)]. \quad (7.67)$$

□

Der Beweis des nachfolgenden Korollars 7.11 stellt eine exemplarische Anwendung von Satz 7.10 dar. Es sei angemerkt, dass ein alternativer Beweis von Korollar 7.11, der ohne die im vorliegenden Abschnitt 7.4 behandelten entscheidungstheoretischen Konzepte auskommt, in [San10a] veröffentlicht wurde.

**Korollar 7.11.** Für zwei Ereignisse  $B_1, B_2$  mit  $B_1 \subset U, B_2 \subset Z$  gilt gerade  $P(B_1|d) \geq P(B_2|d)$ , wenn die Schranken der zugehörigen Wahrscheinlichkeitsintervalle gemäß Gleichung (7.20) die folgende Beziehung erfüllen

$$r(B_2 \cap \bar{U}) \leq l(B_1) - l(B_2 \cap U) . \quad (7.68)$$

Korollar 7.11 kann als (globales) Entscheidungsproblem mit Aktionenmenge  $A = \{B_1, B_2\}$  und Nutzenfunktion

$$u(B, z) = \mathbf{1}_B(z) = \begin{cases} 1, & \text{falls } z \in B, \\ 0, & \text{falls } z \notin B, \end{cases} \quad (7.69)$$

aufgefasst werden. Es gilt damit für  $B \in \{B_1, B_2\}$  nämlich gerade die Beziehung  $E_{p(z|d)}[u(B, z)] = P(B|d)$ .

*Beweis.* Im Folgenden kann ohne Beschränkung der Allgemeinheit angenommen werden, dass  $B_1 \cap B_2 = \emptyset$  gilt. Gilt dies nicht, so ergibt sich Korollar 7.11 nämlich aus einem Vergleich der globalen A-Posteriori-Wahrscheinlichkeiten von  $B_1 \setminus (B_1 \cap B_2)$  und  $B_2 \setminus (B_1 \cap B_2)$ .

Nachfolgend wird gezeigt, dass  $B_2$  von  $B_1$  dominiert wird, wenn die in Korollar 7.11 formulierte Bedingung erfüllt ist. Es gelte ohne Beschränkung der Allgemeinheit  $B_2 = V \cup \left( \bigcup_{j \in J} \mathcal{V}_j \right)$  mit  $V \subseteq U$  und  $J \subseteq \{1, \dots, M\}$ . Das Dominianzkriterium gemäß Gleichung (7.61) liefert dann die Bedingung

$$\max \{ r(B_2 \cap U) - r(B_1), r(B_2 \cap \bar{U}) + l(B_2 \cap U) - l(B_1) \} \leq 0 . \quad (7.70)$$

Der zweite Anteil in der in Gleichung (7.70) angegebenen Menge ist nicht größer als Null, wenn gilt

$$r(B_2 \cap \bar{U}) \leq l(B_1) - l(B_2 \cap U) . \quad (7.71)$$

Weiter ist die Bedingung  $r(B_2 \cap U) - r(B_1) \leq 0$  erfüllt, wenn die Beziehung  $l(B_2 \cap U) - l(B_1) \leq 0$  gilt. Da  $r(B_2 \cap \bar{U}) \geq 0$  ist, garantiert die Bedingung gemäß Gleichung (7.71) somit, dass  $B_2$  von  $B_1$  dominiert wird.  $\square$

Der Fall  $B_2 \subset U \setminus B_1$  zeigt, dass das in Gleichung (7.60) angegebene Dominanzkriterium eine hinreichende, aber nicht notwendige Bedingung für Dominanz darstellt: In diesem Fall gilt (mit der Festlegung gemäß Gleichung (7.58)) für  $B \in \{B_1, B_2\}$  gerade  $I_{W_{LF}}[u_{LF}(B, \nu_i)] = [l(B), r(B)]$ . Gemäß Korollar 7.11 wird  $B_2$  von  $B_1$  dominiert, wenn gilt  $l(B_2) \leq l(B_1)$ . Dies ist eine schwächere Bedingung, als sie Gleichung (7.60) mit  $r(B_2) \leq l(B_1)$  vorgibt.

Auch in Fällen, in denen es auf Basis des lokalen Bayes'schen Modells nicht möglich ist, eine Aktion  $a \in A$  zu identifizieren, welche einen maximalen erwarteten Nutzen im Hinblick auf alle möglichen (globalen) A-Posteriori-Verteilungen  $p(z|d) \in \mathfrak{P}_z$  besitzt, kann eine ausreichend gute (globale) Entscheidungsfindung unter Umständen möglich sein: Der Entscheider kann unter Umständen in der Lage sein, eine Aktion  $a \in A$  auszuwählen, welche möglicherweise nicht optimal, aber dennoch ausreichend gut im Hinblick auf die konkrete Aufgabenstellung, ist.

Um einen derartigen Ansatz umzusetzen, können sog. Regret-Werte betrachtet werden. Der Regret  $R(a)$  einer Aktion  $a \in A$  ist dabei definiert als der im Vergleich zur besten Aktion  $a_{\text{opt}}$  gemäß Gleichung (7.42) maximal fehlende a posteriori erwartete Nutzen, wenn man  $a$  statt  $a_{\text{opt}}$  wählt (vgl. auch [Pre02]).

Der nachfolgende Satz stellt ein Mittel zur Einschätzung des Regrets der möglichen Aktionen  $a \in A$  auf Basis eines lokalen Bayes'schen Modells dar.

**Satz 7.12.** *Hinsichtlich der A-Posteriori-Verteilung  $p(z|d)$  sei gerade bekannt, dass diese in der Menge  $\mathfrak{P}_z$  aller Wahrscheinlichkeitsverteilungen auf  $Z$ , welche konsistent mit Gleichung (7.2) und Gleichung (7.8) sind, enthalten ist. Dann ist für jede Aktion  $a^* \in A$  der (globale) Regret  $R_{LF}(a^*)$  von  $a^*$  wie folgt nach oben beschränkt:*

$$R_{LF}(a^*) \leq \max \left\{ R_{p(z|d,U)}(a^*), \right. \\ \left. (1 - \beta) \max_{a \in A} \max_{1 \leq i \leq M} \{u(a, \nu_i) - u(a^*, \nu_i)\} + \beta R_{p(z|d,U)}(a^*) \right\}. \quad (7.72)$$

Dabei ist  $R_{p(z|d,U)}(a^*)$  der Regret von  $a^*$  im lokalen Bayes'schen Modell, d. h.

$$R_{p(z|d,U)}(a^*) = \max_{a \in A} E_{p(z|d,U)}[u(a,z)] - E_{p(z|d,U)}[u(a^*,z)] . \quad (7.73)$$

*Beweis.* Es gilt

$$R_{LF}(a^*) = \max_{a \in A} \max_{p \in W_{LF}} \{E_p[u_{LF}(a, \nu_i)] - E_p[u_{LF}(a^*, \nu_i)]\} . \quad (7.74)$$

Die Menge in Gleichung (7.74) entspricht im Wesentlichen der Menge in Gleichung (7.64). Satz 7.12 ergibt sich damit in nahezu analoger Weise, wie bei der Maximierung der Menge in Gleichung (7.64) vorgegangen wurde. Führt man eine zusätzliche Maximierung bezüglich  $a \in A$  durch und berücksichtigt, dass diese subadditiv ist, so führt dies unter Berücksichtigung der Definition von  $R_{p(z|d,U)}(a^*)$  direkt auf die in Gleichung (7.72) formulierte Abschätzung für  $R_{LF}(a^*)$ .  $\square$

### 7.4.3 Verbesserung der Entscheidungsgrundlage

Ist für den Entscheider keine adäquate (globale) Entscheidungsfindung auf Basis des lokalen Bayes'schen Modells möglich, so kann er, wiederum auch dem konzeptionell verwandten Grundgedanken des Lazy Decision Makings folgend, prinzipiell die gemäß Lemma 7.6 vorliegende lineare partielle Information weiter präzisieren.

Eine Möglichkeit für eine derartige Präzisierung besteht darin, die lokale Bayes'sche Fusion auf eine Obermenge des (bisherigen) lokalen Kontexts  $U$  ausweiten. Um dieses Vorgehen exemplarisch zu beschreiben, sei angenommen, dass  $U$  um  $\nu_M$  zu  $U_L := U \cup \{\nu_M\}$  erweitert wird und dass für alle  $z \in \bar{U}_L$  dann die Bedingung  $l(d|z) \leq \delta_L$  erfüllt ist. Dabei sei  $\delta_L$  eine Schranke, welche  $\delta_L \leq \delta$  erfüllt, und  $\delta$  sei die gemäß der Annahme in Satz 7.1 für alle  $z \in \bar{U}$  geltende Schranke.

Hierdurch wird dann  $W_{LF}$  gemäß Gleichung (7.47) wie folgt präzisiert: Die Ungleichung  $0 \leq p^M$  wird zur Gleichung

$$p^M = p^{M+1} \frac{p(\nu_M|d, U_L)}{1 - p(\nu_M|d, U_L)} \quad (7.75)$$

und die Ungleichung  $p^{M+1} \geq \beta$  wird zu

$$p^{M+1} \geq \frac{\int_{U_L} l(d|z) p(z|U_L) dz}{\int_{U_L} l(d|z) p(z|U_L) dz + \left(\frac{1}{P(U_L)} - 1\right) \delta_L} (\geq \beta). \quad (7.76)$$

Die Ausweitung von  $U$  zu  $U_L$  führt also auf lineare partielle Information, welche mit einer (präzisierten) Teilmenge von  $W_{LF}$  gemäß Gleichung (7.47) korrespondiert. Dies kann dazu führen, dass die Länge der Intervalle für den (globalen) a posteriori erwarteten Nutzen von Aktionen abnimmt, mehr (globale) Dominanzbeziehungen zwischen Aktionen identifizierbar sind und sich die (globalen) Regret-Werte von Aktionen verringern.

Um das Entscheidungsproblem unter der neuen linearen partiellen Information auszuwerten, ist es sinnvoll, die resultierende Notation wieder an die zuvor vorliegende Notation anzupassen. Setzt man dementsprechend neu

$$\begin{aligned} M_L &:= M - 1, \\ Z_{LF} &:= \{\nu_1, \dots, \nu_{M_L+1}\} \quad \text{mit} \quad \nu_{M_L+1} := U_L, \\ \bar{p} &:= (p^1, \dots, p^{M_L+1})^\top \\ &\quad \text{mit } p^i := p(\nu_i|d) \text{ für } i \in \{1, \dots, M_L\} \text{ und } p^{M_L+1} := P(U_L|d), \quad (7.77) \end{aligned}$$

so sind die Ergebnisse aus Abschnitt 7.4.1 und Abschnitt 7.4.2 direkt anwendbar, wenn man  $M$  durch  $M_L$  ersetzt.

Eine andere Möglichkeit für eine Präzisierung besteht darin, die zusätzliche Annahme, dass die A-Posteriori-Wahrscheinlichkeit  $P(\bar{U}|d)$  des Komplements  $\bar{U}$  des lokalen Kontexts  $U$  nicht vollständig auf einen kleinen Anteil von  $\bar{U}$  konzentriert ist, zu machen. Um die lineare partielle Information dementsprechend zu präzisieren, können für  $i \in \{1, \dots, M\}$  nicht triviale obere Schranken bezüglich  $p^i$  gesetzt werden. Dieses Vorgehen führt dann ebenfalls auf eine (präzisierte) Teilmenge von  $W_{LF}$  gemäß Gleichung (7.47).

Macht man z. B. die Annahme, dass für ein  $k \in \{2, \dots, M\}$  gilt, dass

$$p^i \leq \frac{1 - \beta}{k} \quad (7.78)$$

ist, so ändert sich die Struktur der Ecken des mit der linearen partiellen Information korrespondierenden konvexen Polyeders. Als Konsequenz müssen die Formeln in Gleichung (7.56), Gleichung (7.57), Gleichung (7.61) und Gleichung (7.72) entsprechend angepasst werden: Die minimalen und maximalen Werte des Nutzens von Aktionen und die Nutzendifferenzen von Aktionen bezüglich  $\bar{U}$  werden dabei ersetzt durch die arithmetischen Mittel der entsprechenden  $k$  niedrigsten und höchsten Werte. Der Fall  $k = M$  korrespondiert mit der Annahme, dass entweder  $P(\bar{U}|d) = 0$  oder  $p^i = (1 - \beta)/M$  für  $i \in \{1, \dots, M\}$  gilt.

#### 7.4.4 Berechnungsbeispiel

Der Übersichtlichkeit halber wird im vorliegenden Beispiel wieder angenommen, dass der Umfang von  $Z$  eher gering ist. Konkret gelte  $Z = \{\nu_1, \dots, \nu_6\}$ . Weiter sei angenommen, dass zwei Informationsquellen vorliegen. Die von ihnen gelieferten Informationsbeiträge  $d_1$  und  $d_2$  seien bedingt unabhängig gegeben  $z$ . Die A-Priori-Verteilung  $p(z)$  und die individuellen Likelihood-Funktionen  $l(d_1|z)$  sowie  $l(d_2|z)$  seien wie in Tabelle 7.1 angegeben.

**Tabelle 7.1:** A-Priori-Verteilung und Likelihood-Funktionen.

	$\nu_1$	$\nu_2$	$\nu_3$	$\nu_4$	$\nu_5$	$\nu_6$
$p(z)$	4/15	2/15	2/15	3/20	4/15	1/20
$l(d_1 z)$	1/20	1/40	1/28	1/36	1/40	1/2
$l(d_2 z)$	1/80	1/40	1/36	1/36	1/4	1/12

Zur Festlegung des lokalen Kontexts  $U$  werde die (gleiche) Schranke  $\epsilon$  an die standardisierten Likelihood-Funktionen angelegt. Ein möglicher Wert  $z \in Z$  wird also ignoriert im lokalen Bayes'schen Modell, d. h. nicht in den lokalen Kontext  $U$  aufgenommen, wenn für  $s = 1$  und  $s = 2$  gilt

$$l(d_s|z) \leq \delta_s \quad \text{mit } \delta_s = \epsilon \cdot \max_{z \in Z} l(d_s|z). \quad (7.79)$$

Die Festlegung  $\epsilon = 1/8$  liefert  $\delta_1 = 1/16$  und  $\delta_2 = 1/32$  und somit den lokalen Kontext  $U = \{\nu_5, \nu_6\}$ . Mit  $\delta_1 \cdot \delta_2 = 1/512$  liefert Korollar 7.2 die untere Schranke  $\tilde{\beta} \approx 0,7375$  für die A-Posteriori-Wahrscheinlichkeit  $P(U|d)$  des lokalen Kontexts  $U$ . Gleichung (7.20) liefert damit das folgende Intervall-Schema für die durch  $p(z|d)$  verkörperte A-Posteriori-Wahrscheinlichkeit der Elementarereignisse:

$$p(z|d) \in [l(z), r(z)] \approx \begin{cases} [0, 0,2625] & \text{für } z \in \bar{U} = \{\nu_1, \nu_2, \nu_3\}, \\ [0,3278, 0,4444] & \text{für } z = \nu_5, \\ [0,4097, 0,5556] & \text{für } z = \nu_6. \end{cases} \quad (7.80)$$

Es gelten  $r(z) \leq l(\nu_6)$  für  $z \in \bar{U}$  und  $l(\nu_5) \leq l(\nu_6)$ . Aus Korollar 7.11 folgt damit, dass  $p(\nu_6|d) \geq p(z|d)$  für alle  $z \in Z$  gilt. Der mögliche Wert  $\nu_6$  von  $z$  konnte damit also als globaler Maximum-A-Posteriori-Schätzwert identifiziert werden.

Weiter sei nun angenommen, dass eine Menge möglicher Aktionen durch  $A = \{a_1, a_2, a_3\}$  gegeben ist. Deren Nutzen in Abhängigkeit vom Wert von  $z$  sei durch die in Tabelle 7.2 angegebene Nutzenfunktion  $u(a, z)$  quantifiziert.

**Tabelle 7.2:** Nutzenfunktion.

	$\nu_1$	$\nu_2$	$\nu_3$	$\nu_4$	$\nu_5$	$\nu_6$
$u(a_1, z)$	1	4	3	4	4	9
$u(a_2, z)$	3	8	4	3	4	4
$u(a_3, z)$	2	3	3	3	1	10



Man berechnet  $E_{p(z|d,U)}[u(a_1,z)] = 61/9$ ,  $E_{p(z|d,U)}[u(a_2,z)] = 4$  und  $E_{p(z|d,U)}[u(a_3,z)] = 6$ .

Bestimmt man für  $j \in \{1,2,3\}$  weiter die minimalen und maximalen Werte von  $u(a_j,z)$  für  $z \in \bar{U}$ , so liefert Satz 7.9 näherungsweise

$$E_{p(z|d)}[u(a|z)] \in \begin{cases} [5,2611, 6,7778] & \text{für } a = a_1, \\ [3,7375, 5,0500] & \text{für } a = a_2, \\ [4,9501, 6,0000] & \text{für } a = a_3. \end{cases} \quad (7.81)$$

Aus Gleichung (7.81) erkennt man direkt, dass die Aktion  $a_1$  global gesehen definitiv besser ist als die Aktion  $a_2$  (vgl. auch Gleichung (7.60)). Bestimmt man für  $z \in \bar{U}$  den maximalen Wert der Nutzendifferenz  $u(a_3,z) - u(a_1,z)$ , so liefert das Dominanzkriterium aus Satz 7.10 zusätzlich das Ergebnis, dass die Aktion  $a_1$  weiter auch die Aktion  $a_3$  dominiert. Insgesamt folgt also, dass sichergestellt ist, dass die Aktion  $a_1$  den maximalen a posteriori erwarteten Nutzen im globalen Bayes'schen Modell aufweist.

Um aufzuzeigen, dass es nicht in allen Fällen möglich ist, auf Basis eines lokalen Bayes'schen Modells wie beschrieben eine Aktion zu identifizieren, welche sicher global gesehen optimal ist, wird das vorliegende Beispiel nun wie folgt modifiziert: Die Menge der möglichen Aktionen sei nun gegeben durch  $A = \{a_2, a_3\}$  und es gelte (anders als in Tabelle 7.2 angegeben) nun  $u(a_2, v_1) = 8$ .

Diese Änderungen haben keine Auswirkung auf die Intervalle für den (globalen) a posteriori erwarteten Nutzen bezüglich  $E_{p(z|d)}[u(a_2,z)]$  und  $E_{p(z|d)}[u(a_3,z)]$  – obwohl sich der tatsächlich vorliegende Wert von  $E_{p(z|d)}[u(a_2,z)]$  natürlich geändert hat durch die Modifizierung von  $u(a_2,z)$ . Nimmt man an, dass im globalen Bayes'schen Modell  $p(v_1|d) = 1 - \beta$  gilt, so erhält man  $E_{p(z|d)}[u(a_2,z)] \approx 5,0500$  und  $E_{p(z|d)}[u(a_3,z)] \approx 4,9501$ . Unter dieser Annahme gilt also, dass die Aktion  $a_2$  besser als die Aktion  $a_3$  ist. Nimmt man dagegen an, dass im globalen Bayes'schen Modell  $p(v_4|d) = 1 - \beta$  gilt, so erhält man  $E_{p(z|d)}[u(a_2,z)] \approx 3,7375$  und  $E_{p(z|d)}[u(a_3,z)] \approx 5,2125$ . Unter dieser Annahme gilt also entgegengesetzt, dass die Aktion  $a_3$  besser

als die Aktion  $a_2$  ist. Der im lokalen Bayes'schen Modell verfügbare Informationsstand im Hinblick auf  $z$  reicht also definitiv nicht aus, um festzustellen, ob eine der beiden Aktionen im globalen Bayes'schen Modell einen höheren a posteriori erwarteten Nutzen aufweist bzw. um welche der beiden Aktionen es sich dabei handelt.

Im Hinblick auf die Einschätzung des Regrets liefert die exemplarische Anwendung von Satz 7.12 für die Aktion  $a_3$  näherungsweise  $R_{LF}(a_3) \leq 0,0999$ . In Abhängigkeit von der konkret vorliegenden Aufgabenstellung könnte diese obere Schranke für den Regret der Aktion  $a_3$  ausreichend gering sein, um dem Entscheider zu erlauben, sich direkt für  $a_3$  zu entscheiden.

Ist ihm dies nicht möglich, so könnte man den lokalen Kontext  $U$  wie in Abschnitt 7.4.3 dargestellt ausweiten. Beispielsweise könnte man dazu die untere Schranke  $\epsilon$ , welche an die standardisierten Likelihood-Funktionen angelegt wird (vgl. Gleichung (7.79)), absenken. Setzt man z. B. konkret  $\epsilon = 1/10$ , so enthält der lokale Kontext  $U$  zusätzlich zu  $v_5$  und  $v_6$  noch die möglichen Werte  $v_3$  und  $v_4$ . Man berechnet  $\delta_1 \cdot \delta_2 = 1/800$  und Korollar 7.2 liefert als untere Schranke für die A-Posteriori-Wahrscheinlichkeit  $P(U|d)$  des (nun umfassenderen) lokalen Kontexts  $U$  gerade  $\tilde{\beta} \approx 0,8888$ . Man berechnet weiter  $E_{p(z|d,U)}[u(a_2,z)] \approx 3,9711$  und  $E_{p(z|d,U)}[u(a_3,z)] \approx 5,8139$ . Satz 7.9 liefert damit dann näherungsweise

$$E_{p(z|d)}[u(a|z)] \in \begin{cases} [3,9711, 4,4191] & \text{für } a = a_2, \\ [5,3898, 5,8139] & \text{für } a = a_3. \end{cases} \quad (7.82)$$

Es ist nun also definitiv bekannt, dass im globalen Bayes'schen Modell der a posteriori erwartete Nutzen von Aktion  $a_3$  größer ist als der a posteriori erwartete Nutzen von Aktion  $a_2$  (vgl. auch Gleichung (7.60)).

## 8 Zusammenfassung und Ausblick

### 8.1 Zusammenfassung

In der vorliegenden Arbeit werden neue Ansätze zur Bewältigung Bayes'scher Fusion beim Vorliegen eines umfangreichen, ggf. recht hochdimensionalen Zielgrößenraums eingeführt. Diese erlauben es, die Bayes'sche Fusion möglichst konzentriert auf die Anteile des Wertebereichs der Größen des Interesses durchzuführen, in denen sich tatsächlich Aufgabenrelevantes abspielt.

In Rahmen der Arbeit werden diese Ansätze formal mathematisch formalisiert und analysiert. Die Arbeit zeigt insbesondere auf, dass die anvisierte Einnahme eines lokalen Standpunkts bei der Bayes'schen Fusion konsistent in den Bayes'schen Kalkül integrierbar ist und stellt letztlich das hierfür notwendige Werkzeug in konzeptioneller und methodischer Hinsicht zur Verfügung. Als oftmals besonders vielversprechend für die lokale Bayes'sche Fusion heterogener Informationsquellen erweist sich ein Ansatz, welcher eine vollständige Einschränkung des probabilistischen Modells auf den lokalen Kontext vornimmt. Für die Umsetzung der lokalen Bayes'schen Fusion kann gerade dann auch auf das Konzept für die agentenbasierte Fusionsarchitektur zurückgegriffen werden.

Nachfolgend werden die wissenschaftlichen Ergebnisse, welche die vorliegende Arbeit in Bezug auf zentrale Themenkomplexe liefert, in komprimierter und in Bezug auf die zugrundeliegenden mathematischen Details abstrahierter Form zusammengefasst:

#### **Repräsentation und Propagation lokaler Informationsstände**

Die Arbeit zeigt auf, dass die anvisierte Einnahme eines lokalen Standpunkts

bei der Bayes'schen Fusion, d. h. die Konzentration der Fusion auf einen lokalen Kontext, auf Basis lokaler Bayes'scher Modelle, welche aus mathematischer Hinsicht auch tatsächlich valide sind, erfolgen kann. Sie liefert dabei auch eine Antwort auf die Frage, wie dies genau möglich ist. Es werden so letztlich probabilistische Modelle erarbeitet, für welche die theoretische Fundiertheit der Bayes'schen Fusionsmethodik erhalten bleibt. Dadurch steht dann das gesamte Spektrum der durch die Bayes'sche Theorie für die Informationsfusion bereitgestellten Konzepte und Verfahren grundsätzlich weiterhin zur Verfügung. Konkret werden in der Arbeit zwei verschiedene Ansätze, Informationsstände lokal zu repräsentieren und durch die (lokale) Einbeziehung der von den Informationsquellen gelieferten Beiträge weiterzuentwickeln, formalisiert und grundlegend untersucht. Zusätzlich wird aufgezeigt, wie genau sich diese Modelle ausgestalten, wenn die lokale Bayes'sche Fusion kooperativ durch unterschiedliche Agenten, welche jeweils lokale Informationsstände in Bezug auf bestimmte Anteile des lokalen Kontexts besitzen und weiterentwickeln, vorgenommen wird, und was die Voraussetzungen dafür sind, damit derartige lokale Informationsstände überhaupt zu einem gemeinsamen Resultat zusammengeführt werden können.

### **Kriterien zur Festlegung des lokalen Kontexts**

Der Gedanke, die lokale Bayes'sche Fusion auf Basis vorab ermittelter Auffälligkeiten im Zielgrößenraum auf einen lokalen Kontext zu konzentrieren, ist ausgehend vom Konzept für die agentenbasierte Fusionsarchitektur zwar intuitiv nachvollziehbar, die Antwort auf die Frage, wie sich dieser Gedanke formal gesehen umsetzen lässt, ist aber nicht trivial. Im Rahmen der Arbeit werden die Bezüge zwischen der Bayes'schen Inferenz und weiteren, im Bereich der induktiven Statistik verfügbaren Inferenzkonzepten sowie die Bezüge der Bayes'schen Theorie zu informationstheoretischen Konzepten und Ansätzen herausgearbeitet. Auf dieser Basis werden konkret Prinzipien aus der Likelihood-Inferenz und Ansätze aus der Informationstheorie als vielversprechend identifiziert, aufgegriffen und im Hinblick auf ihre Anwendung für die lokale Bayes'sche Fusion weiterentwickelt, um letztlich abzuleiten, welche Kriterien sie im Hinblick auf die Festlegung des lokalen Kontexts liefern. Ein wesentlicher Beitrag der Arbeit besteht dabei insbesondere darin, dass auch beleuchtet wird, inwieweit und wie genau diese Ansätze es ermöglichen, die

Adäquatheit des lokalen Kontexts und damit auch die der lokalen Bayes'schen Modelle quantitativ zu bewerten und wo die jeweiligen Grenzen diesbezüglich liegen.

### **Einbettung in den globalen Kontext**

Die vorliegende Arbeit liefert dem genauen Ziel der Informationsfusion entsprechend unterschiedliche Antworten auf die Frage, wie sich die aus der lokalen Bayes'schen Fusion, welche durch Einschränkung auf den lokalen Kontext erfolgt, ergebende partielle Information in Hinblick auf die Größen des Interesses unter Einbeziehung zusätzlich verfügbarer Information darüber, wie der lokale Kontext gebildet wurde, in Bezug zur (lokal unbekannt) globalen A-Posteriori-Verteilung setzen lässt. Es wird ein Wahrscheinlichkeitsintervall-Schema formuliert, anhand dessen sich lokale und globale A-Posteriori-Wahrscheinlichkeiten genauer in Beziehung setzen lassen. Weiter wird, wieder unter Berücksichtigung des engen Bezugs der Bayes'schen Theorie zur Informationstheorie, eine analytische Bestimmung der globalen (Maximum-Entropie) A-Posteriori-Verteilung, welche global gesehen den auf Basis des lokalen Ansatzes resultierenden Informationsstand in möglichst objektiver Weise repräsentiert, vorgenommen und untersucht. Als Alternative wird unter Zugrundelegung der Theorie linearer partieller Information und von Prinzipien des Lazy Decision Making untersucht, wie sich das Wahrscheinlichkeitsintervall-Schema methodisch gesehen direkt in die Lösung eines Entscheidungsproblems einbringen lässt und welche Schlüsse sowie Möglichkeiten zur Verbesserung der Informationsgrundlage auf dieser Basis möglich sind.

### **Beiträge zur Fusion in Verbänden**

Insbesondere (jedoch nicht ausschließlich) wenn die Informationsfusion in Verbänden umgesetzt wird, kann die Herausforderung bei der Bayes'schen Fusion nicht nur darin bestehen, dass der Aufwand zur Berechnung (und gegebenenfalls auch zur weiteren Auswertung) der A-Posteriori-Verteilung nicht tragbar ist. Vielmehr können auch bereits die Bestimmung und Übermittlung der zu ihrer Berechnung erforderlichen Größen problematisch, möglicherweise sogar nicht realisierbar, sein. Das bei der (nicht lokalen) Bayes'schen Fusion umgesetzte Vorgehen, in Form der A-Posteriori-Verteilung jeweils stets zuerst

ein maximal umfassendes Fusionsresultat zu berechnen, widerspricht zudem auch konzeptionell dem für ein optimales Ressourcenmanagement notwendigen Prinzip, dass gerade in Verbänden Information letztlich nur dort und in der Form bereitgestellt werden sollte, wie sie auch tatsächlich benötigt wird. Die erarbeiteten Ansätze zur lokalen Bayes'schen Fusion liefern letztlich gerade auch Möglichkeiten, diese Herausforderungen zu adressieren.

## 8.2 Ausblick

In einer zunehmend vernetzten Welt wächst der Bedarf nach Systemen bzw. Systemverbänden, die in einer definierten Weise interoperabel miteinander interagieren. Dieser Themenkomplex war lange kein Fokus des Forschungsgebiets Informationsfusion und zentrale Konzepte und Modelle, wie z. B. das JDL Fusionsmodell, wurden unabhängig hiervon entwickelt. Die vorliegende Arbeit greift wichtige Aspekte im Hinblick auf die Umsetzung der Informationsfusion in Verbänden auf und zeigt insbesondere auch das Potential der lokalen Bayes'schen Fusion für derartige Aufgabenstellungen. Die Arbeit kann aber keine vollständigere Abhandlung des Themenkomplexes der Informationsfusion in Verbänden leisten, so dass eine vertiefte Betrachtung in Folgearbeiten als zielführend zu erachten ist.

Die software-technische Präzisierung und letztlich Umsetzung der agentenbasierten Fusionsarchitektur, welche sicher auch ein wichtiges Thema für Folgearbeiten darstellt, könnte dann ggf. auch gerade im Hinblick auf ihre Umsetzung in Verbänden erfolgen. Ihre konkrete Umsetzung sollte dann auch unter Berücksichtigung von Aspekten der IT-Sicherheit und ggf. juristischer Aspekte erfolgen.

Im Rahmen der Arbeit zeigte sich, vor allem im Hinblick auf die Möglichkeit, die Güte lokaler Bayes'scher Modelle zu bewerten und explizite Rückschlüsse auf das globale Bayes'sche Modell vorzunehmen, dass es problematisch sein kann, wenn der lokale Kontext allzu gering ausfällt. Diese Beobachtung steht letztlich konträr zum Wunsch, den Umfang des lokalen Kontexts möglichst gering zu halten, um den Aufwand für die lokale Bayes'sche Fusion so ge-

ring wie möglich auszugestalten. Eine Weiterentwicklung der in der vorliegenden Arbeit erarbeiteten Ansätze zur lokalen Bayes'schen Fusion, indem man diese mit dem Prinzip konjugierter Verteilungsfamilien bzw. mit Verfahren zur Approximation der A-Posteriori-Verteilung bzw. von Kenngrößen derselben kombiniert, ist gerade auch vor dem Hintergrund dieser Tatsache als zielführend zu erachten. Dies gilt gerade auch bei einer verteilten Umsetzung der lokalen Bayes'schen Fusion. So ist es insbesondere zu erwarten, dass sich die im lokalen Bayes'schen Modell benötigten probabilistischen Größen deutlich besser durch überschaubare konjugierte Verteilungsfamilien approximieren lassen als ihre Pendanten im globalen Bayes'schen Modell. Bei der Umsetzung entsprechender Ansätze könnte man u. U. auch gezielt auf Ansätze zur Bayes'schen Robustheitsanalyse zurückgreifen.

Zumindest für bestimmte Aufgabenstellungen ist auch der im Rahmen der Arbeit zwar erarbeitete, aber nicht in den Fokus gestellte Ansatz der Lokalität durch Vergrößerung durchaus als vielversprechend zu erachten. Es wäre ein weiteres Thema für Folgearbeiten, diesen gerade in Bezug auf spezielle Aufgabenstellungen der Informationsfusion hinsichtlich der Möglichkeiten, die mit ihm einhergehenden Schwierigkeiten zu adressieren, zu untersuchen und ihn weiterzuentwickeln.





## Literatur

- [Ait04] AITKEN, Colin G. G. and TARONI, Franco: *Statistics and the Evaluation of Evidence for Forensic Scientists*. 2. Aufl. John Wiley & Sons, Ltd, 2004. DOI: 10.1002/0470011238 (siehe S. 60, 73, 177).
- [Aki02] AKITA, Richard M.: „User based data fusion approaches“. In: *2002 5th International Conference on Information Fusion (FUSION)*. IEEE, 2002. DOI: 10.1109/ICIF.2002.1020987 (siehe S. 2).
- [Ald05] ALDOSARI, Saeed A. and MOURA, José M. F.: „Distributed Detection in Sensor Networks: Connectivity Graph and Small World Networks“. In: *Conference Record of the Thirty-Ninth Asilomar Conference on Signals, Systems and Computers, 2005*. IEEE, 2005. DOI: 10.1109/ACSSC.2005.1599738 (siehe S. 48).
- [And19] ANDERSEN, Stig: *Technical Report: A preliminary Process Model for Investigation*. Mai 2019. DOI: 10.31235/osf.io/z4wma (siehe S. 105).
- [Asc93] ASCHWANDEN, Peter Franz: „Experimenteller Vergleich von Korrelationskriterien in der Bildanalyse“. Dissertation. 1993. DOI: 10.3929/ETHZ-A-000904762 (siehe S. 163).
- [Asu07] ASUNCION, Arthur and NEWMAN, David J.: *UCI Machine Learning Repository*. University of California, Irvine, School of Information und Computer Sciences, 2007. URL: <http://www.ics.uci.edu/~mlern/MLRepository.html> (siehe S. 185).
- [Aug98] AUGUSTIN, Thomas: *Optimale Tests bei Intervallwahrscheinlichkeit*. Vandenhoeck und Ruprecht, 1998 (siehe S. 238).
- [Aza07] AZAD, Pedram; GOCKEL, Tilo and DILLMANN, Rüdiger: *Computer Vision: Das Praxisbuch*. Elektor-Verlag, 2007 (siehe S. 169).

- [Bar04] BARTELBORTH, Thomas: „Wofür sprechen die Daten?“ In: *Journal for General Philosophy of Science* 35 (2004), S. 13–40. DOI: 10.1023/b:jgps.0000035150.90850.c1 (siehe S. 71).
- [Bar20] BARBU, Adrian and ZHU, Song-Chun: „Data Driven Markov Chain Monte Carlo“. In: *Monte Carlo Methods*. Hrsg. von BARBU, Adrian and ZHU, Song-Chun. Springer, 2020, S. 211–280. DOI: 10.1007/978-981-13-2971-5\_8 (siehe S. 103).
- [Bed00] BEDWORTH, Mark and O'BRIEN, Jane: „The Omnibus model: a new model of data fusion?“ In: *IEEE Aerospace and Electronic Systems Magazine* 15.4 (2000), S. 30–36. DOI: 10.1109/62.839632 (siehe S. 30).
- [Beh13] BEHRENDTS, Ehrhard: *Elementare Stochastik*. Vieweg+Teubner Verlag, 2013. DOI: 10.1007/978-3-8348-2331-1 (siehe S. 124).
- [Bei03] BEICHEL, Frank E. and MONTGOMERY, Douglas C., Hrsg.: *Teubner-Taschenbuch der Stochastik : Wahrscheinlichkeitstheorie, stochastische Prozesse, mathematische Statistik*. Vieweg+Teubner Verlag, 2003. DOI: 10.1007/978-3-322-80067-1 (siehe S. 60, 64).
- [Bel15] BELKIN, Andrey: „World Modeling for Intelligent Autonomous Systems“. Dissertation. Karlsruher Institut für Technologie (KIT), 2015. DOI: 10.5445/IR/1000060148 (siehe S. 1).
- [Ber03] BERNARDO, José M.: „Bayesian Statistics“. In: *UNESCO Encyclopedia of Life Support Systems (EOLSS), Probability and Statistics*. Hrsg. von VIERTL, Reinhard. EOLSS Publishers, 2003 (siehe S. 62).
- [Ber04] BERNARDO, José M. and SMITH, Adrian F. M.: *Bayesian theory*. John Wiley & Sons, Inc., 2004. DOI: 10.1002/9780470316870 (siehe S. 62, 63, 65, 67, 76, 86, 92, 95, 96, 201, 202, 217, 237).
- [Ber85] BERGER, James O.: *Statistical Decision Theory and Bayesian Analysis*. 2. Aufl. New York: Springer, 1985. DOI: 10.1007/978-1-4757-4286-2 (siehe S. 53, 56, 76, 89, 92, 94, 217).

- [Bey06a] BEYERER, Jürgen; HEIZMANN, Michael and SANDER, Jennifer: „Fusion: an agent based architecture for fusion of heterogeneous information and data“. In: *Proc. SPIE 6242, Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2006, 62420Q (18 April 2006)*. SPIE, 2006. DOI: 10.1117/12.665938 (siehe S. 7, 54, 104).
- [Bey06b] BEYERER, Jürgen; SANDER, Jennifer and WERLING, Stefan: „Fusion heterogener Informationsquellen“. In: *Informationsfusion in der Mess- und Sensortechnik*. Hrsg. von BEYERER, Jürgen; PUENTE LEÓN, Fernando and SOMMER, Klaus-Dieter. Universitätsverlag Karlsruhe, 2006, S. 21–37 (siehe S. 24, 54, 55, 75, 84, 156).
- [Bey07] BEYERER, Jürgen; SANDER, Jennifer and WERLING, Stefan: „Bayes’sche Methodik zur lokalen Fusion heterogener Informationsquellen (Bayesian Methodology for the Local Fusion of Heterogeneous Information Sources)“. In: *tm – Technisches Messen* 74.3 (2007), S. 103–111. DOI: 10.1524/teme.2007.74.3.103 (siehe S. 4, 14, 20, 23, 63, 104, 156, 160, 165).
- [Bey08] BEYERER, Jürgen; HEIZMANN, Michael; SANDER, Jennifer and GHETA, Ioana: „Bayesian methods for image fusion“. In: *Image Fusion: Algorithms and Applications*. Hrsg. von STATHAKI, Tania. Academic Press, 2008, S. 157–192. DOI: 10.1016/b978-0-12-372529-5.00003-2 (siehe S. 18, 29, 88, 186).
- [Bey12] BEYERER, Jürgen; PUENTE LEÓN, Fernando and FRESE, Christian: *Automatische Sichtprüfung: Grundlagen, Methoden und Praxis der Bildgewinnung und Bildauswertung*. Springer, 2012. DOI: 10.1007/978-3-642-23966-3 (siehe S. 163, 168).
- [Bey99] BEYERER, Jürgen: „Verfahren zur quantitativen statistischen Bewertung von Zusatzwissen in der Meßtechnik“. Habilitationsschrift. Universität Karlsruhe (TH), 1999 (siehe S. 14, 60, 63, 76, 78, 83, 202, 235).

- [Bie06] BIERMANN, J.: „Remarks on resource management in intelligence“. In: *2006 9th International Conference on Information Fusion (FUSION)*. IEEE, 2006. DOI: 10.1109/icif.2006.301792 (siehe S. 40, 42).
- [Bla02] BLASCH, Erik P. and PLANO, Susan: „JDL level 5 fusion model: user refinement issues and applications in group tracking“. In: *Proc. SPIE 4729, Signal Processing, Sensor Fusion, and Target Recognition XI, (31 July 2002)*. 2002. DOI: 10.1117/12.477612 (siehe S. 31).
- [Bla08] BLASCH, Erik: „Introduction to Level 5 Fusion: The Role of the User“. In: *Handbook of Multisensor Data Fusion*. Hrsg. von LIGGINS II, Martin; HALL, David and LINAS, James. 2. Aufl. 2008, S. 503–535 (siehe S. 2).
- [Bla12] BLASCH, Erik; VALIN, Pierre; JOUSSELME, Anne-Laure; LAMBERT, Dale and BOSSÉ, Éloi: „Top ten trends in High-Level Information Fusion“. In: *2012 15th International Conference on Information Fusion (FUSION)*. IEEE, 2012 (siehe S. 33, 34).
- [Bla14] BLASCH, Erik; YANG, Chun and KADAR, Ivan: „Summary of tracking and identification methods“. In: *Proc. SPIE 9091, Signal Processing, Sensor/Information Fusion, and Target Recognition XXIII, 909104 (20 June 2014)*. SPIE, 2014. DOI: 10.1117/12.2050260 (siehe S. 70).
- [Bla17] BLASCH, Erik: „JDL Model (III) Updates for an Information Management Enterprise“. In: *Multisensor Data Fusion*. Hrsg. von FOURATI, Hassen and INIEWSKI, Krzysztof. CRC Press, 2017, S. 55–73. DOI: 10.1201/b18851-4 (siehe S. 31).
- [Bla19] BLASCH, Erik; KADAR, Ivan; GREWE, Lynne L.; STEVENSON, Garrett; MAJUMDER, Uttam K. and CHONG, Chee-Yee: „Deep learning in AI and information fusion panel discussion“. In: *Proc. SPIE 11018, Signal Processing, Sensor/Information Fusion, and Target Recognition XXVIII, 110180Q (7 May 2019)*. 2019. DOI: 10.1117/12.2519230 (siehe S. 33).

- [Blu02] BLUME, Jeffrey D.: „Likelihood methods for measuring statistical evidence“. In: *Statistics in Medicine* 21.17 (2002), S. 2563–2599. DOI: 10.1002/sim.1216 (siehe S. 71, 73, 177, 179).
- [Blu11] BLUME, Jeffrey D.: „Likelihood and its Evidential Framework“. In: *Philosophy of Statistics*. Hrsg. von BANDYOPADHYAY, Prasanta S. and FORSTER, Malcolm R. Elsevier, 2011, S. 493–511. DOI: 10.1016/b978-0-444-51862-0.50014-9 (siehe S. 178, 212, 213).
- [Bod06] BODENDORF, Freimut: Daten- und Wissensmanagement. Springer, 2006. DOI: 10.1007/3-540-28682-9 (siehe S. 21, 22).
- [Bos07a] BOSSE, Eloi; ROY, Jean and WARK, Steve: Concepts, models, and tools for information fusion. Artech House, 2007 (siehe S. 2, 38).
- [Bos07b] BOSTRÖM, Henrik; ANDLER, Sten F.; BROHEDE, Marcus; JOHANSSON, Ronnie; KARLSSON, Alexander; LAERE, Joeri van; NIKLASSON, Lars; NILSSON, Marie; PERSSON, Anne and ZIEMKE, Tom: On the Definition of Information Fusion as a Field of Research. Techn. Ber. HS-IKI -TR-07-006. Skövde: Institutionen för kommunikation och information, 2007. URL: <https://nbn-resolving.org/urn:nbn:se:kth:diva-221462> (siehe S. 18).
- [Bou16] BOUVRY, Pascal; CHAUMETTE, Serge; DANOY, Gregoire; GUERRINI, Gilles; JURQUET, Gilles; KUWERTZ, Achim; MÜLLER, Wilmuth; ROSALIE, Martin and SANDER, Jennifer: „Using heterogeneous multilevel swarms of UAVs and high-level data fusion to support situation management in surveillance scenarios“. In: *2016 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*. IEEE, 2016. DOI: 10.1109/mfi.2016.7849525 (siehe S. 1).
- [Bou17] BOUVRY, Pascal; SEGOR, Florian; CHAUMETTE, Serge; DANOY, Grégoire; GUERRINI, Gilles; JURQUET, Gilles; KUWERTZ, Achim; MÜLLER, Wilmuth; ROSALIE, Martin and SANDER, Jennifer: „ASIMUT project: Aid to SIuation Management based on MUlti-modal, MUltiUAVs, MUltilevel acquisition Techniques“. In: *Proceedings of the 3rd Workshop on Micro Aerial Vehicle Networks*,

- Systems, and Applications - DroNet '17*. ACM Press, 2017, S. 17–20. DOI: 10.1145/3086439.3086445 (siehe S. 1).
- [Bou96] BOUTILIER, Craig; FRIEDMAN, Nir; GOLDSZMIDT, Moises and KOLLER, Daphne: „Context-Specific Independence in Bayesian Networks“. In: *Proceedings of the 12th Conference on Uncertainty in Artificial Intelligence (UAI1996)*. 1996, S. 115–123 (siehe S. 145).
- [Bov09] BOVENS, Luc: „Judy Benjamin is a Sleeping Beauty“. In: *Analysis* 70.1 (2009), S. 23–26. DOI: 10.1093/analys/anp127 (siehe S. 230).
- [Boy04] BOYD, Stephen and VANDENBERGHE, Lieven: *Convex Optimization*. Cambridge University Press, 2004 (siehe S. 227).
- [Bre06] BREITER, Andreas and LIGHT, Daniel: „Data for School Improvement: Factors for designing effective information systems to support decision-making in schools“. In: *Educational Technology & Society* 9.3 (2006), S. 206–217 (siehe S. 21).
- [Bré99] BRÉMAUD, Pierre: *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*. Springer, 1999. DOI: 10.1007/978-1-4757-3124-8 (siehe S. 102).
- [Bry04] BRYANT, David J.: „Modernizing Our Cognitive Model“. In: *Command and Control Research and Technology Symposium 2004, San Diego, CA, June 15-17, 2004*. Command und Control Research Program (U.S.), 2004 (siehe S. 39).
- [Bur08] BURGER, Wilhelm and BURGE, Mark James: *Digital Image Processing: An Algorithmic Introduction using Java*. Springer, 2008 (siehe S. 163).
- [Cap07] CAPPE, Olivier; GODSILL, Simon J. and MOULINES, Eric: „An Overview of Existing Methods and Recent Advances in Sequential Monte Carlo“. In: *Proceedings of the IEEE* 95.5 (2007), S. 899–924. DOI: 10.1109/jproc.2007.893250 (siehe S. 97).
- [Car09] CARLIN, Bradley P. and LOUIS, Thomas A.: *Bayesian Methods for Data Analysis*. 3. Aufl. Chapman & Hall/CRC, 2009 (siehe S. 81, 93, 95–97, 99, 102).

- [Cas13] CASTANEDO, Federico: „A Review of Data Fusion Techniques“. In: *The Scientific World Journal* 2013, Article ID 704504 (2013). DOI: 10.1155/2013/704504 (siehe S. 42, 44).
- [Cat17] CATANO, Victor and GAUGER, Jeffery: „Information Fusion: Intelligence Centers and Intelligence Analysis“. In: *Information Sharing in Military Operations*. Hrsg. von GOLDENBERG, Irina; SOETERS, Joseph and DEAN, Waylon H. Springer, 2017, S. 17–34. DOI: 10.1007/978-3-319-42819-2\_2 (siehe S. 3, 30).
- [Cha90] CHANG, Kuo-Chu and FUNG, Robert: „Refinement and Coarsening of Bayesian Networks“. In: *Proceedings of the 6th Annual Conference on Uncertainty in Artificial Intelligence (UAI1990)*. 1990, S. 435–446 (siehe S. 137).
- [Cop04] COPPIN, Ben: *Artificial Intelligence Illuminated*. Jones und Bartlett Publishers, Inc, 2004 (siehe S. 107–109).
- [Cov91] COVER, Thomas M. and THOMAS, Joy A.: *Elements of Information Theory*. Wiley, 1991 (siehe S. 81, 203, 204).
- [Cox61] COX, Richard T.: *The algebra of probable inference*. Johns Hopkins Press, 1961 (siehe S. 62).
- [Cre04] CRESSIE, Noel; RICHARDSON, Sylvia and JAUSSENT, Isabelle: „Ecological Bias: Use of Maximum-Entropy Approximations“. In: *Australian & New Zealand Journal of Statistics* 46.2 (2004), S. 233–255. DOI: 10.1111/j.1467-842x.2004.00327.x (siehe S. 76).
- [Cum05] CUMMINGS, Mary L. and BRUNI, Sylvain: „Collaborative Human-Computer Decision Making in Network Centric Warfare“. In: *TCP HUM TP7 Workshop on Aerospace Human Factors Issues in Network-Centric Warfare*. Salisbury, UK, 2005 (siehe S. 2).
- [Das01] DASARATHY, Belur V.: „Information Fusion – what, where, why, when, and how?“ In: *Information Fusion* 2.2 (2001), S. 75–76. DOI: 10.1016/s1566-2535(01)00032-x (siehe S. 19).
- [Das08] DAS, Subrata Kumar: *High-Level Data Fusion*. Artech House, 2008 (siehe S. 4, 16, 30, 33, 37).

- [Das10] DAS, Subrata: „Agent-based information fusion“. In: *Information Fusion* 11.3 (2010), S. 216–219. DOI: 10.1016/j.inffus.2010.01.008 (siehe S. 107).
- [Das97] DASARATHY, Belur V.: „Sensor fusion potential exploitation–innovative architectures and illustrative applications“. In: *Proceedings of the IEEE* 85.1 (1997), S. 24–38. DOI: 10.1109/5.554206 (siehe S. 1, 14, 29, 34–36).
- [De 18] DE MARTINO, Andrea and DE MARTINO, Daniele: „An introduction to the maximum entropy approach and its application to inference problems in biology“. In: *Heliyon* 4.4 (e00596 2018). DOI: 10.1016/j.heliyon.2018.e00596 (siehe S. 76).
- [DeG70] DEGROOT, Morris H: *Optimal Statistical Decisions*. McGraw-Hill, 1970 (siehe S. 56).
- [Dem97] DEMPSTER, A. P.: „The direct use of likelihood for significance testing“. In: *Statistics and Computing* 7.4 (1997), S. 247–252. DOI: 10.1023/a:1018598421607 (siehe S. 178).
- [Dip05] DIPPOLD, Rolf; MEIER, Andreas; SCHNIDER, Walter and SCHWINN, Klaus: *Unternehmensweites Datenmanagement*. 4. Aufl. Vieweg+Teubner Verlag, 2005. DOI: 10.1007/978-3-322-86870-1 (siehe S. 22).
- [Döb18] DÖBEL, I.; LEIS, M.; MOLINA VOGELSANG, M.; WELZ, J.; NEUSTROEV, D.; PETZKA, H.; RIEMER, A.; PÜPING, S.; VOSS, A. and WEGELE, M.: *Maschinelles Lernen: Eine Analyse zu Kompetenzen, Forschung und Anwendung*. Studie, erstellt im Kontext eines vom Bundesministerium für Bildung und Forschung (BMBF) geförderten Projekts. München: Fraunhofer-Gesellschaft, 2018. URL: <http://publica.fraunhofer.de/dokumente/N-497408.html> (siehe S. 25).
- [Dom97] DOMINGOS, Pedro and PAZZANI, Michael: „On the Optimality of the Simple Bayesian Classifier under Zero-One Loss“. In: *Machine Learning* 29.2–3 (1997), S. 103–130. DOI: 10.1023/a:1007413511361 (siehe S. 85, 186).



- [Dou11] DOUVEN, Igor and ROMEIJN, Jan-Willem: „A New Resolution of the Judy Benjamin Problem“. In: *Mind* 120.479 (2011), S. 637–670. DOI: 10.1093/mind/fzr051 (siehe S. 230).
- [Du19] DU, Mengnan; LIU, Ninghao and HU, Xia: „Techniques for interpretable machine learning“. In: *Communications of the ACM* 63.1 (2019), S. 68–77. DOI: 10.1145/3359786 (siehe S. 25).
- [Dua19] DUA, Dheeru and GRAFF, Casey: UCI Machine Learning Repository. University of California, Irvine, School of Information and Computer Sciences, 2019. URL: <http://archive.ics.uci.edu/ml> (siehe S. 185).
- [Edw63] EDWARDS, Ward; LINDMAN, Harold and SAVAGE, Leonard J.: „Bayesian statistical inference for psychological research.“ In: *Psychological Review* 70.3 (1963), S. 193–242. DOI: 10.1037/h0044139 (siehe S. 83).
- [Edw92] EDWARDS, Anthony W. F.: Likelihood: Expanded Edition. Johns Hopkins University Press, 1992 (siehe S. 62, 68, 71, 74, 177).
- [Ekd06] EKDAHL, Magnus and KOSKI, Timo: „Bounds for the Loss in Probability of Correct Classification Under Model Based Approximation“. In: *The Journal of Machine Learning Research* 7 (2006), S. 2449–2480 (siehe S. 85).
- [Elm07] ELMENREICH, Wilfried: „A Review on System Architectures for Sensor Fusion Applications“. In: *Software Technologies for Embedded and Ubiquitous Systems. SEUS 2007. Lecture Notes in Computer Science, vol 4761*. Hrsg. von YUNMOOK NAH, Roman Obermaisser ans; PUSCHNER, Peter and RAMMIG, Franz J. Springer, 2007, S. 547–559. DOI: 10.1007/978-3-540-75664-4\_57 (siehe S. 30).
- [End95] ENDSLEY, Mica R.: „Toward a Theory of Situation Awareness in Dynamic Systems“. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37 (1995), S. 32–64. DOI: 10.1518/001872095779049543 (siehe S. 36, 37).

- [Eng08] ENGELMANN, Florian and GROßMANN, Christoph: „Was wissen wir über Information?“ In: *Daten- und Informationsqualität*. Hrsg. von HILDEBRAND, K.; GEBAUER, M.; HINRICHS, H. and MIELKE, M. Vieweg+Teubner Verlag, 2008, S. 3–24. DOI: 10.1007/978-3-8348-9266-9\_1 (siehe S. 21).
- [Ess17] ESSENDORFER, Barbara; KERTH, Christian and ZASCHKE, Christian: „Adaptation of interoperability standards for cross domain usage“. In: *Proc. SPIE 10207, Next-Generation Analyst V, 102070E (3 May 2017)*. SPIE, 2017. DOI: 10.1117/12.2262092 (siehe S. 30).
- [Ess18a] ESSENDORFER, Barbara; HOFFMANN, Almuth; SANDER, Jennifer and KUWERTZ, Achim: „Integrating Coalition Shared Data in a System Architecture for High Level Information Management“. In: *Proc. SPIE 10802, Counterterrorism, Crime Fighting, Forensics, and Surveillance Technologies II, 108020F (8 October 2018)*. SPIE, 2018. DOI: 10.1117/12.2501861 (siehe S. 3, 41).
- [Ess18b] ESSENDORFER, Barbara; KUWERTZ, Achim and SANDER, Jennifer: „Distributed Information Management through Coalition Shared Data“. In: *NATO STO-MP-IST-160 Big Data and Artificial Intelligence for Military Decision Making, Bordeaux, France, 30th May to 1st June 2018*. 2018 (siehe S. 19).
- [Ess19] ESSENDORFER, Barbara; SANDER, Jennifer; NISTOR, Marian Sorin; HOFFMANN, Almuth and PICKL, Stefan: „Distributed Data and Information Management for Crisis Forecasting and Management“. In: *Human Systems Engineering and Design II. Advances in Intelligent Systems and Computing (AISC)*, Bd. 1026. Springer, 2019, S. 813–819. DOI: 10.1007/978-3-030-27928-8\_124 (siehe S. 3, 30).
- [Fis01] FISCHER, Gerd: *Analytische Geometrie: Eine Einführung für Studienanfänger*. 7. Aufl. Vieweg, 2001 (siehe S. 241, 242).
- [Fit08] FITELSON, Branden and THOMASON, Neil: „Bayesians sometimes cannot ignore even very implausible theories (even ones that have not yet been thought of)“. In: *The Australasian Journal of Logic* 6 (2008). DOI: 10.26686/ajl.v6i0.1791 (siehe S. 135).

- [Flo90] FLORENS, Jean Pierre; MOUCHART, Michel and ROLIN, Jean-Marie: Elements of Bayesian Statistics. Marcel Dekker, Inc., 1990 (siehe S. 121, 122).
- [Foo13] Foo, Pek Hui and NG, Gee Wah: „High-level Information Fusion: An Overview“. In: *Journal of Advances in Information Fusion (JAIF)* 8 (1 2013), S. 33–72 (siehe S. 30).
- [Gal17] GALAVOTTI, Maria Carla: „The Interpretation of Probability: Still an Open Issue?“ In: *Philosophies* 2 (3 2017). Article Nr. 20. DOI: 10.3390/philosophies2030020 (siehe S. 60).
- [Gam06] GAMERMAN, Dani and LOPES, Hedibert F.: Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference. 2. Aufl. Chapman & Hall/CRC, 2006 (siehe S. 95–102).
- [Geo07] GEORGI, Hans-Otto: Stochastik : Einführung in die Wahrscheinlichkeitstheorie und Statistik. 3. Aufl. De Gruyter, 2007. DOI: 10.1515/9783110206777 (siehe S. 58, 59, 96).
- [Ghe08a] GHETA, Ioana; HEIZMANN, Michael and BEYERER, Jürgen: „Bayesian fusion of multivariate image to obtain depth information“. In: *2008 11th International Conference on Information Fusion (FUSION)*. IEEE, 2008 (siehe S. 160).
- [Ghe08b] GHETA, Ioana; HEIZMANN, Michael and BEYERER, Jürgen: „Fusion kombinierter Stereo- und Fokusserien zur 3D-Rekonstruktion (Fusion of Combined Stereo and Focus Series for 3D Reconstruction)“. In: *tm - Technisches Messen* 75.7-8/2008 (2008). DOI: 10.1524/teme.2008.0883 (siehe S. 27).
- [Ghe11] GHETA, Ioana: „Fusion multivariater Bildserien am Beispiel eines Kamera-Arrays“. Dissertation. Karlsruher Institut für Technologie (KIT), 2011. DOI: 10.5445/KSP/1000023232 (siehe S. 157).
- [Gho06] GHOSH, Jayanta K.; DELAMPADY, Mohan and SAMANTA, Tapas: An Introduction to Bayesian Analysis: Theory and Methods. Springer, 2006. DOI: 10.1007/978-0-387-35433-0 (siehe S. 68, 70).

- [Gil10] GILLIES, Duncan; THORNLEY, David and BISDIKIAN, Chatschik: „Probabilistic Approaches to Estimating the Quality of Information in Military Sensor Networks“. In: *The Computer Journal* 53.5 (2010), S. 493–502. DOI: 10.1093/comjnl/bxp047 (siehe S. 14).
- [Gon02] GONZALEZ, Rafael C. and WOODS, Richard E.: *Digital Image Processing*. 2. Aufl. Prentice Hall, 2002 (siehe S. 168, 169).
- [Goo59] GOOD, Irving J.: „Kinds of Probability“. In: *Science* 129.3347 (1959), S. 443–447. DOI: 10.1126/science.129.3347.443 (siehe S. 60).
- [Gor09] GORDON, Marvin: „Lokale Bayes’sche Fusion von Vorwissen und Multispektral-Bildern zur Detektion und Klassifikation von Fahrzeugen“. Studienarbeit. Karlsruher Institut für Technologie (KIT), 2009 (siehe S. 156, 159).
- [Gou08] GOUSSEV, Igor: „Lokale Bayes’sche Fusion von Vorwissen und Bildinformation zur Detektion und Klassifikation von Fahrzeugen“. Studienarbeit. Universität Karlsruhe (TH), 2008 (siehe S. 156).
- [Gou10] GOUSSEV, Igor: „Merkmalsgestützte lokale Bayes’sche Fusion zur Detektion und Klassifikation von Fahrzeugen in einer Szene“. Diplomarbeit. Karlsruher Institut für Technologie (KIT), 2010 (siehe S. 156, 159).
- [Gui77] GUIASU, Silviu: *Information Theory with Applications*. McGraw-Hill Inc., 1977 (siehe S. 77, 78).
- [Hac65] HACKING, Ian: *Logic of Statistical Inference*. Cambridge University Press, 1965 (siehe S. 71).
- [Háj19] HÁJEK, Alan: „Interpretations of Probability“. In: *The Stanford Encyclopedia of Philosophy (Fall 2019 Edition)*. Hrsg. von ZALTA, Edward N. Metaphysics Research Lab, Stanford University, 2019. URL: <https://plato.stanford.edu/archives/fall2019/entries/probability-interpret/> (siehe S. 60).

- [Hal01] HALL, David L. and STEINBERG, Alan N.: „Dirty Secrets in Multisensor Data Fusion“. In: *Handbook of Multisensor Data Fusion*. Hrsg. von HALL, David L. and LLINAS, James. 1. Aufl. CRC Press, 2001. Kap. 2, S. 2-1–2-19 (siehe S. 32).
- [Hal04] HALL, David L. and McMULLEN, Sonya A. H.: *Mathematical Techniques in Multisensor Data Fusion*. 2. Aufl. Artech House, 2004 (siehe S. 4, 13).
- [Hal13] HALL, David L.: „Perspectives on Distributed Data Fusion“. In: *Distributed Data Fusion for Network-Centric Operations*. Hrsg. von HALL, David L.; CHONG, Chee-Yee; LLINAS, James and LIGGINS II, Martin. CRC Press, 2013. DOI: 10.1201/b10124 (siehe S. 31).
- [Hal97] HALL, D. L. and LLINAS, J.: „An introduction to multisensor data fusion“. In: *Proceedings of the IEEE* 85.1 (1997), S. 6–23. DOI: 10.1109/5.554205 (siehe S. 1, 27, 28, 42).
- [Han01] HAND, David J. and YU, Keming: „Idiot’s Bayes? Not So Stupid After All?“ In: *International Statistical Review* 69.3 (2001), S. 385–398. DOI: 10.1111/j.1751-5823.2001.tb00465.x (siehe S. 85).
- [Har14] HARTE, John and NEWMAN, Erica A.: „Maximum information entropy: a foundation for ecological theory“. In: *Trends in Ecology & Evolution* 29.7 (2014), S. 384–389. DOI: 10.1016/j.tree.2014.04.009 (siehe S. 76).
- [Hei08] HEIZMANN, Michael: „Aspects of image fusion for automated visual inspection“. In: *Pattern Recognition and Image Analysis* 18.2 (2008), S. 222–230. DOI: 10.1134/S1054661808020065. URL: <http://dx.doi.org/10.1134/S1054661808020065> (siehe S. 1, 14, 26, 27, 29).
- [Hen10] HENZE, Norbert: *Stochastik für Einsteiger: Eine Einführung in die faszinierende Welt des Zufalls*. 8. Aufl. Vieweg+Teubner Verlag, 2010. DOI: 10.1007/978-3-8348-9351-2 (siehe S. 56, 58, 60, 61, 177).

- [Hit03] HITCHCOCK, David B.: „A History of the Metropolis–Hastings Algorithm“. In: *The American Statistician* 57.4 (2003), S. 254–257. DOI: 10.1198/0003130032413 (siehe S. 100).
- [Hob73] HOBSON, Arthur and CHENG, Bin-Kang: „A comparison of the Shannon and Kullback information measures“. In: *Journal of Statistical Physics* 7.4 (1973), S. 301–310. DOI: 10.1007/bf01014906 (siehe S. 78–80).
- [Hof19] HOFFMANN, Almuth; KUWERTZ, Achim and SANDER, Jennifer: „Towards information extraction and semantic world modelling to support information management and intelligence creation in defense coalitions“. In: *Proc. SPIE 11169, Artificial Intelligence and Machine Learning in Defense Applications, 111690G (19 September 2019)*. SPIE, 2019. DOI: 10.1117/12.2533320 (siehe S. 34).
- [Hol18] HOLZINGER, Andreas: „Explainable AI (ex-AI)“. In: *Informatik-Spektrum* 41.2 (2018), S. 138–143. DOI: 10.1007/s00287-018-1102-5 (siehe S. 25).
- [Hub11] HUBER, Marco; KRAUTHAUSEN, Peter and HANEBECK, Uwe D.: „Huber, Marco; Krauthausen, Peter; Hanebeck, Uw“. In: *INFORMATIK 2011 – Informatik schafft Communities*. Hrsg. von HEIß, Hans-Ulrich; PEPPER, Peter; SCHLINGLOFF, Holger and SCHNEIDER, Jörg. Bonn: Gesellschaft für Informatik e.V., 2011 (siehe S. 94, 95).
- [Inc19] INCEOGLU, Arda; KOC, Cagatay; KANAT, Besim Ongun; ERSEN, Mustafa and SARIEL, Sanem: „Continuous Visual World Modeling for Autonomous Robot Manipulation“. In: *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 49.1 (2019), S. 192–205. DOI: 10.1109/tsmc.2017.2787482 (siehe S. 1).
- [Ins00] INSUA, David Ríos and RUGGERI, Fabrizio, Hrsg.: *Robust Bayesian Analysis*. Springer, 2000. DOI: 10.1007/978-1-4612-1306-2 (siehe S. 94, 214).
- [Irl05] IRLE, Albrecht: *Wahrscheinlichkeitstheorie und Statistik: Grundlagen, Resultate, Anwendungen*. 2. Aufl. Vieweg+Teubner Verlag, 2005 (siehe S. 59).

- [Iro97] IRONY, Telba Z. and SINGPURWALLA, Nozer D.: „Non-informative priors do not exist A dialogue with José M. Bernardo“. In: *Journal of Statistical Planning and Inference* 65.1 (1997), S. 159–177. DOI: 10.1016/s0378-3758(97)00074-8 (siehe S. 63).
- [Iuc06] IUCULANO, G.; PELLEGRINI, G. and ZANOBINI, A.: „On the Use of Minimum Cross Entropy Principle and Bayes’ Theorem for the Uncertainty Evaluation in a Measurement Process“. In: *18th IMEKO World Congress 2006: Metrology for a Sustainable Development, Volume 1*. 2006, S. 526–531 (siehe S. 83).
- [Jam17] JAMES, Alex Pappachen and DASARATHY, Belur V.: „A Review of Feature and Data Fusion with Medical Images“. In: *Multisensor Data Fusion: From Algorithms and Architectural Design to Applications*. Hrsg. von FOURATI, Hassen and INIEWSKI, Krzysztof. CRC Press, 2017, S. 491–507. DOI: 10.1201/b18851-27 (siehe S. 1).
- [Jay57] JAYNES, Edwin T.: „Information Theory and Statistical Mechanics“. In: *Physical Review* 106.4 (1957), S. 620–630. DOI: 10.1103/PhysRev.106.620 (siehe S. 81).
- [Jay68] JAYNES, Edwin T.: „Prior Probabilities“. In: *IEEE Transactions on Systems Science and Cybernetics* 4.3 (1968), S. 227–241. DOI: 10.1109/tssc.1968.300117 (siehe S. 76, 78).
- [Jef61] JEFFREYS, Harold: *Theory of Probability*. 3. Aufl. Oxford University Press, 1961 (siehe S. 177).
- [Jen07] JENSEN, Finn V. and NIELSEN, Thomas D.: *Bayesian Networks and Decision Graphs*. 2. Aufl. Springer, 2007. DOI: 10.1007/978-0-387-68282-2 (siehe S. 137).
- [Kap89] KAPUR, Jagat Narain: *Maximum-Entropy Models in Science and Engineering*. New York (NY): Wiley, 1989 (siehe S. 76, 81).
- [Kas95] KASS, Robert E. and RAFTERY, Adrian E.: „Bayes Factors“. In: *Journal of the American Statistical Association* 90.430 (1995), S. 773–795. DOI: 10.1080/01621459.1995.10476572 (siehe S. 177).

- [Kas96] KASS, Robert E. and WASSERMAN, Larry: „The Selection of Prior Distributions by Formal Rules“. In: *Journal of the American Statistical Association* 91.435 (1996), S. 1343–1370. DOI: 10.1080/01621459.1996.10477003 (siehe S. 76).
- [Kha09] KHALEGHI, Bahador; RAZAVI, Saiedeh N.; KHAMIS, Alaa; KARRAY, Fakhreddine O. and KAMEL, Mohamed: „Multisensor data fusion: Antecedents and directions“. In: *2009 3rd International Conference on Signals, Circuits and Systems (SCS)*. IEEE, 2009. DOI: 10.1109/icscs.2009.5412296 (siehe S. 1, 4, 42, 43, 49).
- [Kir19] KIRSTE, Moritz: „Augmented Intelligence – Wie Menschen mit KI zusammen arbeiten“. In: *Künstliche Intelligenz*. Springer, 2019, S. 58–71. DOI: 10.1007/978-3-662-58042-4\_4 (siehe S. 109).
- [Kla99] KLAUS, Ferdinand: „Einführung in Techniken und Methoden der Multisensor-Datenfusion“. Habilitationsschrift. Universität Siegen, 1999. URL: <https://nbn-resolving.org/urn:nbn:de:hbz:467-575> (siehe S. 4).
- [Kle04] KLEIN, Lawrence A.: *Sensor and Data Fusion: A Tool for Information Assessment and Decision Making*. SPIE, 2004. DOI: 10.1117/3.563340 (siehe S. 4, 19, 42, 44, 45, 47, 84).
- [Kli06] KLIR, George J.: *Uncertainty and information : foundations of generalized information theory*. Hoboken, NJ: Wiley-Interscience, 2006. DOI: 10 . 1002 / 0471755575 (siehe S. 77, 78).
- [Koc05] KOCH, Wolfgang: „Multisensordatenfusion als Kernfunktion in Überwachungssystemen (Multiple Sensor Data Fusion as a Key Function in Surveillance Systems)“. In: *it - Information Technology* 47.2 (2005), S. 79–86. DOI: 10.1524/itit.47.2.79.62259 (siehe S. 1).
- [Koc14] KOCH, Wolfgang: *Tracking and Sensor Data Fusion: Methodological Framework and Selected Applications*. Springer, 2014. DOI: 10.1007/978-3-642-39271-9 (siehe S. 1, 2, 68, 69, 174).



- [Kof76] KOFLER, Eduard and MENGES, Günter: Entscheidungen bei unvollständiger Information. Springer, 1976. DOI: 10.1007/978-3-642-95290-6 (siehe S. 238, 239, 241, 242).
- [Kol33] KOLMOGOROFF, A.: Grundbegriffe der Wahrscheinlichkeitsrechnung. Springer, 1933. DOI: 10.1007/978-3-642-49888-6 (siehe S. 57).
- [Krü08] KRÜGER, Max and ZIEGLER, Jürgen: „User-oriented Bayesian identification and its configuration“. In: *2008 11th International Conference on Information Fusion (FUSION)*. IEEE, 2008 (siehe S. 69).
- [Kuk18] KUKKALA, Vipin Kumar; TUNNELL, Jordan; PASRICHA, Sudeep and BRADLEY, Thomas: „Advanced Driver-Assistance Systems: A Path Toward Autonomous Vehicles“. In: *IEEE Consumer Electronics Magazine* 7.5 (2018), S. 18–25. DOI: 10.1109/mce.2018.2828440 (siehe S. 1).
- [Kul59] KULLBACK, Solomon: Information theory and statistics. Wiley, 1959 (siehe S. 78, 79).
- [Kuw13] KUWERTZ, Achim; SANDER, Jennifer; SCHNEIDER, Gerd and ESSENDORFER, Barbara: „An Architectural Framework for ISR Analytics“. In: *Proceedings of the NATO Symposium on Architecture Assessment for NEC (SCI-254), Systems Concepts and Integration Panel, Tallinn, Estonia, 14th to 15th May 2013*. 2013 (siehe S. 34).
- [Kuw17] KUWERTZ, Achim; SANDER, Jennifer; PFIRRMANN, Uwe and DYCK, Sergius: „High-Level information management in Joint ISR based on an object-oriented approach“. In: *2017 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*. IEEE, 2017. DOI: 10.1109/sdf.2017.8126360 (siehe S. 27).
- [Kuw18] KUWERTZ, Achim; MÜHLENBERG, Dirk; SANDER, Jennifer and MÜLLER, Wilmoth: „Applying Knowledge-Based Reasoning for Information Fusion in Intelligence, Surveillance, and Reconnaissance“. In: *Multisensor Fusion and Integration in the Wake of Big Data, Deep Learning and Cyber Physical System: An Edition of*

- the Selected Papers from the 2017 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2017)*. Hrsg. von LEE, Sukhan; KO, Hanseok and OH, Songhwai. Lecture Notes in Electrical Engineering (LNEE), Bd. 501. Springer, 2018, S. 119–139. DOI: 10.1007/978-3-319-90509-9\_7 (siehe S. 1, 4, 34).
- [Kuw19] KUWERTZ, Achim; MOLL, Maximilian; SANDER, Jennifer and PICKL, Stefan: „A Systemic Approach for Early Warning in Crisis Prevention and Management“. In: *Human Systems Engineering and Design II. Advances in Intelligent Systems and Computing (AISC)*, Bd. 1026. Springer, 2019, S. 517–522. DOI: 10.1007/978-3-030-27928-8\_78 (siehe S. 3, 34).
- [Kuy82] KUYKENDALL, Jack: „The criminal investigative process: Toward a conceptual framework“. In: *Journal of Criminal Justice* 10.2 (1982), S. 131–145. DOI: 10.1016/0047-2352(82)90004-6 (siehe S. 105).
- [Las91] LASKEY, Kathryn Blackmond: „Conflict and Surprise: Heuristics for Model Revision“. In: *Uncertainty Proceedings 1991: Proceedings of the Seventh Conference (1991)*. Hrsg. von D’AMBROSIO, Bruce D.; SMETS, Philippe and BONISSONE, Piero P.. Elsevier, 1991, S. 197–204. DOI: 10.1016/b978-1-55860-203-8.50029-2 (siehe S. 178).
- [Las93] LASKEY, Kathryn Blackmond and LEHNER, Paul Edward: „Towards a theory of small worlds“. In: *1993 (2nd) International Symposium on Uncertainty Modeling and Analysis*. IEEE, 1993. DOI: 10.1109/isuma.1993.366749 (siehe S. 151).
- [Las94] LASKEY, Kathryn Blackmond and LEHNER, Paul Edward: „Metareasoning and the problem of small worlds“. In: *IEEE Transactions on Systems, Man, and Cybernetics* 24.11 (1994), S. 1643–1652. DOI: 10.1109/21.328913 (siehe S. 151).
- [Lew95] LEWIS, J. P.: *Fast Normalized Cross-Correlation*. Industrial Light und Magic, 1995 (siehe S. 168).

- [Lig08] LIGGINS, Martin E. and CHANG, Kuo-Chu: „Distributed Fusion Architectures, Algorithms, and Performance within a Network-Centric Architecture“. In: *Handbook of Multisensor Data Fusion*. Hrsg. von LIGGINS, Martin E.; HALL, David L. and LINAS, James. 2. Aufl. CRC Press, 2008, S. 411–435 (siehe S. 31, 43, 48, 51).
- [Lin56] LINDLEY, D. V.: „On a Measure of the Information Provided by an Experiment“. In: *The Annals of Mathematical Statistics* 27.4 (1956), S. 986–1005. DOI: 10.1214/aoms/1177728069 (siehe S. 202, 203).
- [Lin87] LINDLEY, Dennis V.: „The Probability Approach to the Treatment of Uncertainty in Artificial Intelligence and Expert Systems“. In: *Statistical Science* 2.1 (1987), S. 17–24 (siehe S. 62).
- [Liu02] LIU, Chao-Lin and WELLMAN, Michael P.: „Evaluation of Bayesian networks with flexible state-space abstraction methods“. In: *International Journal of Approximate Reasoning* 30.1 (2002), S. 1–39. DOI: 10.1016/s0888-613x(01)00067-6 (siehe S. 137).
- [Liu98] LIU, Chao-Lin: „State-space abstraction methods for approximate evaluation of Bayesian networks.“ Dissertation. University of Michigan, 1998 (siehe S. 137, 140, 144).
- [Luc05] LUCY, David: *Introduction to Statistics for Forensic Scientists*. John Wiley & Sons, Inc., 2005 (siehe S. 73, 177).
- [Mac10] MACHMER, Timo: „Generierung und Fusion von Umweltwissen für eine wissensbasierte Umwelterfassung“. Dissertation. Karlsruher Institut für Technologie (KIT), 2010 (siehe S. 1).
- [Man94] MANYIKA, James and DURRANT-WHYTE, Hugh: *Data Fusion and Sensor Management: A Decentralized Information-Theoretic Approach*. Ellis Horwood, 1994 (siehe S. 42, 44–47).
- [Mar07] MARIN, Jean-Michel and ROBERT, Christian P.: *Bayesian Core: A Practical Approach to Computational Bayesian Statistics*. Springer, 2007. DOI: 10.1007/978-0-387-38983-7 (siehe S. 98, 102).

- [Mar95] MARTIN, Jerome and CROWLEY, James L.: „Comparison of Correlation Techniques“. In: *Intelligent Autonomous Systems IAS-4: Proceedings of the International Conference Karlsruhe, Germany, March 27-30, 1995*. Hrsg. von REMBOLD, Ulrich; DILLMANN, Rüdiger; HERTZBERGER, Louis O. and KANADE, Takeo. IOS Press, 1995 (siehe S. 163).
- [Mei05] MEINTRUP, David and SCHÄFFLER, Stefan: *Stochastik: Theorie und Anwendungen. Statistik und ihre Anwendungen*. Springer, 2005. DOI: 10.1007/b137972 (siehe S. 126).
- [Men20] MENG, Tong; JING, Xuyang; YAN, Zheng and PEDRYCZ, Witold: „A survey on machine learning for data fusion“. In: *Information Fusion 57* (2020), S. 115–129. DOI: 10.1016/j.inffus.2019.12.001 (siehe S. 4, 27).
- [Mey18] MEYER, Johannes; MELCHERT, Wolfgang; HARTRUMPF, Matthias; LÄNGLE, Thomas and BEYERER, Jürgen: „SNR-optimized image fusion for transparent object inspection“. In: *Proc. SPIE 10677, Unconventional Optical Imaging, 106770A (24 May 2018)*. SPIE, 2018. DOI: 10.1117/12.2307392 (siehe S. 1).
- [Mey93] MEYN, Sean P. and TWEEDIE, Richard L.: *Markov Chains and Stochastic Stability*. Springer, 1993. DOI: 10.1007/978-1-4471-3267-7 (siehe S. 101).
- [Mit07] MITCHELL, H. B.: *Multi-Sensor Data Fusion*. Springer, 2007. DOI: 10.1007/978-3-540-71559-7 (siehe S. 4, 42, 44, 47).
- [Moh90] MOHAMMAD-DJAFARI, A. and DEMOMENT, G.: „Estimating priors in maximum entropy image processing“. In: *International Conference on Acoustics, Speech, and Signal Processing*. IEEE, 1990. DOI: 10.1109/icassp.1990.115936 (siehe S. 76).
- [Mor17] MORA, André D.; FALCÃO, António J.; MIRANDA, Luís; RIBEIRO, Rita A. and FONSECA, José M.: „A Fuzzy Multicriteria Approach for Data Fusion“. In: *Multisensor Data Fusion*. Hrsg. von FOURATI, Hassen and INIEWSKI, Krzysztof. CRC Press, 2017, S. 109–125. DOI: 10.1201/b18851-7 (siehe S. 30).

- [Mül07] MÜLLER, Markus; KRÜGER, Wolfgang and SAUR, Günter: „Robust image registration for fusion“. In: *Information Fusion* 8.4 (2007), S. 347–353. DOI: 10.1016/j.inffus.2006.05.001 (siehe S. 26).
- [Mül18a] MÜLLER, Wilmuth; REINERT, Frank and PALLMER, Dirk: „Open architecture of a counter UAV system“. In: *Proc. SPIE 10651, Open Architecture/Open Business Model Net-Centric Systems and Defense Transformation 2018, 1065106* (9 May 2018). SPIE, 2018. DOI: 10.1117/12.2305606 (siehe S. 30).
- [Mül18b] MÜLLER, Wilmuth; SANDER, Jennifer; KUWERTZ, Achim and MÜHLENBERG, Dirk: „High-level data fusion component for drone classification and decision support in counter UAV“. In: *Proc. SPIE 10651, Open Architecture/Open Business Model Net-Centric Systems and Defense Transformation 2018, 106510F* (9 May 2018). SPIE, 2018. DOI: 10.1117/12.395578 (siehe S. 30).
- [NAT09] NATO STANDARDIZATION OFFICE (NSO): STANAG 4162, Identification Data Combining Process, Edition 2, February 2009. NATO Standard. 2009 (siehe S. 69).
- [NAT16] NATO STANDARDIZATION OFFICE (NSO): AJP-2.1, Allied Joint Doctrine for Intelligence Procedures, Edition B, Version 1, June 2016. NATO Standard. 2016 (siehe S. 41).
- [NAT19] NATO STANDARDIZATION OFFICE (NSO): AAP-06 Edition 2019, NATO Glossary of Terms and Definitions (Englisch and French), Version 1, November 2019. NATO Standard. 2019 (siehe S. 16, 18, 40).
- [Neu04] NEUMANN, Klaus and MORLOCK, Martin: Operations Research. 2. Aufl. Hanser, 2004 (siehe S. 242).
- [Not16] NOTHWANG, William D.; McCOURT, Michael J.; ROBINSON, Ryan M.; BURDEN, Samuel A. and CURTIS, J. Willard: „The human should be part of the control loop?“ In: *2016 Resilience Week (RWS)*. IEEE, 2016. DOI: 10.1109/RWEEK.2016.7573336 (siehe S. 34).

- [Ntz09] NTZOUFRAS, Ioannis: Bayesian Modeling Using WinBUGS. Wiley, 2009. DOI: 10.1002/9780470434567 (siehe S. 81, 100, 102, 103).
- [Obe07] OBERMAIER, Michael and AUGUSTIN, Thomas: „Luceno’s discretization method and its application in decision making under ambiguity“. In: *Proceedings of the Fifth International Symposium on Imprecise Probabilities and their Applications (ISIPTA '07)*. SIP-TA, 2007, S. 327–336 (siehe S. 238).
- [Par06] PARIS, Jeff B.: The Uncertain Reasoner’s Companion: A Mathematical Perspective. Cambridge University Press, 2006 (siehe S. 79).
- [Pre02] PRESSER, Gero: „Lazy Decision Making – Entscheiden durch zielgerichtetes Präzisieren der Wahrscheinlichkeitsinformation“. Dissertation. Universität Dortmund, 2002. DOI: 10.17877/DE290R-14789 (siehe S. 237, 238, 242–244, 248).
- [Rai61] RAIFFA, Howard and SCHLAIFER, Robert: Applied statistical decision theory. Division of Research, Graduate School of Business Administration, Harvard University, 1961 (siehe S. 94).
- [Ral07] RALESCU, Anca and MINOH, Michihiko: „Measuring Proximity between Heterogeneous Data“. In: *2007 IEEE International Fuzzy Systems Conference (FUZZ-IEEE)*. IEEE, 2007. DOI: 10.1109/fuzzy.2007.4295603 (siehe S. 16).
- [Reh96] REHÄUSER, Jakob and KRČMAR, Helmut: „Wissensmanagement im Unternehmen“. In: *Wissensmanagement*. Hrsg. von SCHREYÖGG, Georg and CONRAD, Peter. de Gruyter, 1996 (siehe S. 22).
- [Rév17] RÉVAY, M. and LÍŠKA, M.: „OODA loop in command & control systems“. In: *2017 Communication and Information Technologies (KIT)*. IEEE, 2017. DOI: 10.23919/kit.2017.8109463 (siehe S. 38).
- [Ris01] RISH, Irina: „An Empirical Study of the Naïve Bayes Classifier“. In: *IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence*. Bd. 3. 22. 2001, S. 41–46 (siehe S. 85).

- [Rob01] ROBERT, Christian P.: The Bayesian choice: from decision-theoretic foundations to computational implementation. 2. Aufl. Springer, 2001. DOI: 10.1007/0-387-71599-1 (siehe S. 76, 92–94, 96).
- [Rob70] ROBBINS, Herbert: „Statistical Methods Related to the Law of the Iterated Logarithm“. In: *The Annals of Mathematical Statistics* 41.5 (1970), S. 1397–1409. DOI: 10.1214/aoms/1177696786 (siehe S. 179).
- [Rot16] ROTH, Florian and HERZOG, Michel: „Strategische Krisenfrüherkennung – Instrumente, Möglichkeiten und Grenzen“. In: *Zeitschrift für Außen- und Sicherheitspolitik* 9.2 (2016), S. 201–211. DOI: 10.1007/s12399-016-0556-2 (siehe S. 3).
- [Row07] ROWLEY, Jennifer: „The wisdom hierarchy: representations of the DIKW hierarchy“. In: *Journal of Information Science* 33.2 (2007), S. 163–180. DOI: 10.1177/0165551506070706 (siehe S. 21).
- [Roy00] ROYALL, Richard: „On the Probability of Observing Misleading Statistical Evidence“. In: *Journal of the American Statistical Association* 95.451 (2000), S. 760–768. DOI: 10.1080/01621459.2000.10474264 (siehe S. 177–179).
- [Roy01] ROY, Jean: „From Data Fusion to Situation Analysis“. In: *Proceedings of the Fourth International Conference on Information Fusion (FUSION)*. 2001 (siehe S. 39).
- [Roy97] ROYALL, Richard M.: *Statistical Evidence: a Likelihood Paradigm*. London: Chapman & Hall, 1997 (siehe S. 62, 71, 72, 177–179).
- [Rüg99] RÜGER, Bernhard: *Test- und Schätztheorie: Band I: Grundlagen*. R. Oldenbourg Verlag, 1999. DOI: 10.1524/9783486599701 (siehe S. 55, 63, 69, 70, 72, 81, 237).
- [Rus07] RUSER, Heinrich and PUENTE LEÓN, Fernando: „Informationsfusion – Eine Übersicht (Information Fusion – An Overview)“. In: *tm – Technisches Messen* 74.3 (2007), S. 93–102. DOI: 10.1524/teme.2007.74.3.93 (siehe S. 26).

- [San06] SANDER, Jennifer and BEYERER, Jürgen: „Fusion agents - realizing Bayesian fusion via a local approach“. In: *2006 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*. IEEE, 2006. DOI: 10.1109/mfi.2006.265619 (siehe S. 121).
- [San07] SANDER, Jennifer and BEYERER, Jürgen: „Local Bayesian fusion realized via an agent based architecture“. In: *INFORMATIK 2007 – Informatik trifft Logistik – Band 2*. Hrsg. von KOSCHKE, Rainer; HERZOG, Otthein; RÖDIGER, Karl-Heinz and RONTALER, Marc. Bonn: Gesellschaft für Informatik e.V., 2007 (siehe S. 121).
- [San08] SANDER, Jennifer and BEYERER, Jürgen: „Decreased complexity and increased problem specificity of Bayesian fusion by local approaches“. In: *2008 11th International Conference on Information Fusion (FUSION)*. IEEE, 2008 (siehe S. 121, 179, 184).
- [San09a] SANDER, Jennifer: „Further Investigation of Focussed Bayesian Fusion“. In: *Proceedings of the 2009 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory*. Hrsg. von BEYERER, Jürgen and HUBER, Marco. Karlsruhe Schriften zur Anthropomatik, Bd. 4. Technischer Bericht IES-2009-03. KIT Scientific Publishing, Karlsruhe, 2009, S. 31–46 (siehe S. 194, 217).
- [San09b] SANDER, Jennifer and BEYERER, Jürgen: „A local approach for Bayesian fusion: Mathematical analysis and agent based conception“. In: *Robotics and Autonomous Systems* 57.3 (2009), S. 259–267. DOI: 10.1016/j.robot.2008.10.005 (siehe S. 121).
- [San09c] SANDER, Jennifer; HEIZMANN, Michael; GOUSSEV, Igor and BEYERER, Jürgen: „A local approach for focussed Bayesian fusion“. In: *Proc. SPIE 7345, Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2009, 73450N (13 April 2009)*. SPIE, 2009. DOI: 10.1117/12.820165 (siehe S. 90, 152, 156, 166, 179, 194).



- [San10a] SANDER, Jennifer: „Use of a Probability Interval Scheme for Focussed Bayesian Fusion“. In: *Proceedings of the 2010 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory*. Hrsg. von BEYERER, Jürgen and HUBER, Marco. Karlsruher Schriften zur Anthropomatik, Bd. 7. Technischer Bericht IES-2010-03. KIT Scientific Publishing, Karlsruhe, 2010, S. 31–46 (siehe S. 224, 237, 246).
- [San10b] SANDER, Jennifer; HEIZMANN, Michael; GOUSSEV, Igor and BEYERER, Jürgen: „Global evaluation of focussed Bayesian fusion“. In: *Proc. SPIE 7710, Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2010, 77100A (28 April 2010)*. SPIE, 2010. DOI: 10.1117/12.850028 (siehe S. 13, 14, 156, 157, 166, 217, 231).
- [San10c] SANDER, Jennifer; KRIEGER, Jonas and BEYERER, Jürgen: „The Importance of Statistical Evidence for Focussed Bayesian Fusion“. In: *KI 2010: Advances in Artificial Intelligence*. Lecture Notes in Computer Science (LNCS), Bd. 6359. Springer, 2010, S. 299–308. DOI: 10.1007/978-3-642-16111-7\_34 (siehe S. 179, 185).
- [San11] SANDER, Jennifer and BEYERER, Jürgen: „Decision theoretic approaches for focussed Bayesian fusion“. In: *INFORMATIK 2011 – Informatik schafft Communities*. Hrsg. von HEIß, Hans-Ulrich; PEPPER, Peter; SCHLINGLOFF, Holger and SCHNEIDER, Jörg. Bonn: Gesellschaft für Informatik e.V., 2011 (siehe S. 237).
- [San12] SANDER, Jennifer; SCHNEIDER, Gerd; ESSENDORFER, Barbara and KUWERTZ, Achim: „ISR analytics: Architectural and methodic concepts“. In: *2012 Workshop on Sensor Data Fusion: Trends, Solutions, Applications (SDF)*. IEEE, 2012. DOI: 10.1109/sdf.2012.6327916 (siehe S. 19).
- [San13] SANDER, Jennifer and BEYERER, Jürgen: „Bayesian fusion: Modeling and application“. In: *2013 Workshop on Sensor Data Fusion: Trends, Solutions, Applications (SDF)*. IEEE, 2013. DOI: 10.1109/sdf.2013.6698254 (siehe S. 91).

- [San14] SANDER, Jennifer and HEIZMANN, Michael: „Embedding the results of focussed Bayesian fusion into a global context“. In: *Proc. SPIE 9121, Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2014, 912103 (22 May 2014)*. SPIE, 2014. DOI: 10.1117/12.2050323 (siehe S. 217, 224).
- [San19] SANDER, Jennifer; ESSENDORFER, Barbara; DYCK, Sergius and PFIRRMANN, Uwe: „Formal reporting for information exploitation and dissemination in Joint ISR“. In: *Proc. SPIE 11015, Open Architecture/Open Business Model Net-Centric Systems and Defense Transformation 2019, 110150Q (30 April 2019)*. SPIE, 2019. DOI: 10.1117/12.2518600 (siehe S. 17).
- [Sav54] SAVAGE, Leonard J.: *The Foundations of Statistics*. John Wiley & Sons, Inc., 1954 (siehe S. 151).
- [Sch06] SCHÖNBEIN, Rainer: „Agenten- und ontologiebasierte Software-Architektur zur interaktiven Bildauswertung“. Dissertation. Universität Karlsruhe (TH), 2006 (siehe S. 2, 4).
- [Sha48] SHANNON, C. E.: „A mathematical theory of communication“. In: *The Bell System Technical Journal* 27.3 (1948), S. 379–423. DOI: 10.1002/j.1538-7305.1948.tb01338.x (siehe S. 76).
- [Sni11] SNIDARO, Lauro; VISENTINI, Ingrid and FORESTI, Gian Luca: „Data Fusion in Modern Surveillance“. In: *Innovations in Defence Support Systems – 3*. Hrsg. von REMAGNINO, Paolo; MONEKOSSO, Dorothy and JAIN, Lakhmi C. Bd. 336. *Studies in Computational Intelligence*. Springer, 2011, S. 1–21. DOI: 10.1007/978-3-642-18278-5\_1 (siehe S. 18).
- [Som09] SOMMER, Klaus-Dieter; KÜHN, Olaf; PUENTE LEÓN, Fernando and SIEBERT, Bernd R. L.: „A Bayesian approach to information fusion for evaluating the measurement uncertainty“. In: *Robotics and Autonomous Systems* 57.3 (2009), S. 339–344. DOI: 10.1016/j.robot.2008.10.011 (siehe S. 1, 76).

- [Ste01] STEINBERG, Allan N. and BOWMAN, Christopher L.: „Revisions to the JDL Data Fusion Model“. In: *Handbook of Multisensor Data Fusion*. Hrsg. von HALL, David L. and LLINAS, James. 1. Aufl. CRC Press, 2001. Kap. 2, S. 2-1–2-19 (siehe S. 31, 32, 36).
- [Ste08] STEINBERG, Allan N. and BOWMAN, Christopher L.: „Revisions to the JDL Data Fusion Model“. In: *Handbook of Multisensor Data Fusion*. Hrsg. von LIGGINS, Martin E.; HALL, David L. and LLINAS, James. 2. Aufl. CRC Press, 2008, S. 45–67 (siehe S. 31–33, 36).
- [Ste99] STEINBERG, Alan N.; BOWMAN, Christopher L. and WHITE, Franklin E.: „Revisions to the JDL data fusion model“. In: *Proc. SPIE 3719, Sensor Fusion: Architectures, Algorithms, and Applications III, (12 March 1999)*. SPIE, 1999. DOI: 10.1117/12.341367 (siehe S. 19, 31).
- [Stü04] STÜKER, Dirk: „Heterogene Sensordatenfusion zur robusten Objektverfolgung im automobilen Straßenverkehr“. Dissertation. Universität Oldenburg, 2004. URL: <https://oops.uni-oldenburg.de/id/eprint/201> (siehe S. 1).
- [The18] THEOBALD, Elke: *Marketing Intelligence: Ein Lehrbuch für die Praxis*. W. Kohlhammer GmbH, 2018 (siehe S. 41).
- [Tot08] TOTH, Gary; KOKAR, Mieczyslaw M.; WALLENIUS, Klas; LASKEY, Kathryn B.; SUDIT, Moises; HULTNER, Michael and KESSLER, Otto: „Higher-Level Information Fusion: Challenges to the Academic Community“. In: *Proceedings of the 11th International Conference on Information Fusion (FUSION)*. IEEE, 2008 (siehe S. 33, 69).
- [Tu02] TU, Zhuowen and ZHU, Song-Chun: „Image segmentation by data-driven Markov chain Monte Carlo“. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24.5 (2002), S. 657–673. DOI: 10.1109/34.1000239 (siehe S. 103).
- [Utk05] UTKIN, Lev and AUGUSTIN, Thomas: „Powerful Algorithms for Decision Making under Partial Prior Information and General

- Ambiguity Attitudes“. In: *Proceedings of the Fourth International Symposium on Imprecise Probabilities and their Applications (ISIPTA '05)*. SIPTA, 2005, S. 349–358 (siehe S. 238).
- [Utk09] UTKIN, Lev and DESTERCKE, Sebastien: „Computing expectations with continuous p-boxes: Univariate case“. In: *International Journal of Approximate Reasoning* 50.5 (2009), S. 778–798. DOI: 10.1016/j.ijar.2009.02.004 (siehe S. 238).
- [Van81] VAN FRAASSEN, Bas C.: „A Problem for Relative Information Minimizers in Probability Kinematics“. In: *The British Journal for the Philosophy of Science* 32.4 (1981), S. 375–379. DOI: 10.1093/bjps/32.4.375 (siehe S. 79, 230).
- [Vio01] VIOLA, Paul and JONES, Michael: „Rapid object detection using a boosted cascade of simple features“. In: *2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2001. DOI: 10.1109/CVPR.2001.990517 (siehe S. 168).
- [Wal06] WALKER, S. G.: „Bayesian Inference via a Minimization Rule“. In: *Sankhyā: The Indian Journal of Statistics* 68.4 (2006), S. 542–553 (siehe S. 199).
- [Wal90] WALTZ, Edward and LLINAS, James: *Multisensor Data Fusion*. Artech House, 1990 (siehe S. 33, 34).
- [Wal91] WALLEY, Peter: *Statistical Reasoning with Imprecise Probabilities*. Chapman & Hall/CRC, 1991 (siehe S. 229, 238).
- [Wal99] WALD, Lucien: „Some terms of reference in data fusion“. In: *IEEE Transactions on Geoscience and Remote Sensing* 37.3 (1999), S. 1190–1193. DOI: 10.1109/36.763269 (siehe S. 19).
- [Wei01] WEICHSELBERGER, Kurt: *Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung, Band 1, Intervallwahrscheinlichkeit als umfassendes Konzept*. Physica-Verlag, 2001. DOI: 10.1007/978-3-642-57583-9 (siehe S. 238).
- [Wei05] WEIß, Gerhard and JAKOB, Ralf: *Agentenorientierte Softwareentwicklung*. Springer, 2005. DOI: 10.1007/b138124 (siehe S. 107).

- [Wei99a] WEISE, Klaus and WÖGER, Wolfgang: Meßunsicherheit und Meßdatenauswertung. Wiley, 1999. DOI: 10.1002/3527602984 (siehe S. 76, 83).
- [Wei99b] WEISS, Gerhard: Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence. The MIT Press, 1999 (siehe S. 108).
- [Whi87] WHITE, Franklin E.: Data Fusion Lexicon, Joint Directors of Laboratories, Technical Panel for C3, Data Fusion Sub-Panel, Naval Ocean Systems Center, San Diego. 1987 (siehe S. 19).
- [Wil80] WILLIAMS, Peter M.: „Bayesian Conditionalisation and the Principle of Minimum Information“. In: *The British Journal for the Philosophy of Science* 31.2 (1980), S. 131–144. DOI: 10.1093/bjps/31.2.131 (siehe S. 79, 203, 205).
- [Woo09] WOOLDRIDGE, Michael: Introduction to Multiagent Systems. 2. Aufl. Wiley, 2009 (siehe S. 107).
- [Wun09] WUNDER, Michael and GROSCHE, Jürgen: Verteilte Führungsinformationssysteme. Springer, 2009. DOI: 10.1007/978-3-642-00509-1 (siehe S. 156).
- [Zar99] ZARNEKOW, Rüdiger: „Softwareagenten“. In: *Softwareagenten und elektronische Kaufprozesse*. Deutscher Universitätsverlag, 1999, S. 9–96. DOI: 10.1007/978-3-322-95205-9\_2 (siehe S. 107).
- [Zas16] ZASCHKE, Christian; ESSENDORFER, Barbara and KERTH, Christian: „Interoperability of heterogeneous distributed systems“. In: *Proc. SPIE 9825, Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security, Defense, and Law Enforcement Applications XV, 98250Q (12 May 2016)*. SPIE, 2016. DOI: 10.1117/12.2223895 (siehe S. 3).
- [Zel02] ZELLNER, Arnold: „Information processing and Bayesian analysis“. In: *Journal of Econometrics* 107.1-2 (2002), S. 41–50. DOI: 10.1016/s0304-4076(01)00112-9 (siehe S. 198).

- [Zel88] ZELLNER, Arnold: „Optimal Information Processing and Bayes’s Theorem“. In: *The American Statistician* 42.4 (1988), S. 278–280. DOI: 10.1080/00031305.1988.10475585 (siehe S. 195–198, 201).
- [Zha01] ZHAO, Tao and NEVATIA, R.: „Car detection in low resolution aerial image“. In: *2001 Eighth IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2001. DOI: 10.1109/ICCV.2001.937593 (siehe S. 166).
- [Zhu00] ZHU, Song-Chun; TU, Zhuowen and ZHANG, Rong: „Effective Bayesian inference by data-driven Markov chain Monte Carlo for object recognition and image segmentation“. In: *Proc. SPIE 4050, Automatic Target Recognition X, (17 August 2000)*. SPIE, 2000. DOI: 10.1117/12.2306148 (siehe S. 103).

## Publikationen

Dieser Abschnitt enthält ein vollständiges Verzeichnis der eigenen Veröffentlichungen. Die Publikationen [19], [23], [30], [32], [36] adressieren speziell die Ressourcen-optimale Sensoreinsatzplanung und die Publikationen [21], [22] adressieren speziell das Management von Open Source Software Lizenzbedingungen in Softwareprojekten. Sie können nur im weiteren Sinne als relevant für die vorliegende Arbeit angesehen werden.

- [1] BEYERER, Jürgen; HEIZMANN, Michael and SANDER, Jennifer: „Fuselets: an agent based architecture for fusion of heterogeneous information and data“. In: *Proc. SPIE 6242, Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2006, 62420Q (18 April 2006)*. SPIE, 2006. DOI: 10.1117/12.665938.
- [2] BEYERER, Jürgen; SANDER, Jennifer and WERLING, Stefan: „Fusion heterogener Informationsquellen“. In: *Informationsfusion in der Mess- und Sensortechnik*. Hrsg. von BEYERER, Jürgen; PUENTE LEÓN, Fernando and SOMMER, Klaus-Dieter. Universitätsverlag Karlsruhe, 2006, S. 21–37.
- [3] SANDER, Jennifer and BEYERER, Jürgen: „Fusion agents - realizing Bayesian fusion via a local approach“. In: *2006 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*. IEEE, 2006. DOI: 10.1109/mfi.2006.265619.
- [4] BEYERER, Jürgen; SANDER, Jennifer and WERLING, Stefan: „Bayes'sche Methodik zur lokalen Fusion heterogener Informationsquellen (Bayesian Methodology for the Local Fusion of Heterogeneous Information Sources)“. In: *tm – Technisches Messen* 74.3 (2007), S. 103–111. DOI: 10.1524/teme.2007.74.3.103.

- [5] SANDER, Jennifer and BEYERER, Jürgen: „Local Bayesian fusion realized via an agent based architecture“. In: *INFORMATIK 2007 – Informatik trifft Logistik – Band 2*. Hrsg. von KOSCHKE, Rainer; HERZOG, Otthein; RÖDIGER, Karl-Heinz and RONTHALER, Marc. Bonn: Gesellschaft für Informatik e.V., 2007.
- [6] BEYERER, Jürgen; HEIZMANN, Michael; SANDER, Jennifer and GHETA, Ioana: „Bayesian methods for image fusion“. In: *Image Fusion: Algorithms and Applications*. Hrsg. von STATHAKI, Tania. Academic Press, 2008, S. 157–192. DOI: 10.1016/b978-0-12-372529-5.00003-2.
- [7] SANDER, Jennifer and BEYERER, Jürgen: „Decreased complexity and increased problem specificity of Bayesian fusion by local approaches“. In: *2008 11th International Conference on Information Fusion (FUSION)*. IEEE, 2008.
- [8] SANDER, Jennifer: „Further Investigation of Focussed Bayesian Fusion“. In: *Proceedings of the 2009 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory*. Hrsg. von BEYERER, Jürgen and HUBER, Marco. Karlsruher Schriften zur Anthropomatik, Bd. 4. Technischer Bericht IES-2009-03. KIT Scientific Publishing, Karlsruhe, 2009, S. 31–46.
- [9] SANDER, Jennifer and BEYERER, Jürgen: „A local approach for Bayesian fusion: Mathematical analysis and agent based conception“. In: *Robotics and Autonomous Systems* 57.3 (2009), S. 259–267. DOI: 10.1016/j.robot.2008.10.005.
- [10] SANDER, Jennifer; HEIZMANN, Michael; GOUSSEV, Igor and BEYERER, Jürgen: „A local approach for focussed Bayesian fusion“. In: *Proc. SPIE 7345, Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2009, 73450N (13 April 2009)*. SPIE, 2009. DOI: 10.1117/12.820165.
- [11] SANDER, Jennifer: „Use of a Probability Interval Scheme for Focused Bayesian Fusion“. In: *Proceedings of the 2010 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion*



- Laboratory*. Hrsg. von BEYERER, Jürgen and HUBER, Marco. Karlsruhe Schriften zur Anthropomatik, Bd. 7. Technischer Bericht IES-2010-03. KIT Scientific Publishing, Karlsruhe, 2010, S. 31–46.
- [12] SANDER, Jennifer; HEIZMANN, Michael; GOUSSEV, Igor and BEYERER, Jürgen: „Global evaluation of focussed Bayesian fusion“. In: *Proc. SPIE 7710, Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2010, 77100A (28 April 2010)*. SPIE, 2010. DOI: 10.1117/12.850028.
- [13] SANDER, Jennifer; KRIEGER, Jonas and BEYERER, Jürgen: „The Importance of Statistical Evidence for Focussed Bayesian Fusion“. In: *KI 2010: Advances in Artificial Intelligence*. Lecture Notes in Computer Science (LNCS), Bd. 6359. Springer, 2010, S. 299–308. DOI: 10.1007/978-3-642-16111-7\_34.
- [14] SANDER, Jennifer and BEYERER, Jürgen: „Decision theoretic approaches for focussed Bayesian fusion“. In: *INFORMATIK 2011 – Informatik schafft Communities*. Hrsg. von HEIß, Hans-Ulrich; PEPPER, Peter; SCHLINGLOFF, Holger and SCHNEIDER, Jörg. Bonn: Gesellschaft für Informatik e.V., 2011.
- [15] SANDER, Jennifer; SCHNEIDER, Gerd; ESSENDORFER, Barbara and KUWERTZ, Achim: „ISR analytics: Architectural and methodic concepts“. In: *2012 Workshop on Sensor Data Fusion: Trends, Solutions, Applications (SDF)*. IEEE, 2012. DOI: 10.1109/sdf.2012.6327916.
- [16] KUWERTZ, Achim; SANDER, Jennifer; SCHNEIDER, Gerd and ESSENDORFER, Barbara: „An Architectural Framework for ISR Analytics“. In: *Proceedings of the NATO Symposium on Architecture Assessment for NEC (SCI-254), Systems Concepts and Integration Panel, Tallinn, Estonia, 14th to 15th May 2013*. 2013.
- [17] SANDER, Jennifer and BEYERER, Jürgen: „Bayesian fusion: Modeling and application“. In: *2013 Workshop on Sensor Data Fusion: Trends, Solutions, Applications (SDF)*. IEEE, 2013. DOI: 10.1109/sdf.2013.6698254.

- [18] SANDER, Jennifer and HEIZMANN, Michael: „Embedding the results of focussed Bayesian fusion into a global context“. In: *Proc. SPIE 9121, Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2014, 912103 (22 May 2014)*. SPIE, 2014. DOI: 10.1117/12.2050323.
- [19] REINERT, Frank and SANDER, Jennifer: „Supporting Sensor Scheduling in Intelligence“. In: *2015 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*. IEEE, 2015. DOI: 10.1109/sdf.2015.7347711 (siehe S. 291).
- [20] BOUVRY, Pascal; CHAUMETTE, Serge; DANOY, Gregoire; GUERRINI, Gilles; JURQUET, Gilles; KUWERTZ, Achim; MÜLLER, Wilmuth; ROSALIE, Martin and SANDER, Jennifer: „Using heterogeneous multilevel swarms of UAVs and high-level data fusion to support situation management in surveillance scenarios“. In: *2016 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*. IEEE, 2016. DOI: 10.1109/mfi.2016.7849525.
- [21] DYCK, Sergius; HAFERKORN, Daniel and SANDER, Jennifer: „An organizational-technical concept to deal with open source software license terms“. In: *20th World Multi-Conference on Systemics, Cybernetics and Informatics, WMSCI 2016*. 2016 (siehe S. 291).
- [22] DYCK, Sergius; HAFERKORN, Daniel and SANDER, Jennifer: „An organizational-technical concept to deal with open source software license terms“. In: *Journal of Systemics, Cybernetics and Informatics* 14.3 (2016), S. 68–73 (siehe S. 291).
- [23] SANDER, Jennifer and REINERT, Frank: „A Computer-Aided Assistance System for Ressource-Optimal Sensor Scheduling in Intelligence, Surveillance, and Reconnaissance“. In: *2016 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*. IEEE, 2016. DOI: 10.1109/mfi.2016.7849464 (siehe S. 291).
- [24] BOUVRY, Pascal; SEGOR, Florian; CHAUMETTE, Serge; DANOY, Grégoire; GUERRINI, Gilles; JURQUET, Gilles; KUWERTZ, Achim; MÜLLER, Wilmuth; ROSALIE, Martin and SANDER, Jennifer: „ASIMUT project: Aid to Situation Management based on MULTimodal, MULTiUAVs,

- MUltilevel acquisition Techniques“. In: *Proceedings of the 3rd Workshop on Micro Aerial Vehicle Networks, Systems, and Applications - DroNet '17*. ACM Press, 2017, S. 17–20. DOI: 10.1145/3086439.3086445.
- [25] KUWERTZ, Achim; SANDER, Jennifer; PFIRRMANN, Uwe and DYCK, Sergius: „High-Level information management in Joint ISR based on an object-oriented approach“. In: *2017 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*. IEEE, 2017. DOI: 10.1109/sdf.2017.8126360.
- [26] MÜLLER, Wilmuth; KUWERTZ, Achim; MÜHLENBERG, Dirk and SANDER, Jennifer: „Semantic information fusion to enhance situational awareness in surveillance scenarios“. In: *2017 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*. IEEE, 2017. DOI: 10.1109/MFI.2017.8170353.
- [27] ESSENDORFER, Barbara; HOFFMANN, Almuth; SANDER, Jennifer and KUWERTZ, Achim: „Integrating Coalition Shared Data in a System Architecture for High Level Information Management“. In: *Proc. SPIE 10802, Counterterrorism, Crime Fighting, Forensics, and Surveillance Technologies II, 108020F (8 October 2018)*. SPIE, 2018. DOI: 10.1117/12.2501861.
- [28] ESSENDORFER, Barbara; KUWERTZ, Achim and SANDER, Jennifer: „Distributed Information Management through Coalition Shared Data“. In: *NATO STO-MP-IST-160 Big Data and Artificial Intelligence for Military Decision Making, Bordeaux, France, 30th May to 1st June 2018*. 2018.
- [29] KUWERTZ, Achim; MÜHLENBERG, Dirk; SANDER, Jennifer and MÜLLER, Wilmuth: „Applying Knowledge-Based Reasoning for Information Fusion in Intelligence, Surveillance, and Reconnaissance“. In: *Multisensor Fusion and Integration in the Wake of Big Data, Deep Learning and Cyber Physical System: An Edition of the Selected Papers from the 2017 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2017)*. Hrsg. von LEE,

- Sukhan; KO, Hanseok and OH, Songhwai. *Lecture Notes in Electrical Engineering (LNEE)*, Bd. 501. Springer, 2018, S. 119–139. DOI: 10.1007/978-3-319-90509-9\_7.
- [30] MÜLLER, Wilmuth; REINERT, Frank; RODENBECK, Roland and SANDER, Jennifer: „Software architecture of an ISR asset planning application“. In: *SERP 2018, International Conference on Software Engineering Research & Practice*. CSREA Press, 2018 (siehe S. 291).
- [31] MÜLLER, Wilmuth; SANDER, Jennifer; KUWERTZ, Achim and MÜHLENBERG, Dirk: „High-level data fusion component for drone classification and decision support in counter UAV“. In: *Proc. SPIE 10651, Open Architecture/Open Business Model Net-Centric Systems and Defense Transformation 2018, 106510F (9 May 2018)*. SPIE, 2018. DOI: 10.1117/12.395578.
- [32] SANDER, Jennifer; RODENBECK, Roland; REINERT, Frank; MÜLLER, Wilmuth and BERUDE, Kim: „Sensor Operation Deployment with Multiple Routes per Asset“. In: *2018 21st International Conference on Information Fusion (FUSION)*. IEEE, 2018. DOI: 10.23919/icif.2018.8455861 (siehe S. 291).
- [33] ESSENDORFER, Barbara; SANDER, Jennifer; NISTOR, Marian Sorin; HOFFMANN, Almuth and PICKL, Stefan: „Distributed Data and Information Management for Crisis Forecasting and Management“. In: *Human Systems Engineering and Design II. Advances in Intelligent Systems and Computing (AISC)*, Bd. 1026. Springer, 2019, S. 813–819. DOI: 10.1007/978-3-030-27928-8\_124.
- [34] HOFFMANN, Almuth; KUWERTZ, Achim and SANDER, Jennifer: „Towards information extraction and semantic world modelling to support information management and intelligence creation in defense coalitions“. In: *Proc. SPIE 11169, Artificial Intelligence and Machine Learning in Defense Applications, 111690G (19 September 2019)*. SPIE, 2019. DOI: 10.1117/12.2533320.
- [35] KUWERTZ, Achim; MOLL, Maximilian; SANDER, Jennifer and PICKL, Stefan: „A Systemic Approach for Early Warning in Crisis Prevention and Management“. In: *Human Systems Engineering and Design*

*II. Advances in Intelligent Systems and Computing (AISC)*, Bd. 1026. Springer, 2019, S. 517–522. DOI: 10.1007/978-3-030-27928-8\_78.

- [36] RODENBECK, Roland; REINERT, Frank; SANDER, Jennifer and JACKWERTH, André: „Applicability of AI methods to JISR resource planning“. In: *Proc. SPIE 11169, Artificial Intelligence and Machine Learning in Defense Applications, 111690E (19 September 2019)*. SPIE, 2019. DOI: 10.1117/12.2533182 (siehe S. 291).
- [37] SANDER, Jennifer; ESSENDORFER, Barbara; DYCK, Sergius and PFIRRMANN, Uwe: „Formal reporting for information exploitation and dissemination in Joint ISR“. In: *Proc. SPIE 11015, Open Architecture/Open Business Model Net-Centric Systems and Defense Transformation 2019, 110150Q (30 April 2019)*. SPIE, 2019. DOI: 10.1117/12.2518600.



## Betreute studentische Arbeiten

Dieser Abschnitt enthält ein vollständiges Verzeichnis der betreuten studentischen Arbeiten. Die Arbeiten [9], [10], [11] adressieren speziell die Ressourcen-optimale Sensoreinsatzplanung. Sie können nur im weiteren Sinne als relevant für die vorliegende Arbeit angesehen werden.

- [1] GOUSSEV, Igor: „Lokale Bayes'sche Fusion von Vorwissen und Bildinformation zur Detektion und Klassifikation von Fahrzeugen“. Studienarbeit. Universität Karlsruhe (TH), 2008.
- [2] MEGREHI, Ghassen: „Informationstheoretisch basierte Methoden zur Aufdeckung von Abhängigkeitsstrukturen für die lokale Bayes'sche Fusion“. Studienarbeit. Universität Karlsruhe (TH), 2008.
- [3] GORDON, Marvin: „Lokale Bayes'sche Fusion von Vorwissen und Multispektral-Bildern zur Detektion und Klassifikation von Fahrzeugen“. Studienarbeit. Karlsruher Institut für Technologie (KIT), 2009.
- [4] FÄHNDRICH, Johannes: „Analyse von Verfahren zur Kombination von Expertenwissen in Form von Wahrscheinlichkeitsverteilungen im Hinblick auf die verteilte lokale Bayes'sche Fusion“. Diplomarbeit. Karlsruher Institut für Technologie (KIT), 2010.
- [5] GOUSSEV, Igor: „Merkmalsgestützte lokale Bayes'sche Fusion zur Detektion und Klassifikation von Fahrzeugen in einer Szene“. Diplomarbeit. Karlsruher Institut für Technologie (KIT), 2010.
- [6] KRIEGER, Jonas: „Agenten zur lokalen Bayes'schen Fusion“. Studienarbeit. Karlsruher Institut für Technologie (KIT), 2010.
- [7] PUTSCH, Daniel: „Komplexitätsreduktion Bayes'scher Fusion“. Seminararbeit. Karlsruher Institut für Technologie (KIT), 2010.

- [8] HÖHNE, Dorothee: „Kriterien zur Informationsqualität für Stakeholder im Kontext Schutz größerer Menschenmengen“. Bachelorarbeit. Hochschule Offenburg (University of Applied Sciences), 2013.
- [9] RODENBECK, Roland: „Operations Research Verfahren zur Sensoreinsatzplanung“. Diplomarbeit. Karlsruher Institut für Technologie (KIT), 2015 (siehe S. 299).
- [10] RIERA SANCHEZ, Ivan: „Simheuristic Algorithms for the Team Orienteering Problem with Time Windows“. Masterarbeit. Universitat Oberta de Catalunya, Spanien, 2016 (siehe S. 299).
- [11] BERUDE, Kim: „Erweiterung eines auf Methoden aus dem Operations Research basierenden Verfahrens zur ressourcenoptimalen Sensoreinsatzplanung um die Möglichkeit des Mehrfacheinsatzes der Assets innerhalb des Planungsintervalls“. Diplomarbeit. TU Bergakademie Freiberg, 2017 (siehe S. 299).



# Abkürzungsverzeichnis

<b>AOR</b>	Area of Interest
<b>C2</b>	Command and Control
<b>DDMCMC</b>	Data-Driven Markov Chain Monte Carlo
<b>DIKW</b>	Data-Information-Knowledge-Wisdom
<b>DSGVO</b>	Datenschutzgrundverordnung
<b>Ex-AI, XAI</b>	Explainable Artificial Intelligence
<b>FFT</b>	fast Fourier transform
<b>GMTI</b>	Ground Moving Target Indicator
<b>HPD</b>	Highest Posterior Density
<b>HUMINT</b>	Human Intelligence
<b>I/O</b>	Input/Output
<b>ICP</b>	Information Conservation Principle
<b>IDCP</b>	Identification Data Combining Process

<b>IMINT</b>	Imagery Intelligence
<b>ISR</b>	Intelligence, Surveillance, and Reconnaissance
<b>JDL</b>	Joint Directors of Laboratories
<b>KKT</b>	Karush-Kuhn-Tucker
<b>LPI</b>	lineare partielle Information
<b>MASINT</b>	Measurement and Signature Intelligence
<b>MCMC</b>	Markov Chain Monte Carlo
<b>NATO</b>	North Atlantic Treaty Organization
<b>nm</b>	Nanometer
<b>OODA</b>	Observe-Orient-Decide-Act
<b>OSINT</b>	Open Source Intelligence
<b>SIGINT</b>	Signals Intelligence
<b>SIR</b>	Sampling/ Importance Resampling
<b>STANAG</b>	(NATO) Standardization Agreement
<b>UAV</b>	Unmanned Aerial Vehicle
<b>UCI</b>	University of California at Irvine





## Karlsruher Schriftenreihe zur Anthropomatik (ISSN 1863-6489)

---

- Band 1** Jürgen Geisler  
**Leistung des Menschen am Bildschirmarbeitsplatz.** 2006  
ISBN 3-86644-070-7
- Band 2** Elisabeth Peinsipp-Byma  
**Leistungserhöhung durch Assistenz in interaktiven Systemen zur Szenenanalyse.** 2007  
ISBN 978-3-86644-149-1
- Band 3** Jürgen Geisler, Jürgen Beyerer (Hrsg.)  
**Mensch-Maschine-Systeme.** 2010  
ISBN 978-3-86644-457-7
- Band 4** Jürgen Beyerer, Marco Huber (Hrsg.)  
**Proceedings of the 2009 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory.** 2010  
ISBN 978-3-86644-469-0
- Band 5** Thomas Usländer  
**Service-oriented design of environmental information systems.** 2010  
ISBN 978-3-86644-499-7
- Band 6** Giulio Milighetti  
**Multisensorielle diskret-kontinuierliche Überwachung und Regelung humanoider Roboter.** 2010  
ISBN 978-3-86644-568-0
- Band 7** Jürgen Beyerer, Marco Huber (Hrsg.)  
**Proceedings of the 2010 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory.** 2011  
ISBN 978-3-86644-609-0
- Band 8** Eduardo Monari  
**Dynamische Sensorselektion zur auftragsorientierten Objektverfolgung in Kameranetzwerken.** 2011  
ISBN 978-3-86644-729-5

- Band 9** Thomas Bader  
**Multimodale Interaktion in Multi-Display-Umgebungen.** 2011  
ISBN 3-86644-760-8
- Band 10** Christian Frese  
**Planung kooperativer Fahrmanöver für kognitive Automobile.** 2012  
ISBN 978-3-86644-798-1
- Band 11** Jürgen Beyerer, Alexey Pak (Hrsg.)  
**Proceedings of the 2011 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory.** 2012  
ISBN 978-3-86644-855-1
- Band 12** Miriam Schleipen  
**Adaptivität und Interoperabilität von Manufacturing Execution Systemen (MES).** 2013  
ISBN 978-3-86644-955-8
- Band 13** Jürgen Beyerer, Alexey Pak (Hrsg.)  
**Proceedings of the 2012 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory.** 2013  
ISBN 978-3-86644-988-6
- Band 14** Hauke-Hendrik Vagts  
**Privatheit und Datenschutz in der intelligenten Überwachung: Ein datenschutzgewährendes System, entworfen nach dem „Privacy by Design“ Prinzip.** 2013  
ISBN 978-3-7315-0041-4
- Band 15** Christian Kühnert  
**Data-driven Methods for Fault Localization in Process Technology.** 2013  
ISBN 978-3-7315-0098-8
- Band 16** Alexander Bauer  
**Probabilistische Szenenmodelle für die Luftbildauswertung.** 2014  
ISBN 978-3-7315-0167-1
- Band 17** Jürgen Beyerer, Alexey Pak (Hrsg.)  
**Proceedings of the 2013 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory.** 2014  
ISBN 978-3-7315-0212-8

- Band 18** Michael Teutsch  
**Moving Object Detection and Segmentation for Remote Aerial Video Surveillance.** 2015  
ISBN 978-3-7315-0320-0
- Band 19** Marco Huber  
**Nonlinear Gaussian Filtering: Theory, Algorithms, and Applications.** 2015  
ISBN 978-3-7315-0338-5
- Band 20** Jürgen Beyerer, Alexey Pak (Hrsg.)  
**Proceedings of the 2014 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory.** 2014  
ISBN 978-3-7315-0401-6
- Band 21** Todor Dimitrov  
**Permanente Optimierung dynamischer Probleme der Fertigungssteuerung unter Einbeziehung von Benutzerinteraktionen.** 2015  
ISBN 978-3-7315-0426-9
- Band 22** Benjamin Kühn  
**Interessengetriebene audiovisuelle Szenenexploration.** 2016  
ISBN 978-3-7315-0457-3
- Band 23** Yvonne Fischer  
**Wissensbasierte probabilistische Modellierung für die Situationsanalyse am Beispiel der maritimen Überwachung.** 2016  
ISBN 978-3-7315-0460-3
- Band 24** Jürgen Beyerer, Alexey Pak (Hrsg.)  
**Proceedings of the 2015 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory.** 2016  
ISBN 978-3-7315-0519-8
- Band 25** Pascal Birnstill  
**Privacy-Respecting Smart Video Surveillance Based on Usage Control Enforcement.** 2016  
ISBN 978-3-7315-0538-9
- Band 26** Philipp Woock  
**Umgebungskartenschätzung aus Sidescan-Sonardaten für ein autonomes Unterwasserfahrzeug.** 2016  
ISBN 978-3-7315-0541-9

- Band 27** Janko Petereit  
**Adaptive State × Time Lattices: A Contribution to Mobile Robot Motion Planning in Unstructured Dynamic Environments.** 2017  
ISBN 978-3-7315-0580-8
- Band 28** Erik Ludwig Krempel  
**Steigerung der Akzeptanz von intelligenter Videoüberwachung in öffentlichen Räumen.** 2017  
ISBN 978-3-7315-0598-3
- Band 29** Jürgen Moßgraber  
**Ein Rahmenwerk für die Architektur von Frühwarnsystemen.** 2017  
ISBN 978-3-7315-0638-6
- Band 30** Andrey Belkin  
**World Modeling for Intelligent Autonomous Systems.** 2017  
ISBN 978-3-7315-0641-6
- Band 31** Chettapong Janya-Anurak  
**Framework for Analysis and Identification of Nonlinear Distributed Parameter Systems using Bayesian Uncertainty Quantification based on Generalized Polynomial Chaos.** 2017  
ISBN 978-3-7315-0642-3
- Band 32** David Münch  
**Begriffliche Situationsanalyse aus Videodaten bei unvollständiger und fehlerhafter Information.** 2017  
ISBN 978-3-7315-0644-7
- Band 33** Jürgen Beyerer, Alexey Pak (Eds.)  
**Proceedings of the 2016 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory.** 2017  
ISBN 978-3-7315-0678-2
- Band 34** Jürgen Beyerer, Alexey Pak and Miro Taphanel (Eds.)  
**Proceedings of the 2017 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory.** 2018  
ISBN 978-3-7315-0779-6
- Band 35** Michael Grinberg  
**Feature-Based Probabilistic Data Association for Video-Based Multi-Object Tracking.** 2018  
ISBN 978-3-7315-0781-9



- Band 36** Christian Herrmann  
**Video-to-Video Face Recognition for Low-Quality Surveillance Data.** 2018  
ISBN 978-3-7315-0799-4
- Band 37** Chengchao Qu  
**Facial Texture Super-Resolution by Fitting 3D Face Models.** 2018  
ISBN 978-3-7315-0828-1
- Band 38** Miriam Ruf  
**Geometrie und Topologie von Trajektorienoptimierung für vollautomatisches Fahren.** 2018  
ISBN 978-3-7315-0832-8
- Band 39** Angelika Zube  
**Bewegungsregelung mobiler Manipulatoren für die Mensch-Roboter-Interaktion mittels kartesischer modellprädiktiver Regelung.** 2018  
ISBN 978-3-7315-0855-7
- Band 40** Jürgen Beyerer and Miro Taphanel (Eds.)  
**Proceedings of the 2018 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory.** 2019  
ISBN 978-3-7315-0936-3
- Band 41** Marco Thomas Gewohn  
**Ein methodischer Beitrag zur hybriden Regelung der Produktionsqualität in der Fahrzeugmontage.** 2019  
ISBN 978-3-7315-0893-9
- Band 42** Tianyi Guan  
**Predictive energy-efficient motion trajectory optimization of electric vehicles.** 2019  
ISBN 978-3-7315-0978-3
- Band 43** Jürgen Metzler  
**Robuste Detektion, Verfolgung und Wiedererkennung von Personen in Videodaten mit niedriger Auflösung.** 2020  
ISBN 978-3-7315-0968-4
- Band 44** Sebastian Bullinger  
**Image-Based 3D Reconstruction of Dynamic Objects Using Instance-Aware Multibody Structure from Motion.** 2020  
ISBN 978-3-7315-1012-3

- Band 45** Jürgen Beyerer, Tim Zander (Eds.)  
**Proceedings of the 2019 Joint Workshop of  
Fraunhofer IOSB and Institute for Anthropomatics,  
Vision and Fusion Laboratory.** 2020  
ISBN 978-3-7315-1028-4
- Band 46** Stefan Becker  
**Dynamic Switching State Systems for Visual Tracking.** 2020  
ISBN 978-3-7315-1038-3
- Band 47** Jennifer Sander  
**Ansätze zur lokalen Bayes'schen Fusion von  
Informationsbeiträgen heterogener Quellen.** 2021  
ISBN 978-3-7315-1062-8



Lehrstuhl für Interaktive Echtzeitsysteme  
Karlsruher Institut für Technologie

Fraunhofer-Institut für Optronik, Systemtechnik  
und Bildauswertung IOSB Karlsruhe

Die Lösung diverser Aufgaben profitiert von der Informationsfusion oder setzt sie sogar voraus. Die Bayes'sche Fusionsmethodik ist anschaulich, fundiert und erfüllt die essentiellen Anforderungen an eine sinnvolle Methodik gerade auch zur Fusion der Beiträge heterogener Informationsquellen. In praktisch relevanten Aufgaben ist jedoch der Wertebereich der Größen des Interesses, die den interessierenden Sachverhalt mathematisch fassen, oft sehr umfangreich und von recht hoher Dimension. In solchen Situationen verursachen Bayes'sche Verfahren in der Regel einen hohen, in der Praxis oft nicht tragbaren Aufwand. In dieser Arbeit werden neuartige Ansätze zur Bewältigung Bayes'scher Fusion formuliert und grundlegend untersucht. Ausgangspunkt ist eine Analogie zu kriminalistischen Ermittlungen, bei denen zunächst zu verfolgende Spuren identifiziert werden und bei denen Spuren anschließend auch nur detailliert in Bezug auf einen lokalen Ausschnitt des Gesamtzusammenhangs untersucht werden. Zur Umsetzung lokaler Bayes'scher Fusion kann auf das Konzept für eine agentenbasierte Fusionsarchitektur zurückgegriffen werden.

ISSN 1863-6489  
ISBN 978-3-7315-1062-8

