

# Statistische Inferenz in invarianten graphischen Modellen mit Normalverteilungsannahme

Dissertation  
zur Erlangung des Doktorgrades  
des Fachbereichs Mathematik  
der Universität Augsburg



Vorgelegt von  
Andreas Käußl

Augsburg, September 2012

Erstgutachter: Prof. Dr. Friedrich Pukelsheim; Universität Augsburg

Zweitgutachter: Prof. Mathias Drton, PhD; University of Chicago

Termin der mündlichen Prüfung: 07. November 2012

## **Zusammenfassung**

Die vorliegende Arbeit beschäftigt sich mit statistischer Inferenz in invarianten graphischen Modellen mit Normalverteilungsannahme. Hierbei handelt es sich um Verteilungsfamilien, die zwei Arten von Restriktionen berücksichtigen. Zum einen spiegeln alle Modellverteilungen eine gewisse Abhängigkeitsstruktur wider, die von einem Graphen vorgegeben wird. Zum anderen werden die von einer endlichen Gruppe vorgegebenen Symmetrien zwischen den beteiligten Variablen berücksichtigt. Zunächst werden bekannte Ergebnisse für diese Modelle vorgestellt, unter anderem hinreichende Bedingungen für die Existenz des Maximum-Likelihood-Schätzers der Kovarianzmatrix sowie dessen explizite Form. Anschließend wird mit Hilfe einer Verallgemeinerung der klassischen Wishart-Verteilung die Verteilung des Maximum-Likelihood-Schätzers ermittelt und anhand eines Beispiels veranschaulicht. Außerdem wird die berechnete Verteilung dazu verwendet, Likelihood-Quotienten-Tests für geschachtelte invariante graphische Modelle zu entwickeln.

## Danksagungen

Allen voran möchte ich mich bei meinem Betreuer Prof. Friedrich Pukelsheim ganz herzlich bedanken, der mir seit meinem dritten Studiensemester mit Rat und Tat zur Seite stand. Gerade bei fachlichen Fragen hatte er immer ein offenes Ohr und nahm sich die Zeit, alles ausführlich zu diskutieren und zu erklären. Doch auch darüber hinaus gab er mir viele wertvolle Hilfestellungen, sei es bei der Themensuche für die vorliegende Arbeit, bei der Bewerbung für ein Stipendium oder bei der Vermittlung eines Forschungsaufenthalts in den USA.

Auch Prof. Mathias Drton, meinem Gastgeber in Chicago und gleichzeitig Zweitgutachter dieser Dissertation, möchte ich auf diesem Weg ein riesiges Dankeschön aussprechen. Auch er war mir sowohl bei der Organisation meines Aufenthalts als auch in fachlicher Hinsicht eine sehr große Hilfe. Denn ohne das Wissen über graphische Modelle, das ich mir während unserer zahlreichen Diskussionen aneignen konnte, wäre die vorliegende Arbeit in ihrer heutigen Form nicht möglich gewesen.

Ebenfalls bedanken will ich mich bei Dr. Thomas Klein, der mir sowohl bei der Themensuche als auch während der Arbeit an der Dissertation mit vielen wertvollen Bemerkungen und Anregungen unter die Arme gegriffen hat.

Auch bei allen Mitwirkenden des Studiengangs „TopMath“ will ich mich an dieser Stelle bedanken, da sie sich bei allen organisatorischen Fragen um eine schnelle Lösung bemühten.

Besonderer Dank gebührt allerdings meinen Eltern, meinen Großeltern und meiner Schwester, die mich während meiner gesamten Ausbildung auf jede nur erdenkliche Weise unterstützt haben. Ich kann mich glücklich schätzen, eine Familie hinter mir zu wissen, die keine meiner Entscheidungen anzweifelte und die mir in jeder noch so schwierigen Situation zur Seite stand und auch in Zukunft stehen wird.

Ein letztes Danke – und wie so oft kommt auch hier das Wichtigste zum Schluss – geht an eine wundervolle Frau. Ohne sie wäre mir die Bewältigung der Herausforderungen der letzten Jahre erheblich schwieriger gefallen, ja manchmal wohl unmöglich gewesen. Liebe Kathi, danke für dein Verständnis für all meine Entscheidungen. Danke für deine bedingungslose Unterstützung in allen Situationen. Und vor allem Danke für deine liebevolle Art.

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
1.1	Invariante graphische Modelle . . . . .	1
1.2	Resultate der Arbeit . . . . .	3
1.3	Notationen für den Umgang mit Matrizen . . . . .	6
<b>2</b>	<b>Invariante graphische Modelle mit Normalverteilungsannahme</b>	<b>7</b>
2.1	Azyklisch gemischte Graphen . . . . .	7
2.2	Graphische Modelle mit Normalverteilungsannahme . . . . .	11
2.3	Permutationen, Orbits und Automorphismen . . . . .	17
2.4	Einbindung von Invarianzrestriktionen in graphische Modelle . . . . .	23
<b>3</b>	<b>Verallgemeinerte Riesz- und Wishart-Verteilungen</b>	<b>37</b>
3.1	Matrixwertige Normalverteilungen . . . . .	37
3.2	Klassische Wishart-Verteilungen . . . . .	40
3.3	Verallgemeinerte Cholesky-Zerlegungen . . . . .	44
3.4	Verallgemeinerte Riesz-Verteilungen . . . . .	47
3.5	Verallgemeinerte Wishart-Verteilungen . . . . .	52
<b>4</b>	<b>Maximum-Likelihood-Schätzer der Kovarianzmatrix</b>	<b>56</b>
4.1	Graphische Modelle ohne Invarianzrestriktionen . . . . .	56
4.2	Graphische Modelle mit Invarianzrestriktionen . . . . .	58
<b>5</b>	<b>Verteilung der Schätzer in invarianten graphischen Modellen</b>	<b>61</b>
5.1	Verteilungen der Cholesky-Parameter . . . . .	62
5.2	Verteilung des Maximum-Likelihood-Schätzers . . . . .	72
5.3	Beispiel zur Maximum-Likelihood-Schätzung . . . . .	89
<b>6</b>	<b>Likelihood-Quotienten-Tests für invariante graphische Modelle</b>	<b>96</b>
6.1	Testprobleme in invarianten graphischen Modellen . . . . .	96
6.2	Die Verteilung der Likelihood-Quotienten-Teststatistik . . . . .	100
<b>7</b>	<b>Abschließende Bemerkungen und Ausblick</b>	<b>106</b>

## Abbildungsverzeichnis

Abb. 1:	Graphentheoretische Konzepte . . . . .	11
Abb. 2:	Automorphismen über gerichteten Graphen . . . . .	19
Abb. 3:	Symmetrien in den Elternboxen einer Box . . . . .	32
Abb. 4:	Implizite boxeninterne Symmetrien . . . . .	63
Abb. 5:	$G$ -reduzierte Versionen eines azyklisch gemischten Graphen bezüglich verschiedener Invarianzgruppen . . . . .	73
Abb. 6:	Kanonische Wahl der $G$ -reduzierten Version eines azyklisch gemischten Graphen . . . . .	77
Abb. 7:	Modifikationen eines azyklisch gemischten Graphen . . . . .	87
Abb. 8:	Beispiel zur Verteilung des Maximum-Likelihood-Schätzers . . . . .	89
Abb. 9:	Vergleich zulässiger Invarianzgruppen für verschiedene Graphen . . . . .	98

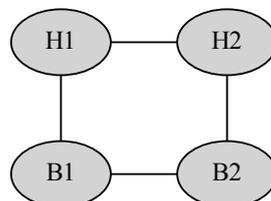
# 1 Einleitung

## 1.1 Invariante graphische Modelle

Eine wichtige Zielstellung in der multivariaten Statistik ist es, die Abhängigkeiten der an einem Datensatz beteiligten Variablen zu verstehen. Aus dem Wunsch, diese Analyse auf möglichst effiziente und intuitive Weise durchzuführen, haben sich seit den 1980er Jahren die sogenannten graphischen Modelle entwickelt. Deren Grundidee ist es, die Abhängigkeitsbeziehungen des Modells mit Hilfe eines Graphen darzustellen und diesen Graphen eindeutig mit einer Verteilungsfamilie zu identifizieren. Als Einstiegslektüre in die Theorie der graphischen Modelle sind allen voran die Bücher von Lauritzen [18] und Whittaker [25] zu nennen. Das Buch von Edwards [14] dagegen legt den Fokus etwas mehr auf die rechnerbasierte Analyse multivariater Datensätze mit Hilfe von graphischen Modellen.

Sind die zugrunde liegenden Variablen kontinuierlich, so wird meist unterstellt, dass sie einer multivariaten Normalverteilung folgen. Diese Annahme führt zu einer aus mathematischer Sicht sehr ästhetischen Theorie, da beispielsweise bedingte Unabhängigkeitsbeziehungen direkt im Inversen der Kovarianzmatrix abgelesen werden können. Dadurch werden unter anderem die Berechnung von Maximum-Likelihood-Schätzern, die Bestimmung der asymptotischen Verteilung der Schätzer und die Durchführung von Likelihood-Quotienten-Tests ermöglicht. Gerade letztere sind auch von praktischer Bedeutung. Sie erlauben es, für einen gegebenen Datensatz ein Modell zu finden, das die Struktur der Daten ausreichend gut beschreibt, ohne jedoch übermäßig kompliziert zu sein.

Einen bekannten Datensatz, der mit Hilfe solcher Methoden analysiert worden ist, findet man in Beispiel 8.4.1 in Whittaker [25]. Für ihn wurden 25 erwachsene Brüderpaare untersucht und jeweils sowohl die Kopfhöhen der beiden Brüder notiert ( $H_1$  und  $H_2$ ), als auch ihre Kopfbreiten ( $B_1$  und  $B_2$ ). Verschiedene Methoden zur Modellauswahl legen nahe, dass das zum folgenden Graphen korrespondierende graphische Modell die Daten am Besten beschreibt.



Dieser Graph soll die Abhängigkeiten der einzelnen Variablen beschreiben. So machen die beiden Kanten  $H_1 - B_1$  und  $H_2 - B_2$  deutlich, dass die Kopfhöhe und die Kopfbreite einer Person miteinander korreliert sind. Es gibt aber auch einen Zusammenhang zwischen den beiden Kopfgrößen eines Geschwisterpaars. Deutlich wird dies durch die Kanten  $H_1 - H_2$  und  $B_1 - B_2$ . Die erstere zeigt, dass die Kopfhöhe des einen Bruders einen Einfluss auf die Kopfhöhe des anderen Bruders hat. Selbiges gilt für die Kante  $B_1 - B_2$  und die Kopfbreiten des Bruderpaars. Schließlich lassen sich aber nicht nur aus den im Graphen vorhandenen Kanten Erkenntnisse gewinnen, sondern auch fehlende Kanten liefern Informationen über die Struktur des Datensatzes. Die Tatsache, dass im Modellgraphen keine Kante zwischen  $H_1$  und  $B_2$  auftaucht, deutet beispielsweise darauf hin, dass die Kopfhöhe des erstens Bruders keinen direkten Einfluss auf die Kopfbreite des zweiten Bruders hat und umgekehrt. Stattdessen ist es für die Erklärung von  $H_1$  ausreichend, die Variablen  $B_1$  und  $H_2$  zu kennen, da im Graphen aus  $H_1$  nur die beiden Kanten  $H_1 - B_1$  und  $H_1 - H_2$  heraus laufen. In einem statistischen Kontext spricht man davon, dass die Variablen  $H_1$  und  $B_2$  bedingt unabhängig sind gegeben den Größen  $B_1$  und  $H_2$ .

Sowohl der zugrunde liegende Datensatz als auch die Intuition legen allerdings nahe, dass das beschriebene graphische Modell nicht die gesamte in den Daten verborgene Struktur aufdecken kann. Unter anderem liefert der Datensatz keinen Anhaltspunkt darauf, dass der ältere Bruder meist einen größeren Kopf hat als der jüngere Bruder. Bei der Datenerhebung ist es also unerheblich, welcher der beiden Brüder zuerst gemessen wird. Solche Symmetrieaspekte sollten natürlich auch bei der Modellierung berücksichtigt werden. Im konkreten Beispiel bedeutet dies, dass die Verteilung des ursprünglichen Zufallsvektors und die des permutierten Zufallsvektors identisch sein sollen,

$$\mathcal{L}[(H_1, B_1, H_2, B_2)^T] = \mathcal{L}[(H_2, B_2, H_1, B_1)^T].$$

Hier und im Folgenden meint  $\mathcal{L}(X)$  die Verteilung der Zufallsvariablen  $X$ .

Allgemeine Überlegungen dazu, wie solche Symmetriestrukturen in ein statistisches Modell eingebunden werden können, wurden schon seit Mitte der 1970er Jahre angestellt, hier ist insbesondere der Aufsatz von Andersson [2] zu nennen. Einen guten Überblick über die frühe Forschung zum Thema Symmetrien in der Statistik bietet außerdem der Kommentar von Perlman [22]. Wie sich die Einbindung von Invarianzen speziell für graphische Modelle bewerkstelligen lässt und welchen Einfluss dies auf die statistische Inferenz hat, wird beispielsweise in den Aufsätzen von Andersson und Madsen [4] und Madsen [20] untersucht.

Gerade aus heutiger Sicht ist Wissen über die einem Datensatz innewohnenden Symmetrien aber nicht der einzige Grund, warum Invarianzbedingungen in ein statistisches Modell aufgenommen werden. In biologischen oder ökonomischen Anwendungen sind die untersuchten Datensätze oft sehr hochdimensional, was zu einer hohen Komplexität der Modelle führt. In solch einer Situation kann die Berücksichtigung von Symmetrien die Anzahl der Modellparameter unter Umständen erheblich reduzieren. So werden Modelle analysierbar, die zuvor aufgrund ihrer Komplexität nicht effizient untersucht werden konnten. Dieser Aspekt von invarianten Modellen wird beispielsweise in Shah und Chandrasekaran [24] sowohl aus theoretischer Sicht als auch anhand von numerischen Studien analysiert.

Wegen dieser potentiellen Komplexitätsreduktion stellen Symmetrien demnach ein mächtiges Hilfsmittel der statistischen Modellierung dar. Allerdings ist Vorsicht geboten, da die eingebundenen Invarianzen nicht die eigentliche Struktur des Modells zerstören sollten. Ist beispielsweise bekannt, dass es zwischen zwei Variablen einen kausalen Zusammenhang gibt, das heißt, die erste Variable hat einen Einfluss auf die zweite, aber nicht umgekehrt, so dürfen diese beiden Faktoren natürlich nicht vertauscht werden. Daher ist es sinnvoll, in den Modellgraphen auch gerichtete Kanten zuzulassen und nur solche Symmetrien zu verwenden, die die kausale Struktur des Graphen respektieren. Wie dies aus technischer Sicht möglich ist, wird in Madsen [20] erläutert.

In diesem letztgenannten Aufsatz werden aber nicht nur invariante graphische Modelle definiert, sondern auch deren Eigenschaften näher untersucht. So werden notwendige und hinreichende Bedingungen für die Existenz des Maximum-Likelihood-Schätzers in diesen Modellen entwickelt und die explizite Form der Schätzer angegeben. Ungeklärt bleibt allerdings die Frage nach der Verteilung des Maximum-Likelihood-Schätzers in einem invarianten graphischen Modell. Dieses Defizit soll daher von der vorliegenden Arbeit behoben werden.

## 1.2 Resultate der Arbeit

Als Kandidat für diese Verteilung kommen gewisse Verallgemeinerungen der klassischen Wishart-Verteilung in Frage. In der Literatur gibt es bereits zahlreiche unterschiedliche Ansätze für solch eine Verallgemeinerung, sie alle haben jedoch eine Gemeinsamkeit: Während es sich bei der klassischen Wishart-Verteilung um ein Wahrscheinlichkeitsmaß über dem ganzen Kegel der positiv definiten Matrizen handelt, sind die verallgemeinerten Verteilungen auf einer gewissen Teilmenge dieses Kegels konzentriert. Ein erster Schritt in dieser Forschungsrichtung wurde von Dawid und Lauritzen [9] gemacht. Sie beschreiben in ihrem Aufsatz die Verteilung des Maximum-Likelihood-Schätzers der Ko-

varianzmatrix in einem ungerichteten graphischen Modell. Ihr Ansatz ermöglicht es allerdings nicht, diese Verteilung in Form einer Dichtefunktion anzugeben. Eine weitere, sehr große Verteilungsfamilie wurde von Andersson und Wojnar [6] entwickelt. Sie zeigen, wie eine verallgemeinerte Wishart-Verteilung über einem beliebigen homogenen Kegel, das heißt einem Kegel, dessen Automorphismengruppe transitiv wirkt, definiert werden kann. In diese Verteilungsfamilie fällt auch ein großer Teil der in Letac und Massam [19] entwickelten Wahrscheinlichkeitsmaße. Allerdings nutzen die letztgenannten Autoren die Struktur graphischer Modelle, so dass die Beschreibung der Verteilungen intuitiver ist als die allgemeinere Beschreibung der Verteilung über homogenen Kegeln. Die aktuellste Arbeit zu verallgemeinerten Wishart-Verteilungen stammt von Andersson und Klein [3]. Ihnen gelingt es, eine sehr große Verteilungsfamilie mit Hilfe von Lebesgue-Dichten auf intuitive Weise auszudrücken. Insbesondere zeigt sich, dass die Wahrscheinlichkeitsmaße von Letac und Massam [19] in die Verteilungsfamilie von Andersson und Klein [3] fallen. Des Weiteren enthält der letztgenannte Aufsatz auch Verteilungen, die nicht von Andersson und Wojnar [6] abgedeckt werden, da Parameterräume graphischer Modelle im Allgemeinen keine homogenen Kegel sind.

Welche dieser verallgemeinerten Wishart-Verteilungen bei der Maximum-Likelihood-Schätzung in invarianten graphischen Modellen eine Rolle spielen, ist allerdings noch nicht bekannt und soll im Folgenden geklärt werden. Um diese Modelle in praktischen Anwendungen benutzen zu können, ist allein die Kenntnis der Maximum-Likelihood-Schätzer und deren Verteilung jedoch noch nicht ausreichend. Daher soll im Anschluss auch die Verteilung der Teststatistik eines Likelihood-Quotienten-Tests ermittelt werden, so dass bei zwei gegebenen invarianten graphischen Modellen getestet werden kann, welches der beiden die Struktur der Daten besser beschreibt.

Dazu werden in Kapitel 2 zunächst verschiedene Konzepte aus der Graphentheorie und der Gruppentheorie erläutert, die es im Anschluss ermöglichen, invariante graphische Modelle formell zu definieren. Dabei besteht der erste wichtige Arbeitsschritt in der vorliegenden Dissertation darin, eine passende Beschreibung invarianter graphischer Modelle zu entwickeln. Aus zwei Gründen ist die in Madsen [20] gewählte Definition nämlich ungeeignet für die hier verfolgten Ziele. Zum einen beruht sie auf fortgeschrittenen Konzepten der Gruppentheorie, was zwar aus technischer Sicht zu einer eleganten Theorie führt, aber wenig intuitiv ist. Zum anderen sind zwar die Parameterräume aus Andersson und Klein [3] und Madsen [20] als geometrische Objekte identisch, allerdings werden in den beiden Aufsätzen unterschiedliche Beschreibungen der Modelle verwendet. Um die verallgemeinerten Riesz-Verteilungen aus Andersson und Klein [3] anwenden zu können müssen also die invarianten graphischen Modelle aus Madsen [20] auf eine andere Weise

dargestellt werden. Details zu der angesprochenen alternativen Definition folgen in Kapitel 2.4. Dort wird sich auch zeigen, dass diese Darstellung die Struktur der Modelle besser zum Vorschein bringt und die Beschreibung der gesuchten Verteilungen vereinfacht.

In Kapitel 3 werden einige matrixwertige Verteilungsfamilien vorgestellt, unter anderem gewisse Verallgemeinerungen der klassischen Wishart-Verteilung. Später wird sich zeigen, dass eben diese Verteilungen bei der Maximum-Likelihood-Schätzung in invarianten graphischen Modellen auftreten. Um den Bogen zurück zur statistischen Anwendung zu spannen, werden anschließend in Kapitel 4 bekannte Ergebnisse verwendet, um den Maximum-Likelihood-Schätzer der Kovarianzmatrix in solchen Modellen explizit zu beschreiben. Allgemein sind die Kapitel 3 und 4 hauptsächlich der Aufgabe gewidmet, bekannte Ergebnisse aus der älteren und jüngeren Forschung zusammenzutragen und an die hiesige Notation anzupassen.

Dies ermöglicht es schließlich, den zentralen Teil der Dissertation, Kapitel 5, in einer sehr kompakten Weise darzustellen. Dieser Abschnitt beschäftigt sich damit, die Verteilung des Maximum-Likelihood-Schätzers der Kovarianzmatrix in invarianten graphischen Modellen zu berechnen. Aufgrund der Struktur des Modellgraphen reicht es dabei prinzipiell aus, die Verteilungen und Abhängigkeiten gewisser Teilmatrizen des Maximum-Likelihood-Schätzers zu bestimmen, was in den Theoremen 5.5 und 5.6 geschieht. In manchen Situationen kann es allerdings auch nützlich sein, die Verteilung des gesamten Kovarianzschätzers zu kennen. Daher widmet sich Theorem 5.13 der Berechnung eben dieser Verteilung. Es sei hervorgehoben, dass sich keine der im Literaturverzeichnis aufgeführten Veröffentlichungen mit dieser oder ähnlichen Fragestellungen beschäftigt und Kapitel 5 somit vollkommen selbstständig erarbeitet wurde.

In Kapitel 6 werden die bis dahin ermittelten Ergebnisse verwendet, um Likelihood-Quotienten-Tests zwischen invarianten graphischen Modellen zu entwickeln. Dazu wird zunächst beschrieben, welche Gestalt die zugrunde liegenden Modellgraphen und Invarianzgruppen haben müssen, um wohldefinierte Hypothesentests zu ermöglichen. Anschließend wird bewiesen, dass die korrespondierende Teststatistik einer Verteilung folgt, die der sogenannten Box-Familie angehört, siehe Theorem 6.3 und Korollar 6.4. Dies ermöglicht es, konkrete Likelihood-Quotienten-Tests durchzuführen, da für solche Verteilungen sehr akkurate Näherungsmethoden verfügbar sind. Wie Kapitel 5 wurde auch das gesamte Kapitel 6 eigenständig entwickelt und stellt somit einen Beitrag zur aktuellen Forschung dar.

In Kapitel 7 folgen schließlich einige abschließende Bemerkungen und ein kleiner Ausblick. Unter anderem wird erläutert, wie eine alternative Klasse graphischer Modelle definiert werden kann, indem nicht bedingte Unabhängigkeiten zwischen den Variablen

angenommen werden, sondern marginale Unabhängigkeiten. In diesem Zusammenhang sei insbesondere auf die Mitwirkung des Autors bei den Aufsätzen Drton, Fox und Käuffl [10] und [12] hingewiesen.

### 1.3 Notationen für den Umgang mit Matrizen

Seien  $n \in \mathbb{N}$  und  $p \in \mathbb{N}$  natürliche Zahlen. Die Menge aller  $n \times p$ -Matrizen mit reellen Einträgen wird bezeichnet mit  $\mathbb{R}^{n \times p}$ . Sei  $S \in \mathbb{R}^{n \times p}$  solch eine Matrix, sowie  $i \in \{1, \dots, n\}$  und  $j \in \{1, \dots, p\}$ . Die Komponente in der  $i$ -ten Zeile und  $j$ -ten Spalte von  $S$  wird bezeichnet mit  $S_{ij} \in \mathbb{R}$ . Seien außerdem zwei Indexmengen  $A \subseteq \{1, \dots, n\}$  und  $B \subseteq \{1, \dots, p\}$  vorgegeben. Dann wird die Teilmatrix von  $S$ , die zu diesen Indexmengen korrespondiert, mit  $S_{A,B}$  bezeichnet, das heißt

$$S_{A,B} = (S_{ij})_{i \in A, j \in B} \in \mathbb{R}^{A \times B}.$$

Hier und im Folgenden wird oft auf die Kardinalitätsstriche verzichtet, wenn aus dem Kontext eindeutig hervorgeht, dass nicht eine Menge, sondern ihre Mächtigkeit gemeint ist. So meint die Bezeichnung  $\mathbb{R}^{A \times B}$  hier den linearen Raum  $\mathbb{R}^{|A| \times |B|}$ .

Ist  $p = n$ , das heißt  $S \in \mathbb{R}^{n \times n}$  ist eine quadratische Matrix, so wird die zur Indexmenge  $A \subseteq \{1, \dots, n\}$  korrespondierende Teilmatrix von  $S$  abgekürzt mit  $S_A := S_{A,A}$ . Solche Indexoperationen werden per Konvention grundsätzlich vor dem Transponieren oder Invertieren einer Matrix durchgeführt, das heißt

$$S_A^T := (S_A)^T \quad \text{und} \quad S_A^{-1} := (S_A)^{-1}.$$

Soll die Matrix  $S$  sowohl transponiert als auch invertiert werden, so wird dies abgekürzt mit  $S^{-T} = (S^{-1})^T = (S^T)^{-1}$ . Die  $n \times n$ -Identitätsmatrix wird mit  $I_n$  bezeichnet.

Eine wichtige Matrizenoperation, die in der multivariaten Statistik häufig auftritt, ist das *Kronecker-Produkt* zweier Matrizen  $S \in \mathbb{R}^{a \times b}$  und  $T \in \mathbb{R}^{c \times d}$  für  $a, b, c, d \in \mathbb{N}$ . Dieses ist definiert als die  $ac \times bd$ -Matrix  $S \otimes T$ , die in Blockmatrixform als

$$S \otimes T := \begin{pmatrix} S_{11}T & \cdots & S_{1b}T \\ \vdots & \ddots & \vdots \\ S_{a1}T & \cdots & S_{ab}T \end{pmatrix} \in \mathbb{R}^{ac \times bd}$$

geschrieben werden kann. Jeder Block in  $S \otimes T$  besteht also aus dem korrespondierenden skalaren Eintrag in  $S$ , multipliziert mit der kompletten Matrix  $T$ .

## 2 Normale invariante graphische Modelle

Im folgenden Kapitel sollen die Grundkonzepte der graphischen Modellierung mit Invarianzen erläutert werden. Ausführlichere Einführungen zu graphischen Modellen findet man beispielsweise in den Büchern von Lauritzen [18] oder Whittaker [25]. Die Berücksichtigung von Symmetriestrektionen wird unter anderem in Andersson und Madsen [4] und Madsen [20] behandelt.

### 2.1 Azyklisch gemischte Graphen

Zunächst werden die notwendigen graphentheoretischen Begrifflichkeiten definiert und erläutert. Die meisten der folgenden Definitionen stammen aus Andersson und Klein [3, Kapitel 4,5] sowie Lauritzen [18, Kapitel 2]. In der letztgenannten Quelle findet man auch eine ausführlichere Einführung in die für graphische Modelle relevante Graphentheorie.

Ein *gemischter Graph*  $\mathcal{V} = (V, D, U)$  ist ein Tupel aus einer Menge von Knoten  $V = \{1, \dots, p\}$ , einer Menge von *gerichteten Kanten*  $D \subseteq V \times V$  und einer Menge von *ungerichteten Kanten*  $U \subseteq V \times V$ . Für  $\alpha, \beta \in V$  und  $\alpha \neq \beta$  bedeutet dies insbesondere, dass  $(\alpha, \beta) \in D$  bereits  $(\beta, \alpha) \notin D$  impliziert, während aus  $(\alpha, \beta) \in U$  schon  $(\beta, \alpha) \in U$  folgt. Eine gerichtete Kante  $(\alpha, \beta) \in D$  wird im Folgenden anschaulicher mit  $\alpha \rightarrow \beta \in D$  bezeichnet, während eine ungerichtete Kante  $(\alpha, \beta) \in U$  durch  $\alpha - \beta \in U$  symbolisiert wird. Die *Eltern*  $pa(\beta)$  von  $\beta$  in  $\mathcal{V}$  sind diejenigen Knoten, von denen aus eine gerichtete Kante zu  $\beta$  läuft,  $pa(\beta) := \{\alpha \in V \mid \alpha \rightarrow \beta \in D\}$ . Dagegen bezeichnet man die Knoten, die durch eine ungerichtete Kante mit  $\beta$  verbunden sind, als dessen *Nachbarn*  $nb(\beta)$ , das heißt  $nb(\beta) := \{\alpha \in V \mid \alpha - \beta \in U\}$ . Dabei wird angenommen, dass in  $\mathcal{V}$  keine Schleifen auftreten, das heißt  $\alpha \notin pa(\alpha)$  und  $\alpha \notin nb(\alpha)$  für alle Knoten  $\alpha \in V$ . Außerdem werden nur *einfache Graphen* untersucht, das heißt zwischen  $\alpha \in V$  und  $\beta \in V$  gibt es entweder keine Kante oder die gerichtete Kante  $\alpha \rightarrow \beta$  oder die gerichtete Kante  $\alpha \leftarrow \beta$  oder die ungerichtete Kante  $\alpha - \beta$ .

Für eine Menge von Knoten  $A \subseteq V$  besteht der *induzierte Teilgraph*  $\mathcal{V}_A = (V_A, D_A, U_A)$  aus der Knotenmenge  $V_A = A$  und den Kanten, die nur zwischen Knoten aus  $A$  verlaufen, das heißt  $D_A = D \cap (A \times A)$  und  $U_A = U \cap (A \times A)$ .

Ein *Pfad* der Länge  $k$  in  $\mathcal{V}$  ist eine Sequenz von Knoten  $\alpha_0, \dots, \alpha_k \in V$ , so dass  $\alpha_{\ell-1} \rightarrow \alpha_\ell \in D$  oder  $\alpha_{\ell-1} - \alpha_\ell \in U$  für alle  $\ell = 1, \dots, k$ . Ist  $\alpha_0 = \alpha_k$ , das heißt der Anfangs- und der Endpunkt des Pfades sind identisch, so spricht man von einem *Zykel* der Länge  $k$ . Enthält ein Pfad nur ungerichtete Kanten, so spricht man von einem *ungerichteten Pfad*. Sind auch gerichtete Kanten vorhanden, so handelt es sich um einen *semi-gerichteten Pfad*. Dieselbe Terminologie gilt auch für Zykel. Ein gemischter Graph  $\mathcal{V}$ , der keinen semi-

gerichteten Zykel enthält, wird *azyklisch gemischter Graph* genannt. Besitzt  $\mathcal{V}$  zusätzlich keine ungerichteten Kanten, so handelt es sich um einen *azyklisch gerichteten Graphen*. In solch einem Graphen  $\mathcal{D} = (V, D, \emptyset)$  heißen diejenigen Knoten  $\beta \in V$ , die von einem festen Knoten  $\alpha$  nicht über einen semi-gerichteten Pfad zu erreichen sind, die *Nicht-Nachkommen*  $nd(\alpha)$  von  $\alpha$ . Enthält ein gemischter Graph dagegen keine gerichteten Kanten, so spricht man von einem *ungerichteten Graphen*  $\mathcal{U} = (V, \emptyset, U)$ .

Ein Graph  $\mathcal{V}$ , in dem jeder Knoten  $\alpha$  mit jedem anderen Knoten  $\beta \neq \alpha$  durch eine gerichtete oder ungerichtete Kante verbunden ist, wird *vollständig* genannt. Ist  $\mathcal{U}$  ein ungerichteter Graph, so heißt eine inklusionsmaximale Teilmenge von Knoten  $C \subseteq V$ , die einen vollständigen Teilgraphen  $\mathcal{U}_C$  induziert, eine *Clique* von  $\mathcal{U}$ . Die Menge aller Cliques von  $\mathcal{U}$  wird mit  $\mathcal{C}(\mathcal{U})$  bezeichnet. Des Weiteren heißt  $\mathcal{U}$  *triangulierbar* oder *zerlegbar*, wenn jeder Zykel der Länge  $k \geq 4$  eine Sehne enthält, das heißt zwei im Zykel nicht-aufeinanderfolgende Knoten, die in  $\mathcal{U}$  verbunden sind. Solche Graphen sind gerade aus statistischer Sicht interessant, da die dadurch induzierten Modelle nützliche Eigenschaften haben, vergleiche beispielsweise Lauritzen [18, Kapitel 5.3]. Seien des Weiteren drei Teilmengen von Knoten  $A, B, S \subseteq V$  gegeben. Dann werden  $A$  und  $B$  im ungerichteten Graphen  $\mathcal{U}$  von  $S$  getrennt, wenn jeder Pfad von einem Knoten in  $A$  zu einem Knoten in  $B$  durch mindestens einen Knoten aus  $S$  läuft.

Das *Gerüst*  $\mathcal{V}^\sim = (V, \emptyset, U^\sim)$  eines azyklisch gemischten Graphen  $\mathcal{V} = (V, D, U)$  entsteht, wenn alle gerichteten Kanten in  $\mathcal{V}$  durch ungerichtete ersetzt werden. Mit anderen Worten ist  $U^\sim = \{\alpha - \beta \mid \alpha - \beta \in U \text{ oder } \alpha \rightarrow \beta \in D\}$ . Eine *Zusammenhangskomponente* oder *Box* von  $\mathcal{V}$  ist eine maximale Teilmenge  $B \subseteq V$  von Knoten, in denen jeder Knoten  $\alpha \in B$  mit jedem anderen Knoten  $\beta \in B$  über einen ungerichteten Pfad verbunden ist. Um hervorzuheben, dass es sich bei einer Teilmenge von Knoten  $B \subseteq V$  um eine Box handelt, wird diese im Folgenden in eckige Klammern gesetzt,  $[B] \subseteq V$ . Die Menge aller Boxen von  $\mathcal{V}$  sei bezeichnet mit  $\mathcal{B}(\mathcal{V})$ . Hierbei handelt es sich folglich um eine Teilmenge der Potenzmenge von  $V$ , die die Knotenmenge  $V$  partitioniert,  $V = \cup_{[B] \in \mathcal{B}(\mathcal{V})} [B]$ .

Zu einem gegebenen azyklisch gemischten Graphen  $\mathcal{V} = (V, D, U)$  wird der *Graph der Boxen*  $GoB(\mathcal{V}) = (V^B, D^B, \emptyset)$  definiert als der folgende azyklisch gerichtete Graph: Jeder Knoten repräsentiert eine Box von  $\mathcal{V}$ , das heißt  $V^B = \{[B_1], \dots, [B_k]\}$  mit  $k := |\mathcal{B}(\mathcal{V})|$ . Außerdem ist für  $i \neq j$  die gerichtete Kante  $[B_i] \rightarrow [B_j] \in D^B$  genau dann vorhanden, wenn es einen Knoten  $\alpha \in [B_i]$  und einen Knoten  $\beta \in [B_j]$  gibt, so dass die Kante  $\alpha \rightarrow \beta$  in  $\mathcal{V}$  auftaucht. Der Graph der Boxen  $GoB(\mathcal{V})$  entsteht also, indem jede Box von  $\mathcal{V}$  zu einem Knoten zusammengefasst wird und die durch die gerichteten Kanten von  $\mathcal{V}$  vorgegebene Abhängigkeitsstruktur auf den neuen Graphen übertragen wird. Eine Box  $[B] \in \mathcal{B}(\mathcal{V})$  heißt *maximal*, wenn im Graph der Boxen  $GoB(\mathcal{V})$  keine Kante aus dem korrespondierenden Knoten herauszeigt, das heißt  $[B] \rightarrow [B'] \notin D^B$  für alle  $[B'] \in \mathcal{B}(\mathcal{V})$ .

Eine *Immoralität* in  $\mathcal{V}$  ist eine dreielementige Teilmenge von Knoten  $A = \{\alpha, \gamma, \beta\} \subseteq V$ , so dass der induzierte Teilgraph  $\mathcal{V}_A$  die Form  $\alpha \rightarrow \gamma \leftarrow \beta$  hat. Der Name stammt daher, dass die Knoten  $\alpha$  und  $\beta$  zwar ein gemeinsames Kind  $\gamma$  haben, allerdings nicht verheiratet, das heißt durch eine Kante verbunden, sind. Induziert  $A$  dagegen einen Teilgraphen der Form  $\alpha - \gamma \leftarrow \beta$ , so spricht man von einer *Flagge* in  $\mathcal{V}$ . Im Folgenden werden oft azyklisch gemischte Graphen untersucht, die die folgenden Annahmen erfüllen, vergleiche Andersson und Klein [3, Kapitel 4].

**Annahme 2.1 (Bedingungen an azyklisch gemischte Graphen)**

Sei  $\mathcal{V} = (V, D, U)$  ein azyklisch gemischter Graph, der die folgenden Bedingungen erfüllt:

(A1) Für jede Box  $[B] \in \mathcal{B}(\mathcal{V})$  sei der induzierte Teilgraph  $\mathcal{V}_{[B]}$  vollständig.

(A2)  $\mathcal{V}$  enthalte keine Immoralitäten oder Fahnen.

Im Folgenden habe  $\mathcal{V}$  immer  $p = |V|$  Knoten und  $k = |\mathcal{B}(\mathcal{V})|$  Boxen.

Diese Annahmen führen zum einen dazu, dass zwei Knoten  $\alpha$  und  $\beta$ , die ein gemeinsames Kind  $\gamma$  haben, entweder in der gleichen Box liegen oder durch eine gerichtete Kante verbunden sein müssen. Andernfalls entstünde eine Immoralität  $\alpha \rightarrow \gamma \leftarrow \beta$ . Zum anderen implizieren die Annahmen, dass sobald ein Knoten  $\alpha$  aus der Box  $[B]$  ein Kind  $\beta$  in der Box  $[B']$  hat, jeder weitere Knoten  $\gamma \in [B']$  auch ein Kind von  $\alpha$  ist. Sonst entstünde wegen Annahme (A1) eine Flagge  $\alpha \rightarrow \beta - \gamma$ . Folglich haben alle Knoten  $\beta \in [B']$  dieselbe Elternmenge  $\text{pa}(\beta)$  und diese wird im Folgenden auch mit  $\langle B' \rangle$  bezeichnet.

Des Weiteren treten im Gerüst  $\mathcal{V}^\sim$  eines Annahme 2.1 erfüllenden azyklisch gemischten Graphen  $\mathcal{V}$  keine sehnlosen Zyklen der Länge  $k \geq 4$  auf. Folglich ist  $\mathcal{V}^\sim$  zerlegbar. Dies motiviert die folgende Definition: Ist  $\mathcal{U}$  ein zerlegbarer ungerichteter Graph und  $\mathcal{V}$  ein Annahme 2.1 erfüllender azyklisch gemischter Graph mit  $\mathcal{V}^\sim = \mathcal{U}$ , so heißt  $\mathcal{V}$  die *Darstellung von  $\mathcal{U}$  (als azyklisch gemischter Graph)*. Da jeder zerlegbare Graph  $\mathcal{U}$  durch Umwandeln von ungerichteten Kanten in gerichtete Kanten in einen azyklisch gerichteten Graphen verwandelt werden kann, vergleiche Andersson, Madigan und Perlman [7], gibt es für jeden zerlegbaren Graphen  $\mathcal{U}$  mindestens eine Darstellung  $\mathcal{V}$ . Im Allgemeinen ist diese Darstellung allerdings nicht eindeutig, beispielsweise hat schon der einfache Graph  $\mathcal{U} = \alpha - \beta$  die Darstellungen  $\alpha - \beta$ ,  $\alpha \rightarrow \beta$  und  $\alpha \leftarrow \beta$ .

Gerade um später den Umgang mit Blockmatrizen anschaulicher zu machen, ist es nützlich, die Knoten in einem beliebigen azyklisch gemischten Graphen auf eine gewisse kanonische Weise zu nummerieren. Die folgende Beschreibung stellt keine Einschränkung dar, sondern nur ein Hilfsmittel zur Verbesserung der Anschaulichkeit.

**Annahme 2.2 (Intrinsische Nummerierung der Knoten)**

Sei ein azyklisch gemischter Graph  $\mathcal{V}$  gegeben. Nummeriere dann dessen  $k$  Boxen und  $p$  Knoten auf die folgende Weise:

- (i) Bestimme den Graph der Boxen  $GoB(\mathcal{V})$ . Nummeriere die Boxen so, dass dort alle gerichteten Kanten von einer Box mit kleinerem Index zu einer Box mit größerem Index zeigen.
- (ii) Gehe nun die Boxen  $[B_1], \dots, [B_k]$  in dieser Reihenfolge durch. Die Knoten innerhalb einer Box dürfen beliebig nummeriert werden. Knoten, die aus einer Box mit höherem Index stammen, müssen aber immer einen höheren Index bekommen als Knoten, die in einer Box mit niedrigerem Index liegen.

**Matrizenindizierung für einen vorgegebenen azyklisch gemischten Graphen**

Sei ein azyklisch gemischter Graph  $\mathcal{V} = (V, D, U)$  mit Boxen  $\mathcal{B}(\mathcal{V}) = \{[B_1], \dots, [B_k]\}$  vorgegeben. Beachte, dass sowohl  $[B_i]$  als auch  $\langle B_i \rangle$  für alle  $i = 1, \dots, k$  jeweils eine Teilmenge von  $V = \{1, \dots, p\}$  darstellt. Mit den Notationen aus Kapitel 1.3 von Seite 6 sind damit für alle  $[A], [B] \in \mathcal{B}(\mathcal{V})$  die folgenden Submatrizen einer Matrix  $W \in \mathbb{R}^{p \times p}$  wohldefiniert,

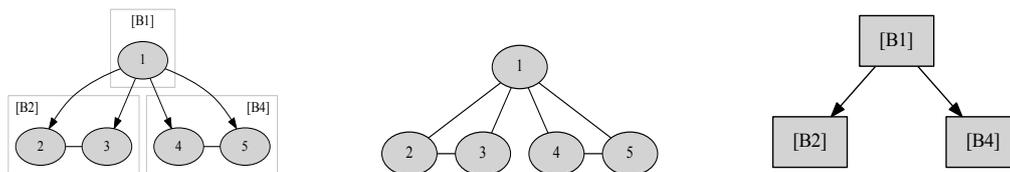
$$\begin{aligned} W_{[A],[B]} &:= W_{[A],[B]} \in \mathbb{R}^{[A] \times [B]}, & W_{\langle A \rangle, \langle B \rangle} &:= W_{\langle A \rangle, \langle B \rangle} \in \mathbb{R}^{\langle A \rangle \times \langle B \rangle}, \\ W_{[B]} &:= W_{[B],[B]} \in \mathbb{R}^{[B] \times [B]}, & W_{[B]} &:= W_{[B], \langle B \rangle} \in \mathbb{R}^{[B] \times \langle B \rangle}, \\ W_{\langle B \rangle} &:= W_{\langle B \rangle, [B]} \in \mathbb{R}^{\langle B \rangle \times [B]}, & W_{\langle B \rangle} &:= W_{\langle B \rangle, \langle B \rangle} \in \mathbb{R}^{\langle B \rangle \times \langle B \rangle}. \end{aligned}$$

Wie bereits erwähnt wird der Übersichtlichkeit halber oft auf die Kardinalitätsstriche verzichtet, wenn die Mächtigkeit einer Box  $|[B]|$  oder ihrer Eltern  $|\langle B \rangle|$  auftaucht. Mit  $\mathbb{R}^{[B] \times [B]}$  ist also beispielsweise  $\mathbb{R}^{|[B]| \times |[B]|}$  gemeint, und so weiter.

**Beispiel 2.3 (Graphentheoretische Konzepte)**

Abschließend sollen die bisher eingeführten Definitionen anhand eines Beispiels erläutert werden. Sei dazu  $\mathcal{V} = (V, D, U)$  der azyklisch gemischte Graph aus Abbildung 1a. Hier ist  $V = \{1, 2, 3, 4, 5\}$ ,  $D = \{1 \rightarrow j \mid j = 2, \dots, 5\}$  und  $U = \{2-3, 4-5\}$ . Folglich gibt es  $p = |V| = 5$  Knoten und  $k = |\mathcal{B}(\mathcal{V})| = 3$  Boxen, nämlich  $[B_1] = \{1\}$ ,  $[B_2] = \{2, 3\}$  und  $[B_4] = \{4, 5\}$ .<sup>1</sup>

<sup>1</sup>Der Übersichtlichkeit halber werden in vielen Beispielen dieser Arbeit die Boxen jeweils nicht von 1 bis  $k$  durchnummeriert. Stattdessen wird eine Box  $[B_i]$  benannt nach dem kleinsten Knoten  $i$ , den sie enthält.


 (a) Azyklisch gemischter Graph  $\mathcal{V}$  (b) Zerlegbares Gerüst  $\mathcal{U} = \mathcal{V}^\sim$  (c) Graph der Boxen  $GoB(\mathcal{V})$ 

**Abbildung 1:** Ein azyklisch gemischter Graph  $\mathcal{V}$  (links). Sein Gerüst  $\mathcal{U} = \mathcal{V}^\sim$  entsteht durch Ersetzen aller gerichteten Kanten durch ungerichtete Kanten (Mitte). Die kausale Abhängigkeit der Boxen von  $\mathcal{V}$  ist im Graph der Boxen  $GoB(\mathcal{V})$  kodiert (rechts).

Den induzierten Graph der Boxen sieht man in Abbildung 1c. Hier erkennt man auch, dass die beiden Boxen  $[B_2]$  und  $[B_4]$  in  $\mathcal{V}$  maximal sind. Außerdem erfüllt  $\mathcal{V}$  die Annahme 2.1. Daher ist das Gerüst  $\mathcal{U} := \mathcal{V}^\sim$  zerlegbar, siehe Abbildung 1b. Umgekehrt ist  $\mathcal{V}$  eine Darstellung von  $\mathcal{U}$  als azyklisch gemischter Graph. Sei schließlich  $W = (w_{ij}) \in \mathbb{R}^{5 \times 5}$  eine symmetrische Matrix. Mit der oben beschriebenen Indizierungstechnik ergeben sich unter Beachtung von  $\langle B_2 \rangle = \{1\} = \langle B_4 \rangle$  die Partitionierungen

$$W = \begin{pmatrix} W_{[B_1]} & W_{[B_1, B_2]} & W_{[B_1, B_4]} \\ W_{[B_2, B_1]} & W_{[B_2]} & W_{[B_2, B_4]} \\ W_{[B_4, B_1]} & W_{[B_4, B_2]} & W_{[B_4]} \end{pmatrix} = \begin{pmatrix} W_{[B_1]} & W_{[B_2]}^T & W_{[B_4]}^T \\ W_{[B_2]} & W_{[B_2]} & W_{[B_4, B_2]}^T \\ W_{[B_4]} & W_{[B_4, B_2]} & W_{[B_4]} \end{pmatrix}.$$

Die auftretenden Teilmatrizen haben dabei die komponentenweise Darstellung

$$\begin{aligned} W_{[B_1]} &= (w_{11}), & W_{[B_2]} &= \begin{pmatrix} w_{22} & w_{23} \\ w_{23} & w_{33} \end{pmatrix}, & W_{[B_4]} &= \begin{pmatrix} w_{44} & w_{45} \\ w_{45} & w_{55} \end{pmatrix}, \\ W_{[B_2]} &= \begin{pmatrix} w_{21} \\ w_{31} \end{pmatrix}, & W_{[B_4]} &= \begin{pmatrix} w_{41} \\ w_{51} \end{pmatrix}, & W_{[B_4, B_2]} &= \begin{pmatrix} w_{42} & w_{43} \\ w_{52} & w_{53} \end{pmatrix}. \end{aligned}$$

## 2.2 Graphische Modelle mit Normalverteilungsannahme

Um einen Bogen zur Statistik zu spannen, werden die vorgestellten graphentheoretischen Konzepte dazu verwendet, statistische Modelle zu definieren, die eine gewisse Abhängigkeitsstruktur der beteiligten Variablen widerspiegeln. Dabei wird weitestgehend dem Ansatz aus Andersson und Klein [3] gefolgt.

### Parameterräume graphischer Modelle

Sei ein ungerichteter Graph  $\mathcal{U} = (V, \emptyset, U)$  über den Knoten  $V = \{1, \dots, p\}$  vorgegeben. Außerdem sei  $S(\mathcal{U})$  die Menge aller symmetrischen  $p \times p$ -Matrizen, für die der Eintrag  $S_{\alpha\beta}$  verschwindet, falls die korrespondierende Kante  $\alpha - \beta$  nicht in  $\mathcal{U}$  auftaucht,

$$S(\mathcal{U}) := \{W \in \mathbb{R}^{p \times p} \text{ symmetrisch} \mid \alpha - \beta \notin U \Rightarrow W_{\alpha\beta} = 0\}.$$

Hierbei handelt es sich offensichtlich um einen linearen Unterraum des Vektorraums der symmetrischen  $p \times p$ -Matrizen. Der Kegel aller symmetrischen und positiv definiten  $p \times p$ -Matrizen sei bezeichnet mit  $PD_p$ . Definiere dann die drei Mengen

$$\begin{aligned} PD^0(\mathcal{U}) &:= PD_p \cap S(\mathcal{U}) = \{K \in PD_p \mid \alpha - \beta \notin U \Rightarrow K_{\alpha\beta} = 0\}, \\ PD(\mathcal{U}) &:= \{\Sigma \in PD_p \mid \Sigma^{-1} \in PD^0(\mathcal{U})\} \text{ und} \\ P(\mathcal{U}) &:= \{\Omega \in S(\mathcal{U}) \mid \Omega_{C,C} \in PD_{|C|} \text{ für alle Cliques } C \in \mathcal{C}(\mathcal{U})\}. \end{aligned} \tag{2.1}$$

Eine fehlende Kante in  $\mathcal{U}$  impliziert damit einen gewissen Nulleintrag in allen Matrizen der Menge  $PD^0(\mathcal{U})$ . Letztere ist ein konvexer Kegel und eine offene Teilmenge des Unterraums  $S(\mathcal{U})$ . Da die Matrizen in  $P(\mathcal{U})$  dieselbe vorgegebene Nullstruktur wie die Matrizen in  $PD^0(\mathcal{U})$  haben, ist auch  $P(\mathcal{U})$  ein konvexer Kegel und eine offene Teilmenge von  $S(\mathcal{U})$ . Beachte allerdings, dass  $P(\mathcal{U})$  im Allgemeinen keine Teilmenge des Kegels der positiv definiten Matrizen  $PD_p$  ist, sondern auch Matrizen  $\Omega \notin PD_p$  enthalten kann.<sup>2</sup> Die Geometrie des Parameterraums  $PD(\mathcal{U})$  dagegen ist komplizierter als die der anderen beiden. Im Allgemeinen handelt es sich hierbei nicht um einen konvexen Kegel, sondern nur um eine offene Untermannigfaltigkeit von  $S(\mathcal{U})$ , siehe Andersson und Klein [3, Seite 792]. Auf die statistische Bedeutung der definierten Mengen wird später ab Seite 13 noch genauer eingegangen.

Es sei jedoch schon einmal hervorgehoben, dass es sich bei „dem graphischen Modell“ eigentlich nicht um die Menge  $PD(\mathcal{U})$  handelt, sondern um die korrespondierende Verteilungsfamilie  $\{\mathcal{N}_p(0, \Sigma) \mid \Sigma \in PD(\mathcal{U})\}$ , wobei  $\mathcal{N}_p(0, \Sigma)$  die zentrierte  $p$ -variate Normalverteilung zur Streuungsmatrix  $\Sigma$  meint. Da die hier untersuchten Modelle aber alle identifizierbar sind, das heißt die Parametrisierungsabbildung  $\Sigma \mapsto \mathcal{N}_p(0, \Sigma)$  ist bijektiv, werden im Folgenden sowohl die Verteilungsfamilie als auch der Parameterraum als „graphisches Modell“ bezeichnet.

<sup>2</sup>Ein einfaches Beispiel ist die Matrix  $\begin{pmatrix} 4 & 3 & 0 \\ 3 & 4 & 3 \\ 0 & 3 & 4 \end{pmatrix}$ . Sie ist zwar indefinit, liegt aber für  $\mathcal{U} = 1 - 2 - 3$  in  $P(\mathcal{U})$ , da die Teilmatrix  $\begin{pmatrix} 4 & 3 \\ 3 & 4 \end{pmatrix}$  positiv definit ist.

Zunächst soll geklärt werden, wie die drei obigen Parameterräume zusammenhängen. Per Definition ist die klassische Matrizeninvertierung eine bijektive Abbildung zwischen  $PD(\mathcal{U})$  und  $PD^0(\mathcal{U})$ . Des Weiteren ist das Invertieren einer Matrix gemäß der Cramerschen Regel eine rationale Abbildung der ursprünglichen Matrix. Folglich ist die Invertierung über den positiv definiten Matrizen stetig differenzierbar, so dass  $PD(\mathcal{U})$  und  $PD^0(\mathcal{U})$  diffeomorph sind. Einen ähnlichen Zusammenhang gibt es zwischen  $PD(\mathcal{U})$  und  $P(\mathcal{U})$ . Betrachte dazu die Abbildung

$$\pi : \mathbb{R}^{p \times p} \rightarrow S(\mathcal{U}),$$

$$W \mapsto \pi(W) \quad \text{mit} \quad \pi(W)_{\alpha\beta} = \begin{cases} W_{\alpha\beta} & \text{falls } \alpha = \beta \text{ oder } \alpha - \beta \in U \\ 0 & \text{sonst} \end{cases}. \quad (2.2)$$

Anschaulich streicht  $\pi$  somit alle Einträge aus der Matrix  $W \in \mathbb{R}^{p \times p}$ , die nicht zum vorgegebenen Graphen  $\mathcal{U}$  passen. Beachte insbesondere, dass  $\pi$  somit vom Graphen  $\mathcal{U}$  abhängt. Es zeigt sich, dass  $\pi$  einen Diffeomorphismus zwischen den Parameterräumen  $PD(\mathcal{U})$  und  $P(\mathcal{U})$  induziert.

**Lemma 2.4 (Diffeomorphismus zwischen  $PD(\mathcal{U})$  und  $P(\mathcal{U})$ )**

Sei  $\mathcal{U}$  ein zerlegbarer Graph und  $\pi$  wie in Gleichung (2.2) definiert. Dann gilt:

- (i) Das Bild der Einschränkung von  $\pi$  auf  $PD(\mathcal{U})$  ist  $P(\mathcal{U})$ .
- (ii) Die Abbildung  $\pi : PD(\mathcal{U}) \rightarrow P(\mathcal{U})$  ist ein Diffeomorphismus.

**Beweis**

Behauptung (i) und die Bijektivität von  $\pi : PD(\mathcal{U}) \rightarrow P(\mathcal{U})$  folgen aus Andersson und Klein [3, Proposition 3.1]. Außerdem sind die Komponenten von  $\pi$  jeweils stetig differenzierbar, so dass  $\pi$  insgesamt ebenfalls stetig differenzierbar und somit diffeomorph ist. □

Die explizite Form der Umkehrabbildung  $\pi^{-1} : P(\mathcal{U}) \rightarrow PD(\mathcal{U})$  wird in Theorem 3.13 auf Seite 45 angegeben.

**Statistische Bedeutung der definierten Parameterräume**

Sei wieder ein zerlegbarer Graph  $\mathcal{U} = (V, \emptyset, U)$  vorgegeben. Jeder der  $p$  Knoten  $\alpha \in V$  soll mit einer univariaten Zufallsgröße  $X_\alpha$  identifiziert werden. Der Graph  $\mathcal{U}$  soll die Abhängigkeitsstruktur der Verteilung  $\mathbb{P}$  des so entstehenden  $p$ -variaten Zufallsvektors  $X = (X_\alpha)_{\alpha \in V}$  repräsentieren. Dies geschieht mit Hilfe der *globalen Markov-Eigenschaft*.

Sie besagt, dass für je zwei Teilmengen  $A, B \subseteq V$ , die von einer dritten Menge  $S \subseteq V$  in  $\mathcal{U}$  getrennt werden, die Zufallsvektoren  $X_A$  und  $X_B$  unabhängig sind gegeben  $X_S$ ,

$$\forall A, B, S \subseteq V: \quad S \text{ trennt } A \text{ und } B \text{ in } \mathcal{U} \Rightarrow X_A \perp X_B \mid X_S \ [\mathbb{P}].$$

Genauere Ausführungen zur globalen Markov-Eigenschaft und die Definition der bedingten Unabhängigkeit findet man beispielsweise bei Lauritzen [18, Kapitel 3]. Zwei wichtige Eigenschaften der bedingten Unabhängigkeit werden jedoch im Laufe der Arbeit mehrmals benötigt und daher im folgenden Lemma zusammengefasst.

**Lemma 2.5 (Grundlegende Eigenschaften der bedingten Unabhängigkeit)**

Seien  $X, Y, Z$  und  $W$  Zufallsgrößen und die Verteilung des Zufallsvektors  $(X, Y, Z, W)$  sei bezeichnet mit  $\mathbb{P}$ . Außerdem sei  $h$  eine messbare Funktion über dem Grundraum von  $X$ . Dann gelten die folgenden Aussagen.

- (i) Aus  $X \perp Y \mid Z \ [\mathbb{P}]$  und  $W = h(X)$  folgt  $W \perp Y \mid Z \ [\mathbb{P}]$ .
- (ii) Die Beziehungen  $X \perp Y \mid Z \ [\mathbb{P}]$  und  $X \perp W \mid \{Y, Z\} \ [\mathbb{P}]$  gelten genau dann, wenn  $X \perp \{W, Y\} \mid Z \ [\mathbb{P}]$  ist.
- (iii) Hängt die bedingte Verteilung  $\mathcal{L}(X|Y, Z)$  nur von  $Z$  ab, so ist  $X \perp Y \mid Z \ [\mathbb{P}]$  und  $\mathcal{L}(X|Y, Z) = \mathcal{L}(X|Z)$ .

**Beweis**

Alle drei Aussagen folgen direkt aus der Definition der bedingten Unabhängigkeit in Lauritzen [18, Seite 28] und den Gleichungen (C2), (C4) und (3.4) in Lauritzen [18, Seite 29]. □

Nun sei der obige Zufallsvektor  $X$  als zentriert normalverteilt angenommen, das heißt  $X \sim \mathcal{N}_p(0, \Sigma)$ , wobei  $\mathcal{N}_p(0, \Sigma)$  wieder die vektorwertige Normalverteilung mit  $p$  Komponenten zum Mittelwertvektor  $0 \in \mathbb{R}^p$  und zur Streuungsmatrix  $\Sigma \in PD_p$  meint.<sup>3</sup> Dann lässt sich die Menge aller möglichen Kovarianzmatrizen  $\Sigma$ , für die die korrespondierende Verteilung  $\mathcal{N}_p(0, \Sigma)$  die globale Markoveigenschaft erfüllt, auf einfache Weise charakterisieren. Es handelt sich gerade um die oben definierte Menge  $PD(\mathcal{U})$ , vergleiche beispielsweise Lauritzen [18, Kapitel 5.1.3]. Dementsprechend ist  $PD^0(\mathcal{U})$  die Menge aller

---

<sup>3</sup>Die Annahme, dass der Mittelwertvektor verschwinden soll, begründet sich darin, dass der Fokus dieser Arbeit auf der Abhängigkeitsstruktur der beobachteten Variablen liegt. Auf diese hat bei normalverteilten Daten der Mittelwertparameter  $\xi \in \mathbb{R}^p$  keinen Einfluss. Daher ist für die im Folgenden untersuchten Fragestellungen der Fall  $\xi = 0$  ausreichend. Für andere Probleme kann es jedoch durchaus nötig sein, den Mittelwert zu untersuchen. In diesem Fall können die Techniken aus Andersson und Perlman [5, Kapitel 6] adaptiert werden, um die Modelle aus der vorliegenden Arbeit zu Modellen mit nicht-trivialem Mittelwertparameter zu verallgemeinern, vergleiche auch Madsen [20, Seite 1178].

Konzentrationsmatrizen  $K$ , für die die korrespondierende Normalverteilung  $\mathcal{N}_p(0, K^{-1})$  die globale Markov-Eigenschaft erfüllt. Mit diesem Hintergrund lässt sich nun auch die Menge  $P(\mathcal{U})$  besser veranschaulichen. Sie ist nämlich diffeomorph zur Menge der markovschen Streuungsmatrizen  $PD(\mathcal{U})$ . Außerdem ist der Diffeomorphismus  $\pi$  von einer sehr einfachen Form, er streicht nur Einträge aus  $\Sigma \in PD(\mathcal{U})$ , die zu nicht-vorhandenen Kanten in  $\mathcal{U}$  korrespondieren. Mit anderen Worten gibt es zu jeder Matrix  $\Omega \in P(\mathcal{U})$  genau eine zugehörige markovsche Streuungsmatrix  $\Sigma \in PD(\mathcal{U})$  und diese entsteht durch Auffüllen der strukturellen Nulleinträge in  $\Omega$ . Alle Einträge  $\Sigma_{\alpha\beta}$ , die zu einer nicht-vorhandenen Kante  $\alpha - \beta \notin \mathcal{U}$  korrespondieren, sind folglich schon eindeutig bestimmt durch die restlichen Komponenten  $\Sigma_{\gamma\delta}$  mit  $\gamma = \delta$  oder  $\gamma - \delta \in \mathcal{U}$ . Dies bedeutet, dass eine Matrix  $\Omega \in P(\mathcal{U})$  den vollkommen identischen Informationsgehalt hat wie die kompliziertere Matrix  $\Sigma = \pi^{-1}(\Omega) \in PD(\mathcal{U})$ . Dies wird später bei der Maximum-Likelihood-Schätzung in Kapitel 4 von Nutzen sein, da nicht mehr die ganze Streuungsmatrix  $\Sigma$  geschätzt werden muss, sondern ein Schätzer der projizierten Matrix  $\Omega = \pi(\Sigma)$  ausreicht.

Abschließend soll ein Blick auf die Dimensionen der drei Parameterräume als Teilmenge des Unterraums  $S(\mathcal{U})$  geworfen werden. Da alle drei diffeomorph zueinander und jeweils offene Teilmengen von  $S(\mathcal{U})$  sind, entspricht ihre Dimension einfach der Dimension von  $S(\mathcal{U})$ . Letztere bestimmt sich zu  $p + |\mathcal{U}|$ , sie kann also gerade in Situationen mit vielen Variablen sehr groß werden. Oft sind solche Modelle dann nicht mehr mit vertretbarem Aufwand analysierbar, da eine hohe Dimensionalität des Parameterraums beispielsweise mit einer hohen Laufzeit bei der Durchführung eines Likelihood-Quotienten-Tests einhergehen kann. Eine Möglichkeit, die Dimension von  $P(\mathcal{U})$  zu verringern, ist die Berücksichtigung von Symmetrien, was in Kapitel 2.4 näher untersucht wird.

Um diese Einbindung von Invarianzen zu bewerkstelligen, wird im Folgenden der Ansatz aus dem Aufsatz von Madsen [20] verwendet. Leider sind die Rahmenbedingungen dort anders formuliert als in den Kapiteln 2.1 und 2.2. Daher wird sich der abschließende Teil des aktuellen Kapitels damit beschäftigen, wie die Ausgangssituation in Madsen [20] in der Notation der vorliegenden Arbeit beschrieben werden kann.

### Multivariat gerichtete Graphen vs. univariat gemischte Graphen

Madsen verwendet in seinem Aufsatz keine gemischten Graphen mit univariaten Knoten, sondern gerichtete Graphen mit multivariaten Knoten. Sei daher  $\mathcal{D} = (V_{\mathcal{D}}, D_{\mathcal{D}}, \emptyset)$  ein azyklisch gerichteter Graph ohne Immoralitäten mit Knotenmenge  $V_{\mathcal{D}} = \{1, \dots, k\}$ . Außerdem sei  $\mathcal{B} = \{B_1, \dots, B_k\}$  eine Partitionierung der Menge  $V = \{1, \dots, p\}$  in  $k$  disjunkte Mengen. Dann erfüllt eine positiv definite Matrix  $\Sigma \in PD_p$  die *gerichtete Markov-Eigenschaft bezüglich  $\mathcal{D}$  und  $\mathcal{B}$* , falls für jeden zentriert normalverteilten Zufallsvektor  $X$

mit Streuungsmatrix  $\Sigma$  die (bedingten) Unabhängigkeitsrestriktionen

$$\{X_{B_i}\} \perp\!\!\!\perp \{X_{B_j}\}_{j \in \text{nd}_{\mathcal{D}}(i) \setminus \text{pa}_{\mathcal{D}}(i)} \mid \{X_{B_\ell}\}_{\ell \in \text{pa}_{\mathcal{D}}(i)} \quad \forall i = 1, \dots, k$$

halten, vergleiche Madsen [20, Seite 1153]. Damit gibt der Graph  $\mathcal{D}$  die Abhängigkeitsstruktur eines  $p$ -variaten Zufallsvektors wieder, wobei jeder Knoten  $i$  in  $\mathcal{D}$  zu einer multivariaten Zufallsgröße mit  $|B_i|$  Komponenten korrespondiert. Die Menge aller Streuungsmatrizen  $\Sigma \in PD_p$ , die die gerichtete Markov-Eigenschaft bezüglich  $\mathcal{D}$  und  $\mathcal{B}$  erfüllen, sei bezeichnet mit  $\mathcal{M}(\mathcal{D}, \mathcal{B})$ .

Als nächstes wird ein gemischter Graph mit univariaten Knoten konstruiert, der die gleiche Abhängigkeitsstruktur wie  $\mathcal{D}$  repräsentiert. Setze dazu wieder  $V = \{1, \dots, p\}$ , das heißt der neue Graph wird  $p$  Knoten enthalten. Verbinde dann alle Knoten, die in derselben Menge  $B \in \mathcal{B}$  liegen, durch ungerichtete Kanten. Schließlich sollen im neuen Graphen zwei Knoten genau dann durch eine gerichtete Kante verbunden sein, wenn die korrespondierenden multivariaten Knoten in  $\mathcal{D}$  durch eine gerichtete Kante verbunden sind. Man setzt also

$$U = \{\alpha - \beta \mid \alpha, \beta \in B \text{ für ein } B \in \mathcal{B}\} \quad \text{und}$$

$$D = \{\alpha \rightarrow \beta \mid \alpha \in B_i, \beta \in B_j, i \rightarrow j \in D_{\mathcal{D}}\}.$$

Der so konstruierte gemischte Graph  $\mathcal{V}_{\mathcal{D}, \mathcal{B}} = (V, D, U)$  heißt der vom azyklisch gerichteten Graphen  $\mathcal{D}$  und der Partitionierung  $\mathcal{B}$  induzierte azyklisch gemischte Graph  $\mathcal{V}_{\mathcal{D}, \mathcal{B}}$ .

Diese Konstruktion bringt die Notation aus der vorliegenden Arbeit und die Rahmenbedingungen aus Madsen [20] in Einklang. Für einen gegebenen azyklisch gerichteten Graphen  $\mathcal{D}$  und eine gegebene Partitionierung  $\mathcal{B}$  wie oben setze dazu  $\mathcal{V} := \mathcal{V}_{\mathcal{D}, \mathcal{B}}$  und  $\mathcal{U} := \mathcal{V}^\sim$ . Dann sind per Konstruktion der in Madsen [20] verwendete Parameterraum  $\mathcal{M}(\mathcal{D}, \mathcal{B})$  und das in Kapitel 2.2 definierte Modell  $PD(\mathcal{U})$  identisch, vergleiche auch Andersson, Madigan und Perlman [7]. Außerdem sei hervorgehoben, dass die Definition von  $\mathcal{V} = \mathcal{V}_{\mathcal{D}, \mathcal{B}}$  dazu führt, dass die Partitionierung  $\mathcal{B}$  gerade der Menge der Boxen von  $\mathcal{V}$  entspricht, das heißt  $\mathcal{B} = \mathcal{B}(\mathcal{V})$ . Zudem wird die kausale Struktur von  $\mathcal{V}$  durch  $\mathcal{D} = \text{GoB}(\mathcal{V})$  schon eindeutig festgelegt. Im Folgenden sollen azyklisch gemischte Graphen der Form  $\mathcal{V}_{\mathcal{D}, \mathcal{B}}$  näher untersucht werden. Die essentiellen Eigenschaften dieser Graphen werden in der folgenden Annahme zusammengefasst.

**Annahme 2.6 (Verschärfte Bedingungen an azyklisch gemischte Graphen)**

Sei  $\mathcal{V} = (V, D, U)$  ein azyklisch gemischter Graph, der die folgenden Bedingungen erfüllt:

- (A1) Für jede Box  $[B] \in \mathcal{B}(\mathcal{V})$  sei der induzierte Teilgraph  $\mathcal{V}_{[B]}$  vollständig.
- (A2) Der Graph  $\mathcal{V}$  enthalte keine Immoralitäten.
- (A3) Für alle Boxen  $[A], [B] \in \mathcal{B}(\mathcal{V})$  soll gelten: Falls es Knoten  $\alpha \in [A]$  und  $\alpha' \in [B]$  mit  $\alpha \rightarrow \alpha' \in D$  gibt, so ist  $\beta \rightarrow \beta' \in D$  für alle  $\beta \in [A]$  und alle  $\beta' \in [B]$ .

Im Folgenden habe  $\mathcal{V}$  immer  $p = |V|$  Knoten und  $k = |\mathcal{B}(\mathcal{V})|$  Boxen.

Beachte, dass in einem azyklisch gemischtem Graphen  $\mathcal{V}$ , der diese Annahme erfüllt, keine Flaggen auftreten können. Annahme 2.6 ist damit eine stärkere Forderung an  $\mathcal{V}$  als Annahme 2.1 auf Seite 9. Beachte außerdem, dass ein Graph der Form  $\mathcal{V}_{\mathcal{D}, \mathcal{B}}$  die Forderungen (A1) bis (A3) erfüllt. Auch hier sei hervorgehoben, dass es zu jedem zerlegbaren Graphen  $\mathcal{U}$  einen azyklisch gemischtem Graphen  $\mathcal{V}$  mit Gerüst  $\mathcal{V}^{\sim} = \mathcal{U}$  gibt, der Annahme 2.6 erfüllt. Wie bereits erwähnt, lässt sich nämlich jeder zerlegbare Graph durch Umwandeln von ungerichteten Kanten in gerichtete Kanten zu einem azyklisch gerichteten Graphen machen, siehe Andersson, Madigan und Perlman [7].

Des Weiteren lassen sich die Cliques des Gerüsts  $\mathcal{U} = \mathcal{V}^{\sim}$  von  $\mathcal{V}$  aus Annahme 2.6 charakterisieren, denn aufgrund von (A1) bis (A3) ist  $[B] \cup \langle B \rangle$  für alle  $[B] \in \mathcal{B}(\mathcal{V})$  eine vollständige Knotenmenge in  $\mathcal{U}$ . Umgekehrt muss sich jede Clique  $C$  von  $\mathcal{U}$  in der Form  $[B] \cup \langle B \rangle$  darstellen lassen. Die Menge aller Cliques von  $\mathcal{U}$  lässt sich also beschreiben durch

$$\mathcal{C}(\mathcal{U}) = \left\{ [B] \cup \langle B \rangle \mid [B] \in \mathcal{B}(\mathcal{V}) \text{ und } \nexists [B'] \neq [B] : [B] \cup \langle B \rangle \subsetneq [B'] \cup \langle B' \rangle \right\}. \quad (2.3)$$

**2.3 Permutationen, Orbits und Automorphismen**

Bevor im nächsten Abschnitt erläutert wird, wie gewisse Invarianz-Restriktionen in die graphischen Modelle aus Kapitel 2.2 eingebunden werden können, werden zunächst einige Konzepte aus der Gruppentheorie eingeführt.

**Orthogonale Transformationen und Permutationen**

Im Folgenden wird für  $\ell \in \mathbb{N}$  die Menge aller orthogonalen  $\ell \times \ell$ -Matrizen bezeichnet mit  $Orth(\ell)$ . Außerdem sei  $Sym(\ell)$  die Menge aller Permutationen vom Grad  $\ell$ . Es ist also

$$Orth(\ell) = \{ \rho \in \mathbb{R}^{\ell \times \ell} \mid \rho \rho^T = I_\ell = \rho^T \rho \},$$

$$Sym(\ell) = \{ \sigma \mid \sigma : \{1, \dots, \ell\} \rightarrow \{1, \dots, \ell\} \text{ bijektiv} \}.$$

Schließlich sei  $Perm(\ell)$  die Menge aller  $\ell \times \ell$ -Permutationsmatrizen. Insbesondere sind  $Perm(\ell)$  und  $Sym(\ell)$  isomorph, der kanonische Isomorphismus ist gegeben durch

$$\varrho_\ell : Sym(\ell) \xrightarrow{\cong} Perm(\ell), \quad \sigma \xrightarrow{\cong} \varrho_\ell(\sigma) = \sum_{j=1}^{\ell} e_{\sigma(j)} e_j^T = \begin{pmatrix} | & & | \\ e_{\sigma(1)} & \cdots & e_{\sigma(\ell)} \\ | & & | \end{pmatrix}, \quad (2.4)$$

wobei  $e_j$  den  $j$ -ten Einheitsvektor im  $\mathbb{R}^\ell$  meint.

### Beispiel 2.7 (Permutationen und Permutationsmatrizen)

Die symmetrische Gruppe  $Sym(3)$  zum Grad  $\ell = 3$  ist

$$Sym(3) = \{\text{id}_3, (132), (213), (312), (231), (321)\}.$$

Hier und im Folgenden wird die übliche Einzeilen-Notation für Permutationen verwendet. Damit meint  $\sigma = (132)$  beispielsweise die Permutation  $\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}$ , das heißt  $\sigma(1) = 1$ ,  $\sigma(2) = 3$  und  $\sigma(3) = 2$ . Des Weiteren meint  $\text{id}_\ell = (1 \dots \ell)$  das neutrale Element der Gruppe  $Sym(\ell)$ . Mit der obigen Definition ergeben sich beispielsweise die korrespondierenden Permutationsmatrizen

$$\varrho(\text{id}_3) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \varrho((132)) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \varrho((312)) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

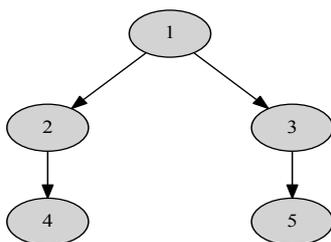
### Permutationen der Knoten und Permutationen der Boxen

Sei  $\mathcal{V}$  ein azyklisch gemischter Graph mit  $p$  Knoten und  $k$  Boxen. Später werden sowohl die endliche Gruppe  $Sym(p) \cong Perm(p)$  als auch die endliche Gruppe  $Sym(k) \cong Perm(k)$  eine wichtige Rolle spielen. Anschaulich lässt sich ein Gruppenelement  $\sigma \in Sym(p)$  und die korrespondierende Permutationsmatrix  $\varrho_p(\sigma) \in Perm(p)$  als eine Vertauschung der Knoten des Graphen  $\mathcal{V}$  interpretieren. Ebenso permutieren  $\tau \in Sym(k)$  und  $\varrho_k(\tau) \in Perm(k)$  die Boxen des Graphen  $\mathcal{V}$ .

Für solche Boxen-Permutationen wird es sich als nützlich erweisen, eine intuitive Notation einzuführen. Seien dazu die Boxen von  $\mathcal{V}$  durchnummeriert gemäß  $\mathcal{B}(\mathcal{V}) = \{[B_1], \dots, [B_k]\}$ . Für eine Permutation der Boxen  $\tau \in Sym(k)$  wird im Folgenden oft die Schreibweise

$$[\tau(B_i)] := [B_{\tau(i)}] \in \mathcal{B}(\mathcal{V}) \quad \text{für alle } i = 1, \dots, k \quad (2.5)$$

verwendet. Wenn also von der Box  $[\tau(B)]$  die Rede ist, so ist die Box  $[B_j]$  mit dem Index  $j = \tau(i)$  gemeint, wenn  $[B]$  selbst den Index  $i$  hat,  $[B] = [B_i]$ .



**Abbildung 2:** Ein azyklisch gemischter Graph  $\mathcal{V}$ , für den die Knotenpermutation  $(13254)$  ein Automorphismus ist. Die Permutationen  $(12354)$  und  $(21345)$  dagegen sind keine Automorphismen über  $\mathcal{V}$ , da sie die kausale Struktur zerstören.

Ein wichtiges Konzept im Zusammenhang mit Permutationen der Boxen sind sogenannte Automorphismen über gerichteten Graphen.

**Definition 2.8 (Automorphismen über gerichteten Graphen)**

Eine Permutation  $\tau \in \text{Sym}(\ell)$  heißt Automorphismus über einem azyklisch gerichteten Graphen  $\mathcal{D} = (V', D', \emptyset)$  mit  $V' = \{1, \dots, \ell\}$ , falls

$$i \rightarrow j \in D' \Leftrightarrow \tau(i) \rightarrow \tau(j) \in D' \quad \text{für alle } i, j = 1, \dots, \ell$$

gilt. Die Menge aller Automorphismen über  $\mathcal{D}$  wird bezeichnet mit  $\text{Aut}(\mathcal{D}) \subseteq \text{Sym}(\ell)$ .

Ein Automorphismus über dem Graphen  $\mathcal{D}$  kann somit anschaulich die Knoten von  $\mathcal{D}$  vertauschen, muss dabei aber dessen kausale Struktur erhalten.

**Beispiel 2.9 (Automorphismen über gerichteten Graphen)**

Sei  $\mathcal{D} = (V', D', \emptyset)$  der azyklisch gerichtete Graph aus Abbildung 2. Dann ist beispielsweise die Permutation  $\tau_1 = (13254)$  ein Automorphismus über  $\mathcal{D}$ . Dagegen ist  $\tau_2 = (12354)$  kein Automorphismus, da zwar  $2 \rightarrow 4$  in  $D'$  ist, zwischen den permutierten Knoten  $\tau_2(2) = 2$  und  $\tau_2(4) = 5$  verläuft jedoch keine Kante. Ebenso ist  $\tau_3 = (21345)$  kein Automorphismus, da der Pfeil  $2 = \tau_3(1) \rightarrow \tau_3(2) = 1$  nicht in  $\mathcal{D}$  auftaucht.

**Lemma 2.10 (Gruppeneigenschaft von  $\text{Aut}(\mathcal{D})$ )**

Sei  $\mathcal{D} = (V', D', \emptyset)$  ein azyklisch gerichteter Graph über  $\ell$  Knoten. Dann ist  $\text{Aut}(\mathcal{D})$  eine Untergruppe von  $\text{Sym}(\ell)$ .

### Beweis

Seien  $\tau_1, \tau_2 \in \text{Aut}(\mathcal{D})$  und seien  $i, j \in \{1, \dots, \ell\}$ . Insbesondere liegen dann  $\tau_2(i)$  und  $\tau_2(j)$  in  $\{1, \dots, \ell\}$ , so dass

$$i \rightarrow j \in D' \Leftrightarrow \tau_2(i) \rightarrow \tau_2(j) \in D' \Leftrightarrow \tau_1(\tau_2(i)) \rightarrow \tau_1(\tau_2(j)) \in D'$$

gilt. Folglich ist  $\tau_1 \circ \tau_2 \in \text{Aut}(\mathcal{D})$ , das heißt die Menge der Automorphismen über  $\mathcal{D}$  ist abgeschlossen. Des Weiteren ist für  $\tau \in \text{Aut}(\mathcal{D})$  auch  $\tau^{-1} \in \text{Aut}(\mathcal{D})$ , da

$$\tau^{-1}(i) \rightarrow \tau^{-1}(j) \in D' \Leftrightarrow i = \tau(\tau^{-1}(i)) \rightarrow \tau(\tau^{-1}(j)) = j \in D'$$

direkt aus der Automorphismeigenschaft von  $\tau$  folgt. Damit ist die Behauptung gezeigt, da  $\text{id}_\ell$  das neutrale Element von  $\text{Aut}(\mathcal{D})$  ist und die Assoziativität von der symmetrischen Gruppe  $\text{Sym}(\ell)$  geerbt wird. □

Betrachte speziell einen azyklisch gemischten Graphen  $\mathcal{V}$  mit  $k$  Boxen und dessen Graph der Boxen  $\text{GoB}(\mathcal{V}) = (V^B, D^B, \emptyset)$ . Die Menge aller Automorphismen über  $\text{GoB}(\mathcal{V})$  besteht somit aus allen Permutationen  $\tau \in \text{Sym}(k)$ , die die Bedingung

$$[A] \rightarrow [B] \in D^B \Leftrightarrow [\tau(A)] \rightarrow [\tau(B)] \in D^B \quad \text{für alle } [A], [B] \in V^B = \mathcal{B}(\mathcal{V})$$

erfüllen, siehe Gleichung (2.5).

### Orbits und Stabilisatoren

Die folgenden Konzepte aus der klassischen Gruppentheorie werden auch in der statistischen Anwendung eine wichtige Rolle spielen.

#### Definition 2.11 (Orbits und Stabilisatoren)

Sei  $H \subseteq \text{Sym}(\ell)$  eine Gruppe von Permutationen und  $i \in \{1, \dots, \ell\}$  fest. Dann ist das Orbit  $\text{Orb}(i)$  von  $i$  unter  $H$  definiert als

$$\text{Orb}(i) := \{\tau(i) \mid \tau \in H\}.$$

Die Menge

$$\text{Sta}(i) := \{\tau \in H \mid \tau(i) = i\}$$

heißt Stabilisatormenge  $\text{Sta}(i)$  von  $i$  unter  $H$ .

Beachte, dass sowohl das Orbit als auch die Stabilisatormenge von  $i$  implizit von der Gruppe  $H$  abhängen, auch wenn dies aus der Notation nicht hervorgeht. Das Orbit von  $i$  besteht aus allen Elementen  $j \in \{1, \dots, \ell\}$ , für die es eine Permutation  $\tau \in H$  gibt, die  $i$  auf  $j$  abbildet. Die Stabilisatormenge von  $i$  wiederum enthält alle Permutationen der Gruppe  $H$ , die  $i$  unberührt lassen.

Später werden vor allem im Zusammenhang mit Permutationen der Boxen auftretende Orbits von Interesse sein. Sei daher  $\mathcal{V}$  ein azyklisch gemischter Graph mit  $k = |\mathcal{B}(\mathcal{V})|$  Boxen, sowie  $H \subseteq \text{Sym}(k)$  eine Gruppe. Das Orbit einer Box  $[B] \in \mathcal{B}(\mathcal{V})$  ergibt sich dann zu

$$\text{Orb}[B] = \{[\tau(B)] \mid \tau \in H\} = \{[\tau^{-1}(B)] \mid \tau \in H\}.$$

Die letzte Gleichheit folgt aus der Gruppeneigenschaft von  $H$ . Das Orbit von  $[B]$  ist also die Menge aller Boxen  $[A]$ , die von einer Permutation  $\tau \in H$  auf  $[B]$  abgebildet werden. Ebenso ist  $\text{Sta}[B] = \{\tau \in H \mid [\tau(B)] = [B]\}$  die Stabilisatormenge der Box  $[B]$ .

Eine interessante Kenngröße in diesem Zusammenhang ist, wie oft eine Box  $[A] \in \mathcal{B}(\mathcal{V})$  unter  $H$  auf eine weitere Box  $[B] \in \mathcal{B}(\mathcal{V})$  abgebildet wird. Setze dazu

$$\kappa_{[B]}([A]) := \left| \{ \tau \in H \mid [\tau(A)] = [B] \} \right| \begin{cases} = 0 & \text{falls } [A] \notin \text{Orb}[B] \\ > 0 & \text{falls } [A] \in \text{Orb}[B] \end{cases}.$$

Das folgende Lemma besagt, dass jede Box aus dem Orbit von  $[B]$  gleich oft auf  $[B]$  abgebildet wird.

**Lemma 2.12 (Verhalten von  $\kappa_{[B]}$  über dem Orbit von  $[B]$ )**

Sei  $\mathcal{V}$  ein azyklisch gemischter Graph mit  $k$  Boxen. Seien außerdem  $[B] \in \mathcal{B}(\mathcal{V})$  eine feste Box und  $H \subseteq \text{Sym}(k)$  eine Gruppe. Dann ist

$$\kappa_{[B]}([A_1]) = \kappa_{[B]}([A_2]) \quad \text{für alle } [A_1], [A_2] \in \text{Orb}[B].$$

**Beweis**

Seien  $[A_1], [A_2] \in \text{Orb}[B]$  beliebig aber fest. Wähle dann  $\tau_1, \tau_2 \in H$  so, dass  $[\tau_i(A_i)] = [B]$  für  $i = 1, 2$ . Betrachte die Abbildung  $h : H \rightarrow H$  mit  $h(\tau) := \tau \circ \tau_1^{-1} \circ \tau_2$ . Sie ist wohldefiniert und bijektiv mit Umkehrabbildung  $\tau \mapsto \tau_2^{-1} \circ \tau_1 \circ \tau$ . Außerdem ist per Definition

$$[h(\tau)(A_2)] = [\tau \circ \tau_1^{-1} \circ \tau_2(A_2)] = [\tau \circ \tau_1^{-1}(B)] = [\tau(A_1)] \quad \text{für alle } \tau \in H.$$

Mit anderen Worten gibt es zu jeder Permutation  $\tau \in H$ , die  $[A_1]$  auf  $[B]$  abbildet, genau eine Permutation  $h(\tau) \in H$ , die  $[A_2]$  auf  $[B]$  abbildet. Dies zeigt die Behauptung.  $\square$

Dieses Lemma ermöglicht die folgende Definition.

**Definition 2.13 (Orbitzahl einer Box)**

Sei  $\mathcal{V}$  ein azyklisch gemischter Graph mit  $k$  Boxen und  $H \subseteq \text{Sym}(k)$  eine Gruppe. Wegen Lemma 2.12 ist die Orbitzahl  $\kappa_{[B]}$  einer Box  $[B] \in \mathcal{B}(\mathcal{V})$  wohldefiniert,

$$\kappa_{[B]} := \kappa_{[B]}([A]) \text{ für ein } [A] \in \text{Orb}[B].$$

Das folgende nützliche Korollar ist eine direkte Konsequenz aus Lemma 2.12.

**Korollar 2.14 (Orbitzahl und Stabilisatormenge)**

Sei  $\mathcal{V}$  ein azyklisch gemischter Graph mit  $k$  Boxen und  $H \subseteq \text{Sym}(k)$  eine Gruppe. Dann gilt für alle Boxen  $[B] \in \mathcal{B}(\mathcal{V})$  der Zusammenhang

$$|H| = |\text{Orb}[B]| \cdot \kappa_{[B]}.$$

Insbesondere folgt daraus, dass die Orbitzahl von  $[B]$  gerade der Kardinalität der Stabilisatormenge von  $[B]$  entspricht,

$$\kappa_{[B]} = |\text{Sta}[B]|.$$

**Beweis**

Die Elemente in  $H$  können gezählt werden, indem alle Boxen  $[A] \in \mathcal{B}(\mathcal{V})$  durchlaufen werden und jeweils addiert wird, wie oft  $[A]$  von einer Permutation  $\tau \in H$  auf die feste Box  $[B]$  abgebildet wird. Damit ist

$$|H| = \sum_{[A] \in \mathcal{B}(\mathcal{V})} \kappa_{[B]}([A]) = \sum_{[A] \in \text{Orb}[B]} \kappa_{[B]}([A]) = |\text{Orb}[B]| \cdot \kappa_{[B]}.$$

Bei der letzten Gleichheit wurde Lemma 2.12 verwendet. Die zweite Behauptung folgt schließlich direkt aus dem Orbit-Stabilisator-Theorem. Dieses besagt, dass für alle Boxen  $[B]$  die Beziehung  $|H| = |\text{Orb}[B]| \cdot |\text{Sta}[B]|$  gilt, siehe Sepanski [23, Theorem 2.65].  $\square$

Definiere nun für eine Gruppe  $H \subseteq \text{Sym}(k)$  und  $[A], [B] \in \mathcal{B}(\mathcal{V})$  die Relation  $\sim$  durch

$$[A] \sim [B] \iff [A] \in \text{Orb}[B].$$

Beachte, dass auch die Relation  $\sim$  implizit von der Gruppe  $H$  abhängt. Es ist eine leichte Übung nachzurechnen, dass es sich hierbei um eine Äquivalenzrelation handelt, man benötigt dazu nur die grundlegenden Eigenschaften einer Gruppe. Per Definition sind die

Äquivalenzklassen der Relation  $\sim$  gerade durch die Orbits von  $H$  gegeben. Die Quotientenmenge von  $\mathcal{B}(\mathcal{V})$  bezüglich  $\sim$  stellt sich also dar gemäß

$$\mathcal{B}(\mathcal{V})/\sim = \{Orb[B] \mid [B] \in \mathcal{B}(\mathcal{V})\}. \quad (2.6)$$

Beachte, dass es sich bei einem Element dieser Quotientenmenge um eine Menge von Boxen handelt. Später wird es von Nutzen sein, diese Sichtweise etwas zu ändern und jedes Orbit mit einer Repräsentantenbox zu identifizieren.

**Definition 2.15 (Repräsentantenmenge)**

Eine Box  $[A] \in Orb[B]$  heißt Repräsentant des Orbits von  $[B] \in \mathcal{B}(\mathcal{V})$ . Wähle nun für jedes Orbit  $Orb[B] \in \mathcal{B}(\mathcal{V})/\sim$  einen beliebigen Repräsentanten aus und nenne ihn  $[B]_{\sim}$ . Eine Repräsentantenmenge  $\mathcal{B}(\mathcal{V})_{\sim}$  von  $\mathcal{V}$  bezüglich  $H$  ist dann definiert als

$$\mathcal{B}(\mathcal{V})_{\sim} := \{[B]_{\sim} \mid Orb[B] \in \mathcal{B}(\mathcal{V})/\sim\}.$$

Die Quotientenmenge  $\mathcal{B}(\mathcal{V})/\sim$  und die Repräsentantenmenge  $\mathcal{B}(\mathcal{V})_{\sim}$  haben somit prinzipiell denselben Informationsgehalt, beide können als Menge der Orbits interpretiert werden. Während in  $\mathcal{B}(\mathcal{V})/\sim$  jedoch die kompletten Orbits als Mengen von Boxen auftauchen, enthält  $\mathcal{B}(\mathcal{V})_{\sim}$  für jedes Orbit nur einen Repräsentanten. Insbesondere ist ein Element von  $\mathcal{B}(\mathcal{V})_{\sim}$  somit eine einzelne Box. Beachte schließlich, dass die Repräsentantenmenge eine Partitionierung der Boxenmenge  $\mathcal{B}(\mathcal{V})$  liefert, denn es ist

$$\bigsqcup_{[B]_{\sim} \in \mathcal{B}(\mathcal{V})_{\sim}} Orb([B]_{\sim}) = \mathcal{B}(\mathcal{V}).$$

## 2.4 Einbindung von Invarianzrestriktionen in graphische Modelle

Nun sollen diese gruppentheoretischen Hilfsmittel dazu verwendet werden, gewisse Invarianz-Restriktionen in die graphischen Modelle aus Kapitel 2.2 einzubinden. Zu diesem Zweck muss zunächst geklärt werden, welche Arten von Invarianzen hierfür zulässig sind. Anschließend können dann die Parameterräume eines invarianten graphischen Modells mit Normalverteilungsannahme definiert werden. Der im folgenden Kapitel erläuterte Ansatz ist dem Aufsatz von Madsen [20] nachempfunden.

### Zulässige und unzulässige Invarianzen

Sei  $\mathcal{V}$  ein Annahme 2.6 erfüllender azyklisch gemischter Graph mit Gerüst  $\mathcal{U} = \mathcal{V}^{\sim}$ . Sei des Weiteren  $G$  eine endliche Untergruppe von  $Orth(p)$ , den orthogonalen  $p \times p$ -Matrizen. Enthält  $G$  nur Permutationsmatrizen,  $G \subseteq Perm(p)$ , dann vertauscht die Kongruenz-

transformation  $W \mapsto \rho W \rho^T$  für alle Gruppenelemente  $\rho \in G$  die Zeilen und Spalten der Matrix  $W \in \mathbb{R}^{p \times p}$  gemäß der zu  $\rho$  korrespondierenden  $p$ -wertigen Permutation.

Die Matrix  $W$  heißt *G-invariant*, wenn für alle Gruppenelemente  $\rho \in G$  die Gleichheit  $\rho W \rho^T = W$  gilt. Damit ist auch klar, wie die von  $G$  vorgegebenen Symmetrien in ein graphisches Modell eingebunden werden können: Alle Parametermatrizen müssen nicht nur die vom Graphen vorgegebenen bedingten Unabhängigkeitsbeziehungen erfüllen, sondern auch  $G$ -invariant sein. Dies kann zu einer erheblichen Reduktion der Dimension des Parameterraums führen. Diese Reduktion muss jedoch auf sinnvolle Weise geschehen, das heißt, die eingebauten Symmetrien müssen die Struktur des Modellgraphen berücksichtigen. Genauer gesagt muss zunächst spezifiziert werden, welches Aussehen die Gruppe  $G$  haben darf, um sowohl die kausalen als auch die wechselseitigen Abhängigkeitsbeziehungen des Graphen  $\mathcal{V}$  zu respektieren.

Dazu soll zunächst motiviert werden, welche Einschränkungen die Gruppe  $G$  erfüllen muss. Um diese Motivation möglichst anschaulich zu gestalten, sei zunächst  $G \subseteq \text{Perm}(p)$ , so dass sich jedes Gruppenelement  $\rho \in G$  als Permutation der Knoten von  $\mathcal{V}$  auffassen lässt. Anschließend wird erläutert, wie sich diese Motivation auf eine endliche Untergruppe der orthogonalen Matrizen  $\text{Orth}(p)$  verallgemeinern lässt.

Eine Art von Permutationen  $\rho \in G$  vertauscht nur Knoten innerhalb einer Box. Für solche Symmetrien muss keine Einschränkung getroffen werden. Da in  $\mathcal{V}$  der von einer Box  $[B] \in \mathcal{B}(\mathcal{V})$  induzierte Teilgraph  $\mathcal{V}_{[B]}$  vollständig ist, bleibt die Abhängigkeitsstruktur der Knoten innerhalb der Box auch nach der Vertauschung erhalten.

Eine zweite Art der Symmetrie vertauscht ganze Boxen miteinander. Diese Permutationen werden nur mit Einschränkungen zugelassen, da darauf geachtet werden muss, dass die kausale Abhängigkeitsstruktur des Graphen  $\mathcal{V}$  mit den Symmetrien zusammenpasst. Das bedeutet, dass eine gerichtete Kante zwischen zwei Knoten  $\alpha \rightarrow \beta$  genau dann auftritt, wenn sie auch zwischen den permutierten Knoten auftaucht.

Die dritte und letzte Symmetrieart lässt zwei Knoten die Box wechseln, das heißt, sie vertauscht Knoten unterschiedlicher Boxen, ohne die ganzen Boxen zu permutieren. Solche Symmetrien werden von vorneherein ausgeschlossen. Man sollte nämlich immer daran denken, dass die Symmetrien in den meisten Anwendungen ein technisches Hilfsmittel darstellen. Die Struktur, die das Modell hauptsächlich wiedergeben soll, stammt dagegen vom Graphen  $\mathcal{V}$  her. Daher macht es keinen Sinn, Permutationen zuzulassen, die diese Struktur zerstören würden.

Auch wenn die Gruppe  $G$  nicht nur Permutationsmatrizen, sondern auch allgemeinere orthogonale Matrizen enthalten darf, ändert sich an dieser Motivation nicht viel. Es sind weiterhin nur Transformationen  $\rho$  erlaubt, die die Graphen- und insbesondere die Boxen-

Struktur respektieren. Dies bedeutet, dass auf der Ebene der Boxen nur Vertauschungen zulässig sind und dass solch eine Vertauschung nur ganze Boxen aufeinander abbilden darf. Es ist weiterhin nicht zulässig, dass ein Knoten seine ursprüngliche Box verlässt. Innerhalb einer Box dagegen sind nicht nur Permutationen der Knoten erlaubt, sondern beliebige orthogonale Transformationen. Wie diese bisher eher vagen Ideen konkretisiert werden können, zeigt der nächste Abschnitt.

### Definition der zulässigen Gruppe

Sei die Boxenmenge von  $\mathcal{V}$  nummeriert gemäß  $\mathcal{B}(\mathcal{V}) = \{[B_1], \dots, [B_k]\}$ . Ein wichtiges Hilfsmittel auf dem Weg zur Konstruktion der Menge aller zulässigen Invarianzen wird die Abbildung

$$\delta : Orth(p) \rightarrow \mathbb{R}^{k \times k} \quad \text{mit} \quad \delta(\rho)_{ij} = \begin{cases} 0 & \text{falls } \rho_{[B_i, B_j]} = 0 \\ 1 & \text{sonst} \end{cases} \quad (2.7)$$

sein, die implizit vom Graphen  $\mathcal{V}$  abhängt. Um sie zu interpretieren, sei  $\rho \in Perm(p)$  eine Permutationsmatrix. Dann ist die Komponente  $\delta(\rho)_{ij}$  genau dann gleich Eins, wenn es einen Knoten in der Box  $[B_j]$  gibt, der von der zu  $\rho$  korrespondierenden Permutation auf einen Knoten in der Box  $[B_i]$  abgebildet wird, siehe Gleichung (2.4). Folglich lässt  $\rho$  genau dann keinen Knoten die Box wechseln, wenn  $\delta(\rho)$  eine  $k \times k$ -Permutationsmatrix ist.

Auch wenn  $\rho \in Orth(p)$  eine allgemeine orthogonale Matrix ist, ändert sich daran nichts. Die Matrix  $\delta(\rho)$  ist genau dann eine Permutationsmatrix, wenn die Invarianzen der Gruppe  $G$  die Boxenstruktur respektieren. Dies führt zu der folgenden Definition.

### Definition 2.16 (Induzierte Boxenpermutation)

Sei  $\mathcal{V}$  ein azyklisch gemischter Graph wie in Annahme 2.6. Außerdem sei  $\rho \in Orth(p)$ . Falls die Matrix  $\delta(\rho) \in \mathbb{R}^{k \times k}$  eine Permutationsmatrix ist, so sei die zugehörige Permutation bezeichnet mit  $\tau_\rho \in Sym(k)$ . Mit anderen Worten ist

$$i = \tau_\rho(j) \Leftrightarrow \delta(\rho)_{ij} = 1 \quad \text{für alle } i, j \in \{1, \dots, k\}.$$

Diese Bedingung ist wegen Gleichung (2.4) äquivalent zu

$$\varrho_k(\tau_\rho) = \delta(\rho) \quad \text{bzw.} \quad \tau_\rho = \varrho_k^{-1}(\delta(\rho)).$$

Anschaulich beschreibt die Matrix  $\delta(\rho)$  folglich, was die Knoten-Transformation  $\rho \in Orth(p)$  auf der Ebene der Boxen bewirkt. Falls  $\rho$  die Boxenstruktur von  $\mathcal{V}$  erhält, das

heißt, falls  $\delta(\rho)$  eine  $k \times k$ -Permutationsmatrix ist, so beschreiben  $\delta(\rho)$  und die zugehörige Permutation  $\tau_\rho$ , auf welche Weise die Boxen von  $\mathcal{V}$  vertauscht werden.

Als nächstes muss darauf geachtet werden, dass alle Gruppenelemente  $\rho \in G$  die kausale Struktur des Graphen  $\mathcal{V} = (V, D, U)$  berücksichtigen. Dazu sei daran erinnert, dass ein Annahme 2.6 erfüllender azyklisch gemischter Graph  $\mathcal{V}$  schon eindeutig durch seinen Graph der Boxen  $GoB(\mathcal{V}) = (V^B, D^B, \emptyset)$  (und die Kenntnis der Größe der Boxen) bestimmt ist. Dessen kausale Struktur bleibt per Definition genau dann unangetastet, wenn die induzierten Permutationen  $\tau_\rho$  für  $\rho \in G$  jeweils Automorphismen über  $GoB(\mathcal{V})$  sind. Dies führt zu der folgenden Definition.

**Definition 2.17 (Gruppe der von  $\mathcal{V}$  zugelassenen Symmetrien)**

Sei  $\mathcal{V}$  ein Annahme 2.6 erfüllender azyklisch gemischter Graph. Dann ist die Gruppe der von  $\mathcal{V}$  zugelassenen Symmetrien  $Orth(\mathcal{V})$  definiert als

$$Orth(\mathcal{V}) := \left\{ \rho \in Orth(p) \mid \delta(\rho) \in Perm(k) \text{ und } \tau_\rho \in Aut[GoB(\mathcal{V})] \right\}.$$

Später wird in Lemma 2.20 gezeigt, dass  $Orth(\mathcal{V})$  eine Gruppe ist und somit ihren Namen verdient hat. Es sei hervorgehoben, dass eine Symmetrie  $\rho \in Orth(\mathcal{V})$  nach Konstruktion keine zwei Boxen vertauschen kann, die eine unterschiedliche Anzahl von Knoten beinhalten. Andernfalls wäre aufgrund der Bijektivität von  $\rho$  die induzierte Matrix  $\delta(\rho)$  keine Permutationsmatrix. Beachte des Weiteren, dass orthogonale Transformationen innerhalb einer Box  $[B]$  nur deshalb ohne Einschränkungen zugelassen werden, weil der induzierte Teilgraph der Box  $\mathcal{V}_{[B]}$  aufgrund von Annahme 2.6 vollständig ist. Wäre dies nicht der Fall, so müsste man auch hier auf die Struktur Acht geben und analog zu den Automorphismen über gerichteten Graphen noch Automorphismen über ungerichteten Graphen definieren.

Bevor einige nützliche Lemmas zu den zugelassenen Symmetrien folgen, sollen diese erstmal anhand eines Beispiels veranschaulicht werden.

**Beispiel 2.18 (Zulässige Symmetrien)**

Sei  $\mathcal{V}$  der azyklisch gemischte Graph aus Abbildung 1a von Seite 11. Er erfüllt offensichtlich Annahme 2.6. Bei der Bestimmung der von  $\mathcal{V}$  zugelassenen Invarianzen fällt zunächst auf, dass eine Transformation  $\rho \in Orth(\mathcal{V})$  den Knoten 1 nie mit einem anderen Knoten vertauschen darf, da sonst die kausale Struktur von  $\mathcal{V}$  verletzt würde. Des Weiteren ist beispielsweise die zur Permutation  $\sigma = (12435)$  korrespondierende Matrix  $\rho = \varrho_5(\sigma)$  nicht

zulässig, da die induzierte Matrix

$$\delta(\rho) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

keine Permutationsmatrix ist. Dies liegt darin begründet, dass  $\sigma$  die Knoten 3 und 4 die Box wechseln lässt. Mit Überlegungen dieser Art ergibt sich schließlich, dass die zulässige Gruppe  $Orth(\mathcal{V})$  unter anderem die zu den Permutationen

$$\{\text{id}_5, (13245), (12354), (13254), (14523), (15423), (14532), (15432)\}$$

korrespondierenden Permutationsmatrizen enthält. Weitere Beispiele für zulässige orthogonale Matrizen  $\rho \in Orth(\mathcal{V})$  erhält man durch orthogonale Transformationen innerhalb der Boxen, beispielsweise sind

$$\rho_1 = \begin{pmatrix} 1 & & & & \\ & \sqrt{3}/2 & -1/2 & & \\ & 1/2 & \sqrt{3}/2 & & \\ & & & 1 & 0 \\ & & & 0 & 1 \end{pmatrix}, \quad \rho_2 = \begin{pmatrix} 1 & & & & \\ & 1/\sqrt{2} & 1/\sqrt{2} & & \\ & 1/\sqrt{2} & -1/\sqrt{2} & & \\ & & & 1 & 0 \\ & & & 0 & 1 \end{pmatrix}, \quad \dots$$

zulässig. Als induzierte Boxenpermutationen ergeben sich hier

$$\tau_{\rho_1} = ([B_1][B_2][B_4]) \quad \text{und} \quad \tau_{\rho_2} = ([B_1][B_4][B_2]).$$

Das folgende Lemma liefert eine nähere Beschreibung einer zulässigen orthogonalen Matrix  $\rho \in Orth(\mathcal{V})$ . Außerdem werden einige Identitäten angegeben, die mit diesen Matrizen zusammenhängen und später von Nutzen sein werden.

**Lemma 2.19 (Zulässige orthogonale Matrizen)**

Sei  $\mathcal{V}$  ein Annahme 2.6 erfüllender azyklisch gemischter Graph mit Gerüst  $\mathcal{U}$ . Außerdem sei  $\rho \in Orth(\mathcal{V})$  eine zulässige Symmetrie. Dann gilt:

- (i) Die Matrix  $\rho$  ist durch ihre Blöcke  $\{\rho_{[\tau_\rho(B), B]}\}_{[B] \in \mathcal{B}(\mathcal{V})}$  schon eindeutig bestimmt, denn es gilt für alle Boxen  $[A], [B] \in \mathcal{B}(\mathcal{V})$ :

$$\rho_{[A, B]} \begin{cases} \in Orth([B]) & \text{falls } [A] = [\tau_\rho(B)] \text{ bzw. } [B] = [\tau_\rho^{-1}(A)] \\ = 0 & \text{sonst} \end{cases} .$$

(ii) Für die transponierte Matrix von  $\rho$  gilt für alle Boxen  $[A], [B] \in \mathcal{B}(\mathcal{V})$ :

$$(\rho^T)_{[B,A]} = \begin{cases} (\rho_{[A,B]})^T & \text{falls } [A] = [\tau_\rho(B)] \text{ bzw. } [B] = [\tau_\rho^{-1}(A)] \\ 0 & \text{sonst} \end{cases}.$$

(iii) Für eine Matrix  $S \in \mathbb{R}^{p \times p}$  und  $[A], [B] \in \mathcal{B}(\mathcal{V})$  ist außerdem

$$(\rho S \rho^T)_{[A,B]} = \rho_{[A, \tau_\rho^{-1}(A)]} S_{[\tau_\rho^{-1}(A), \tau_\rho(B)]} \left( \rho_{[B, \tau_\rho^{-1}(B)]} \right)^T.$$

(iv) Für eine Matrix  $W \in \mathbb{R}^{p \times p}$  gilt

$$W \in P(\mathcal{U}) \Leftrightarrow \rho W \rho^T \in P(\mathcal{U}).$$

### Beweis

Die Teilmatrix  $\rho_{[B_i, B_j]}$  ist genau dann gleich Null, wenn  $\delta(\rho)_{ij}$  verschwindet, siehe Gleichung (2.7). Dies ist wiederum genau dann der Fall, wenn  $i \neq \tau_\rho(j)$  ist, siehe Definition 2.16. Außerdem müssen alle Teilmatrizen der Form  $\rho_{[\tau_\rho(B), B]}$  selbst wieder orthogonale Matrizen sein, andernfalls wäre das Produkt  $\rho \rho^T$  nicht die Identitätsmatrix. Damit folgen (i) und (ii). Teil (iii) wiederum folgt aus

$$\begin{aligned} (\rho S \rho^T)_{[A,B]} &= \rho_{[A, V]} \cdot S \cdot (\rho^T)_{V, [B]} = \\ &= \left( 0, \dots, \rho_{[A, \tau_\rho^{-1}(A)]}, \dots, 0 \right) \cdot S \cdot \left( 0, \dots, \left( \rho_{[B, \tau_\rho^{-1}(B)]} \right)^T, \dots, 0 \right)^T \end{aligned}$$

unter Verwendung von (i) und (ii). Für die letzte Behauptung sei  $W \in P(\mathcal{U})$ , so dass zu zeigen ist, dass  $W$  und  $\rho W \rho^T$  dieselben strukturellen Nullblöcke aufweisen. Wegen Teil (iii) verschwindet die Teilmatrix  $(\rho W \rho^T)_{[A,B]}$  genau dann, wenn die Teilmatrix  $W_{[\tau_\rho^{-1}(A), \tau_\rho^{-1}(B)]}$  verschwindet. Dies ist genau dann der Fall, wenn  $[A] \neq [B]$  ist und die Boxen  $[\tau_\rho^{-1}(A)]$  und  $[\tau_\rho^{-1}(B)]$  im Graph der Boxen  $GoB(\mathcal{V})$  nicht durch eine Kante verbunden sind. Da  $\tau_\rho^{-1}$  aber ein Automorphismus über  $GoB(\mathcal{V})$  ist, ist letztere Bedingung äquivalent dazu, dass  $[A]$  und  $[B]$  in  $GoB(\mathcal{V})$  nicht durch eine Kante verbunden sind, siehe Definition 2.8. Folglich hat  $\rho W \rho^T$  dieselben vorgegebenen Nullblöcke wie  $W$  selbst und damit folgt Aussage (iv). Einen alternativen Beweis für (iv) findet man beispielsweise auch in Madsen [20, Proposition 6.1].  $\square$

Das nächste Lemma liefert eine nachträgliche Begründung für den Namen von  $Orth(\mathcal{V})$ , vergleiche Definition 2.17.

**Lemma 2.20 (Gruppeneigenschaft von  $Orth(\mathcal{V})$ )**

Sei  $\mathcal{V}$  ein Annahme 2.6 erfüllender azyklisch gemischter Graph.

(i) Für alle  $\rho_1, \rho_2 \in Orth(\mathcal{V})$  gilt:  $\delta(\rho_1\rho_2) = \delta(\rho_1)\delta(\rho_2)$ .

(ii) Für alle  $\rho_1, \rho_2 \in Orth(\mathcal{V})$  gilt:  $\tau_{\rho_1\rho_2} = \tau_{\rho_1} \circ \tau_{\rho_2}$ .

(iii)  $Orth(\mathcal{V})$  ist eine Untergruppe von  $Orth(p)$ .

**Beweis**

Seien  $\rho_1, \rho_2 \in Orth(\mathcal{V})$  und  $i, j \in \{1, \dots, k\}$ . Dann ist

$$[\delta(\rho_1)\delta(\rho_2)]_{ij} = \delta(\rho_1)_{i, \{1, \dots, k\}} \delta(\rho_2)_{\{1, \dots, k\}, j}.$$

Da  $\delta(\rho_1)$  und  $\delta(\rho_2)$  Permutationsmatrizen sind, das heißt in jeder Zeile und jeder Spalte taucht genau eine Eins auf, ist dieser Ausdruck genau dann gleich Eins, wenn

$$\delta(\rho_1)_{i\ell} = 1 = \delta(\rho_2)_{\ell j}$$

für ein  $\ell \in \{1, \dots, k\}$ . Wegen Definition 2.16 muss in diesem Fall  $\ell = \tau_{\rho_1}^{-1}(i) = \tau_{\rho_2}(j)$  sein.

Auf der anderen Seite ist  $\delta(\rho_1\rho_2)_{ij}$  genau dann gleich Eins, wenn  $(\rho_1\rho_2)_{[B_i, B_j]}$  nicht verschwindet, siehe Gleichung (2.7). Diese Matrix wiederum ist identisch mit

$$(\rho_1)_{[B_i], V} (\rho_2)_{V, [B_j]} = \begin{cases} (\rho_1)_{[B_i, \tau_{\rho_1}^{-1}(B_i)]} (\rho_2)_{[\tau_{\rho_2}(B_j), B_j]} & \text{falls } [\tau_{\rho_1}^{-1}(B_i)] = [\tau_{\rho_2}(B_j)] \\ 0 & \text{sonst} \end{cases},$$

siehe Lemma 2.19(i). Folglich ist  $\delta(\rho_1\rho_2)_{ij}$  genau dann gleich Eins, wenn  $\tau_{\rho_1}^{-1}(i) = \tau_{\rho_2}(j)$  ist. Damit sind  $[\delta(\rho_1)\delta(\rho_2)]_{ij}$  und  $\delta(\rho_1\rho_2)_{ij}$  für dieselben Indexpaare  $(i, j)$  gleich Eins und Behauptung (i) folgt.

Wegen Definition 2.16 ist  $\tau_{\rho_1\rho_2} = \varrho_k^{-1}(\delta(\rho_1\rho_2))$ . Zusammen mit Teil (i) folgt also

$$\tau_{\rho_1\rho_2} = \varrho_k^{-1}(\delta(\rho_1)\delta(\rho_2)) = \varrho_k^{-1}(\delta(\rho_1)) \circ \varrho_k^{-1}(\delta(\rho_2)) = \tau_{\rho_1} \circ \tau_{\rho_2}.$$

Dabei wird insbesondere verwendet, dass  $\varrho_k$  ein Isomorphismus ist, siehe Gleichung (2.4) auf Seite 18.

Insbesondere folgt aus Teil (i) und der Gruppeneigenschaft von  $Perm(k)$ , dass  $\delta(\rho_1\rho_2)$  ebenfalls in  $Perm(k)$  liegt. Ebenso impliziert Teil (ii) und die Gruppeneigenschaft von  $Aut[GoB(\mathcal{V})]$  (siehe Lemma 2.10), dass  $\tau_{\rho_1\rho_2}$  in  $Aut[GoB(\mathcal{V})]$  liegt. Folglich ist  $\rho_1\rho_2$  in  $Orth(\mathcal{V})$ , siehe Definition 2.17. Damit ist  $Orth(\mathcal{V})$  abgeschlossen bezüglich Multiplikation. Da die Assoziativität und das neutrale Element  $I_p$  von  $Orth(p)$  geerbt werden, bleibt

nur zu zeigen, dass für  $\rho \in Orth(\mathcal{V})$  auch  $\rho^{-1}$  in  $Orth(\mathcal{V})$  liegt. Aus Lemma 2.19(ii) folgt

$$\delta(\rho^{-1}) = \delta(\rho^T) = \delta(\rho)^T \in Perm(k).$$

Diese Gleichung impliziert zudem  $\tau_{\rho^{-1}} = (\tau_\rho)^{-1}$ , siehe Definition 2.16. Da  $Aut[GoB(\mathcal{V})]$  eine Gruppe ist, ist damit auch  $\tau_{\rho^{-1}}$  in  $Aut[GoB(\mathcal{V})]$  und die Behauptung folgt.  $\square$

Später wird oft die Teilmatrix von  $\rho \in Orth(\mathcal{V})$ , die zu den Eltern  $\langle B \rangle$  einer Box  $[B]$  korrespondiert, von Interesse sein. Da es sich bei  $\langle B \rangle$  unter Umständen nicht um eine einzelne Box, sondern um eine disjunkte Vereinigung mehrerer Boxen handeln kann, ist hier etwas Vorsicht geboten.

**Lemma 2.21 (Symmetrien in den Elternboxen)**

Sei  $\mathcal{V}$  ein Annahme 2.6 erfüllender azyklisch gemischter Graph. Sei  $[B] \in \mathcal{B}(\mathcal{V})$  eine Box, deren Eltern gegeben sind durch  $\langle B \rangle = [A_1] \dot{\cup} \dots \dot{\cup} [A_m]$  mit  $[A_i] \in \mathcal{B}(\mathcal{V})$  für alle  $i = 1, \dots, m$ . Sei außerdem  $\rho \in Orth(\mathcal{V})$  mit korrespondierender Boxenpermutation  $\tau_\rho \in Sym(k)$ . Dann gilt:

(i) Die Eltern der Box  $[\tau_\rho(B)]$  sind gegeben durch

$$[\tau_\rho(A_1)] \dot{\cup} \dots \dot{\cup} [\tau_\rho(A_m)] =: \langle \tau_\rho(B) \rangle.$$

(ii) Die Teilmatrix  $\rho_{\langle \tau_\rho(B), B \rangle}$  hat die Blockdiagonalgestalt

$$\rho_{\langle \tau_\rho(B), B \rangle} = \begin{pmatrix} \rho_{[\tau_\rho(A_1), A_1]} & & & \\ & \ddots & & \\ & & \rho_{[\tau_\rho(A_m), A_m]} & \\ & & & \end{pmatrix}.$$

**Beweis**

Da  $\rho \in Orth(\mathcal{V})$  ist, ist  $\tau_\rho$  per Definition 2.17 ein Automorphismus über dem Graph der Boxen  $GoB(\mathcal{V})$ . Insbesondere taucht also für alle  $i = 1, \dots, m$  die Kante  $[A_i] \rightarrow [B]$  im Graph der Boxen genau dann auf, wenn auch die Kante  $[\tau_\rho(A_i)] \rightarrow [\tau_\rho(B)]$  auftaucht. Daraus folgt Behauptung (i).

Teil (ii) ist trivial, wenn  $[B]$  nur eine Elternbox hat, das heißt  $m = 1$ . Sei also  $m > 1$ . Da  $\mathcal{V}$  Annahme 2.6 erfüllt, muss der zu den Elternboxen von  $[B]$  korrespondierende Teil des Graph der Boxen  $GoB(\mathcal{V})_{\langle B \rangle}$  ein vollständiger azyklisch gerichteter Graph sein. Wäre beispielsweise die Kante  $[A_i] \rightarrow [A_j]$  mit  $i < j$  nicht vorhanden, so entstünde die Immoralität  $[A_i] \rightarrow [B] \leftarrow [A_j]$  in  $GoB(\mathcal{V})$ . Folglich würden auch in  $\mathcal{V}$  Immoralitäten

auftauchen, was im Widerspruch zu Annahme 2.6 steht. Damit dürfen innerhalb von  $\langle B \rangle$  keine Vertauschungen stattfinden, da sonst die kausale Struktur von  $GoB(\mathcal{V})$  zerstört würde. Mit anderen Worten ist  $\rho_{[A_i, A_j]} = 0$  für alle  $i \neq j$ . Falls  $\langle B \rangle = \langle \tau_\rho(B) \rangle$  ist, das heißt  $\rho$  lässt die Elternboxen von  $[B]$  unberührt, zeigt dies die Behauptung, denn es ist

$$\rho_{\langle \tau_\rho(B), B \rangle} = \rho_{\langle B, B \rangle} = \begin{pmatrix} \rho_{[A_1, A_1]} & \cdots & \rho_{[A_1, A_m]} \\ \vdots & \ddots & \vdots \\ \rho_{[A_m, A_1]} & \cdots & \rho_{[A_m, A_m]} \end{pmatrix} = \begin{pmatrix} \rho_{[\tau_\rho(A_1), A_1]} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \rho_{[\tau_\rho(A_m), A_m]} \end{pmatrix}.$$

Ist dagegen  $\langle B \rangle \neq \langle \tau_\rho(B) \rangle$ , so müssen  $\langle B \rangle$  und  $\langle \tau_\rho(B) \rangle$  disjunkt sein. Falls eine Box  $[A_i] \subseteq \langle B \rangle$  auch in  $\langle \tau_\rho(B) \rangle$  auftauchen würde, so müsste nämlich aufgrund der Automorphismeigenschaft von  $\tau_\rho$  schon  $\langle B \rangle = \langle \tau_\rho(B) \rangle$  sein, da sonst die kausale Struktur des vollständigen Graphen  $GoB(\mathcal{V})_{\langle B \rangle}$  zerstört würde. Dadurch lässt sich  $\rho_{\langle \tau_\rho(B), B \rangle}$  darstellen als

$$\rho_{\langle \tau_\rho(B), B \rangle} = \begin{pmatrix} \rho_{[\tau_\rho(A_1), A_1]} & \cdots & \rho_{[\tau_\rho(A_1), A_m]} \\ \vdots & \ddots & \vdots \\ \rho_{[\tau_\rho(A_m), A_1]} & \cdots & \rho_{[\tau_\rho(A_m), A_m]} \end{pmatrix}.$$

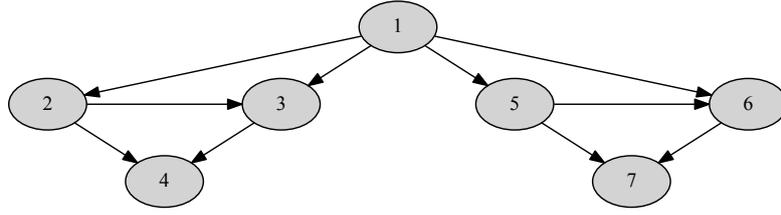
Die Behauptung folgt dann direkt aus Lemma 2.19(i).  $\square$

### Beispiel 2.22 (Symmetrien in den Elternboxen)

Betrachte den Graphen  $\mathcal{V}$  in Abbildung 3. Seine Boxen sind alle einelementig, das heißt  $[B_i] = \{i\}$  für alle  $i = 1, \dots, 7$ . Damit ist  $\langle B_4 \rangle = [B_2] \dot{\cup} [B_3]$  und  $\langle B_7 \rangle = [B_5] \dot{\cup} [B_6]$ . Beachte insbesondere, dass wie im Beweis von Lemma 2.21 beschrieben, die von den Eltern induzierten Teilgraphen  $GoB(\mathcal{V})_{\langle B_4 \rangle}$  und  $GoB(\mathcal{V})_{\langle B_7 \rangle}$  vollständig sind. In diesem Beispiel bedeutet dies, dass weder die Kante  $2 \rightarrow 3$  noch die Kante  $5 \rightarrow 6$  fehlen darf, da sonst die Immoralitäten  $2 \rightarrow 4 \leftarrow 3$  und  $5 \rightarrow 7 \leftarrow 6$  entstünden.

Da im Graphen  $\mathcal{V}$  sehr viele kausale Abhängigkeiten kodiert sind, ist neben der Identität  $\text{id}_7$  nur eine Permutation der Knoten zulässig, nämlich  $\sigma = (1567234) \in \text{Sym}(7)$ . Alle anderen Permutationen der Knoten sind keine Automorphismen. Die zu  $\sigma$  korrespondierende Permutationsmatrix ist

$$\rho = \varrho_7(\sigma) = \begin{pmatrix} 1 & & & & & & \\ & 1 & & & & & \\ & & 1 & & & & \\ & & & 1 & & & \\ & 1 & & & & & \\ & & 1 & & & & \\ & & & & 1 & & \end{pmatrix} \in \text{Perm}(7).$$



**Abbildung 3:** Ein azyklisch gemischter Graph  $\mathcal{V}$ , dessen einziger nicht-trivialer Automorphismus die Knotenpermutation (1567234) ist. Man erkennt, dass aufgrund der Vertauschung der Knoten 4 und 7 auch deren Eltern 2,3 und 5,6 vertauscht werden müssen.

Des Weiteren sind die Eltern der Box  $[\tau_\rho(B_4)] = [B_7]$  gegeben durch

$$\langle \tau_\rho(B_4) \rangle = \langle B_7 \rangle = [B_5] \cup [B_6] = [\tau_\rho(B_2)] \cup [\tau_\rho(B_3)],$$

siehe Lemma 2.21(i). Damit ist auch

$$\begin{aligned} \rho_{\langle \tau_\rho(B_4), B_4 \rangle} &= \rho_{[B_5] \cup [B_6], [B_2] \cup [B_3]} = \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \rho_{[B_5], [B_2]} & 0 \\ 0 & \rho_{[B_6], [B_3]} \end{pmatrix} = \begin{pmatrix} \rho_{[\tau_\rho(B_2)], [B_2]} & 0 \\ 0 & \rho_{[\tau_\rho(B_3)], [B_3]} \end{pmatrix}, \end{aligned}$$

siehe Lemma 2.21(ii).

Das folgende Korollar ist eine direkte Konsequenz aus den Lemmata 2.19 und 2.21 und wird später bei der Bestimmung der Verteilung des Maximum-Likelihood-Schätzers in Kapitel 5 sehr nützlich sein.

**Korollar 2.23 (Wirkung zulässiger orthogonaler Matrizen auf eine Box)**

Sei  $\mathcal{V}$  ein Annahme 2.6 erfüllender azyklisch gemischter Graph und  $\rho \in \text{Orth}(\mathcal{V})$  eine zulässige orthogonale Matrix. Sei außerdem  $S \in \mathbb{R}^{p \times p}$  und  $[B] \in \mathcal{B}(\mathcal{V})$  eine Box. Dann gilt:

$$\begin{aligned} (\rho S \rho^T)_{[B]} &= \rho_{[B, \tau_\rho^{-1}(B)]} S_{[\tau_\rho^{-1}(B)]} \left( \rho_{[B, \tau_\rho^{-1}(B)]} \right)^T, \\ (\rho S \rho^T)_{\langle B \rangle} &= \rho_{\langle B, \tau_\rho^{-1}(B) \rangle} S_{\langle \tau_\rho^{-1}(B) \rangle} \left( \rho_{\langle B, \tau_\rho^{-1}(B) \rangle} \right)^T, \\ (\rho S \rho^T)_{\langle B \rangle} &= \rho_{\langle B, \tau_\rho^{-1}(B) \rangle} S_{\langle \tau_\rho^{-1}(B) \rangle} \left( \rho_{\langle B, \tau_\rho^{-1}(B) \rangle} \right)^T. \end{aligned}$$

**Beweis**

Die erste Gleichheit ist ein Spezialfall von Lemma 2.19(iii). Für die verbleibenden zwei Gleichungen sei  $\langle B \rangle = [A_1] \dot{\cup} \dots \dot{\cup} [A_m]$  mit  $[A_i] \in \mathcal{B}(\mathcal{V})$  für alle  $i = 1, \dots, m$ . Dann ist

$$\begin{aligned} (\rho S \rho^T)_{[B]} &= [(\rho S \rho^T)_{[B, A_1]}, \dots, (\rho S \rho^T)_{[B, A_m]}] = \\ &= \left[ \dots, \rho_{[B, \tau_\rho^{-1}(B)]} S_{[\tau_\rho^{-1}(B), \tau_\rho^{-1}(A_i)]} \left( \rho_{[A_i, \tau_\rho^{-1}(A_i)]} \right)^T, \dots \right], \quad i = 1, \dots, m. \end{aligned}$$

Dies vereinfacht sich weiter zu

$$\begin{aligned} (\rho S \rho^T)_{[B]} &= \rho_{[B, \tau_\rho^{-1}(B)]} \left[ \dots, S_{[\tau_\rho^{-1}(B), \tau_\rho^{-1}(A_i)]}, \dots \right] \begin{pmatrix} \rho_{[A_1, \tau_\rho^{-1}(A_1)]} & & \\ & \ddots & \\ & & \rho_{[A_m, \tau_\rho^{-1}(A_m)]} \end{pmatrix}^T = \\ &= \rho_{[B, \tau_\rho^{-1}(B)]} S_{[\tau_\rho^{-1}(B)]} \left( \rho_{\langle B, \tau_\rho^{-1}(B) \rangle} \right)^T, \end{aligned}$$

wobei mehrfach Lemma 2.19(iii) sowie Lemma 2.21 verwendet wurde. Die dritte Gleichheit aus der Behauptung folgt vollkommen analog.  $\square$

**Invariante graphische Normalverteilungsmodelle**

Sei jetzt ein Annahme 2.6 erfüllender azyklisch gemischter Graph  $\mathcal{V} = (V, D, U)$  mit Gerüst  $\mathcal{U} := \mathcal{V}^\sim$  gegeben. Des Weiteren sei  $G \subseteq \text{Orth}(\mathcal{V})$  eine endliche Gruppe von zulässigen Symmetrien. Dann werden die von  $G$  vorgegebenen Invarianzen in das von  $\mathcal{U}$  induzierte graphische Modell eingebunden, indem die Parameterräume aus Gleichung (2.1) von Seite 12 gemäß

$$\begin{aligned} PD(\mathcal{U}, G) &:= \{ \Sigma \in PD(\mathcal{U}) \mid \rho \Sigma \rho^T = \Sigma \text{ für alle } \rho \in G \}, \\ PD^0(\mathcal{U}, G) &:= \{ K \in PD^0(\mathcal{U}) \mid \rho K \rho^T = K \text{ für alle } \rho \in G \}, \\ P(\mathcal{U}, G) &:= \{ \Omega \in P(\mathcal{U}) \mid \rho \Omega \rho^T = \Omega \text{ für alle } \rho \in G \}, \end{aligned} \tag{2.8}$$

modifiziert werden. Um diese Definition zu verstehen, sei  $X \sim \mathcal{N}_p(0, \Sigma)$  ein normalverteilter Zufallsvektor mit  $\Sigma \in PD(\mathcal{U}, G)$ . Dann gilt für jedes Gruppenelement  $\rho \in G$ , dass die transformierte Zufallsgröße  $\rho X$  derselben Verteilung wie  $X$  folgt, da  $\rho \Sigma \rho^T$  und  $\Sigma$  identisch sind. Das folgende Lemma zeigt, dass sich der Diffeomorphismus  $\pi : PD(\mathcal{U}) \rightarrow P(\mathcal{U})$  aus Lemma 2.4 von Seite 13 auf die neuen Parameterräume überträgt.

**Lemma 2.24 (Diffeomorphismus zwischen  $PD(\mathcal{U}, G)$  und  $P(\mathcal{U}, G)$ )**

Sei  $\mathcal{V}$  ein Annahme 2.6 erfüllender azyklisch gemischter Graph und  $\mathcal{U} = \mathcal{V}^\sim$  sein Gerüst. Sei außerdem  $f_\rho(S) := \rho S \rho^T$  für alle  $S \in \mathbb{R}^{p \times p}$  und alle  $\rho \in \text{Orth}(\mathcal{V})$ . Dann gilt:

(i) Die Funktionen  $\pi : \mathbb{R}^{p \times p} \rightarrow S(\mathcal{U})$  aus Gleichung (2.2) von Seite 13 und  $f_\rho : \mathbb{R}^{p \times p} \rightarrow \mathbb{R}^{p \times p}$  kommutieren für alle  $\rho \in \text{Orth}(\mathcal{V})$ , das heißt

$$\pi(\rho S \rho^T) = \rho \pi(S) \rho^T \quad \text{für alle } S \in \mathbb{R}^{p \times p}.$$

(ii) Sei  $G \subseteq \text{Orth}(\mathcal{V})$  eine zulässige Gruppe. Dann ist die Einschränkung von  $\pi$  aus Lemma 2.4 von Seite 13 auf  $\pi : PD(\mathcal{U}, G) \rightarrow P(\mathcal{U}, G)$  ein Diffeomorphismus.

**Beweis**

(i) Für eine beliebige zulässige Symmetrie  $\rho \in \text{Orth}(\mathcal{V})$  sind die einzigen nicht-verschwindenden Blöcke der Matrix  $\pi(\rho S \rho^T)$  die Teilmatrizen der Form  $(\rho S \rho^T)_{[B]}$  und  $(\rho S \rho^T)_{[B]}$  für  $[B] \in \mathcal{B}(\mathcal{V})$ , siehe Gleichung (2.2) auf Seite 13. Damit sind zwei Dinge zu zeigen: Zum einen müssen diese Teilmatrizen und die entsprechenden Blöcke in  $\rho \pi(S) \rho^T$  identisch sein. Zum anderen muss  $\rho \pi(S) \rho^T$  dieselbe Nullstruktur wie  $\pi(\rho S \rho^T)$  aufweisen, das heißt in  $S(\mathcal{U})$  liegen.

Sei dazu  $[B] \in \mathcal{B}(\mathcal{V})$  eine Box und  $\rho \in \text{Orth}(\mathcal{V})$ . Damit ist  $\tau_\rho^{-1}$  ein Automorphismus über  $GoB(\mathcal{V})$  und da  $\mathcal{V}$  Annahme 2.6 erfüllt ist  $[\tau_\rho^{-1}(B)] \cup \langle \tau_\rho^{-1}(B) \rangle$  vollständig in  $\mathcal{U}$ . Insbesondere sind also  $S_{[\tau_\rho^{-1}(B)]} = \pi(S)_{[\tau_\rho^{-1}(B)]}$  und  $S_{\langle \tau_\rho^{-1}(B) \rangle} = \pi(S)_{\langle \tau_\rho^{-1}(B) \rangle}$ . Wegen Korollar 2.23 ist dann

$$\begin{aligned} (\rho S \rho^T)_{[B]} &= \rho_{[B, \tau_\rho^{-1}(B)]} S_{[\tau_\rho^{-1}(B)]} \left( \rho_{[B, \tau_\rho^{-1}(B)]} \right)^T = \\ &= \rho_{[B, \tau_\rho^{-1}(B)]} \pi(S)_{[\tau_\rho^{-1}(B)]} \left( \rho_{[B, \tau_\rho^{-1}(B)]} \right)^T = (\rho \pi(S) \rho^T)_{[B]} \quad \text{und} \\ (\rho S \rho^T)_{\langle B \rangle} &= \rho_{[B, \tau_\rho^{-1}(B)]} \pi(S)_{\langle \tau_\rho^{-1}(B) \rangle} \left( \rho_{\langle B, \tau_\rho^{-1}(B) \rangle} \right)^T = (\rho \pi(S) \rho^T)_{\langle B \rangle}. \end{aligned}$$

Sei außerdem  $[A] \neq [B]$  eine weitere Box, die in  $GoB(\mathcal{V})$  nicht durch eine gerichtete Kante mit  $[B]$  verbunden ist. Wegen Lemma 2.19(iii) ist dann

$$(\rho \pi(S) \rho^T)_{[B, A]} = \rho_{[B, \tau_\rho^{-1}(B)]} \pi(S)_{[\tau_\rho^{-1}(B), \tau_\rho^{-1}(A)]} \rho_{[A, \tau_\rho^{-1}(A)]} = 0,$$

da  $\tau_\rho^{-1}$  ein Automorphismus ist und somit auch  $[\tau_\rho^{-1}(A)]$  und  $[\tau_\rho^{-1}(B)]$  in  $GoB(\mathcal{V})$  nicht verbunden sein dürfen. Folglich verschwindet die entsprechende Teilmatrix von  $\pi(S)$  und Behauptung (i) folgt.

Beachte, dass sich daraus insbesondere auch die Kommutativität von  $f_\rho$  und der (eingeschränkten) Umkehrabbildung  $\pi^{-1} : P(\mathcal{U}) \rightarrow PD(\mathcal{U})$  ergibt. Für  $\Omega \in P(\mathcal{U})$  mit korres-

pondierendem  $\Sigma = \pi^{-1}(\Omega) \in PD(\mathcal{U})$ , siehe Lemma 2.4 auf Seite 13, ist nämlich

$$\pi^{-1}(\rho\Omega\rho^T) = \pi^{-1}(\rho\pi(\Sigma)\rho^T) = \pi^{-1}(\pi(\rho\Sigma\rho^T)) = \rho\Sigma\rho^T = \rho\pi^{-1}(\Omega)\rho^T. \quad (2.9)$$

(ii) Da  $\pi$  als Abbildung  $PD(\mathcal{U}) \rightarrow P(\mathcal{U})$  ein Diffeomorphismus ist, siehe Lemma 2.4, muss nur noch gezeigt werden, dass  $\pi$  die Teilmenge  $PD(\mathcal{U}, G)$  gerade auf  $P(\mathcal{U}, G)$  abbildet.

Sei dazu  $\Sigma \in PD(\mathcal{U}, G)$  fest. Für alle  $\rho \in G$  folgt wegen Teil (i) dann

$$\rho\pi(\Sigma)\rho^T = \pi(\rho\Sigma\rho^T) = \pi(\Sigma),$$

das heißt  $\pi(\Sigma)$  liegt in  $P(\mathcal{U}, G)$ . Andersherum ist für ein festes  $\Omega \in P(\mathcal{U}, G)$  die Matrix  $\Sigma = \pi^{-1}(\Omega) \in PD(\mathcal{U})$  gegeben. Da wegen Gleichung (2.9) auch  $\pi^{-1} : P(\mathcal{U}) \rightarrow PD(\mathcal{U})$  mit  $f_\rho$  vertauscht, folgt

$$\rho\Sigma\rho^T = \rho\pi^{-1}(\Omega)\rho^T = \pi^{-1}(\rho\Omega\rho^T) = \pi^{-1}(\Omega) = \Sigma,$$

das heißt  $\Sigma$  liegt in  $PD(\mathcal{U}, G)$ . Folglich wird die Menge  $PD(\mathcal{U}, G)$  von  $\pi$  gerade auf den Kegel  $P(\mathcal{U}, G)$  abgebildet und die Behauptung folgt.  $\square$

### Beispiel 2.25 (Invariante graphische Modelle)

Abschließend soll die Einbindung von Invarianzen in ein graphisches Modell anhand eines Beispiels veranschaulicht werden. Seien dazu  $\mathcal{V}$  und  $\mathcal{U}$  wieder wie in Abbildung 1 auf Seite 11. Außerdem sei  $G = \{\text{id}_5, \rho = \varrho_5((14523))\}$ . Damit ist  $G \subseteq \text{Orth}(\mathcal{V})$  eine zulässige Gruppe, vergleiche Beispiel 2.18. Sei schließlich eine Matrix  $\Omega \in P(\mathcal{U})$  gegeben. Es soll nun geklärt werden, welche Bedingungen  $\Omega$  erfüllen muss, um sogar in  $P(\mathcal{U}, G)$  zu liegen. Betrachte dazu

$$\Omega = \begin{pmatrix} \omega_{11} & \omega_{12} & \omega_{13} & \omega_{14} & \omega_{15} \\ \omega_{12} & \omega_{22} & \omega_{23} & 0 & 0 \\ \omega_{13} & \omega_{23} & \omega_{33} & 0 & 0 \\ \omega_{14} & 0 & 0 & \omega_{44} & \omega_{45} \\ \omega_{15} & 0 & 0 & \omega_{45} & \omega_{55} \end{pmatrix} \quad \text{und} \quad \rho\Omega\rho^T = \begin{pmatrix} \omega_{11} & \omega_{14} & \omega_{15} & \omega_{12} & \omega_{13} \\ \omega_{14} & \omega_{44} & \omega_{45} & 0 & 0 \\ \omega_{15} & \omega_{45} & \omega_{55} & 0 & 0 \\ \omega_{12} & 0 & 0 & \omega_{22} & \omega_{23} \\ \omega_{13} & 0 & 0 & \omega_{23} & \omega_{33} \end{pmatrix}.$$

Diese beiden Matrizen sollen identisch sein. Folglich lässt sie der Parameterraum  $P(\mathcal{U}, G)$  beschreiben durch

$$P(\mathcal{U}, G) = \left\{ \left( \begin{pmatrix} \omega_{11} & \omega_{12} & \omega_{13} & \omega_{12} & \omega_{13} \\ \omega_{12} & \omega_{22} & \omega_{23} & 0 & 0 \\ \omega_{13} & \omega_{23} & \omega_{33} & 0 & 0 \\ \omega_{12} & 0 & 0 & \omega_{22} & \omega_{23} \\ \omega_{13} & 0 & 0 & \omega_{23} & \omega_{33} \end{pmatrix} \mid \begin{pmatrix} \omega_{11} & \omega_{12} & \omega_{13} \\ \omega_{12} & \omega_{22} & \omega_{23} \\ \omega_{13} & \omega_{23} & \omega_{33} \end{pmatrix} \in PD_3 \right\}.$$

Die Dimension des Parameterraums ist damit von elf Parametern auf sechs Parameter zurückgegangen. Bevor in Kapitel 4 damit begonnen wird, Maximum-Likelihood-Schätzer in den definierten Modellen zu bestimmen, werden in Kapitel 3 zunächst die theoretischen Grundlagen für verschiedene matrixwertige Verteilungsfamilien gelegt. Diese werden später benötigt, um die Verteilung des Maximum-Likelihood-Schätzers der Kovarianzmatrix zu bestimmen.

### 3 Verallgemeinerte Riesz- und Wishart-Verteilungen

Da in der multivariaten Statistik häufig mehrere Beobachtungen von vektorwertigen Zufallsvariablen betrachtet werden, treten dort auf natürliche Weise matrixwertige Zufallsgrößen auf. Das folgende Kapitel ist Verteilungen eben solcher Zufallsmatrizen gewidmet. Die ersten beiden Unterkapitel beruhen größtenteils auf dem Buch von Kollo und von Rosen [17], während der Rest des Abschnitts an den Aufsatz von Andersson und Klein [3] angelehnt ist.

#### 3.1 Matrixwertige Normalverteilungen

##### Definition der matrixwertigen Normalverteilung

Seien  $X_1, \dots, X_n \sim \mathcal{N}_p(\xi_i, \Sigma)$  homoskedastische,  $p$ -variater normalverteilte Zufallsgrößen, wobei  $\xi_i \in \mathbb{R}^p$  für alle  $i = 1, \dots, n$  und  $\Sigma \in PD_p$  ist. Oft ist es sinnvoll, diese Daten nicht in  $n$  verschiedenen Vektoren zu speichern, sondern sie in eine  $n \times p$ -Matrix zu schreiben. Dazu definiert man die *Datenmatrix*  $X \in \mathbb{R}^{n \times p}$ , indem die Beobachtung  $X_i$  in die  $i$ -te Zeile von  $X$  gesetzt wird. Die Spalten von  $X$  enthalten dann jeweils  $n$  Beobachtungen einer univariaten Zufallsgröße, sie werden mit  $X^{(1)}, \dots, X^{(p)} \in \mathbb{R}^n$  bezeichnet. Damit ist

$$X = \begin{pmatrix} - & X_1^T & - \\ & \vdots & \\ - & X_n^T & - \end{pmatrix} = \begin{pmatrix} | & & | \\ X^{(1)} & \dots & X^{(p)} \\ | & & | \end{pmatrix} \in \mathbb{R}^{n \times p}. \quad (3.1)$$

Nun soll die Verteilung der Matrix  $X$  beschrieben werden. Der Erwartungswertparameter ist einfach zu bestimmen, er besteht aus einer  $n \times p$ -Matrix  $\mu$ , die in der  $i$ -ten Zeile den Mittelwertvektor  $\xi_i \in \mathbb{R}^p$  stehen hat. Etwas komplizierter ist die Darstellung der Kovarianzstruktur, da zweierlei Zusammenhänge modelliert werden müssen: Zum einen die Abhängigkeiten zwischen den einzelnen Komponenten in einer Beobachtung  $X_i$ . Dies geschieht mit Hilfe der Matrix  $\Sigma \in PD_p$ . Zum anderen müssen aber auch die Abhängigkeiten der Zeilen von  $X$ , das heißt der einzelnen Stichproben, berücksichtigt werden. Sei dazu  $\Phi \in PD_n$  eine weitere positiv definite Matrix. Dann wird die *matrixwertige Normalverteilung*  $\mathcal{N}_{n \times p}(\mu, \Phi \otimes \Sigma)$  definiert durch

$$X \sim \mathcal{N}_{n \times p}(\mu, \Phi \otimes \Sigma) \Leftrightarrow \text{vec}(X) \sim \mathcal{N}_{np}(\text{vec}(\mu), \Phi \otimes \Sigma). \quad (3.2)$$

Dabei ist  $\text{vec}(\cdot)$  der klassische Vektorisierungsoperator, der die Zeilen einer Matrix untereinander in einen Vektor schreibt, das heißt  $\text{vec}(X) = (X_1^T, \dots, X_n^T)^T \in \mathbb{R}^{np}$ . Des Weiteren ist  $\Phi \otimes \Sigma$  das Kronecker-Produkt aus  $\Phi$  und  $\Sigma$  und folglich eine  $np \times np$ -Matrix.

Die so definierte matrixwertige Normalverteilung hat die Lebesgue-Dichte

$$\phi_X : \mathbb{R}^{n \times p} \rightarrow [0, 1], \quad N \mapsto \frac{1}{\pi^{np} \det(\Phi)^p \det(\Sigma)^n} e^{-\text{tr}(\Phi^{-1}(N-\mu)\Sigma^{-1}(N-\mu)^T)},$$

siehe Kollo und von Rosen [17, Seite 193]. Etwas Vorsicht ist geboten bei der Verteilung der transponierten Matrix  $X^T$ . Da hier die Rollen von Zeilen und Spalten vertauscht werden, ergibt sich mit derselben Argumentation wie oben

$$X^T \sim \mathcal{N}_{p \times n}(\mu^T, \Sigma \otimes \Phi). \quad (3.3)$$

Abschließend sei hervorgehoben, dass es sich bei der hier definierten matrixwertigen Normalverteilung um einen Spezialfall der Verteilungsfamilie in Eaton [13] handelt, wo Normalverteilungen über allgemeinen Vektorräumen definiert werden.

### Eigenschaften der matrixwertigen Normalverteilung

Die folgenden Eigenschaften werden bei der Bestimmung der Verteilung des Maximum-Likelihood-Schätzers in invarianten graphischen Modellen von Nutzen sein. Größtenteils sind sie aus dem Buch von Kollo und von Rosen [17] übernommen, daher wird anstatt eines Beweises auf die entsprechende Stelle dort verwiesen.

Eine nützliche Eigenschaft wird sein, dass die Normalverteilung unter linearen Transformationen erhalten bleibt.

#### Theorem 3.1 (Lineare Transformationen normalverteilter Zufallsmatrizen)

Sei  $X \sim \mathcal{N}_{n \times p}(\mu, \Phi \otimes \Sigma)$  für  $\mu \in \mathbb{R}^{n \times p}$ ,  $\Phi \in PD_n$  und  $\Sigma \in PD_p$ . Seien des Weiteren  $A \in \mathbb{R}^{a \times n}$ ,  $B \in \mathbb{R}^{p \times b}$  und  $\xi \in \mathbb{R}^{a \times b}$  für  $a, b \in \mathbb{N}$ . Dann gilt:

$$AXB - \xi \sim \mathcal{N}_{a \times b}(A\mu B - \xi, (A\Phi A^T) \otimes (B^T \Sigma B)).$$

#### Beweis

Die Verteilung von  $Y := AXB$  ergibt sich gemäß Theorem 2.2.2 aus Kollo und von Rosen [17] zu  $\mathcal{N}_{a \times b}(\mu', \Phi' \otimes \Sigma')$  mit  $\mu' = A\mu B$ ,  $\Phi' = A\Phi A^T$  und  $\Sigma' = B^T \Sigma B$ . Per Definition der matrixwertigen Normalverteilung ist dann

$$\text{vec}(Y - \xi) = \text{vec}(Y) - \text{vec}(\xi) \sim \mathcal{N}_{ab}(\text{vec}(\mu') - \text{vec}(\xi), \Phi' \otimes \Sigma')$$

und die Behauptung folgt aus der Linearität des Vektorisierungs-Operators.  $\square$

Als direkte Konsequenz lässt sich die Verteilung der Skalierung einer normalverteilten Zufallsmatrix bestimmen.

**Korollar 3.2 (Skalierung einer normalverteilten Zufallsmatrix)**

Sei  $X \sim \mathcal{N}_{n \times p}(\mu, \Phi \otimes \Sigma)$  für  $\mu \in \mathbb{R}^{n \times p}$ ,  $\Phi \in PD_n$  und  $\Sigma \in PD_p$ . Für einen beliebigen Skalar  $\alpha \in \mathbb{R} \setminus \{0\}$  ist dann

$$\alpha X \sim \mathcal{N}_{n \times p}(\alpha \mu, \alpha^2(\Phi \otimes \Sigma)).$$

**Beweis**

Verwende Theorem 3.1 mit  $A = \alpha I_n$ ,  $B = I_p$  und  $\xi = 0$ . □

Unter bestimmten Voraussetzungen können zwei Transformationen derselben Zufallsmatrix stochastisch unabhängig sein, wie das folgende Theorem zeigt.

**Theorem 3.3 (Unabhängigkeit spezieller Transformationen)**

Sei  $X \sim \mathcal{N}_{n \times p}(0, \Phi \otimes \Sigma)$  für  $\Phi \in PD_n$  und  $\Sigma \in PD_p$ . Außerdem seien  $Q \in \mathbb{R}^{n \times n}$  und  $P \in \mathbb{R}^{n \times q}$  für ein  $q \geq 1$ . Dann sind die Zufallsgrößen  $X^T Q X$  und  $X^T P$  genau dann unabhängig, wenn

$$P^T \Phi Q^T \Phi = 0 \in \mathbb{R}^{q \times n} \quad \text{und} \quad P^T \Phi Q \Phi = 0 \in \mathbb{R}^{q \times n}.$$

**Beweis**

Siehe Kollo und von Rosen [17, Theorem 2.2.4(iv)]. □

Des Weiteren ist die Verteilung von Summen unabhängiger, normalverteilter Zufallsmatrizen von Interesse.

**Theorem 3.4 (Faltung normalverteilter Zufallsmatrizen)**

Seien  $\mu_j \in \mathbb{R}^{n \times p}$ , sowie  $\Sigma, \Sigma_j \in PD_p$  und  $\Phi, \Phi_j \in PD_n$  für alle  $j = 1, \dots, \ell$ . Außerdem seien  $X_j \sim \mathcal{N}_{n \times p}(\mu_j, \Phi_j \otimes \Sigma)$  und  $Y_j \sim \mathcal{N}_{n \times p}(\mu_j, \Phi \otimes \Sigma_j)$  so, dass  $X_1, \dots, X_\ell$  unabhängig und  $Y_1, \dots, Y_\ell$  unabhängig sind. Schließlich seien  $A, A_j \in \mathbb{R}^{a \times n}$  und  $B, B_j \in \mathbb{R}^{p \times b}$  für  $a, b \in \mathbb{N}$ . Dann gilt:

$$\begin{aligned} \sum_{j=1}^{\ell} A_j X_j B &\sim \mathcal{N}_{a \times b} \left( \sum_{j=1}^{\ell} A_j \mu_j B, \left( \sum_{j=1}^{\ell} A_j \Phi_j A_j^T \right) \otimes (B^T \Sigma B) \right), \\ \sum_{j=1}^{\ell} A Y_j B_j &\sim \mathcal{N}_{a \times b} \left( \sum_{j=1}^{\ell} A \mu_j B_j, (A \Phi A^T) \otimes \left( \sum_{j=1}^{\ell} B_j^T \Sigma_j B_j \right) \right). \end{aligned}$$

**Beweis**

Siehe Kollo und von Rosen [17, Theorem 2.2.3]. □

Wie bei der vektorwertigen Normalverteilung lässt sich die gemeinsame Verteilung zweier normaler Zufallsmatrizen bestimmen, wenn diese unabhängig sind.

**Lemma 3.5 (Gemeinsame Verteilung bei unabhängigen Zufallsmatrizen)**

Seien  $X_i \sim \mathcal{N}_{n_i \times p}(\mu_i, I_{n_i} \otimes \Sigma)$  mit  $n_i \in \mathbb{N}$ ,  $\mu_i \in \mathbb{R}^{n_i \times p}$  für  $i = 1, 2$  und  $\Sigma \in PD_p$ . Außerdem seien  $X_1$  und  $X_2$  unabhängig. Dann ist

$$X := \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}_{(n_1+n_2) \times p} \left( \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, I_{n_1+n_2} \otimes \Sigma \right).$$

**Beweis**

Aufgrund der Unabhängigkeit von  $X_1$  und  $X_2$  ist

$$\text{vec}(X) = \begin{pmatrix} \text{vec}(X_1) \\ \text{vec}(X_2) \end{pmatrix} \sim \mathcal{N}_{(n_1+n_2)p} \left( \text{vec} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} I_{n_1} \otimes \Sigma & 0 \\ 0 & I_{n_2} \otimes \Sigma \end{pmatrix} \right)$$

Die Behauptung folgt dann aus  $I_{n_1+n_2} = \begin{pmatrix} I_{n_1} & 0 \\ 0 & I_{n_2} \end{pmatrix}$  und der Definition der matrixwertigen Normalverteilung, siehe Gleichung (3.2).  $\square$

Eine weitere matrixwertige Verteilung, die eng mit der Normalverteilung zusammenhängt, ist die klassische Wishart-Verteilung, die im folgenden Kapitel vorgestellt wird.

### 3.2 Klassische Wishart-Verteilungen

**Definition der klassischen Wishart-Verteilung**

Sei jetzt  $X \sim \mathcal{N}_{n \times p}(0, I_n \otimes \Sigma)$  für  $\Sigma \in PD_p$  und  $I_n$  die  $n \times n$ -Einheitsmatrix. Solch eine Datenmatrix tritt insbesondere auf, wenn  $n$  unabhängige, zentrierte, homoskedastisch normalverteilte Zufallsvektoren beobachtet werden. Betrachte nun die induzierte Zufallsmatrix

$$W := X^T X = \begin{pmatrix} | & & | \\ X_1 & \cdots & X_n \\ | & & | \end{pmatrix} \begin{pmatrix} - & X_1^T & - \\ & \vdots & \\ - & X_n^T & - \end{pmatrix} = \sum_{i=1}^n X_i X_i^T \in \mathbb{R}^{p \times p}.$$

Sie folgt einer *klassischen Wishart-Verteilung*  $\mathcal{W}_p(n, \Sigma)$  mit  $n \in \mathbb{N}$  Freiheitsgraden und Streuungsparameter  $\Sigma \in PD_p$ . Beachte, dass es sich für  $n \geq p$  bei den Wishart-Verteilungen um Wahrscheinlichkeitsmaße über dem Kegel der positiv definiten Matrizen  $PD_p$  handelt. Im Folgenden sei daher immer vorausgesetzt, dass  $n \geq p$  gilt. Insbesondere ist somit  $W$  mit Wahrscheinlichkeit Eins positiv definit.

Die Dichtefunktion  $f_W : PD_p \rightarrow [0, 1]$  einer klassisch Wishart-verteiltern Zufallsvariable  $W \sim \mathcal{W}_p(n, \Sigma)$  mit  $n \geq p$  findet man in Kollo und von Rosen [17, Theorem 2.4.6]. Sie

ergibt sich zu

$$f_W(Z) = \frac{\pi^{-\frac{p(p-1)}{4}} 2^{-\frac{np}{2}} \det(Z)^{\frac{n-p-1}{2}}}{\prod_{i=1}^p \Gamma\left(\frac{n-i+1}{2}\right) \det(\Sigma)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2} \operatorname{tr}(\Sigma^{-1}Z)\right\} \quad \forall Z \in PD_p. \quad (3.4)$$

Beachte, dass diese Dichte die Definition der klassischen Wishart-Verteilung sogar für nicht-natürliche Freiheitsgrade  $n \in (p-1, \infty)$  ermöglicht. Liegt  $n$  außerhalb dieses Intervalls, so nimmt das Argument der Gamma-Funktion negative Werte an und die Dichte ist nicht mehr wohldefiniert. Zwar lassen sich mit Hilfe von charakteristischen Funktionen sogenannte singuläre Wishart-Verteilungen für Freiheitsgrade  $n \in \{0, \dots, p-1\}$  definieren, siehe Andersson und Klein [3, Seite 789]. Solche Verteilungen sind jedoch technisch sehr schwer zu handhaben und werden daher in der vorliegenden Arbeit nicht betrachtet.

### Eigenschaften der klassischen Wishart-Verteilung

Das erste Theorem beschäftigt sich mit der Verteilung der Summe von unabhängigen Wishart-verteilten Zufallsgrößen.

#### Theorem 3.6 (Faltung von Wishart-verteilten Zufallsmatrizen)

Seien  $W_1, \dots, W_\ell$  unabhängige Zufallsmatrizen mit  $W_i \sim \mathcal{W}_p(n_i, \Sigma)$ , wobei  $\Sigma \in PD_p$  und  $n_i \geq p$  für  $i = 1, \dots, \ell$  ist. Dann gilt

$$\sum_{i=1}^{\ell} W_i \sim \mathcal{W}_p\left(\sum_{i=1}^{\ell} n_i, \Sigma\right).$$

#### Beweis

Siehe Kollo und von Rosen [17, Theorem 2.4.1]. □

Auch die Wishart-Verteilung bleibt unter linearen Transformationen erhalten, wie das folgende Theorem zeigt.

#### Theorem 3.7 (Lineare Transformationen Wishart-verteilter Zufallsmatrizen)

Sei  $W \sim \mathcal{W}_p(n, \Sigma)$  mit  $\Sigma \in PD_p$  und  $n \geq p$ , sowie  $A \in \mathbb{R}^{a \times p}$ . Dann ist

$$AWA^T \sim \mathcal{W}_a(n, A\Sigma A^T).$$

#### Beweis

Siehe Kollo und von Rosen [17, Theorem 2.4.2]. □

Wieder ist die Verteilung einer skalierten Wishart-Matrix eine direkte Folgerung.

**Korollar 3.8 (Skalierung einer Wishart-verteilten Zufallsmatrix)**

Sei  $W \sim \mathcal{W}_p(n, \Sigma)$  mit  $\Sigma \in PD_p$  und  $n \geq p$ , sowie  $\alpha > 0$ . Dann ist

$$\alpha W \sim \mathcal{W}_p(n, \alpha \Sigma).$$

**Beweis**

Verwende Theorem 3.7 mit  $A = \sqrt{\alpha}I_p$ . □

Beachte, dass im Gegensatz zur Normalverteilung der Skalar  $\alpha$  positiv sein muss, da der Träger der Verteilung von  $\alpha W$  andernfalls nicht mehr der Kegel der positiv definiten Matrizen wäre.

Eine weitere interessante Zufallsgröße ist die Determinante einer Wishart-verteilten Matrix  $W$ . Bei der Entwicklung von Likelihood-Quotienten-Tests in Kapitel 6 werden insbesondere die Momente einer solchen Zufallsvariablen  $\det(W)$  eine Rolle spielen.

**Theorem 3.9 (Momente der Determinante einer Wishart-Matrix)**

Sei  $W \sim \mathcal{W}_p(n, \Sigma)$  für  $n \geq p$  und  $\Sigma \in PD_p$ . Dann ist

$$\mathbb{E}[\det(W)^\alpha] = 2^{\alpha p} \det(\Sigma)^\alpha \prod_{i=1}^p \frac{\Gamma(\frac{n-i+1}{2} + \alpha)}{\Gamma(\frac{n-i+1}{2})} \quad \text{für alle } \alpha > -\frac{n-p+1}{2}.$$

**Beweis**

**I.** Sei zunächst  $Y$  eine  $(0, \infty)$ -wertige Zufallsvariable, die einer  $\chi^2$ -Verteilung mit  $m \in \mathbb{N}$  Freiheitsgraden folgt. Dann ist

$$\mathbb{E}[Y^\alpha] = 2^\alpha \frac{\Gamma(\frac{m}{2} + \alpha)}{\Gamma(\frac{m}{2})} \int_0^\infty \frac{1}{2^{(\alpha+\frac{m}{2})}\Gamma(\frac{m}{2} + \alpha)} y^{(\alpha+\frac{m}{2})-1} e^{-\frac{y}{2}} dy = 2^\alpha \frac{\Gamma(\frac{m}{2} + \alpha)}{\Gamma(\frac{m}{2})} \quad (3.5)$$

für alle  $\alpha > -m/2$ . Diese Einschränkung an  $\alpha$  muss getroffen werden, da sonst das Argument der Gamma-Funktion negativ wäre.

**II.** Betrachte nun die Cholesky-Zerlegung  $\Sigma = AA^T$  einer positiv definiten Matrix  $\Sigma$ . Seien des Weiteren  $V \sim \mathcal{W}_p(n, I_p)$  und  $W \sim \mathcal{W}_p(n, \Sigma)$  zwei Wishart-verteilte Zufallsmatrizen mit identischen Freiheitsgraden  $n \geq p$ . Wegen Theorem 3.7 ist

$$AVA^T \sim \mathcal{W}_p(n, AI_pA^T) = \mathcal{W}_p(n, \Sigma),$$

das heißt  $AVA^T$  und  $W$  sind identisch verteilt. Wegen Korollar 2.4.4.1 in Kollo und von Rosen [17] lässt sich  $V$  in der Form  $V = UU^T$  darstellen. Dabei ist  $U$  eine Zufallsmatrix in unterer Dreiecksform, deren Komponenten alle unabhängig voneinander sind und deren

quadrierte Diagonalelemente jeweils einer  $\chi^2$ -Verteilung folgen,

$$U_{ii}^2 \sim \chi_{n-i+1}^2 \quad \text{für alle } i = 1, \dots, p.$$

Insbesondere lassen sich mit Hilfe von Gleichung (3.5) demnach die Momente der Determinante von  $V$  bestimmen,

$$\mathbb{E}[\det(V)^\alpha] = \mathbb{E}[\det(U)^{2\alpha}] = \prod_{i=1}^p \mathbb{E}[(U_{ii}^2)^\alpha] = 2^{\alpha p} \prod_{i=1}^p \frac{\Gamma\left(\frac{n-i+1}{2} + \alpha\right)}{\Gamma\left(\frac{n-i+1}{2}\right)}$$

für alle  $\alpha > -(n-p+1)/2$ . Die Behauptung folgt dann aus

$$\mathbb{E}[\det(W)^\alpha] = \mathbb{E}[\det(AVA^T)^\alpha] = \mathbb{E}[\det(A)^{2\alpha} \det(V)^\alpha] = \det(\Sigma)^\alpha \mathbb{E}[\det(V)^\alpha]. \quad \square$$

Die nächsten beiden Theoreme zeigen, dass für eine Zufallsmatrix  $X \sim \mathcal{N}_{n \times p}(0, \Phi \otimes \Sigma)$  unter bestimmten Voraussetzungen auch Matrizen der Form  $X^T Q X$  eine Wishart-Verteilung haben können, wobei  $Q \in PD_n$  ist. Falls  $\Phi$  beliebig ist, das heißt die Beobachtungen können abhängig sein, so muss dazu  $Q = \Phi^{-1}$  gewählt werden, siehe Theorem 3.10. Sind die Beobachtungen dagegen unabhängig, das heißt  $\Phi = I_n$ , so ist man bei der Wahl von  $Q$  freier, siehe Theorem 3.11.

**Theorem 3.10 (Wishart-Verteilung bei abhängigen Beobachtungen)**

Sei  $X \sim \mathcal{N}_{n \times p}(0, \Phi \otimes \Sigma)$  mit  $n \geq p$ , sowie  $\Phi \in PD_n$  und  $\Sigma \in PD_p$ . Dann ist

$$X^T \Phi^{-1} X \sim \mathcal{W}_p(n, \Sigma).$$

**Beweis**

Siehe Kollo und von Rosen [17, Theorem 2.4.1]. □

**Theorem 3.11 (Wishart-Verteilungen und orthogonale Projektionen)**

Sei  $X \sim \mathcal{N}_{n \times p}(0, I_n \otimes \Sigma)$  mit  $n \geq p$  und  $\Sigma \in PD_p$ . Sei des Weiteren  $Q \in \mathbb{R}^{n \times n}$  eine symmetrische und idempotente Matrix mit Rang  $rk(Q)$ . Dann ist

$$X^T Q X \sim \mathcal{W}_p(rk(Q), \Sigma).$$

**Beweis**

Siehe Kollo und von Rosen [17, Korollar 2.4.3.1]. □

Beide Theoreme werden bei der Bestimmung der Verteilung der Maximum-Likelihood-Schätzer eine wichtige Rolle spielen. Um diese Schätzer effizient beschreiben zu können, wird im nächsten Kapitel jedoch erst einmal eine an die Struktur des vorgegebenen Graphen  $\mathcal{V}$  angepasste Verallgemeinerung der klassischen Cholesky-Zerlegung entwickelt.

### 3.3 Verallgemeinerte Cholesky-Zerlegungen

Sei zunächst  $W \in PD_p$  eine beliebige positiv definite Matrix. Dann lässt sich  $W$  bekanntermaßen in der Form  $W = (I_p - L)^{-1}D(I_p - L)^{-T}$  darstellen. Dabei ist  $D \in \mathbb{R}^{p \times p}$  eine Diagonalmatrix mit positiven Einträgen und  $L \in \mathbb{R}^{p \times p}$  ist eine strikte untere Dreiecksmatrix, das heißt eine untere Dreiecksmatrix mit Nullen auf der Diagonalen. Diese Darstellung wird als *Cholesky-Zerlegung* von  $W$  bezeichnet.

#### Die verallgemeinerte Cholesky-Zerlegung

Um diese Zerlegung auf die hiesige Situation zu verallgemeinern, sei ein Annahme 2.1 erfüllender azyklisch gemischter Graph  $\mathcal{V} = (V, D, U)$  mit Gerüst  $\mathcal{U} = \mathcal{V}^\sim$  vorgegeben. Außerdem sei  $\Omega \in P(\mathcal{U})$  eine  $p \times p$ -Matrix mit von  $\mathcal{U}$  vorgegebener Nullstruktur. Definiere dann für jede Box  $[B] \in \mathcal{B}(\mathcal{V})$  die beiden induzierten Matrizen

$$\Omega_{[B]\bullet} := \Omega_{[B]} - \Omega_{[B]}\Omega_{\langle B \rangle}^{-1}\Omega_{[B]}^T \in \mathbb{R}^{[B] \times [B]} \quad \text{und} \quad \Omega_{[B]\bullet} := \Omega_{[B]}\Omega_{\langle B \rangle}^{-1} \in \mathbb{R}^{[B] \times \langle B \rangle}. \quad (3.6)$$

Dabei sei daran erinnert, dass per Konvention  $\Omega_{\langle B \rangle}^{-1} = (\Omega_{\langle B \rangle})^{-1}$  und  $\Omega_{[B]}^T = (\Omega_{[B]})^T$  ist. Die Elemente der Menge  $\{\Omega_{[B]\bullet}, \Omega_{[B]\bullet}\}_{[B] \in \mathcal{B}(\mathcal{V})}$  werden im Folgenden als die *Cholesky-Parameter* von  $\Omega$  bezeichnet, da sie eine wichtige Rolle bei der verallgemeinerten Cholesky-Zerlegung von  $\Omega$  spielen werden.

Sei  $[M] \in \mathcal{B}(\mathcal{V})$  eine maximale Box von  $\mathcal{V}$ . Per Definition erfüllt dann der induzierte Teilgraph  $\mathcal{V}_{V \setminus [M]}$  immer noch die Annahme 2.1 und es gilt  $\mathcal{U}_{V \setminus [M]} = (\mathcal{V}_{V \setminus [M]})^\sim$ . Mit anderen Worten ist  $\mathcal{V}_{V \setminus [M]}$  eine Darstellung von  $\mathcal{U}_{V \setminus [M]}$  als azyklisch gemischter Graph. Proposition 5.2 aus Andersson und Klein [3] zeigt dann, dass die Abbildung

$$\begin{aligned} P(\mathcal{U}) &\rightarrow P(\mathcal{U}_{V \setminus [M]}) \times \mathbb{R}^{[M] \times \langle M \rangle} \times PD_{[M]}, \\ \Omega &\mapsto (\Omega_{V \setminus [M]}, \Omega_{[M]\bullet}, \Omega_{[M]\bullet}) \end{aligned}$$

bijektiv ist. Induktive Anwendung dieser Aussage liefert das folgende Theorem.

#### Theorem 3.12 (Bijektivität von Parametermatrix und Cholesky-Parametern)

Sei  $\mathcal{U}$  ein zerlegbarer Graph mit Darstellung  $\mathcal{V}$ . Dann ist die Abbildung

$$\begin{aligned} P(\mathcal{U}) &\rightarrow \prod_{[B] \in \mathcal{B}(\mathcal{V})} \left( \mathbb{R}^{[B] \times \langle B \rangle} \times PD_{[B]} \right), \\ \Omega &\mapsto \left\{ \Omega_{[B]\bullet}, \Omega_{[B]\bullet} \right\}_{[B] \in \mathcal{B}(\mathcal{V})} \end{aligned}$$

bijektiv.

Dieses Theorem besagt insbesondere, dass alle Informationen einer Matrix  $\Omega \in P(\mathcal{U})$  schon in ihren Cholesky-Parametern gespeichert sind. Bei der verallgemeinerten Cholesky-Zerlegung handelt es sich nun um eine Beschreibung, wie die Matrix  $\Omega$  aus ihren Cholesky-Parametern rekonstruiert werden kann.

**Theorem 3.13 (Verallgemeinerte Cholesky-Zerlegung)**

Sei  $\mathcal{U}$  ein zerlegbarer Graph mit Darstellung  $\mathcal{V}$ , sowie  $\Omega \in P(\mathcal{U})$  und  $L, D \in \mathbb{R}^{p \times p}$ . Dabei sei  $L$  eine Blockmatrix mit  $L_{[B]} = \Omega_{[B]\bullet}$  für alle Boxen  $[B] \in \mathcal{B}(\mathcal{V})$ , während alle nicht-spezifizierten Blöcke von  $L$  verschwinden. Bei  $D$  handle es sich um eine Blockdiagonalmatrix mit  $D_{[B]} = \Omega_{[B]\bullet}$  für alle  $[B] \in \mathcal{B}(\mathcal{V})$ . Dann gilt mit der Projektion  $\pi$  aus Gleichung (2.2) von Seite 13 die Identität

$$\pi \left[ (I_p - L)^{-1} D (I_p - L)^{-T} \right] = \Omega$$

und diese Darstellung ist eindeutig.

**Beweis**

Siehe Korollare 5.1 und 5.3 in Andersson und Klein [3]. □

Es sei hervorgehoben, dass  $L$  als strikte untere Blockdreiecksmatrix auftritt, wenn die Knoten von  $\mathcal{V}$  gemäß Annahme 2.2 von Seite 10 nummeriert sind.

**Symmetrien in den Cholesky-Parametern**

Sei wieder ein Annahme 2.6 erfüllender azyklisch gemischter Graph  $\mathcal{V}$  mit Gerüst  $\mathcal{U} = \mathcal{V}^\sim$  und Boxen  $\mathcal{B}(\mathcal{V}) = \{[B_1], \dots, [B_k]\}$  gegeben. Des Weiteren sei  $G \subseteq Orth(\mathcal{V})$  eine zulässige endliche Gruppe. Insbesondere wird dann von jeder orthogonalen Transformation der Knoten  $\rho \in G$  eine Permutation der Boxen  $\tau_\rho \in Aut[GoB(\mathcal{V})] \subseteq Sym(k)$  induziert, siehe Definition 2.17 auf Seite 26. Eine weitere interessante Fragestellung im Zusammenhang mit den Cholesky-Parametern ist die folgende: Gegeben eine Streuungsmatrix  $\Omega \in P(\mathcal{U})$ , lässt sich dann an den Cholesky-Parametern  $\{\Omega_{[B]\bullet}, \Omega_{[B]\bullet}\}_{[B] \in \mathcal{B}(\mathcal{V})}$  ablesen, ob  $\Omega$  die von  $G$  vorgegebenen Symmetrien respektiert, das heißt ob  $\Omega$  sogar in  $P(\mathcal{U}, G)$  liegt? Eine Antwort darauf liefert das folgende Theorem.

**Theorem 3.14 (Symmetrien in den Cholesky-Parametern)**

Seien  $\mathcal{V}, \mathcal{U}, G$  wie oben beschrieben und  $\Omega \in P(\mathcal{U})$ . Dann ist  $\Omega \in P(\mathcal{U}, G)$  genau dann, wenn für alle Transformationen  $\rho \in G$  und alle Boxen  $[B] \in \mathcal{B}(\mathcal{V})$  gilt:

$$\begin{aligned} \Omega_{[\tau_\rho(B)]\bullet} &= \rho_{[\tau_\rho(B), B]} \Omega_{[B]\bullet} \left( \rho_{[\tau_\rho(B), B]} \right)^T \quad \text{und} \\ \Omega_{[\tau_\rho(B)]\bullet} &= \rho_{[\tau_\rho(B), B]} \Omega_{[B]\bullet} \left( \rho_{[\tau_\rho(B), B]} \right)^T. \end{aligned}$$

**Beweis**

Siehe Madsen [20, Theorem 6.1]. □

Das Theorem besagt also, dass eine Matrix  $\Omega \in P(\mathcal{U})$  genau dann die von  $G$  vorgegebenen Symmetrien respektiert, wenn ihre Cholesky-Parameter dies ebenfalls tun.

**Das  $\lambda$ -Inverse einer positiv definiten Matrix**

Später wird es von Nutzen sein, einen flexibleren Invertierungsbegriff zur Verfügung zu haben als die klassische Matrizeninvertierung. Sei dazu wieder  $\mathcal{U}$  ein zerlegbarer Graph mit Darstellung  $\mathcal{V}$  und  $\Omega \in P(\mathcal{U})$ . Dann lässt sich  $\Omega$  gemäß Theorem 3.13 darstellen als

$$\Omega = \pi \left[ (I_p - L)^{-1} \text{Diag} \left[ \Omega_{[B]\bullet} \mid [B] \in \mathcal{B}(\mathcal{V}) \right] (I_p - L)^{-T} \right],$$

wobei  $L$  wie in Theorem 3.13 definiert sei. Hier und im Folgenden meint  $\text{Diag}[\cdot]$  eine Blockdiagonalmatrix mit den angegebenen Matrizen auf der Diagonalen. Sei nun  $\lambda = (\lambda_{[B]}) \in \mathbb{R}^{|\mathcal{B}(\mathcal{V})|}$  ein Vektor, der für jede Box von  $\mathcal{V}$  einen reellen Wert enthält. Definiere dann die Matrix

$$\Omega^{-\lambda} := (I_p - L)^T \text{Diag} \left[ \lambda_{[B]} \Omega_{[B]\bullet}^{-1} \mid [B] \in \mathcal{B}(\mathcal{V}) \right] (I_p - L). \quad (3.7)$$

Dann liegt  $\Omega^{-\lambda}$  in  $PD^0(\mathcal{U})$ , siehe Andersson und Klein [3, Proposition 5.1]. Die Matrix  $\Omega^{-\lambda}$  wird das  $\lambda$ -Inverse von  $\Omega$  genannt. Beachte, dass sie von der konkreten Darstellung  $\mathcal{V}$  des zerlegbaren Graphen  $\mathcal{U}$  abhängt, auch wenn sich dies nicht an der Notation erkennen lässt. Bezüglich welcher Darstellung  $\mathcal{V}$  von  $\mathcal{U}$  das  $\lambda$ -Inverse gebildet wird, wird jedoch immer aus dem Kontext klar sein. Weitere Eigenschaften des  $\lambda$ -Inversen werden im folgenden Lemma zusammengefasst.

**Lemma 3.15 (Eigenschaften des  $\lambda$ -Inversen)**

Sei  $\mathcal{U}$  ein zerlegbarer Graph mit Darstellung  $\mathcal{V}$ , sowie  $\lambda \in \mathbb{R}^{|\mathcal{B}(\mathcal{V})|}$  und  $\beta > 0$ . Außerdem seien  $\Omega \in P(\mathcal{U})$  und  $\Sigma \in PD(\mathcal{U})$ . Dann gilt:

(i) Falls  $\lambda = (\beta, \dots, \beta)^T$ :  $\Omega^{-\lambda} = \beta (\pi^{-1}(\Omega))^{-1}$ .

(ii)  $\Omega^{-\alpha\lambda} = \alpha \Omega^{-\lambda}$  für alle  $\alpha > 0$ .

(iii)  $(\alpha\Omega)^{-\lambda} = \frac{1}{\alpha} \Omega^{-\lambda}$  für alle  $\alpha > 0$ .

**Beweis**

Folgt direkt aus  $\text{Diag}[\alpha\Omega_{[B]\bullet}] = \alpha \text{Diag}[\Omega_{[B]\bullet}]$  und  $(\alpha\Omega)_{[B]\bullet} = \alpha\Omega_{[B]\bullet}$ . □

### 3.4 Verallgemeinerte Riesz-Verteilungen

Mit den bereits entwickelten Hilfsmitteln lässt sich jetzt die verallgemeinerte Riesz-Verteilung definieren, vergleiche Andersson und Klein [3, Definition 10.1].

**Definition 3.16 (Die verallgemeinerte Riesz-Verteilung)**

Sei ein Annahme 2.1 erfüllender azyklisch gemischter Graph  $\mathcal{V} = (V, D, U)$  mit Gerüst  $\mathcal{U} = \mathcal{V}^\sim$  gegeben. Sei des Weiteren  $\Omega \in P(\mathcal{U})$  und  $\lambda \in \mathbb{R}^{|\mathcal{B}(\mathcal{V})|}$  mit

$$\lambda_{[B]} > \frac{[B] + \langle B \rangle - 1}{2} \quad \text{für alle } [B] \in \mathcal{B}(\mathcal{V}). \quad (3.8)$$

Dann hat die verallgemeinerte Riesz-Verteilung  $\mathbb{R}_p(\Omega, \lambda, \mathcal{V})$  zur Darstellung  $\mathcal{V}$  mit Erwartungsparameter  $\Omega$  und Formparameter  $\lambda$  die Lebesgue-Dichte

$$\mathbb{r}_p(\Omega, \lambda, \mathcal{V}; W) := h(\lambda, \mathcal{V}) \prod_{[B] \in \mathcal{B}(\mathcal{V})} \left\{ \frac{\det(W_{[B] \bullet})^{\lambda_{[B]} - \frac{[B] + \langle B \rangle + 1}{2}}}{\det(W_{(B)})^{\frac{[B]}{2}}} \det(\Omega_{[B] \bullet})^{-\lambda_{[B]}} \right\} e^{-\text{tr}(\Omega^{-\lambda} W)}$$

mit  $h(\lambda, \mathcal{V}) := \pi^{\frac{p - \dim(P(\mathcal{U}))}{2}} \prod_{[B] \in \mathcal{B}(\mathcal{V})} \frac{(\lambda_{[B]})^{[B] \cdot \lambda_{[B]}}}{\prod_{i=1}^{[B]} \Gamma\left(\lambda_{[B]} - \frac{\langle B \rangle + i - 1}{2}\right)}$ .

Damit ist die verallgemeinerte Riesz-Verteilung  $\mathbb{R}_p(\Omega, \lambda, \mathcal{V})$  ein Wahrscheinlichkeitsmaß über  $P(\mathcal{U})$ .

Dass es sich bei  $\mathbb{r}_p(\Omega, \lambda, \mathcal{V}; W)$  um eine Lebesgue-Dichte über  $P(\mathcal{U})$  handelt, wird in Andersson und Klein [3, Kapitel 7] gezeigt. Des Weiteren trägt der Erwartungsparameter  $\Omega$  seinen Namen zu Recht, denn der Erwartungswert einer  $\mathbb{R}_p(\Omega, \lambda, \mathcal{V})$ -verteilten Zufallsgröße ist gerade  $\Omega$ , siehe Andersson und Klein [3, Proposition 10.1]. Es sei nochmal hervorgehoben, dass die verallgemeinerte Riesz-Verteilung nicht nur von den beiden Parametern  $\Omega$  und  $\lambda$  abhängt, sondern dass auch die Darstellung  $\mathcal{V}$  von  $\mathcal{U}$  eine entscheidende Rolle spielt. Hat der Formparameter  $\lambda$  dagegen identische Einträge, das heißt  $\lambda = (\beta, \dots, \beta)^T$  für ein  $\beta > 0$ , so wird sich später zeigen, dass die Darstellung  $\mathcal{V}$  keinen Einfluss auf die Verteilung  $\mathbb{R}_p(\Omega, \lambda, \mathcal{V})$  hat, siehe Theorem 3.23. Diese Beobachtung wird unter anderem bei der Beschreibung der Verteilung des Maximum-Likelihood-Schätzers der Kovarianzmatrix von Bedeutung sein, vergleiche Bemerkung 5.16.

Das wichtigste Hilfsmittel bei der Bestimmung dieser Verteilung wird allerdings der folgende Satz sein.

**Theorem 3.17 (Hauptsatz zur verallgemeinerten Riesz-Verteilung)**

Sei  $\mathcal{V}$  ein Annahme 2.1 erfüllender azyklisch gemischter Graph mit maximaler Box  $[M]$  und Gerüst  $\mathcal{U} = \mathcal{V}^\sim$ . Seien außerdem  $\Omega \in P(\mathcal{U})$  und  $\lambda \in \mathbb{R}^{|\mathcal{B}(\mathcal{V})|}$  wie in Gleichung (3.8). Für eine  $P(\mathcal{U})$ -wertige Zufallsmatrix  $W$  gilt in diesem Fall:  $W$  folgt genau dann einer  $\mathbb{R}_p(\Omega, \lambda, \mathcal{V})$ -Verteilung, wenn die folgenden vier Eigenschaften erfüllt sind.

$$(i) \quad W_{[M]\bullet} \perp \{W_{[M]\bullet}, W_{V \setminus [M]}\}.$$

$$(ii) \quad \mathcal{L}(W_{[M]\bullet}) = \mathcal{W}_{[M]} \left( 2\lambda_{[M]} - \langle M \rangle, \frac{1}{2\lambda_{[M]}} \Omega_{[M]\bullet} \right).$$

$$(iii) \quad \mathcal{L}(W_{[M]\bullet} | W_{V \setminus [M]}) = \mathcal{N}_{[M] \times \langle M \rangle} \left( \Omega_{[M]\bullet}, \left( \frac{1}{2\lambda_{[M]}} \Omega_{[M]\bullet} \right) \otimes W_{\langle M \rangle}^{-1} \right).$$

$$(iv) \quad \mathcal{L}(W_{V \setminus [M]}) = \mathbb{R}_{V \setminus [M]}(\Omega_{V \setminus [M]}, \lambda_{V \setminus [M]}, \mathcal{V}_{V \setminus [M]}).$$

Dabei meint  $\mathcal{W}_{[M]}$  die in Kapitel 3.2 eingeführte klassische Wishartverteilung über  $PD_{[M]}$ ,  $\mathcal{N}_{[M] \times \langle M \rangle}$  die matrixwertige Normalverteilung über  $\mathbb{R}^{[M] \times \langle M \rangle}$  aus Kapitel 3.1 und  $\mathbb{R}_{V \setminus [M]}$  die oben definierte verallgemeinerte Riesz-Verteilung über  $P(\mathcal{U}_{V \setminus [M]})$ .

**Beweis**

Siehe Andersson und Klein [3, Proposition 10.2]. Etwas Vorsicht ist geboten bei der Parametrisierung der klassischen Wishart-Verteilung. Näheres dazu folgt in Kapitel 3.5. Beachte aber insbesondere, dass die obige klassische Wishart-Verteilung wohldefiniert ist. Aufgrund der Wahl von  $\lambda$  in Gleichung (3.8) ist nämlich  $2\lambda_{[M]} - \langle M \rangle$  größer als  $[M] - 1$  und stellt somit einen zulässigen Freiheitsgrad dar, siehe Gleichung (3.4) auf Seite 41 und die Bemerkung darunter.  $\square$

**Korollar 3.18 (Unabhängigkeiten in der allgemeinen Riesz-Verteilung)**

Sei  $W$  eine verallgemeinert Riesz-verteilte Zufallsgröße wie in Theorem 3.17. Dann gilt für eine maximale Box  $[M]$  die bedingte Unabhängigkeitsbeziehung

$$W_{[M]\bullet} \perp W_{R, V \setminus [M]} \mid W_{\langle M \rangle} \quad \text{mit } R := V \setminus \{[M] \cup \langle M \rangle\}.$$

**Beweis**

Beachte, dass die Zufallsmatrix  $W_{V \setminus [M]}$  aus den Teilen  $\{W_{R,R}, W_{R,\langle M \rangle}, W_{\langle M \rangle}\}$  besteht. Da die bedingte Verteilung von  $W_{[M]\bullet}$  gegeben  $W_{V \setminus [M]}$  nur von  $W_{\langle M \rangle}$  abhängt, siehe Theorem 3.17 (iii), folgt direkt aus der Definition der bedingten Unabhängigkeit in Lauritzen [18, Seite 28], dass  $W_{[M]\bullet}$  unabhängig ist von  $\{W_{R,R}, W_{R,\langle M \rangle}\}$  gegeben  $W_{\langle M \rangle}$ . Die Teile  $\{W_{R,R}, W_{R,\langle M \rangle}\}$  ergeben zusammen die Zufallsmatrix  $W_{R, V \setminus [M]}$  und die Behauptung folgt.  $\square$

Sei nun  $\mathcal{A} \subseteq \mathcal{B}(\mathcal{V})$  eine Menge von Boxen von  $\mathcal{V}$ . Die Menge  $\mathcal{A}$  heißt *Vorfahrenmenge*, falls für alle Boxen  $[B] \in \mathcal{A}$  auch die Box  $[A]$  zu  $\mathcal{A}$  gehört, sobald die Kante  $[A] \rightarrow [B]$  im Graph der Boxen  $GoB(\mathcal{V})$  auftaucht. Anschaulich muss also  $\mathcal{A}$  für jede Box  $[B] \in \mathcal{A}$  auch alle deren Vorfahren im Graph der Boxen  $GoB(\mathcal{V})$  enthalten. Außerdem sei  $A := \cup_{[B] \in \mathcal{A}} [B]$  die Menge der an  $\mathcal{A}$  beteiligten Knoten. Beachte, dass  $\mathcal{V}_A$  eine Darstellung von  $\mathcal{U}_A$  als azyklisch gemischter Graph ist, falls  $\mathcal{A}$  eine Vorfahrenmenge ist. Das folgende Theorem liefert eine Aussage über die Randverteilungen der verallgemeinerten Riesz-Verteilung.

**Theorem 3.19 (Randverteilung einer verallgemeinerten Riesz-Verteilung)**

Sei  $\mathcal{V}$  ein Annahme 2.1 erfüllender azyklisch gemischter Graph mit Gerüst  $\mathcal{U} = \mathcal{V}^\sim$ , sowie  $\Omega \in P(\mathcal{U})$  und  $\lambda \in \mathbb{R}^{|\mathcal{B}(\mathcal{V})|}$  wie in Gleichung (3.8). Des Weiteren sei  $W \sim \mathbb{R}_p(\Omega, \lambda, \mathcal{V})$  und  $\mathcal{A} \subseteq \mathcal{B}(\mathcal{V})$  sei eine Vorfahrenmenge. Außerdem sei  $A := \cup_{[B] \in \mathcal{A}} [B]$  die Menge der an  $\mathcal{A}$  beteiligten Knoten. Dann hat die  $P(\mathcal{U}_A)$ -wertige Zufallsmatrix  $W_A$  die Verteilung

$$W_A \sim \mathbb{R}_A(\Omega_A, \lambda_A, \mathcal{V}_A).$$

**Beweis**

Siehe Andersson und Klein [3, Proposition 11.1]. □

Auch für die verallgemeinerte Riesz-Verteilung lässt sich eine Faltungsformel entwickeln, wie das folgende Theorem zeigt.

**Theorem 3.20 (Faltung Riesz-verteilter Zufallsmatrizen)**

Sei  $\mathcal{V}$  ein Annahme 2.1 erfüllender azyklisch gemischter Graph mit Gerüst  $\mathcal{U} = \mathcal{V}^\sim$ , sowie  $\Omega \in P(\mathcal{U})$  und  $\lambda \in \mathbb{R}^{|\mathcal{B}(\mathcal{V})|}$  wie in Gleichung (3.8). Des Weiteren seien  $W_1, \dots, W_m$  unabhängig mit  $W_i \sim \mathbb{R}_p(\Omega_i, \lambda_i, \mathcal{V})$  für alle  $i = 1, \dots, m$ . Außerdem seien die  $\lambda$ -Inversen aller Erwartungsparameter konstant, das heißt für alle  $i$  ist  $\Omega_i^{-\lambda_i} = \Lambda$  für ein  $\Lambda \in PD^0(\mathcal{U})$ . Dann ist

$$W_+ := \sum_{i=1}^m W_i \sim \mathbb{R}_p(\tilde{\Omega}, \lambda_+, \mathcal{V}) \quad \text{mit} \quad \lambda_+ := \sum_{i=1}^m \lambda_i \in \mathbb{R}^{|\mathcal{B}(\mathcal{V})|} \quad \text{und} \quad \Lambda = \tilde{\Omega}^{-\lambda_+}.$$

**Beweis**

Siehe Andersson und Klein [3, Proposition 11.3]. Beachte, dass die Aussage dort in einer anderen Parametrisierung formuliert ist. Die hier angeführte Version ergibt sich unter Beachtung der Umparametrisierung  $R_p(\Lambda, \lambda, \mathcal{V}) = \mathbb{R}_p(\tilde{\Omega}, \lambda, \mathcal{V})$  mit  $\tilde{\Omega}^{-\lambda} = \Lambda$ . □

Um die Verteilung von  $W_+$  zu bestimmen, muss folglich zunächst das System  $\Lambda = \tilde{\Omega}^{-\lambda_+}$  nach  $\tilde{\Omega} \in P(\mathcal{U})$  aufgelöst werden. Stelle dazu  $\Omega_1$  in der Form aus Theorem 3.13 dar,

$$\Omega_1 = \pi \left[ (I_p - L_1)^{-1} \text{Diag} [(\Omega_1)_{[B]\bullet} \mid [B] \in \mathcal{B}(\mathcal{V})] (I_p - L_1)^{-T} \right].$$

Definiere dann  $\tilde{\lambda} \in \mathbb{R}^{|\mathcal{B}(\mathcal{V})|}$  über  $\tilde{\lambda}_{[B]} := (\lambda_+)_{[B]} / (\lambda_1)_{[B]}$  für alle  $[B] \in \mathcal{B}(\mathcal{V})$ . Gewichte dann die Diagonaleinträge in der verallgemeinerten Cholesky-Zerlegung von  $\Omega_1$  mit den Einträgen von  $\tilde{\lambda}$ , das heißt, setze

$$\tilde{\Omega} = \pi \left[ (I_p - L_1)^{-1} \text{Diag} [\tilde{\lambda}_{[B]} (\Omega_1)_{[B]\bullet} \mid [B] \in \mathcal{B}(\mathcal{V})] (I_p - L_1)^{-T} \right].$$

Dann ist per Definition des  $\lambda$ -Inversen aus Gleichung (3.7)

$$\tilde{\Omega}^{-\lambda_+} = (I_p - L_1)^T \text{Diag} [(\lambda_+)_{[B]} \tilde{\lambda}_{[B]}^{-1} (\Omega_1)_{[B]\bullet}^{-1} \mid [B] \in \mathcal{B}(\mathcal{V})] (I_p - L_1) = \Omega_1^{-\lambda_1} = \Lambda.$$

Dies zeigt, dass man den Erwartungsparameter  $\tilde{\Omega}$  der Riesz-Verteilung von  $W_+$  bekommt, indem man in der verallgemeinerten Cholesky-Zerlegung von  $\Omega_1$  den zur Box  $[B]$  gehörenden Diagonalblock mit  $\tilde{\lambda}_{[B]} = (\lambda_+)_{[B]} / (\lambda_1)_{[B]}$  gewichtet.

Etwas einfacher gestaltet sich die Bestimmung der Verteilung von  $W_+$ , wenn die beteiligten Zufallsvariablen nicht nur unabhängig, sondern auch identisch verteilt sind.

**Korollar 3.21 (Faltung bei identisch verteilten Zufallsmatrizen)**

Seien  $W_1, \dots, W_m$  wie in Theorem 3.20 mit  $\Omega_i = \Omega \in P(\mathcal{U})$  und  $\lambda_i = \lambda \in \mathbb{R}^{|\mathcal{B}(\mathcal{V})|}$  für alle  $i = 1, \dots, m$ . Dann ist

$$W_+ := \sum_{i=1}^m W_i \sim \mathbb{R}_p(m\Omega, m\lambda, \mathcal{V}).$$

**Beweis**

Verwende Theorem 3.20 und beachte, dass  $\lambda_+ = m\lambda$  ist. Die Behauptung folgt dann wegen  $(m\Omega)^{-m\lambda} = \Omega^{-\lambda}$ , siehe Lemma 3.15.  $\square$

Als nächstes soll gezeigt werden, dass eine skalierte Riesz-verteilte Zufallsgröße selbst wieder Riesz-verteilt ist.

**Theorem 3.22 (Skalierte verallgemeinerte Riesz-Verteilungen)**

Sei  $\mathcal{V}$  ein Annahme 2.1 erfüllender azyklisch gemischter Graph mit Gerüst  $\mathcal{U} = \mathcal{V}^\sim$ , sowie  $\Omega \in P(\mathcal{U})$  und  $\lambda \in \mathbb{R}^{|\mathcal{B}(\mathcal{V})|}$  wie in Gleichung (3.8). Des Weiteren sei  $W \sim \mathbb{R}_p(\Omega, \lambda, \mathcal{V})$  und  $\alpha > 0$ . Dann ist

$$\alpha W \sim \mathbb{R}_p(\alpha\Omega, \lambda, \mathcal{V}).$$

**Beweis**

Da  $P(\mathcal{U})$  ein konvexer Kegel ist, sind sowohl  $W$  als auch  $\alpha W$  fast sicher in  $P(\mathcal{U})$ . Setze  $m := \dim(P(\mathcal{U}))$  und betrachte  $W$  nicht mehr als  $p \times p$ -Matrix, sondern als  $m$ -elementigen Vektor. Der Übersichtlichkeit halber sei dieser Vektor bezeichnet mit  $w$ . Betrachte nun die Bijektion  $g : w \mapsto \alpha w$ . Als Jacobi-Matrix von  $g^{-1}$  an der Stelle  $\alpha w$  ergibt sich  $(Dg^{-1})(\alpha w) = \frac{1}{\alpha} I_m$ . Des Weiteren zeigt eine einfache Rechnung die Identität  $\mathbb{r}_p(\Omega, \lambda, \mathcal{V}; \frac{1}{\alpha} Z) = \alpha^m \mathbb{r}(\alpha \Omega, \lambda, \mathcal{V}; Z)$ . Eine Dichtetransformation der Dichte  $f_W$  von  $W$  mit  $g$  liefert demnach

$$f_{\alpha W}(Z) = f_W\left(\frac{1}{\alpha} Z\right) \det[(Dg^{-1})(\alpha z)] = \mathbb{r}_p(\alpha \Omega, \lambda, \mathcal{V}; Z). \quad \square$$

Das letzte Theorem dieses Abschnitts zeigt eine nützliche Eigenschaft aller verallgemeinerten Riesz-Verteilungen über dem Kegel  $P(\mathcal{U})$ , die einen Formparametervektor mit identischen Einträgen aufweisen.

**Theorem 3.23 (Riesz-Verteilungen mit identischen Formparametern)**

Sei  $\mathcal{U} = (V, \emptyset, U)$  ein zerlegbarer Graph und  $\mathcal{V}_1, \mathcal{V}_2$  zwei Darstellungen von  $\mathcal{U}$  als azyklisch gemischter Graph. Seien  $\lambda_1 = (\beta, \dots, \beta)^T \in \mathbb{R}^{|\mathcal{B}(\mathcal{V}_1)|}$  und  $\lambda_2 = (\beta, \dots, \beta)^T \in \mathbb{R}^{|\mathcal{B}(\mathcal{V}_2)|}$  für ein  $\beta > ([B] + \langle B \rangle - 1)/2$  für alle  $[B] \in \mathcal{B}(\mathcal{V}_1)$  und alle  $[B] \in \mathcal{B}(\mathcal{V}_2)$ . Außerdem sei  $\Omega \in P(\mathcal{U})$ . Dann gilt die folgende Identität von Verteilungen,

$$\mathbb{R}_{|\mathcal{V}_1|}(\Omega, \lambda_1, \mathcal{V}_1) = \mathbb{R}_{|\mathcal{V}_2|}(\Omega, \lambda_2, \mathcal{V}_2).$$

**Beweis**

Der Beweis beruht auf einer Idee aus Bemerkung 8.1 in Andersson und Klein [3]. Definiere  $\Lambda := \Omega^{-\lambda_1} \in PD^0(\mathcal{U})$  wie in Gleichung (3.7) als das  $\lambda_1$ -Inverse von  $\Omega$  bezüglich des Graphen  $\mathcal{V}_1$ . Die Matrix  $\Lambda$  liefert eine Umparametrisierung der verallgemeinerten Riesz-Dichte bezüglich des Lebesgue-Maßes. Die Dichte lässt sich in der Form

$$\mathbb{r}_{|\mathcal{V}_1|}(\Lambda, \lambda_1, \mathcal{V}_1; W) = g_{\mathcal{V}_1}(\lambda_1, W) \left( \prod_{[B] \in \mathcal{B}(\mathcal{V}_1)} \det(\Lambda_{[B]\bullet})^{\lambda_1[B]} \right) e^{-\text{tr}(\Lambda W)}$$

darstellen, wobei die Funktion  $g_{\mathcal{V}_1}$  nicht von  $\Lambda$  abhängt, vergleiche Andersson und Klein [3, Definition 8.1]. Betrachte nun die multivariate momenterzeugende Funktion einer

Riesz-verteiltern Zufallsgröße  $W \sim \mathbb{R}_{|\mathcal{V}|}(\Omega, \lambda_1, \mathcal{V}_1)$ . Diese bestimmt sich zu

$$\begin{aligned} \mathbb{E}\left[e^{\operatorname{tr}(SW)}\right] &= \int_{P(\mathcal{U})} g_{\mathcal{V}_1}(\lambda_1, W) \left( \prod_{[B] \in \mathcal{B}(\mathcal{V}_1)} \det(\Lambda_{[B]\bullet})^{\lambda_1[B]} \right) e^{-\operatorname{tr}[(\Lambda-S)W]} dW \\ &= \prod_{[B] \in \mathcal{B}(\mathcal{V}_1)} \frac{\det(\Lambda_{[B]\bullet})^{\lambda_1[B]}}{\det((\Lambda-S)_{[B]\bullet})^{\lambda_1[B]}} = \frac{\det(\Lambda)^\beta}{\det(\Lambda-S)^\beta} \end{aligned}$$

für alle  $S \in S(\mathcal{U})$  mit  $\Lambda - S \in PD^0(\mathcal{U})$ . Hierbei seien die Cholesky-Parameter von  $\Lambda \in PD^0(\mathcal{U})$  analog zu den Cholesky-Parametern der Matrix  $\Omega \in P(\mathcal{U})$  definiert, vergleiche Gleichung (3.6) und die Bemerkung unter Korollar 5.1 in Andersson und Klein [3]. Die vorletzte Gleichheit folgt dann, da über eine verallgemeinerte Riesz-Dichte zur Parametermatrix  $\Lambda - S$  integriert wird, während sich die letzte Identität aus der Konstanzheit der Einträge des Formparameters  $\lambda_1$  ergibt.

Damit hängt die momenterzeugende Funktion von  $W$  in keiner Weise von der Darstellung  $\mathcal{V}_1$  ab. Genauer gesagt ergibt sich für eine  $\mathbb{R}_{|\mathcal{V}|}(\Lambda, \lambda_2, \mathcal{V}_2)$ -verteilte Zufallsmatrix exakt dieselbe momenterzeugende Funktion. Resultat 5.1 auf Seite 100 in Monahan [21] liefert dann die Identität

$$\mathbb{R}_{|\mathcal{V}|}(\Lambda, \lambda_1, \mathcal{V}_1) = \mathbb{R}_{|\mathcal{V}|}(\Lambda, \lambda_2, \mathcal{V}_2).$$

Um die Behauptung zu zeigen, muss daher nur noch die Reparametrisierung auf  $\Lambda$  rückgängig gemacht werden. Dazu sei an Lemma 3.15(i) auf Seite 46 erinnert. Aus diesem folgt

$$\Lambda = \Omega^{-\lambda_1} = \beta (\pi^{-1}(\Omega))^{-1} = \Omega^{-\lambda_2}$$

und dies impliziert die Behauptung. Es sei noch einmal hervorgehoben, dass das  $\lambda_1$ -Inverse der Matrix  $\Omega$  bezüglich des Graphen  $\mathcal{V}_1$  definiert ist, während das  $\lambda_2$ -Inverse von  $\Omega$  bezüglich  $\mathcal{V}_2$  definiert ist.  $\square$

### 3.5 Verallgemeinerte Wishart-Verteilungen

Sei ein zerlegbarer Graph  $\mathcal{U}$  gegeben. Wie bereits erwähnt, ist im Allgemeinen eine Darstellung  $\mathcal{V}$  von  $\mathcal{U}$  als azyklisch gemischter Graph nicht eindeutig. Andersson und Klein stellten sich in [3, Kapitel 13] daher die Frage, ob es unter all diesen Darstellungen eine kanonische Darstellung  $\mathcal{V}_{\mathcal{U}}$  gibt, die von keinerlei Wahlen abhängt. Sie nennen den Graphen  $\mathcal{V}_{\mathcal{U}}$  die *intrinsische Darstellung* von  $\mathcal{U}$ , zeigen, dass er für jeden zerlegbaren Graphen  $\mathcal{U}$  existiert, und geben eine Konstruktionsvorschrift für ihn an.

### Die intrinsische Darstellung von $\mathcal{U}$

Sei zunächst  $\mathcal{V} = (V, D, U)$  ein beliebiger azyklisch gemischter Graph. Dann ist die Relation

$$\forall \alpha, \beta \in V, \alpha \neq \beta: \quad \alpha <_{\mathcal{V}} \beta \iff \alpha - \beta \in U \quad \text{und} \quad \text{nb}(\beta) \setminus \{\alpha\} \subsetneq \text{nb}(\alpha) \setminus \{\beta\}.$$

eine Halbordnung auf den Knoten von  $\mathcal{V}$ . Ein Knoten  $\alpha$  wird genau dann von einem anderen Knoten  $\beta$  bezüglich  $<_{\mathcal{V}}$  dominiert, wenn zwei Eigenschaften erfüllt sind. Erstens müssen die beiden durch eine ungerichtete Kante verbunden sein, das heißt Knoten aus verschiedenen Boxen von  $\mathcal{V}$  können mit  $<_{\mathcal{V}}$  nicht verglichen werden. Zweitens muss jeder Nachbar von  $\beta$  auch ein Nachbar von  $\alpha$  sein und zusätzlich muss  $\alpha$  mindestens einen Nachbar mehr haben als  $\beta$ .

Mit dieser Halbordnung lässt sich die intrinsische Darstellung von  $\mathcal{U}$  konstruieren, siehe Andersson und Klein [3, Algorithmus 13.1].

#### Algorithmus 3.24 (Konstruktion der intrinsischen Darstellung)

Sei im Folgenden immer  $\mathcal{V}_j = (V, D_j, U_j)$ . Setze zunächst  $j := 0$  und  $\mathcal{V}_0 := \mathcal{U}$ .

- (i) Falls der ungerichtete Teil  $(V, U_j)$  eine disjunkte Vereinigung von vollständigen Graphen ist, setze  $\mathcal{V}_{\mathcal{U}} := \mathcal{V}_j$  und höre auf. Sonst zu Schritt (ii).
- (ii) Für alle Knoten  $\alpha, \beta \in V$  mit  $\alpha <_{\mathcal{V}_j} \beta$ , ersetze die ungerichtete Kante  $\alpha - \beta$  durch die gerichtete Kante  $\alpha \rightarrow \beta$ . Nenne den resultierenden Graphen  $\mathcal{V}'_j$ .
- (iii) Ersetze jede gerichtete Kante  $\alpha \rightarrow \beta$ , die in  $\mathcal{V}'_j$  an einem semi-gerichteten Zykel beteiligt ist, durch die ungerichtete Kante  $\alpha - \beta$ . Nenne den resultierenden Graphen  $\mathcal{V}_{j+1}$ .
- (iv) Erhöhe  $j$  um Eins und gehe zurück zu Schritt (i).

Andersson und Klein zeigen schließlich, dass dieser Algorithmus in einer endlichen Zahl von Schritten abbricht und immer einen azyklisch gemischten Graphen mit vollständigen Boxen und ohne Immoralitäten oder Flaggen zurückliefert, siehe Kapitel 13 und Lemma 14.1 in Andersson und Klein [3]. Folglich ist  $\mathcal{V}_{\mathcal{U}}$  immer eine Darstellung von  $\mathcal{U}$ , da das Gerüst im Laufe des Algorithmus nie verändert wird.

### Die verallgemeinerte Wishart-Verteilung

Aus der Familie verallgemeinerter Riesz-Verteilungen über dem Kegel  $P(\mathcal{U})$  zu einem festen Parameterpaar  $(\Omega, \lambda)$  wählt man nun diejenige aus, die zur intrinsischen Darstellung  $\mathcal{V}_{\mathcal{U}}$  von  $\mathcal{U}$  korrespondiert und nennt sie die verallgemeinerte Wishart-Verteilung.

**Definition 3.25 (Verallgemeinerte Wishart-Verteilung)**

Sei  $\mathcal{U}$  ein zerlegbarer Graph mit  $p$  Knoten und  $\mathcal{V}_{\mathcal{U}}$  seine intrinsische Darstellung. Sei des Weiteren  $\Omega \in P(\mathcal{U})$  und  $\lambda \in \mathbb{R}^{|\mathcal{B}(\mathcal{V}_{\mathcal{U}})|}$  mit  $\lambda_{[B]} > ([B] + \langle B \rangle - 1)/2$  für alle Boxen  $[B] \in \mathcal{B}(\mathcal{V}_{\mathcal{U}})$ . Dann wird die verallgemeinerte Wishart-Verteilung zum Erwartungsparameter  $\Omega$  und Formparameter  $\lambda$  definiert als

$$\mathbb{W}_p(\Omega, \lambda) := \mathbb{R}_p(\Omega, \lambda, \mathcal{V}_{\mathcal{U}}).$$

Da es sich bei der verallgemeinerten Wishart-Verteilung folglich um einen Spezialfall der verallgemeinerten Riesz-Verteilung handelt, gelten Theorem 3.19, Theorem 3.20, Korollar 3.21 und Theorem 3.22 auch für die verallgemeinerte Wishart-Verteilung. Die letzten drei übertragen sich dabei direkt, indem die Bezeichnung „Riesz“ durch „Wishart“ ersetzt wird.

Etwas Vorsicht ist dagegen bei Theorem 3.19 geboten, denn die Randverteilung einer verallgemeinerten Wishart-Verteilung ist selbst nicht unbedingt eine verallgemeinerte Wishart-Verteilung. Der Grund hierfür ist, dass der induzierte Teilgraph der intrinsischen Darstellung  $(\mathcal{V}_{\mathcal{U}})_A$  nicht zwangsweise identisch ist mit der intrinsischen Darstellung  $\mathcal{V}_{(\mathcal{U}_A)}$  des induzierten Subgraphen  $\mathcal{U}_A$ . Betrachte beispielsweise den zerlegbaren Graphen  $\mathcal{U} = 2 - 1 - 3$  und seine intrinsische Darstellung  $\mathcal{V}_{\mathcal{U}} = 2 \leftarrow 1 \rightarrow 3$ . Wähle nun  $A = \{1, 3\}$ . Dann ist die intrinsische Darstellung von  $\mathcal{U}_A = 1 - 3$  gerade  $\mathcal{V}_{(\mathcal{U}_A)} = 1 - 3$ , während der induzierte Teilgraph  $(\mathcal{V}_{\mathcal{U}})_A = 1 \rightarrow 3$  ist.

**Der Zusammenhang mit der klassischen Wishart-Verteilung**

Bislang ungeklärt ist, ob der Name „verallgemeinerte Wishart-Verteilung“ für das Wahrscheinlichkeitsmaß  $\mathbb{W}_p(\Omega, \lambda)$  überhaupt sinnvoll ist. Die Rechtfertigung hierfür liefert das folgende Lemma.

**Lemma 3.26 (Klassische und verallgemeinerte Wishart-Verteilung)**

Sei  $\mathcal{U}$  der vollständige ungerichtete Graph über  $p$  Knoten. Seien des Weiteren  $\Omega, \Sigma \in P(\mathcal{U}) = PD_p$  und  $\lambda, n \in \mathbb{R}^{|\mathcal{B}(\mathcal{V}_{\mathcal{U}})|} = \mathbb{R}$  mit  $\lambda > (p - 1)/2$  und  $n > p - 1$ . Dann gelten für den Zusammenhang zwischen der verallgemeinerten Wishart-Verteilung  $\mathbb{W}_p(\Omega, \lambda)$  und der klassischen zentrierten Wishart-Verteilung  $\mathcal{W}_p(n, \Sigma)$  die beiden Gleichungen

$$\mathbb{W}_p(\Omega, \lambda) = \mathcal{W}_p\left(2\lambda, \frac{1}{2\lambda}\Omega\right) \quad \text{und} \quad \mathcal{W}_p(n, \Sigma) = \mathbb{W}_p\left(n\Sigma, \frac{n}{2}\right).$$

**Beweis**

Da  $\mathcal{U}$  vollständig ist, ist  $P(\mathcal{U}) = PD_p$ . Außerdem ist  $\mathcal{V}_{\mathcal{U}} = \mathcal{U}$  und es gibt folglich nur eine Box. Die beiden Gleichungen ergeben sich dann durch einfaches Einsetzen in die beiden

Dichten aus Gleichung (3.4) von Seite 41 und Definition 3.16 von Seite 47. Beachte, dass selbst für einen vollständigen Graphen  $\mathcal{U}$  die klassische und die verallgemeinerte Wishart-Verteilung unterschiedlich parametrisiert sind, das heißt  $\mathbb{W}_p(\Omega, \lambda) \neq \mathcal{W}_p(\lambda, \Omega)$ . Dies liegt daran, dass Andersson und Klein [3] eine andere Parametrisierung der klassischen Wishart-Verteilung verwenden als beispielsweise Kollo und von Rosen [17] und Lauritzen [18].  $\square$

Eine klassische Wishart-Verteilung kann somit auch als eine verallgemeinerte Wishart-Verteilung über einem vollständigen Graphen  $\mathcal{U}$  interpretiert werden.

## 4 Maximum-Likelihood-Schätzer der Kovarianzmatrix

Sei im folgenden Kapitel wieder  $\mathcal{V} = (V, D, U)$  ein Annahme 2.6 erfüllender azyklisch gemischter Graph und  $\mathcal{U} := \mathcal{V}^\sim$  sein Gerüst. Seien des Weiteren  $X_1, \dots, X_n$  unabhängig und identisch verteilte Zufallsvektoren, die einer zentrierten Normalverteilung mit unbekannter Streuungsmatrix  $\Sigma \in PD(\mathcal{U})$  folgen. Ein beliebtes Verfahren zur Bestimmung eines Schätzers  $\hat{\Sigma}$  der Kovarianzstruktur ist die Maximum-Likelihood-Methode. Dabei bestimmt man diejenige positiv definite Matrix  $\hat{\Sigma}$ , für die die Log-Likelihood-Funktion

$$\begin{aligned} \ell_S(\Sigma) &= \sum_{i=1}^n \log \left[ (2\pi)^{-\frac{p}{2}} \det(\Sigma)^{-\frac{1}{2}} \exp \left( -\frac{1}{2} \text{tr}(\Sigma^{-1} x_i x_i^T) \right) \right] \\ &\propto -n \log \det(\Sigma) - n \text{tr}(\Sigma^{-1} S) \end{aligned} \quad (4.1)$$

bei gegebener Stichproben-Kovarianzmatrix  $S = \frac{1}{n} \sum_{i=1}^n x_i x_i^T \in \mathbb{R}^{p \times p}$  ihr Maximum annimmt. Es muss allerdings darauf geachtet werden, dass die bestimmte Matrix  $\hat{\Sigma}$  auch eine zulässige Streuungsmatrix darstellt. Mit anderen Worten wird bei der Bestimmung des Maximum-Likelihood-Schätzers im von  $\mathcal{U}$  vorgegebenen graphischen Modell die Funktion (4.1) über dem Grundraum  $PD(\mathcal{U})$  maximiert, siehe Kapitel 4.1. Dieser Abschnitt enthält insbesondere Ergebnisse aus Lauritzen [18] und Andersson und Klein [3]. Werden zusätzlich auch noch Invarianzbedingungen einer Gruppe  $G$  in das Modell eingebaut, so muss die Funktion (4.1) über dem kleineren Grundraum  $PD(\mathcal{U}, G)$  maximiert werden, siehe Kapitel 4.2, welches größtenteils an Madsen [20] angelehnt ist.

### 4.1 Graphische Modelle ohne Invarianzrestriktionen

Seien also  $X_1, \dots, X_n$  unabhängig und identisch verteilt gemäß einer zentrierten Normalverteilung mit Streuungsmatrix  $\Sigma \in PD(\mathcal{U})$ . Es sei nochmal hervorgehoben, dass die Verteilung ebenso über den Parameter  $\Omega = \pi(\Sigma) \in P(\mathcal{U})$  parametrisiert werden kann, vergleiche Lemma 2.4 auf Seite 13. Das folgende Theorem ist ein bekanntes Ergebnis aus der Theorie der graphischen Normalverteilungsmodelle.

#### Theorem 4.1 (ML-Schätzer in graphischen Normalverteilungsmodellen)

Sei  $\mathcal{U}$  ein zerlegbarer Graph mit  $p$  Knoten, sowie  $\Sigma \in PD(\mathcal{U})$  und  $\Omega = \pi(\Sigma) \in P(\mathcal{U})$ . Außerdem seien  $X_1, \dots, X_n \sim \mathcal{N}_p(0, \Sigma)$  unabhängig und  $S := \frac{1}{n} \sum_{i=1}^n X_i X_i^T$  die Stichproben-Kovarianzmatrix. Die Maximum-Likelihood-Schätzer  $\hat{\Omega}$  und  $\hat{\Sigma}$  existieren mit Wahrscheinlichkeit Eins genau dann, wenn  $n \geq |C|$  für alle Cliques  $C$  von  $\mathcal{U}$  ist. In diesem Fall sind beide Schätzer eindeutig und lassen sich beschreiben durch

$$\hat{\Omega} = \pi(S) \in P(\mathcal{U}) \quad \text{und} \quad \hat{\Sigma} = \pi^{-1}(\hat{\Omega}) \in PD(\mathcal{U}).$$

**Beweis**

Die Existenz folgt aus Proposition 5.9 in Lauritzen [18], da  $\mathcal{U}$  zerlegbar ist. Sei des Weiteren  $C$  eine Clique von  $\mathcal{U}$ . Dann besagen Theorem 5.3 und Gleichung (5.15) aus Lauritzen [18], dass für die zu  $C$  korrespondierenden Teilmatrizen

$$\hat{\Sigma}_{C,C} = \hat{\Omega}_{C,C} = S_{C,C}$$

gelten muss. Da die verbleibenden Einträge von  $\hat{\Omega}$  verschwinden, siehe Gleichung (2.1) auf Seite 12, ist  $\hat{\Omega} = \pi(S)$ . Die letzte Behauptung folgt dann direkt aus Lemma 2.4 von Seite 13.  $\square$

Man bekommt den Schätzer für  $\Omega$  folglich durch Streichen von Einträgen in der Stichproben-Kovarianzmatrix  $S$ . Der Schätzer für  $\Sigma$  wiederum ergibt sich dann durch Auffüllen von Einträgen in  $\hat{\Omega}$  gemäß der Abbildung  $\pi^{-1}$ . Es sei hervorgehoben, dass zwar bereits gezeigt wurde, dass  $\pi$  ein Diffeomorphismus ist, allerdings nur als Abbildung  $\pi : PD(\mathcal{U}) \rightarrow P(\mathcal{U})$ . Da im Allgemeinen  $S$  nicht in  $PD(\mathcal{U})$  liegt, sind meist  $\pi^{-1}(\pi(S))$  und  $S$  nicht identisch.

Obwohl der Schätzer  $\hat{\Omega}$  schon seit mehreren Jahrzehnten bekannt ist, konnte seine exakte Verteilung erst 2010 in Andersson und Klein [3] durch eine Lebesgue-Dichte beschrieben werden. Es handelt sich um eine verallgemeinerte Wishart-Verteilung, wie das folgende Theorem zeigt.

**Theorem 4.2 (Verteilung des ML-Schätzers in graphischen Modellen)**

*Seien die Voraussetzungen wie in Theorem 4.1. Falls der Maximum-Likelihood-Schätzer  $\hat{\Omega} \in P(\mathcal{U})$  existiert und eindeutig ist, so folgt er einer verallgemeinerten Wishart-Verteilung über dem Graphen  $\mathcal{U}$  mit Erwartungsparameter  $\Omega \in P(\mathcal{U})$  und Formparameter  $\lambda = \left(\frac{n}{2}, \dots, \frac{n}{2}\right)^T \in \mathbb{R}^{|\mathcal{B}(\mathcal{V}_{\mathcal{U}})|}$ ,*

$$\hat{\Omega} = \pi(S) \sim \mathbb{W}_p(\Omega, \lambda).$$

**Beweis**

Siehe Andersson und Klein [3, Beispiel 17.1].  $\square$

Es sei daran erinnert, dass die verallgemeinerte Wishart-Verteilung  $\mathbb{W}_p(\Omega, \lambda)$  definiert ist als die verallgemeinerte Riesz-Verteilung zu den Parametern  $\Omega$  und  $\lambda$ , bezüglich der intrinsischen Darstellung  $\mathcal{V}_{\mathcal{U}}$ , siehe Definition 3.25 auf Seite 54. Da der hier auftretende Formparameter  $\lambda$  konstante Einträge hat, besagt Theorem 3.23 von Seite 51 allerdings, dass die Wahl der Darstellung von  $\mathcal{U}$  keinen Einfluss auf die Verteilung hat. Mit anderen Worten kann die Verteilung  $\mathbb{W}_p(\Omega, \lambda)$  aus Theorem 4.2 ersetzt werden durch

$\mathbb{R}_p(\Omega, \lambda', \mathcal{V})$ , wobei  $\mathcal{V}$  eine beliebige Darstellung von  $\mathcal{U}$  als azyklisch gemischter Graph und  $\lambda' = (\frac{n}{2}, \dots, \frac{n}{2})^T \in \mathbb{R}^{|\mathcal{B}(\mathcal{V})|}$  ist.

Das nächste Kapitel widmet sich der Frage, wie sich die Maximum-Likelihood-Schätzer ändern, wenn Invarianzbedingungen in das graphische Modell eingebaut werden.

## 4.2 Graphische Modelle mit Invarianzrestriktionen

Sei neben  $\mathcal{V}$  und  $\mathcal{U}$  eine endliche Gruppe  $G \subseteq Orth(\mathcal{V})$  gegeben. Die implizit von der Gruppe  $G$  abhängende Funktion

$$\psi : PD_p \rightarrow PD_p, \quad S \mapsto \frac{1}{|G|} \sum_{\rho \in G} \rho S \rho^T \quad (4.2)$$

wird eine wichtige Rolle bei der Maximum-Likelihood-Schätzung spielen.<sup>4</sup> Eine wichtige Eigenschaft der Funktion  $\psi$  ist, dass sie eine beliebige positiv definite Matrix  $S \in PD_p$  in eine bezüglich  $G$  invariante Matrix  $\psi(S)$  verwandelt, da

$$(\rho') \psi(S) (\rho')^T = \frac{1}{|G|} \sum_{\rho \in G} (\rho \rho') S (\rho \rho')^T = \frac{1}{|G|} \sum_{\rho \in G} \rho S \rho^T = \psi(S) \quad \text{für alle } \rho' \in G$$

direkt aus der Gruppeneigenschaft von  $G$  folgt. Deswegen wird  $\psi$  als *Glättungs-Funktion* bezeichnet, da sie eine beliebige Matrix  $S \in PD_p$  zu einer bezüglich  $G$  invarianten Matrix  $\psi(S)$  „glättet“. Des Weiteren folgt aus Lemma 2.19(iv) auf Seite 27, dass  $\psi$  die vorgegebenen Nullblöcke einer Matrix  $W \in P(\mathcal{U})$  erhält, das heißt  $\psi(W)$  liegt sogar in  $P(\mathcal{U}, G)$ , siehe auch Madsen [20, Proposition 6.1]. Besonders von Interesse wird daher die Einschränkung  $\psi : P(\mathcal{U}) \rightarrow P(\mathcal{U}, G)$  sein.

Seien nun  $X_1, \dots, X_n$  unabhängig und identisch verteilt gemäß einer zentrierten Normalverteilung mit Streuungsmatrix  $\Sigma \in PD(\mathcal{U}, G)$ . Außerdem sei  $\Omega = \pi(\Sigma) \in P(\mathcal{U}, G)$  und  $W := \pi(S)$  die  $p \times p$ -Matrix, die sich durch Streichen der passenden Einträge in der Stichproben-Kovarianzmatrix  $S$  ergibt. Dann liefert das folgende Theorem die Maximum-Likelihood-Schätzer von  $\Omega$  und  $\Sigma$ .

### Theorem 4.3 (ML-Schätzer in invarianten graphischen Modellen)

*Sei  $\mathcal{V}$  ein Annahme 2.6 erfüllender azyklisch gemischter Graph mit Gerüst  $\mathcal{U} = \mathcal{V}^\sim$ . Außerdem sei  $G \subseteq Orth(\mathcal{V})$  eine endliche Gruppe und  $W = \pi(S)$  sei wie oben beschrieben definiert. Die Maximum-Likelihood-Schätzer  $\hat{\Omega}$  und  $\hat{\Sigma}$  der Parametermatrizen  $\Omega \in P(\mathcal{U}, G)$*

---

<sup>4</sup>Es sei hervorgehoben, dass statt einer endlichen Gruppe auch eine kompakte Gruppe  $G \subseteq Orth(\mathcal{V})$  verwendet werden könnte. In diesem Fall wäre die Glättungsfunktion  $\psi$  von der Form  $\psi(S) = \int \rho S \rho^T dH(S)$ , wobei  $H$  das eindeutige normalisierte Haar-Maß auf  $G$  meint, vergleiche Andersson und Madsen [4, Seite 558].

und  $\Sigma \in PD(\mathcal{U}, G)$  existieren und sind eindeutig, falls für alle Boxen  $[B] \in \mathcal{B}(\mathcal{V})$  die Teilmatrix  $\psi(W)_{[B] \cup \langle B \rangle}$  positiv definit ist. In diesem Fall sind die eindeutigen Schätzer gegeben durch

$$\hat{\Omega} = \psi(W) \quad \text{und} \quad \hat{\Sigma} = \pi^{-1}(\psi(W)).$$

**Beweis**

Theorem 7.1, Korollar 7.2 und Bemerkung 7.6 aus Madsen [20] liefern die Existenz der Maximum-Likelihood-Schätzer, falls  $\psi(S)_{[B] \cup \langle B \rangle}$  für alle Boxen  $[B] \in \mathcal{B}(\mathcal{V})$  positiv definit ist. Wegen Lemma 2.24 auf Seite 34 vertauschen  $\pi$  und  $\psi$  und wegen Annahme 2.6 ist  $[B] \cup \langle B \rangle$  vollständig in  $\mathcal{V}$ . Somit ist

$$\psi(W)_{[B] \cup \langle B \rangle} = \psi[\pi(S)]_{[B] \cup \langle B \rangle} = \pi[\psi(S)]_{[B] \cup \langle B \rangle} = \psi(S)_{[B] \cup \langle B \rangle}$$

für alle Boxen  $[B]$ . Im Falle der Existenz bestimmen sich die Maximum-Likelihood-Schätzer der Cholesky-Parameter von  $\Omega = \pi(\Sigma)$  für alle Boxen  $[B] \in \mathcal{B}(\mathcal{V})$  zu

$$\hat{\Omega}_{[B] \bullet} = \psi(S)_{[B] \bullet} = \psi(W)_{[B] \bullet} \quad \text{und} \quad \hat{\Omega}_{[B] \bullet} = \psi(S)_{[B] \bullet} = \psi(W)_{[B] \bullet},$$

siehe Madsen [20, Korollar 7.2]. Folglich haben  $\hat{\Omega}$  und  $\psi(W)$  identische Cholesky-Parameter. Außerdem wurde weiter oben schon gezeigt, dass  $\psi(W)$  in  $P(\mathcal{U})$  liegt. Wegen Theorem 3.12 auf Seite 44 muss daher  $\hat{\Omega} = \psi(W)$  sein. Die letzte Behauptung  $\hat{\Sigma} = \pi^{-1}(\psi(W))$  ist nun eine direkte Konsequenz. □

**Bemerkung 4.4 (Hinreichende Stichprobengröße)**

Es sei hervorgehoben, dass Theorem 4.3 nur ein hinreichendes Kriterium für die Existenz und Eindeutigkeit des Maximum-Likelihood-Schätzers liefert. Wie im Einzelfall anhand der Anzahl  $n$  der Beobachtungen entschieden werden kann, ob der Schätzer fast sicher existiert, wird in Madsen [20, Seite 1173ff] anhand von sieben Beispielen erläutert. Allgemeine Aussagen zur kleinsten Stichprobengröße  $n$ , die die Existenz des Maximum-Likelihood-Schätzers mit Wahrscheinlichkeit Eins sicher stellt, sind bisher nicht bekannt, vergleiche Bemerkung 7.7 in Madsen [20]. Diese Fragestellung soll in der vorliegenden Arbeit jedoch nicht weiter behandelt werden. Stattdessen widmet sich das nächste Kapitel der Frage nach der Verteilung von  $\hat{\Omega}$ .

Es sei aber abschließend hervorgehoben, dass die Forderung

$$n \geq [B] + \langle B \rangle \quad \text{für alle} \quad [B] \in \mathcal{B}(\mathcal{V})$$

hinreichend ist für die fast sichere Existenz des Maximum-Likelihood-Schätzers  $\hat{\Omega}$  im invarianten graphischen Modell  $P(\mathcal{U}, G)$ . Gleichung (2.3) auf Seite 17 zeigt nämlich, dass sich alle Cliques von  $\mathcal{U}$  in der Form  $[B] \cup \langle B \rangle$  darstellen lassen. Folglich erfüllt solch ein Stichprobenumfang  $n$  die Existenzvoraussetzung des Maximum-Likelihood-Schätzers im graphischen Modell  $P(\mathcal{U})$ , siehe Theorem 4.1. Mit anderen Worten ist in diesem Fall die Likelihood-Funktion (4.1) über  $P(\mathcal{U})$  fast sicher von oben beschränkt. Da aber  $P(\mathcal{U}, G)$  ein Teilmodell von  $P(\mathcal{U})$  ist, ist die Likelihood-Funktion auch über  $P(\mathcal{U}, G)$  von oben beschränkt und der Maximum-Likelihood-Schätzer im invarianten graphischen Modell existiert mit Wahrscheinlichkeit Eins.

## 5 Verteilung der Maximum-Likelihood-Schätzer in invarianten graphischen Modellen

Sei wieder  $\mathcal{V} = (V, D, U)$  ein Annahme 2.6 erfüllender azyklisch gemischter Graph und  $\mathcal{U} := \mathcal{V}^\sim$  sein Gerüst, sowie  $G \subseteq \text{Orth}(\mathcal{V})$  eine endliche Gruppe. Außerdem seien die Daten  $X_1, \dots, X_n$  unabhängig und identisch  $\mathcal{N}_p(0, \pi^{-1}(\Omega))$ -verteilt für ein  $\Omega \in P(\mathcal{U}, G)$  und  $S = \frac{1}{n} \sum_{i=1}^n X_i X_i^T$  sei die korrespondierende Stichproben-Kovarianzmatrix. Das Ziel des folgenden Abschnitts ist, die Verteilung des Maximum-Likelihood-Schätzers des Kovarianzparameters  $\hat{\Omega}$  im invarianten graphischen Modell  $P(\mathcal{U}, G)$  zu bestimmen. Um dies zu bewerkstelligen, wird zunächst der Schätzer  $W = \pi(S)$  des rein graphischen Modells  $P(\mathcal{U})$  betrachtet, siehe Theorem 4.1. Von ihm ist bekannt, dass er einer verallgemeinerten Riesz-Verteilung folgt, siehe Theorem 4.2. Des Weiteren bekommt man  $\hat{\Omega} \in P(\mathcal{U}, G)$  durch Anwendung der Glättungsfunktion  $\psi$  auf  $W$ , siehe Theorem 4.3. Die Grundidee dieses Kapitels ist es daher, die bekannte Verteilung von  $W$  zu verwenden, um die Verteilung von  $\psi(W)$  zu bestimmen.

Um diese Technik anwenden zu können, muss zunächst die Existenz der Maximum-Likelihood-Schätzer sichergestellt werden. Für den Stichprobenumfang  $n$  gelte daher ab sofort  $n \geq [B] + \langle B \rangle$  für alle Boxen  $[B] \in \mathcal{B}(\mathcal{V})$ . Dadurch existieren sowohl der Schätzer des graphischen Modells  $W \in P(\mathcal{U})$  als auch der Schätzer des invarianten graphischen Modells  $\psi(W) \in P(\mathcal{U}, G)$ , siehe Bemerkung 4.4. Außerdem folgt  $W$  einer  $\mathbb{R}_p(\Omega, \lambda, \mathcal{V})$ -Verteilung mit  $\lambda = (\frac{n}{2}, \dots, \frac{n}{2})^T$ , siehe Theorem 4.2 und die Bemerkung darunter.

Um die Notation im Folgenden zu vereinfachen, sei ab jetzt  $W$  eine beliebige  $P(\mathcal{U})$ -wertige Zufallsmatrix, die einer  $\mathbb{R}_p(\Omega, \lambda, \mathcal{V})$ -Verteilung mit Parametern  $\Omega \in P(\mathcal{U}, G)$  und  $\lambda \in \mathbb{R}^{|\mathcal{B}(\mathcal{V})|}$  folgt. Dabei habe der Formparameter  $\lambda$  identische Einträge, das heißt

$$\lambda = (\beta, \dots, \beta)^T \quad \text{für ein } \beta > \max_{[B] \in \mathcal{B}(\mathcal{V})} \frac{[B] + \langle B \rangle - 1}{2}. \quad (5.1)$$

Nun soll die Verteilung von  $\psi(W)$  für diese allgemeine Zufallsmatrix  $W$  bestimmt werden. Da die Matrix  $\psi(W) \in P(\mathcal{U}, G)$  schon eindeutig durch ihre Cholesky-Parameter festgelegt wird, siehe Theorem 3.12 auf Seite 44, werden zunächst in Kapitel 5.1 deren Verteilungen bestimmt. Anschließend wird in Abschnitt 5.2 erläutert, wie die gefundenen Verteilungen zu einer gemeinsamen verallgemeinerten Riesz-Verteilung für  $\psi(W)$  zusammengesetzt werden können. Insbesondere wird darauf eingegangen, wie diese Verteilung aussieht, wenn  $W$  die Projektion der Stichproben-Kovarianzmatrix  $S$  ist,  $W = \pi(S)$ . Im letzten Unterkapitel 5.3 werden die gefundenen Ergebnisse anhand eines Beispiels veranschaulicht.

## 5.1 Verteilungen der Cholesky-Parameter

Bevor die Verteilung der Cholesky-Parameter von  $\Omega$  bestimmt werden kann, muss eine weitere Bedingung an die Gruppe  $G$  gestellt werden.

### Annahme 5.1 (Keine boxeninternen Symmetrien)

Sei  $\mathcal{V}$  ein Annahme 2.6 erfüllender azyklisch gemischter Graph und  $G \subseteq Orth(\mathcal{V})$  eine endliche Gruppe. Dann enthält  $G$  *keine boxeninternen Symmetrien* bezüglich  $\mathcal{V}$ , falls  $\rho_{[\tau_\rho(B),B]} = I_{[B]}$  für alle Boxen  $[B] \in \mathcal{B}(\mathcal{V})$  und alle Gruppenelemente  $\rho \in G$  ist.

Beachte, dass wegen Lemma 2.19(i) von Seite 27 jedes Gruppenelement  $\rho \in G$  schon eindeutig durch die Blöcke der Form  $\rho_{[\tau_\rho(B),B]}$  bestimmt ist. Folglich besagt Annahme 5.1, dass zwar ganze Boxen des Graphen vertauscht werden dürfen, solange dessen kausale Struktur erhalten bleibt. Knoten innerhalb einer Box dürfen aber weder permutiert noch anderweitig orthogonal transformiert werden. Trotz dieser Einschränkung handelt es sich bei invarianten graphischen Modellen ohne boxeninterne Symmetrien um eine interessante Modellklasse, wie später noch näher erläutert wird, siehe Bemerkung 5.17. Dort wird auch erklärt, warum diese Annahme getroffen wird.

### Beispiel 5.2 (Implizite boxeninterne Symmetrien)

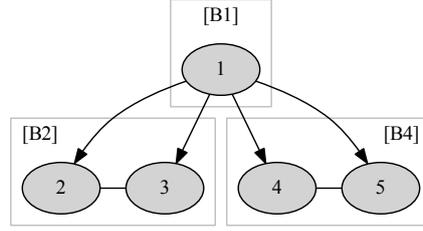
Betrachte den Graphen  $\mathcal{V}$  aus Abbildung 4 zusammen mit der Gruppe  $G = \{I_5, \rho\}$  mit  $\rho = \varrho_5((15432))$ . Insbesondere ist  $G \subseteq Orth(\mathcal{V})$ . Anschaulich werden durch  $\rho$  die Boxen  $[B_2]$  und  $[B_4]$  vertauscht, was selbst unter Annahme 5.1 kein Problem sein sollte. Nun ist aber

$$\rho_{[\tau_\rho(B_2),B_2]} = \rho_{[B_4,B_2]} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \neq I_2.$$

Damit wird die Gruppe  $G$  durch Annahme 5.1 verboten, obwohl nur zwei Boxen vertauscht werden. Diese Beobachtung lässt die Forderung „keine boxeninternen Symmetrien“ zunächst etwas willkürlich erscheinen.

Das beobachtete Verhalten lässt sich jedoch allein durch die ungeschickte Nummerierung der Knoten erklären. Ersetzt man  $\rho$  durch  $\rho' = \varrho((14523))$ , so hat die resultierende Gruppe  $G'$  keine boxeninternen Symmetrien. Trotzdem leistet sie im Grunde weiterhin dasselbe wie die ursprüngliche Gruppe  $G$ , nämlich die Vertauschung der Boxen  $[B_2]$  und  $[B_4]$ . Man kann „implizite boxeninterne Symmetrien“ wie  $\rho$  folglich durch einfaches Umm Nummerieren der Knoten vermeiden.

Um die Verteilung von  $\psi(W)_{[B]\bullet}$  und  $\psi(W)_{[B]\bullet}$  für jede Box  $[B] \in \mathcal{B}(\mathcal{V})$  zu ermitteln, sollen zunächst explizite Formeln für die zu einer Box  $[B]$  und deren Eltern  $\langle B \rangle$  gehörenden Teilmatrizen von  $\psi(W)$  bestimmt werden.



**Abbildung 4:** Ein azyklisch gemischter Graph  $\mathcal{V}$ , für den die Permutation  $\sigma = (15432)$  eine implizite boxeninterne Symmetrie darstellt. Durch  $\sigma$  werden die Knoten 2 und 3 innerhalb der Box  $[B_2]$  und die Knoten 4 und 5 innerhalb der Box  $[B_4]$  vertauscht.

### Lemma 5.3 (Teilmatrizen von $\psi(W)$ )

Sei  $\mathcal{V}$  ein Annahme 2.6 erfüllender azyklisch gemischter Graph mit Gerüst  $\mathcal{U} = \mathcal{V}^\sim$ . Außerdem habe die endliche Gruppe  $G \subseteq \text{Orth}(\mathcal{V})$  keine boxeninternen Symmetrien. Für eine Matrix  $W \in P(\mathcal{U})$  und eine Box  $[B] \in \mathcal{B}(\mathcal{V})$  ist dann

$$\begin{aligned}\psi(W)_{[B]} &= \frac{1}{|\text{Orb}[B]|} \sum_{[A] \in \text{Orb}[B]} \left( W_{[A]\bullet} + W_{[A]\bullet} W_{\langle A \rangle} W_{[A]\bullet}^T \right), \\ \psi(W)_{[B]} &= \frac{1}{|\text{Orb}[B]|} \sum_{[A] \in \text{Orb}[B]} W_{[A]\bullet} W_{\langle A \rangle}, \\ \psi(W)_{\langle B \rangle} &= \frac{1}{|\text{Orb}[B]|} \sum_{[A] \in \text{Orb}[B]} W_{\langle A \rangle}.\end{aligned}$$

### Beweis

Wegen Annahme 5.1, Definition 2.13 von Seite 22 und Korollar 2.23 von Seite 32 ist

$$\psi(W)_{[B]} = \frac{1}{|G|} \sum_{\rho \in G} (\rho W \rho^T)_{[B]} = \frac{1}{|G|} \sum_{\rho \in G} W_{[\tau_\rho^{-1}(B)]} = \frac{1}{|G|} \sum_{[A] \in \text{Orb}[B]} \kappa_{[B]} W_{[A]}.$$

Des Weiteren folgt aus Annahme 5.1 und Lemma 2.21(ii) von Seite 30, dass  $\rho_{(\tau_\rho(B), B)}$  ebenfalls die Identitätsmatrix ist. Korollar 2.23 liefert damit

$$\begin{aligned}\psi(W)_{[B]} &= \frac{1}{|G|} \sum_{\rho \in G} W_{[\tau_\rho^{-1}(B)]} = \frac{\kappa_{[B]}}{|G|} \sum_{[A] \in \text{Orb}[B]} W_{[A]} \text{ und} \\ \psi(W)_{\langle B \rangle} &= \frac{1}{|G|} \sum_{\rho \in G} W_{\langle \tau_\rho^{-1}(B) \rangle} = \frac{\kappa_{[B]}}{|G|} \sum_{[A] \in \text{Orb}[B]} W_{\langle A \rangle}.\end{aligned}$$

Die Beziehung  $|G| = \kappa_{[B]} \cdot |\text{Orb}[B]|$  aus Korollar 2.14 von Seite 22 und die Definition der Cholesky-Parameter aus Gleichung (3.6) von Seite 44 liefern dann die Behauptung.  $\square$

Sei  $[M] \in \mathcal{B}(\mathcal{V})$  nun eine maximale Box in  $\mathcal{V}$ . Definiere dann

$$V_{\text{Orb}} \equiv V_{\text{Orb}}([M]) := V \setminus \left\{ \bigcup_{[A] \in \text{Orb}[M]} [A] \right\} \quad (5.2)$$

als die Menge aller Knoten, die nicht in einer Box im Orbit von  $[M]$  liegen. Besonders von Interesse wird später die zu diesen Knoten korrespondierende Teilmatrix  $W_{V_{\text{Orb}}}$  sein. Beachte insbesondere, dass für jede Orbitbox  $[A] \in \text{Orb}[M]$  die Matrix  $W_{[A]}$  eine Teilmatrix von  $W_{V_{\text{Orb}}}$  ist. Des Weiteren lässt sich die Matrix  $\psi(W)_{V_{\text{Orb}}}$  allein aus der Teilmatrix  $W_{V_{\text{Orb}}}$  bestimmen, das heißt

$$\psi(W)_{V_{\text{Orb}}} = f(W_{V_{\text{Orb}}}) \text{ für eine Funktion } f. \quad (5.3)$$

Um dies einzusehen, seien  $[A], [B] \subseteq V_{\text{Orb}}$  zwei Boxen. Wegen Lemma 2.19(iii) hängt  $(\rho W \rho^T)_{[A,B]}$  für alle  $\rho \in \text{Orth}(\mathcal{V})$  nur von  $W_{[\tau_\rho^{-1}(A), \tau_\rho^{-1}(B)]}$  ab. Da  $[A]$  und  $[B]$  beide außerhalb des Orbits  $\text{Orb}[M]$  liegen, sind auch  $[\tau_\rho^{-1}(A)]$  und  $[\tau_\rho^{-1}(B)]$  nicht in  $\text{Orb}[M]$ . Folglich ist  $W_{[\tau_\rho^{-1}(A), \tau_\rho^{-1}(B)]}$  eine Teilmatrix von  $W_{V_{\text{Orb}}}$  und  $\psi(W)_{[A,B]}$  hängt von  $W$  nur durch  $W_{V_{\text{Orb}}}$  ab.

**Lemma 5.4 (Unabhängigkeiten im Orbit einer maximalen Box)**

Sei ein Annahme 2.6 erfüllender azyklisch gemischter Graph  $\mathcal{V}$  mit Gerüst  $\mathcal{U} = \mathcal{V}^\sim$  und eine endliche Gruppe  $G \subseteq \text{Orth}(\mathcal{V})$  gegeben. Des Weiteren sei  $[M] \in \mathcal{B}(\mathcal{V})$  eine maximale Box von  $\mathcal{V}$  und  $\text{Orb}[M] = \{[A_1], \dots, [A_m]\}$  deren Orbit. Schließlich sei  $W$  eine verallgemeinert Riesz-verteilte Zufallsgröße zur Darstellung  $\mathcal{V}$  mit Parametern  $\Omega \in P(\mathcal{U}, G)$  und  $\lambda \in \mathbb{R}^k$  wie in Gleichung (3.8). Dann gelten für  $i, j \in \{1, \dots, m\}$  die Unabhängigkeiten

$$W_{[A_i]^\bullet} \perp W_{[A_j]^\bullet} \text{ für } i \neq j \quad \text{und} \quad W_{[A_i]^\bullet} \perp W_{[A_j]^\bullet} \text{ für alle } i, j.$$

Des Weiteren halten die bedingten Unabhängigkeiten

$$W_{[A_i]^\bullet} \perp W_{[A_j]^\bullet} \mid W_{V_{\text{Orb}}} \quad \text{für alle } i \neq j.$$

Insgesamt sind also die Zufallsmatrizen

$$W_{[A_1]^\bullet}, W_{[A_1]^\bullet}, \dots, W_{[A_m]^\bullet}, W_{[A_m]^\bullet}$$

alle unabhängig voneinander gegeben die Teilmatrix  $W_{V_{\text{Orb}}}$ .

**Beweis**

**I.** Seien  $i, j \in \{1, \dots, m\}$ . Wegen dem Hauptsatz zur verallgemeinerten Riesz-Verteilung, siehe Theorem 3.17 auf Seite 48, ist  $W_{[A_i]^\bullet}$  unabhängig von  $W_{[A_i]}$  und  $W_{V \setminus [A_i]}$ . Für  $j \neq i$  sind  $[A_j] \cap [A_i] = \emptyset$  und  $\langle A_j \rangle \cap [A_i] = \emptyset$ . Damit lassen sich  $W_{[A_j]^\bullet}$  und  $W_{[A_j]}$  beide aus der Matrix  $W_{V \setminus [A_i]}$  bestimmen, da  $W_{[A_j]}$ ,  $W_{[A_j]}$  und  $W_{\langle A_j \rangle}$  jeweils Teilmatrizen davon sind. Dies zeigt die erste Zeile der Behauptung.

**II.** Setze jetzt  $R_i := V \setminus ([A_i] \cup \langle A_i \rangle)$ . Wegen Korollar 3.18 von Seite 48 ist dann

$$\{W_{R_i}, W_{R_i, \langle A_i \rangle}\} \perp W_{[A_i]^\bullet} \mid W_{\langle A_i \rangle}. \quad (5.4)$$

Nun müssen zwei Fälle unterschieden werden, zunächst sei  $\langle A_i \rangle = \langle A_j \rangle$ . Definiere dann  $R_{ij} := R_i \setminus [A_j]$ . Dann ist

$$W_{R_i} = \begin{pmatrix} W_{R_{ij}} & W_{R_{ij}, [A_j]} \\ & W_{[A_j]} \end{pmatrix} \quad \text{und} \quad W_{R_i, \langle A_i \rangle} = \begin{pmatrix} W_{R_{ij}, \langle A_i \rangle} \\ W_{[A_j], \langle A_i \rangle} \end{pmatrix}.$$

Aus Gleichung (5.4) folgt wegen  $\langle A_i \rangle = \langle A_j \rangle$  also

$$\{W_{R_{ij}}, W_{R_{ij}, [A_j]}, W_{[A_j]}, W_{R_{ij}, \langle A_j \rangle}, W_{[A_j]}\} \perp W_{[A_i]^\bullet} \mid W_{\langle A_j \rangle}.$$

Beachte schließlich, dass  $V_{Orb} \subseteq R_{ij} \cup \langle A_j \rangle$  ist und sich  $W_{V_{Orb}}$  somit aus Teilen von  $W_{R_{ij}}$  und  $W_{R_{ij}, \langle A_j \rangle}$ , sowie  $W_{\langle A_j \rangle}$  zusammensetzt. Wegen den elementaren Eigenschaften der bedingten Unabhängigkeit, siehe Lemma 2.5 auf Seite 14, folgt damit

$$W_{[A_j]} \perp W_{[A_i]^\bullet} \mid W_{V_{Orb}} \quad \text{und} \quad W_{[A_j]^\bullet} \perp W_{[A_i]^\bullet} \mid W_{V_{Orb}}.$$

Im Fall  $\langle A_j \rangle \neq \langle A_i \rangle$  wird dieselbe Beweistechnik verwendet. Setze zunächst also  $R_{ij} := R_i \setminus ([A_j] \cup \langle A_j \rangle)$ , dann ist

$$W_{R_i} = \begin{pmatrix} W_{R_{ij}} & W_{R_{ij}, \langle A_j \rangle} & W_{R_{ij}, [A_j]} \\ & W_{\langle A_j \rangle} & W_{\langle A_j \rangle, [A_j]} \\ & & W_{[A_j]} \end{pmatrix} \quad \text{und} \quad W_{R_i, \langle A_i \rangle} = \begin{pmatrix} W_{R_{ij}, \langle A_i \rangle} \\ W_{\langle A_j \rangle, \langle A_i \rangle} \\ W_{[A_j], \langle A_i \rangle} \end{pmatrix}.$$

Aus Gleichung (5.4) folgt also die bedingte Unabhängigkeit von

$$\{W_{R_{ij}}, W_{R_{ij}, \langle A_j \rangle}, W_{R_{ij}, [A_j]}, W_{\langle A_j \rangle}, W_{[A_j]}, W_{[A_j]}, W_{R_{ij}, \langle A_i \rangle}, W_{\langle A_j \rangle, \langle A_i \rangle}, W_{[A_j], \langle A_i \rangle}\}$$

und  $W_{[A_i]^\bullet}$  gegeben  $W_{\langle A_i \rangle}$ . Da  $V_{Orb} \subseteq R_{ij} \cup \langle A_i \rangle \cup \langle A_j \rangle$  ist, lässt sich  $W_{V_{Orb}}$  zusammensetzen

zen aus Teilen von

$$W_{R_{ij}}, W_{R_{ij}, \langle A_i \rangle}, W_{R_{ij}, \langle A_j \rangle}, W_{\langle A_j \rangle, \langle A_i \rangle}, W_{\langle A_j \rangle} \text{ und } W_{\langle A_i \rangle}.$$

Daher folgt wie oben mit den elementaren Eigenschaften der bedingten Unabhängigkeit

$$W_{\langle A_j \rangle} \perp\!\!\!\perp W_{\langle A_i \rangle} \bullet \mid W_{V_{Orb}} \quad \text{und} \quad W_{\langle A_j \rangle} \bullet \perp\!\!\!\perp W_{\langle A_i \rangle} \bullet \mid W_{V_{Orb}}$$

und die zweite Zeile der Behauptung ist bewiesen.

**III.** Für die dritte und letzte Behauptung beachte, dass  $W_{\langle A_i \rangle} \bullet$  unabhängig ist von  $W_{V_{Orb}}$ . Dies folgt direkt aus dem Hauptsatz zur verallgemeinerten Riesz-Verteilung und der Inklusion  $V_{Orb} \subseteq V \setminus [A_i]$ . Unter Beachtung von Teil I gelten also

$$W_{\langle A_i \rangle} \bullet \perp\!\!\!\perp \{W_{\langle A_j \rangle} \bullet, W_{V_{Orb}}\} \quad \text{für } i \neq j \quad \text{und} \quad W_{\langle A_i \rangle} \bullet \perp\!\!\!\perp \{W_{\langle A_j \rangle} \bullet, W_{V_{Orb}}\} \quad \text{für alle } i, j.$$

Wegen Lemma 2.5(ii) folgt daraus

$$W_{\langle A_i \rangle} \bullet \perp\!\!\!\perp W_{\langle A_j \rangle} \bullet \mid W_{V_{Orb}} \quad \text{für } i \neq j \quad \text{und} \quad W_{\langle A_i \rangle} \bullet \perp\!\!\!\perp W_{\langle A_j \rangle} \bullet \mid W_{V_{Orb}} \quad \text{für alle } i, j.$$

Der Rest der Behauptung folgt dann durch induktive Anwendung von Teil I und II.  $\square$

Die folgenden zwei Theoreme werden die Verteilungen der Cholesky-Parameter von  $\psi(W)$  sowie deren Abhängigkeiten beschreiben. Dazu wird in Theorem 5.5 zunächst die Verteilung von  $\psi(W)_{[M]} \bullet$  gegeben  $\psi(W)_{V_{Orb}}$  bestimmt, wobei  $[M]$  wieder eine maximale Box meint. Diese Verteilung würde sich zwar auch als Korollar des Beweises von Theorem 5.6 ergeben. Allerdings ist der Beweis von Theorem 5.5 intuitiver und lässt die Struktur des Problems deutlicher hervortreten. Schon aus diesem Grund ist er damit von mathematischem Interesse und wird daher als eigenes Theorem aufgeführt.

**Theorem 5.5 (Die Verteilung von  $\psi(W)_{[M]} \bullet$  gegeben  $\psi(W)_{V_{Orb}}$ )**

Sei  $\mathcal{V}$  ein Annahme 2.6 erfüllender azyklisch gemischter Graph mit Gerüst  $\mathcal{U} = \mathcal{V}^\sim$  und sei  $G \subseteq \text{Orth}(\mathcal{V})$  eine endliche Gruppe ohne boxeninterne Symmetrien. Sei des Weiteren  $W \sim \mathbb{R}_p(\Omega, \lambda, \mathcal{V})$  eine verallgemeinert Riesz-verteilte Zufallsmatrix mit  $\Omega \in P(\mathcal{U}, G)$  und  $\lambda = (\beta, \dots, \beta)^T \in \mathbb{R}^k$  wie in Gleichung (5.1). Dann gilt für jede in  $\mathcal{V}$  maximale Box  $[M]$  die Verteilungsaussage

$$\mathcal{L}(\psi(W)_{[M]} \bullet \mid \psi(W)_{V_{Orb}}) = \mathcal{N}_{[M] \times \langle M \rangle} \left( \Omega_{[M]} \bullet, \frac{1}{2\beta |Orb[M]|} \Omega_{[M]} \bullet \otimes \psi(W)_{\langle M \rangle}^{-1} \right).$$

**Beweis**

Sei  $[A] \in Orb[M]$ , das heißt in  $G$  gibt es eine orthogonale Matrix  $\rho \in G$ , die die Box  $[A]$  auf die Box  $[M]$  abbildet. Damit muss  $|[A]| = |[M]|$  und  $|\langle A \rangle| = |\langle M \rangle|$  sein, da sonst die induzierte Boxenpermutation  $\tau_\rho$  kein Automorphismus wäre. Aus demselben Grund muss auch  $[A]$  maximal in  $\mathcal{V}$  sein. Der Hauptsatz zur verallgemeinerten Riesz-Verteilung, siehe Theorem 3.17 auf Seite 48, liefert dann

$$\mathcal{L}(W_{[A]\bullet} | W_{V \setminus [A]}) = \mathcal{N}_{[M] \times \langle M \rangle} \left( \Omega_{[M]\bullet}, \frac{1}{2\beta} \Omega_{[M]\bullet} \otimes W_{\langle A \rangle}^{-1} \right).$$

Hier wurde insbesondere verwendet, dass  $\Omega_{[A]\bullet} = \Omega_{[M]\bullet}$  und  $\Omega_{[A]\bullet} = \Omega_{[M]\bullet}$  ist, siehe Theorem 3.14 auf Seite 45 und Annahme 5.1. Dann liefert Theorem 3.1 von Seite 38 die bedingte Verteilung der folgenden linearen Transformation von  $W_{[A]\bullet}$ ,

$$\mathcal{L}(W_{[A]\bullet} W_{\langle A \rangle} | W_{V \setminus [A]}) = \mathcal{N}_{[M] \times \langle M \rangle} \left( \Omega_{[M]\bullet} W_{\langle A \rangle}, \frac{1}{2\beta} \Omega_{[M]\bullet} \otimes W_{\langle A \rangle} \right),$$

da  $W_{\langle A \rangle}$  eine Teilmatrix von  $W_{V \setminus [A]}$  ist. Genauer gesagt ist  $W_{\langle A \rangle}$  sogar eine Teilmatrix von  $W_{V_{Orb}}$ , wobei  $V_{Orb}$  wieder wie in Gleichung (5.2) definiert sei. Da die obige Verteilung von der bedingenden Variable  $W_{V \setminus [A]}$  nur durch  $W_{\langle A \rangle}$  abhängt, gilt ebenfalls

$$\mathcal{L}(W_{[A]\bullet} W_{\langle A \rangle} | W_{V_{Orb}}) = \mathcal{N}_{[M] \times \langle M \rangle} \left( \Omega_{[M]\bullet} W_{\langle A \rangle}, \frac{1}{2\beta} \Omega_{[M]\bullet} \otimes W_{\langle A \rangle} \right),$$

siehe Lemma 2.5(iii). Des Weiteren zeigt Lemma 5.4, dass die Matrizen  $\{W_{[A]\bullet}\}_{[A] \in Orb[M]}$  unabhängig voneinander sind gegeben  $W_{V_{Orb}}$ . Die Faltungsformel für matrixwertige Normalverteilungen, siehe Theorem 3.4 von Seite 39, liefert dann die Verteilung

$$\begin{aligned} \mathcal{L} \left( \sum_{[A] \in Orb[M]} W_{[A]\bullet} W_{\langle A \rangle} \middle| W_{V_{Orb}} \right) &= \\ &= \mathcal{N}_{[M] \times \langle M \rangle} \left( \Omega_{[M]\bullet} \sum_{[A] \in Orb[M]} W_{\langle A \rangle}, \frac{1}{2\beta} \Omega_{[M]\bullet} \otimes \sum_{[A] \in Orb[M]} W_{\langle A \rangle} \right). \end{aligned}$$

Aus der Skalierungsformel für Normalverteilungen, siehe Korollar 3.2 auf Seite 39, und der Bilinearität des Kronecker-Produkts folgt dann durch Skalieren mit  $1/|Orb[M]|$  die Verteilung

$$\begin{aligned} \mathcal{L} \left( \psi(W)_{[M]} \middle| W_{V_{Orb}} \right) &= \\ &= \mathcal{N}_{[M] \times \langle M \rangle} \left( \Omega_{[M]\bullet} \psi(W)_{\langle M \rangle}, \frac{1}{2\beta |Orb[M]|} \Omega_{[M]\bullet} \otimes \psi(W)_{\langle M \rangle} \right), \end{aligned}$$

siehe Lemma 5.3. Zudem zeigt Gleichung (5.3), dass sich  $\psi(W)_{V_{Orb}}$  allein aus den Einträgen der Matrix  $W_{V_{Orb}}$  bestimmen lässt. Des Weiteren ist  $\psi(W)_{\langle M \rangle}$  eine Teilmatrix von  $\psi(W)_{V_{Orb}}$  und die obige Verteilung hängt von  $W_{V_{Orb}}$  nur durch  $\psi(W)_{\langle M \rangle}$  ab. Es folgt also

$$\begin{aligned} \mathcal{L}\left(\psi(W)_{[M]} \middle| \psi(W)_{V_{Orb}}\right) &= \\ &= \mathcal{N}_{[M] \times \langle M \rangle} \left( \Omega_{[M] \bullet} \psi(W)_{\langle M \rangle}, \frac{1}{2\beta |Orb[M]|} \Omega_{[M] \bullet} \otimes \psi(W)_{\langle M \rangle} \right). \end{aligned}$$

Da  $\psi(W)_{[M] \bullet} = \psi(W)_{[M]} \psi(W)_{\langle M \rangle}^{-1}$  ist, folgt die Behauptung schließlich durch Rechtsmultiplikation von  $\psi(W)_{\langle M \rangle}^{-1}$  unter Berücksichtigung von Theorem 3.1 auf Seite 38. Beachte insbesondere, dass  $\psi(W)_{\langle M \rangle}$  fast sicher invertierbar ist, da  $\psi(W)$  mit Wahrscheinlichkeit Eins in  $P(\mathcal{U})$  liegt und  $\langle M \rangle$  in  $\mathcal{U}$  vollständig ist.  $\square$

**Theorem 5.6 (Die Verteilung von  $\psi(W)_{[M] \bullet}$ )**

Mit den Voraussetzungen aus Theorem 5.5 ist

$$\mathcal{L}\left(\psi(W)_{[M] \bullet}\right) = \mathcal{W}_{[M]} \left( 2\beta |Orb[M]| - \langle M \rangle, \frac{1}{2\beta |Orb[M]|} \Omega_{[M] \bullet} \right).$$

Außerdem gilt die Unabhängigkeitsbeziehung

$$\psi(W)_{[M] \bullet} \perp \left\{ \psi(W)_{[M] \bullet}, \psi(W)_{V_{Orb}} \right\}.$$

**Beweis**

**I.** Betrachte wie im Beweis von Theorem 5.5 zunächst eine beliebige Box  $[A] \in Orb[M]$  aus dem Orbit von  $[M]$ . Sie ist maximal in  $\mathcal{V}$  und da  $W$  einer  $\mathbb{R}_p(\Omega, \lambda, \mathcal{V})$ -Verteilung folgt, liefert der Hauptsatz zur verallgemeinerten Riesz-Verteilung zusammen mit Lemma 2.5 von Seite 14 die Beziehungen

$$W_{[A] \bullet} \perp \left\{ W_{[A] \bullet}, W_{V_{Orb}} \right\}, \quad (5.5)$$

$$\mathcal{L}\left(W_{[A] \bullet}\right) = \mathcal{W}_{[M]} \left( 2\beta - \langle M \rangle, \frac{1}{2\beta} \Omega_{[M] \bullet} \right) \text{ und} \quad (5.6)$$

$$\mathcal{L}\left(W_{[A] \bullet} \middle| W_{V_{Orb}}\right) = \mathcal{N}_{[M] \times \langle M \rangle} \left( \Omega_{[M] \bullet}, \frac{1}{2\beta} \Omega_{[M] \bullet} \otimes W_{\langle A \rangle}^{-1} \right). \quad (5.7)$$

Hier geht erneut ein, dass die Knotenanzahl in jeder Box  $[A]$  aus dem Orbit von  $[M]$  identisch sein muss. Selbiges gilt für die Elternmengen  $\langle A \rangle$ . Außerdem sind die Cholesky-Parameter  $\Omega_{[A] \bullet}$  und  $\Omega_{\langle A \rangle}$  für jede Box  $[A] \in Orb[M]$  dieselben, siehe Theorem 3.14 auf Seite 45.

Wegen Gleichung (5.6) und der Definition der klassischen Wishart-Verteilung in Kapitel 3.2 ab Seite 40 muss es also zu jeder Orbitbox  $[A] \in \text{Orb}[M]$  eine matrixwertig normalverteilte Zufallsgröße  $Y_A$  geben, so dass

$$W_{[A]\bullet} = Y_A^T Y_A \quad \text{mit} \quad \mathcal{L}(Y_A) = \mathcal{N}_{2\beta - \langle M \rangle \times [M]} \left( 0, I_{2\beta - \langle M \rangle} \otimes \frac{1}{2\beta} \Omega_{[M]\bullet} \right). \quad (5.8)$$

Beachte insbesondere, dass per Voraussetzung  $\beta > ([M] + \langle M \rangle - 1)/2$  und damit  $2\beta - \langle M \rangle$  positiv ist.

Betrachte nun für alle  $[A] \in \text{Orb}[M]$  die Teilmatrix  $W_{\langle A \rangle}$ . Sie ist fast sicher positiv definit, da die Zufallsmatrix  $W$  mit Wahrscheinlichkeit Eins einen Wert in  $P(\mathcal{U})$  annimmt und die Elternmenge  $\langle A \rangle$  per Annahme 2.6 vollständig in  $\mathcal{U}$  ist. Folglich lässt sich eine klassische Cholesky-Zerlegung auf diese Matrix anwenden, das heißt  $W_{\langle A \rangle} = U_A^T U_A$  für eine obere Dreiecksmatrix  $U_A \in PD_{\langle M \rangle}$ . Der Anschaulichkeit halber wird im Folgenden statt  $U_A$  die Bezeichnung  $\sqrt{W_{\langle A \rangle}}$  verwendet, das heißt  $W_{\langle A \rangle} = \sqrt{W_{\langle A \rangle}}^T \sqrt{W_{\langle A \rangle}}$ . Definiere nun die neue Zufallsmatrix

$$Z_A := \sqrt{W_{\langle A \rangle}} (W_{[A]\bullet} - \Omega_{[M]\bullet})^T. \quad (5.9)$$

Wegen Gleichung (5.7) und den Eigenschaften der matrixwertigen Normalverteilung, siehe Theorem 3.1 und Gleichung (3.3) auf Seite 38, folgt

$$\mathcal{L}(Z_A | W_{V_{\text{Orb}}}) = \mathcal{N}_{\langle M \rangle \times [M]} \left( 0, I_{\langle M \rangle} \otimes \frac{1}{2\beta} \Omega_{[M]\bullet} \right). \quad (5.10)$$

Seien jetzt die Boxen im Orbit von  $[M]$  gemäß  $\text{Orb}[M] = \{[A_1], \dots, [A_m]\}$  durchnummeriert. Wegen Lemma 5.4 sind die Zufallsgrößen  $Y_{A_1}, \dots, Y_{A_m}, Z_{A_1}, \dots, Z_{A_m}$  alle unabhängig voneinander gegeben  $W_{V_{\text{Orb}}}$ , denn per Konstruktion erben  $Y_A$  und  $Z_A$  die Unabhängigkeiten von  $W_{[A]\bullet}$  und  $W_{[A]\bullet}$ . Definiere dann die Zufallsmatrizen

$$X := \sqrt{\frac{2\beta}{|\text{Orb}[M]|}} \begin{pmatrix} Y_{A_1} \\ \vdots \\ Y_{A_m} \\ Z_{A_1} \\ \vdots \\ Z_{A_m} \end{pmatrix} \quad \text{und} \quad \mu := \sqrt{\frac{1}{|\text{Orb}[M]|}} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \sqrt{W_{\langle A_1 \rangle}} \\ \vdots \\ \sqrt{W_{\langle A_m \rangle}} \end{pmatrix},$$

wobei 0 die Nullmatrix der Dimension  $(2\beta - \langle M \rangle) \times \langle M \rangle$  meint und  $m$ -mal wiederholt wird. Damit ist  $X \in \mathbb{R}^{2\beta|\text{Orb}[M]| \times [M]}$  und  $\mu \in \mathbb{R}^{2\beta|\text{Orb}[M]| \times \langle M \rangle}$ . Wegen der (bedingten) Un-

abhängigkeit der Teilmatrizen von  $X$  lässt sich die Verteilung von  $X$  aus den Gleichungen (5.8) und (5.10) bestimmen. Lemma 3.5 auf Seite 40 liefert

$$\mathcal{L}(X|W_{V_{Orb}}) = \mathcal{N}_{2\beta|Orb[M] \times [M]} \left( 0, I_{2\beta|Orb[M]} \otimes \frac{1}{|Orb[M]|} \Omega_{[M]\bullet} \right). \quad (5.11)$$

II. Wegen  $W_{\langle A_i \rangle} = \sqrt{W_{\langle A_i \rangle}}^T \sqrt{W_{\langle A_i \rangle}}$  ergibt sich

$$\mu^T \mu = \frac{1}{|Orb[M]|} \sum_{i=1}^m \sqrt{W_{\langle A_i \rangle}}^T \sqrt{W_{\langle A_i \rangle}} = \psi(W)_{\langle M \rangle}, \quad (5.12)$$

siehe Lemma 5.3. Insbesondere ist  $\mu^T \mu$  damit fast sicher invertierbar. Ebenso ist

$$\begin{aligned} X^T \mu &= \frac{\sqrt{2\beta}}{|Orb[M]|} \sum_{i=1}^m Z_{A_i}^T \sqrt{W_{\langle A_i \rangle}} \\ &= \frac{\sqrt{2\beta}}{|Orb[M]|} \sum_{i=1}^m (W_{[A_i]\bullet} - \Omega_{[M]\bullet}) W_{\langle A_i \rangle} \\ &= \sqrt{2\beta} (\psi(W)_{[M]} - \Omega_{[M]\bullet} \psi(W)_{\langle M \rangle}), \end{aligned} \quad (5.13)$$

siehe Gleichung (5.9) und Lemma 5.3. Schließlich gilt

$$\begin{aligned} X^T X &= \frac{2\beta}{|Orb[M]|} \sum_{i=1}^m \{Y_{A_i}^T Y_{A_i} + Z_{A_i}^T Z_{A_i}\} \\ &= \frac{2\beta}{|Orb[M]|} \sum_{i=1}^m \left\{ W_{[A_i]\bullet} + (W_{[A_i]\bullet} - \Omega_{[M]\bullet}) W_{\langle A_i \rangle} (W_{[A_i]\bullet} - \Omega_{[M]\bullet})^T \right\} \\ &= 2\beta \left( \psi(W)_{[M]} - \psi(W)_{[M]} \Omega_{[M]\bullet}^T - \Omega_{[M]\bullet} \psi(W)_{[M]}^T + \Omega_{[M]\bullet} \psi(W)_{\langle M \rangle} \Omega_{[M]\bullet}^T \right), \end{aligned} \quad (5.14)$$

siehe Gleichungen (5.8) und (5.9), sowie Lemma 5.3.

III. Wegen Theorem 3.1 auf Seite 38 folgt aus Gleichung (5.11) die Verteilung

$$\mathcal{L}(\mu^T X|W_{V_{Orb}}) = \mathcal{N}_{\langle M \rangle \times [M]} \left( 0, \mu^T \mu \otimes \frac{1}{|Orb[M]|} \Omega_{[M]\bullet} \right).$$

Wegen den Gleichungen (5.12) und (5.13) ergibt sich daraus die Verteilung

$$\begin{aligned} \mathcal{L}(\psi(W)_{[M]\bullet} | W_{V_{Orb}}) &= \mathcal{L} \left( \frac{1}{\sqrt{2\beta}} X^T \mu \psi(W)_{\langle M \rangle}^{-1} + \Omega_{[M]\bullet} \middle| W_{V_{Orb}} \right) \\ &= \mathcal{N}_{[M] \times \langle M \rangle} \left( \Omega_{[M]\bullet}, \frac{1}{2\beta |Orb[M]|} \Omega_{[M]\bullet} \otimes \psi(W)_{\langle M \rangle}^{-1} \right). \end{aligned} \quad (5.15)$$

Da diese Verteilung von der bedingenden Variablen  $W_{V_{Orb}}$  nur durch eine Teilmatrix von  $\psi(W)_{V_{Orb}}$  abhängt, ist auch die Verteilung von  $\psi(W)_{[M]\bullet}$  gegeben  $\psi(W)_{V_{Orb}}$  durch den Ausdruck (5.15) bestimmt, siehe Lemma 2.5 und Gleichung (5.3). Mit anderen Worten reproduziert der Ansatz aus dem vorliegenden Beweis gerade die Verteilung aus Theorem 5.5.

**IV.** Setze nun  $Q := I_{2\beta|Orb[M]|} - \mu(\mu^T \mu)^{-1} \mu^T$ . Damit ist  $Q$  offensichtlich symmetrisch und idempotent, so dass sich der Rang zu

$$\begin{aligned} \text{rk}(Q) &= \text{tr}(Q) = \text{tr}(I_{2\beta|Orb[M]|}) - \text{tr}(\mu(\mu^T \mu)^{-1} \mu^T) \\ &= \text{tr}(I_{2\beta|Orb[M]|}) - \text{tr}(I_{\langle M \rangle}) = 2\beta|Orb[M]| - \langle M \rangle \end{aligned}$$

bestimmt. Gleichung (5.11) liefert dann zusammen mit Theorem 3.11 von Seite 43 die Verteilung

$$\mathcal{L}(X^T Q X | W_{V_{Orb}}) = \mathcal{W}_{[M]} \left( 2\beta|Orb[M]| - \langle M \rangle, \frac{1}{|Orb[M]|} \Omega_{[M]\bullet} \right). \quad (5.16)$$

Außerdem ergibt sich unter Verwendung der Gleichungen (5.12), (5.13) und (5.14) die Beziehung

$$\begin{aligned} X^T Q X &= X^T X - X^T \mu (\mu^T \mu)^{-1} \mu^T X \\ &= 2\beta \left( \psi(W)_{[M]} - \psi(W)_{[M]} \Omega_{[M]\bullet}^T - \Omega_{[M]\bullet} \psi(W)_{[M]}^T + \Omega_{[M]\bullet} \psi(W)_{\langle M \rangle} \Omega_{[M]\bullet}^T \right) - \\ &\quad - 2\beta \left( \psi(W)_{[M]} - \Omega_{[M]\bullet} \psi(W)_{\langle M \rangle} \right) \psi(W)_{\langle M \rangle}^{-1} \left( \psi(W)_{[M]} - \Omega_{[M]\bullet} \psi(W)_{\langle M \rangle} \right)^T \\ &= 2\beta \left( \psi(W)_{[M]} - \psi(W)_{[M]} \psi(W)_{\langle M \rangle}^{-1} \psi(W)_{[M]}^T \right) \\ &= 2\beta \psi(W)_{[M]\bullet}. \end{aligned} \quad (5.17)$$

Folglich wird die Verteilung von  $2\beta \psi(W)_{[M]\bullet}$  gegeben  $W_{V_{Orb}}$  ebenfalls durch Gleichung (5.16) beschrieben. Die gesuchte Verteilung ergibt sich dann durch Multiplikation mit  $1/(2\beta)$  unter Beachtung von Korollar 3.8 auf Seite 42 zu

$$\mathcal{L}(\psi(W)_{[M]\bullet} | W_{V_{Orb}}) = \mathcal{W}_{[M]} \left( 2\beta|Orb[M]| - \langle M \rangle, \frac{1}{2\beta|Orb[M]|} \Omega_{[M]\bullet} \right). \quad (5.18)$$

Diese Verteilung hängt in keiner Weise von der bedingenden Variable  $W_{V_{Orb}}$  ab, so dass auch die unbedingte Verteilung  $\mathcal{L}(\psi(W)_{[M]\bullet})$  durch Gleichung (5.18) gegeben ist, siehe Lemma 2.5.

V. Es bleibt nur die behauptete Unabhängigkeitsbeziehung zu zeigen. Beachte dazu, dass aus Gleichung (5.18) die Unabhängigkeit von  $\psi(W)_{[M]\bullet}$  und  $W_{V_{Orb}}$  folgt, siehe Lemma 2.5(iii). Da sich  $\psi(W)_{V_{Orb}}$  direkt aus  $W_{V_{Orb}}$  bestimmen lässt, siehe Gleichung (5.3), ist somit auch

$$\psi(W)_{[M]\bullet} \perp \psi(W)_{V_{Orb}}.$$

Beachte des Weiteren, dass

$$\mu^T Q^T = \mu^T Q = \mu^T - \mu^T \mu (\mu^T \mu)^{-1} \mu^T = 0 \in \mathbb{R}^{\langle M \rangle \times 2\beta |Orb[M]|}$$

ist. Wegen Theorem 3.3 von Seite 39 sind folglich  $X^T Q X$  und  $X^T \mu$  unabhängig gegeben  $W_{V_{Orb}}$ . Aus den Gleichungen (5.13) und (5.17) folgt also

$$\psi(W)_{[M]\bullet} \perp (\psi(W)_{[M]} - \Omega_{[M]\bullet} \psi(W)_{\langle M \rangle}) \mid W_{V_{Orb}}.$$

Da sich  $\psi(W)_{\langle M \rangle}$  direkt aus  $W_{V_{Orb}}$  bestimmen lässt, siehe Gleichung (5.3), ist auch

$$\psi(W)_{[M]\bullet} \perp \psi(W)_{[M]} \mid W_{V_{Orb}} \quad \text{und} \quad \psi(W)_{[M]\bullet} \perp \psi(W)_{[M]\bullet} \mid W_{V_{Orb}}.$$

Aus der Unabhängigkeit von  $\psi(W)_{[M]\bullet}$  und  $W_{V_{Orb}}$  folgt dann schließlich mit Lemma 2.5(ii) die Beziehung

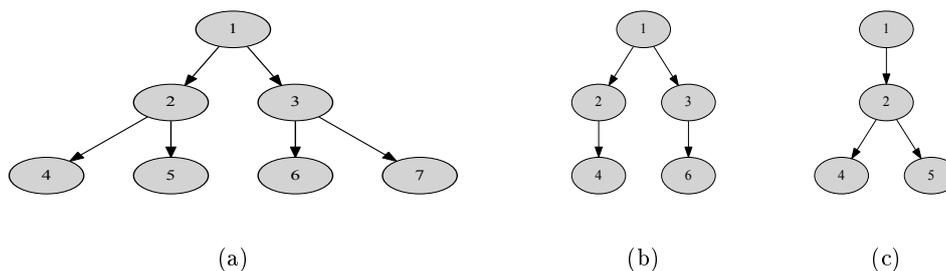
$$\psi(W)_{[M]\bullet} \perp \psi(W)_{[M]\bullet}.$$

Damit ist die Behauptung bewiesen. □

## 5.2 Verteilung des Maximum-Likelihood-Schätzers

Der nächste Abschnitt wird sich der Frage widmen, wie sich die Verteilung von  $\psi(W)$  aus den bereits bestimmten Verteilungen der Cholesky-Parameter ableiten lässt. Es wird sich zeigen, dass  $\psi(W)$  wie auch  $W$  verallgemeinert Riesz-verteilt ist. Allerdings sei hervorgehoben, dass  $\psi(W)$  eine  $P(\mathcal{U}, G)$ -wertige Zufallsvariable ist. Da  $P(\mathcal{U}, G)$  im Allgemeinen eine niedrigere Dimension als  $P(\mathcal{U})$  besitzt, kann  $\psi(W)$  nicht einer verallgemeinerten Riesz-Verteilung bezüglich des ursprünglichen Graphen  $\mathcal{V}$  folgen. Stattdessen muss zunächst ein weiterer Graph  $\mathcal{V}_G$  definiert werden, der die von  $G$  vorgegebenen Symmetrien berücksichtigt.

Sei also wieder  $\mathcal{V} = (V, D, U)$  ein Annahme 2.6 erfüllender azyklisch gemischter Graph und  $G \subseteq Orth(\mathcal{V})$  eine endliche Gruppe ohne Boxensymmetrien. Sei  $\mathcal{B}(\mathcal{V})_{\sim}$  eine Reprä-



**Abbildung 5:** Ein Graph  $\mathcal{V}$  (links) und seine reduzierte Version  $\mathcal{V}_{G_1}$  bezüglich der Gruppe  $G_1 = \{I_7, \varrho_7((1235476))\}$  (Mitte), sowie seine reduzierte Version  $\mathcal{V}_{G_2}$  bezüglich der Gruppe  $G_2 = \{I_7, \varrho_7((1326745))\}$  (rechts).

sentantenmenge bezüglich  $G$  wie in Definition 2.15 von Seite 23. Setze dann

$$V_G := \bigcup_{[B]_{\sim} \in \mathcal{B}(\mathcal{V})_{\sim}} [B]_{\sim} \subseteq V.$$

Damit taucht in  $V_G$  aus jedem Orbit von  $\mathcal{V}$  genau eine Box auf. Des Weiteren sei  $U_G := U \cap (V_G \times V_G)$ , das heißt die ungerichteten Kanten werden einfach aus  $\mathcal{V}$  übernommen. Seien schließlich  $[A]_{\sim}, [B]_{\sim} \in \mathcal{B}(\mathcal{V})_{\sim}$ , sowie  $i \in [A]_{\sim}$  und  $j \in [B]_{\sim}$ . Dann sei die Kante  $i \rightarrow j \in D_G$  genau dann vorhanden, wenn es eine Box  $[B'] \in \text{Orb}[B]_{\sim}$  und einen Knoten  $\ell \in [B']$  gibt, für den die gerichtete Kante  $i \rightarrow \ell \in D$  im ursprünglichen Graphen  $\mathcal{V}$  auftaucht. Der so definierte azyklisch gemischte Graph  $\mathcal{V}_G = (V_G, D_G, U_G)$  wird im Folgenden als *G-reduzierte Version* des Graphen  $\mathcal{V}$  bezeichnet. Beachte, dass der Graph  $\mathcal{V}_G$  insbesondere von der Wahl der Repräsentantenmenge  $\mathcal{B}(\mathcal{V})_{\sim}$  abhängt, auch wenn dies aus der Notation nicht hervorgeht.

Anschaulich werden beim Übergang von  $\mathcal{V}$  zu  $\mathcal{V}_G$  folglich die Boxen eines Orbits zusammengefasst, während die Kanten von  $\mathcal{V}_G$  einfach aus  $\mathcal{V}$  übernommen werden. Bei den gerichteten Kanten muss man darauf achten, welcher Repräsentant eines Orbits gewählt wurde, daher die etwas umständliche Definition von  $D_G$ . Da der ursprüngliche Graph  $\mathcal{V}$  Annahme 2.6 erfüllt, ist dies per Konstruktion auch für den reduzierten Graphen  $\mathcal{V}_G$  der Fall. Des Weiteren ist die Boxenmenge von  $\mathcal{V}_G$  gerade gegeben durch die gewählte Repräsentantenmenge von  $\mathcal{V}$  bezüglich  $G$ , das heißt  $\mathcal{B}(\mathcal{V}_G) = \mathcal{B}(\mathcal{V})_{\sim}$ .

### Beispiel 5.7 (G-reduzierte Version)

Betrachte den Graphen  $\mathcal{V}$  aus Abbildung 5a. Seine Boxen sind alle einelementig, setze daher  $[B_i] := \{i\}$  für  $i = 1, \dots, 7$ . Sei außerdem die Gruppe  $G_1 = \{I_7, \varrho_7((1235476))\}$

gegeben. Die Menge aller Boxenorbits unter  $G_1$  ist dann

$$\mathcal{B}(\mathcal{V})/\sim = \left\{ \{[B_1]\}, \{[B_2]\}, \{[B_3]\}, \{[B_4], [B_5]\}, \{[B_6], [B_7]\} \right\},$$

siehe Gleichung (2.6) auf Seite 23. Dementsprechend kann die Repräsentantenmenge

$$\mathcal{B}(\mathcal{V})_{\sim} = \{[B_1], [B_2], [B_3], [B_4], [B_6]\}$$

gewählt werden. Damit ist  $V_{G_1} = \{1, 2, 3, 4, 6\}$ , sowie  $D_{G_1} = \{1 \rightarrow 2, 1 \rightarrow 3, 2 \rightarrow 4, 3 \rightarrow 6\}$  und  $U_{G_1} = \emptyset$ . Die  $G_1$ -reduzierte Version  $\mathcal{V}_{G_1}$  sieht man in Abbildung 5b.

Ersetze nun  $G_1$  durch  $G_2 = \{I_7, \varrho_7((1326745))\}$ . Diesmal ist die Menge der Orbits gegeben durch

$$\mathcal{B}(\mathcal{V})/\sim = \left\{ \{[B_1]\}, \{[B_2], [B_3]\}, \{[B_4], [B_6]\}, \{[B_5], [B_7]\} \right\}$$

und als Repräsentantenmenge kann

$$\mathcal{B}(\mathcal{V})_{\sim} = \{[B_1], [B_2], [B_4], [B_5]\}$$

gewählt werden. Als reduzierte Knotenmenge ergibt sich folglich  $V_{G_2} = \{1, 2, 4, 5\}$ , während sich die Kantenmengen bestimmen zu  $D_{G_2} = \{1 \rightarrow 2, 2 \rightarrow 4, 2 \rightarrow 5\}$  und  $U_{G_2} = \emptyset$ . Die  $G_2$ -reduzierte Version  $\mathcal{V}_{G_2}$  sieht man in Abbildung 5c.

Ebenso könnte für  $\mathcal{V}_{G_2}$  aber auch die Repräsentantenmenge

$$\mathcal{B}(\mathcal{V})_{\sim} = \{[B_1], [B_2], [B_4], [B_7]\}$$

gewählt werden. Die Struktur des Graphen  $\mathcal{V}_{G_2}$  bleibt dann zwar dieselbe, siehe Abbildung 5c, allerdings muss in dieser Abbildung der Knoten 5 in 7 umbenannt werden.

Es zeigt sich, dass die Wahl der Repräsentantenmenge im letzten Beispiel mit erheblichen technischen Schwierigkeiten einhergehen würde. Der Grund hierfür ist, dass sich die Eltern der Box  $[B_7]$  in  $\mathcal{V}$  und die Eltern von  $[B_7]$  in  $\mathcal{V}_{G_2}$  unterscheiden. Während in  $\mathcal{V}$  nämlich  $[B_3]$  die einzige Elternbox von  $[B_7]$  ist, wird diese Rolle in  $\mathcal{V}_{G_2}$  von  $[B_2]$  übernommen. Diese Tatsache würde die Indizierung von Matrizen unnötig verkomplizieren, da a priori unklar wäre, was beispielsweise mit der Indexmenge „ $\langle B_7 \rangle$ “ gemeint ist.

Da es sich hierbei im Prinzip aber nur um ein Umnummerierungsproblem handelt, lässt sich diese Schwierigkeit durch eine geschickte Wahl der Repräsentantenmenge  $\mathcal{B}(\mathcal{V})_{\sim}$  umgehen, wie das nächste Lemma zeigt.

**Lemma 5.8 (Wahl der Repräsentantenmenge)**

Sei  $\mathcal{V}$  ein Annahme 2.6 erfüllender, zusammenhängender azyklisch gemischter Graph und  $G \subseteq \text{Orth}(\mathcal{V})$  eine endliche Gruppe ohne Boxensymmetrien. Dann gilt:

- (i) Die Repräsentantenmenge  $\mathcal{B}(\mathcal{V})_{\sim}$  kann so gewählt werden, dass die korrespondierende  $G$ -reduzierte Version  $\mathcal{V}_G$  ein induzierter Teilgraph von  $\mathcal{V}$  ist.
- (ii) Mit dieser Wahl von  $\mathcal{B}(\mathcal{V})_{\sim}$  gilt außerdem: Für jede Repräsentantenbox  $[B] \in \mathcal{B}(\mathcal{V})_{\sim}$  sind die Eltern von  $[B]$  in  $\mathcal{V}$  und die Eltern von  $[B]$  in  $\mathcal{V}_G$  identisch.

**Beweis**

(i) Der Beweis des ersten Teils verläuft konstruktiv. Setze dazu zunächst  $\mathcal{R} := \emptyset$ . Es sollen nun iterativ gewisse Boxen zu  $\mathcal{R}$  hinzugefügt werden, so dass am Ende  $\mathcal{R}$  gerade einer Repräsentantenmenge  $\mathcal{B}(\mathcal{V})_{\sim}$  entspricht.

(1) Wähle in  $\mathcal{V}$  eine Box  $[W]$  aus, in die keine gerichtete Kante hineinzeigt. Solch eine Box wird oft als *Wurzel* des Graph der Boxen  $\text{GoB}(\mathcal{V})$  bezeichnet. Beachte insbesondere, dass das Orbit  $\text{Orb}[W]$  nur aus der Box  $[W]$  selbst besteht, da  $\mathcal{V}$  zusammenhängend ist und keine Immoralitäten aufweisen darf. Füge daher  $[W]$  in  $\mathcal{R}$  ein.

(2) Seien  $[A_1], \dots, [A_\ell]$  die Kinder der Wurzelbox  $[W]$  in  $\mathcal{V}$ . Beachte insbesondere, dass  $[A_1]$  nicht im Orbit  $[W]$  liegen kann, da sonst die kausale Struktur von  $\mathcal{V}$  verletzt würde. Füge daher  $[A_1]$  in  $\mathcal{R}$  ein.

(3) Falls es ein  $i \in \{2, \dots, \ell\}$  gibt, für das  $[A_i]$  nicht im Orbit  $\text{Orb}[A_1]$  liegt, so füge die Box  $[A_i]$  der Menge  $\mathcal{R}$  hinzu. Falls es dann immer noch ein  $j \in \{2, \dots, \ell\} \setminus \{i\}$  gibt, für das  $[A_j]$  weder im Orbit  $\text{Orb}[A_1]$  noch im Orbit  $\text{Orb}[A_i]$  liegt, so füge auch die Box  $[A_j]$  der Menge  $\mathcal{R}$  hinzu. Wiederhole diesen Schritt so lange, bis alle Kinderboxen  $[A_1], \dots, [A_\ell]$  im Orbit genau einer Box aus der Menge  $\mathcal{R}$  liegen. Benenne dann die Kinder der Wurzelbox  $[W]$ , die in der Menge  $\mathcal{R}$  liegen, in  $[B_1], \dots, [B_q]$  um,  $q \leq \ell$ .

Beachte, dass jede in Schritt (3) zu  $\mathcal{R}$  hinzugefügte Box ein zulässiger Kandidat für die Repräsentantenmenge  $\mathcal{B}(\mathcal{V})$  ist, da das korrespondierende Orbit noch durch keine weitere Box repräsentiert wird, die bereits in  $\mathcal{R}$  geschrieben wurde. Außerdem gibt es per Konstruktion für jede Kinderbox  $[A]$  von  $[W]$  genau eine Box  $[B_i]$  in  $\mathcal{R}$ ,  $i \in \{1, \dots, q\}$ , so dass  $[A]$  im Orbit  $\text{Orb}[B_i]$  liegt.

(4) Führe nun Schritt (3) für jede der Boxen  $[B_i]$ ,  $i \in \{1, \dots, q\}$ , aus. Das heißt, betrachte alle Kinderboxen von  $[B_i]$  und füge solange Boxen zu  $\mathcal{R}$  hinzu, bis alle dieser Kinderboxen in genau einem von  $\mathcal{R}$  repräsentierten Orbit liegen. Mit anderen Worten übernimmt  $[B_i]$  die Rolle der Wurzelbox  $[W]$  aus Schritt (3). Setze diesen Vorgang induktiv fort, bis nur noch maximale Boxen verbleiben.

Schließlich ist die Menge  $\mathcal{R}$  mit einer Repräsentantenmenge  $\mathcal{B}(\mathcal{V})_{\sim}$  identisch. Die Inklusion  $\mathcal{R} \subseteq \mathcal{B}(\mathcal{V})_{\sim}$  gilt, da im Laufe des Algorithmus nur Boxen zu  $\mathcal{R}$  hinzugefügt werden, die noch nicht im Orbit einer Box aus  $\mathcal{R}$  liegen. Um die Inklusion  $\mathcal{B}(\mathcal{V})_{\sim} \subseteq \mathcal{R}$  einzusehen, nehme an, dass es eine Box  $[B] \in \mathcal{B}(\mathcal{V})_{\sim}$  gibt, die nicht im Orbit einer Box aus  $\mathcal{R}$  liegt. In diesem Fall muss  $[B]$  einen Vorfahren  $[A]$  haben, der Verlauf der Konstruktion nicht in  $\mathcal{R}$  aufgenommen wurde. Dann muss es aber eine Box  $[A'] \in \mathcal{R}$  geben, so dass  $[A]$  im Orbit  $Orb[A']$  liegt. Da jedes Gruppenelement  $\rho \in G$  einen Automorphismus über  $GoB(\mathcal{V})$  induziert, muss es dann aber eine Box  $[B'] \in \mathcal{R}$  geben, die ein Nachfahre von  $[A']$  ist und deren Orbit  $Orb[B']$  unter anderem die Box  $[B]$  enthält. Dies steht im Widerspruch zur Annahme und zeigt die Identität  $\mathcal{R} = \mathcal{B}(\mathcal{V})_{\sim}$ .

Zu guter Letzt sei hervorgehoben, dass im Verlauf des Algorithmus nur Boxen zu  $\mathcal{R}$  hinzugefügt werden, die durch eine gerichtete Kante mit einer Box verbunden sind, die bereits in  $\mathcal{R}$  liegt. Damit ist die von  $\mathcal{R}$  induzierte  $G$ -reduzierte Version  $\mathcal{V}_G$  ein induzierter Teilgraph von  $\mathcal{V}$ ,

$$\mathcal{V}_G = \mathcal{V}_R \quad \text{mit} \quad R = \bigcup_{[B] \in \mathcal{R}} [B].$$

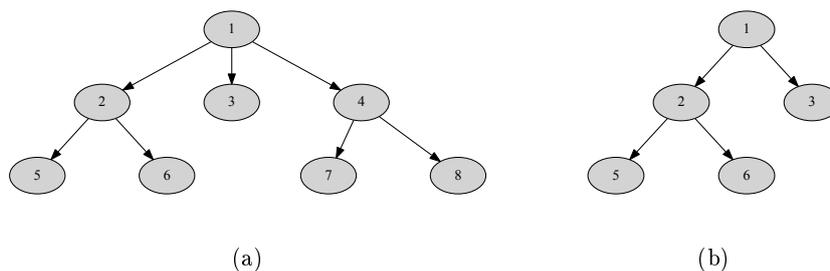
(ii) Für den zweiten Teil des Beweises sei  $\mathcal{B}(\mathcal{V})_{\sim} = \mathcal{R}$  wie gerade konstruiert. Außerdem sei  $[B] \in \mathcal{R}$  eine Repräsentantenbox. Aus Teil (i) folgt, dass jede Elternbox von  $[B]$  in  $\mathcal{V}_G$  auch in  $\mathcal{V}$  eine Elternbox von  $[B]$  ist. Um die umgekehrte Richtung einzusehen, sei  $[A]$  eine Elternbox von  $[B]$  im ursprünglichen Graphen  $\mathcal{V}$ . Dann muss  $[A]$  in  $\mathcal{R}$  liegen, aufgrund der Konstruktion muss nämlich jede Elternbox von  $[B]$  in  $\mathcal{R}$  sein. Damit ist  $[A]$  aber auch in  $\mathcal{V}_G$  eine Elternbox von  $[B]$ .  $\square$

Die obige Argumentation überträgt sich vollkommen analog auf einen nicht zusammenhängenden Graphen  $\mathcal{V}$ , es muss dazu nur Schritt (4) modifiziert werden. Anstatt aufzuhören wenn nur noch maximale Boxen untersucht werden können, muss es erlaubt sein, wieder zu Schritt (1) zurück zu gehen und eine neue Wurzelbox  $[W]$  auszuwählen. Auf diese Weise werden auch Zusammenhangskomponenten von  $\mathcal{V}$  erreicht, die nicht im Orbit der ersten Zusammenhangskomponente liegen. In solch einem Fall kann folglich auch die  $G$ -reduzierte Version  $\mathcal{V}_G$  mehrere Zusammenhangskomponenten haben.

### Beispiel 5.9 (Wahl der Repräsentantenmenge)

Betrachte den Graphen  $\mathcal{V}$  aus Abbildung 6a. Seine Boxen sind einelementig, das heißt  $[B_i] = \{i\}$  für alle  $i = 1, \dots, 8$ . Die einzige Wurzelbox ist  $[B_1]$ , sie wird daher in Schritt (1) des Algorithmus aus Lemma 5.8 zu  $\mathcal{R}$  hinzugefügt.

Anschließend werden ihre Kinderboxen  $[B_2]$ ,  $[B_3]$  und  $[B_4]$  untersucht. In den Schrit-



**Abbildung 6:** Ein Graph  $\mathcal{V}$  (links) und seine  $G$ -reduzierte Version  $\mathcal{V}_G$  bezüglich der von der Permutation (14327856) erzeugten Gruppe  $G$  (rechts). Da die Repräsentantenmenge wie in Lemma 5.8 gewählt wurde, ist der Graph  $\mathcal{V}_G$  ein induzierter Teilgraph von  $\mathcal{V}$ .

ten (2) und (3) werden die beiden Boxen  $[B_2]$  und  $[B_3]$  zu  $\mathcal{R}$  hinzugefügt. Die Box  $[B_4]$  dagegen wird ignoriert, da sie im Orbit  $Orb[B_2]$  liegt.

Im nächsten Iterationsschritt werden die Kinder von  $[B_2]$  untersucht, das heißt die Boxen  $[B_5]$  und  $[B_6]$ . Sie repräsentieren zwei unterschiedliche Orbits und werden daher beide in  $\mathcal{R}$  eingefügt.

Da  $[B_3]$ ,  $[B_5]$  und  $[B_6]$  maximale Boxen sind, bricht der Algorithmus danach ab. Beachte, dass die Boxen  $[B_7]$  und  $[B_8]$  im Laufe der Konstruktion nie untersucht werden. Sie liegen jedoch in den Orbits  $Orb[B_5]$  und  $Orb[B_6]$  und kommen daher als Repräsentanten ohnehin nicht mehr in Frage. Die resultierende  $G$ -reduzierte Version  $\mathcal{V}_G$  sieht man in Abbildung 6b. Dort erkennt man auch, dass  $\mathcal{V}_G$  der von  $\{1, 2, 3, 5, 6\}$  induzierte Teilgraph von  $\mathcal{V}$  ist, siehe Lemma 5.8(i).

Teil (ii) von Lemma 5.8 ist gerade aus technischer Sicht sehr nützlich. Wird die Repräsentantenmenge wie dort beschrieben gewählt, so muss nicht mehr zwischen den Eltern einer Box  $[B]$  in  $\mathcal{V}_G$  und den Eltern von  $[B]$  in  $\mathcal{V}$  unterschieden werden. Damit ist der Ausdruck „ $\langle B \rangle$ “ wohldefiniert. Aus diesem Grund sei im Folgenden die Repräsentantenmenge  $\mathcal{B}(\mathcal{V})$  immer wie in Lemma 5.8 gewählt. Dies erleichtert unter anderem den Beweis des folgenden Lemmas.

**Lemma 5.10 (Invariante graphische Modelle als rein graphische Modelle)**

Sei  $\mathcal{V}$  ein Annahme 2.6 erfüllender azyklisch gemischter Graph und  $G \subseteq Orth(\mathcal{V})$  eine endliche Gruppe ohne boxeninterne Symmetrien. Des Weiteren sei  $\mathcal{V}_G$  die  $G$ -reduzierte Version von  $\mathcal{V}$  aus Lemma 5.8 und die Gerüste seien bezeichnet mit  $\mathcal{U} = \mathcal{V}^\sim$  und  $\mathcal{U}_G = (\mathcal{V}_G)^\sim$ . Dann sind die Parameterräume des invarianten graphischen Modells  $P(\mathcal{U}, G)$  und des graphischen Modells  $P(\mathcal{U}_G)$  diffeomorph.

**Beweis**

Wegen der Definition des Parameterraums  $P(\mathcal{U})$ , siehe Gleichung (2.1) auf Seite 12, ist für eine Matrix  $\Omega \in P(\mathcal{U})$  die Abbildung

$$\Omega \mapsto [\Omega_{[B]}, \Omega_{[B]}]_{[B] \in \mathcal{B}(\mathcal{V})}$$

ein Diffeomorphismus, es werden nämlich nur die nicht-verschwindenden Blöcke von  $\Omega$  in eine Liste geschrieben. Sei jetzt zusätzlich  $\rho\Omega\rho^T = \Omega$  für alle  $\rho \in G$ , das heißt  $\Omega \in P(\mathcal{U}, G)$ , siehe Gleichung (2.8) auf Seite 33. Wegen Korollar 2.23 auf Seite 32 und Annahme 5.1 auf Seite 62 muss für alle Boxen  $[B] \in \mathcal{B}(\mathcal{V})$  gelten, dass

$$\Omega_{[B]} = \Omega_{[\tau_\rho^{-1}(B)]} \quad \text{und} \quad \Omega_{[B]} = \Omega_{[\tau_\rho^{-1}(B)]}$$

für alle  $\rho \in G$  ist. Folglich ist die Abbildung

$$[\Omega_{[B]}, \Omega_{[B]}]_{[B] \in \mathcal{B}(\mathcal{V})} \mapsto [\Omega_{[B]}, \Omega_{[B]}]_{[B] \in \mathcal{B}(\mathcal{V})_\sim}$$

ebenfalls bijektiv, siehe die Konstruktion der Repräsentantenmenge  $\mathcal{B}(\mathcal{V})_\sim$  in Definition 2.15 auf Seite 23. Sie ist sogar ein Diffeomorphismus, da nur die mehrfach auftretenden Parameter aus der Urbildmenge genommen werden.

Definiere nun die Matrix  $\Omega_G \in P(\mathcal{U}_G)$  durch  $(\Omega_G)_{[B]} = \Omega_{[B]}$  und  $(\Omega_G)_{[B]} = \Omega_{[B]}$  für alle Repräsentantenboxen  $[B] \in \mathcal{B}(\mathcal{V})_\sim$ . Beachte, dass an dieser Stelle die Wahl der Repräsentantenmenge aus Lemma 5.8 eingeht. Andernfalls wäre unklar, was mit der Indexmenge „ $[B]$ “ gemeint ist. Auch die Abbildung

$$[\Omega_{[B]}, \Omega_{[B]}]_{[B] \in \mathcal{B}(\mathcal{V})_\sim} \mapsto \Omega_G$$

stellt einen Diffeomorphismus dar, da nur Teilmatrizen aus einer Liste zu einer Matrix zusammengefügt werden.

Durch Komposition der gerade beschriebenen Abbildungen bekommt man schließlich einen Diffeomorphismus zwischen  $P(\mathcal{U}, G)$  und  $P(\mathcal{U}_G)$ , die *Reparametrisierungsabbildung*  $r$  mit

$$r : P(\mathcal{U}, G) \xrightarrow{\cong} P(\mathcal{U}_G), \quad \Omega \xrightarrow{\cong} \Omega_G. \quad (5.19)$$

Insgesamt streicht die Reparametrisierung  $r$  folglich alle mehrfach vorkommenden Blöcke aus der Matrix  $\Omega \in P(\mathcal{U}, G)$  und fügt die verbleibenden Blöcke zu einer Matrix passender Dimension zusammen.  $\square$

**Beispiel 5.11 ( $G$ -reduziertes Modell)**

Betrachte wie in Beispiel 5.7 den Graphen  $\mathcal{V}$  aus Abbildung 5a, sowie die Gruppen  $G_1 = \{I_7, \varrho_7((1235476))\}$  und  $G_2 = \{I_7, \varrho_7((1326745))\}$ . Werden die Repräsentantenmengen wie in Lemma 5.8 gewählt, so ergeben sich die  $G$ -reduzierten Versionen aus den Abbildungen 5b und 5c. Die Gerüste der drei Graphen seien bezeichnet mit  $\mathcal{U} := \mathcal{V}^\sim$ , sowie  $\mathcal{U}_{G_1} := (\mathcal{V}_{G_1})^\sim$  und  $\mathcal{U}_{G_2} := (\mathcal{V}_{G_2})^\sim$ .

Betrachte nun das invariante graphische Modell  $P(\mathcal{U}, G_1)$  und das graphische Modell  $P(\mathcal{U}_{G_1})$ . Es ist

$$P(\mathcal{U}, G_1) = \left\{ \left( \begin{array}{cccccc} \omega_{11} & \omega_{12} & \omega_{13} & & & \\ \omega_{12} & \omega_{22} & & \omega_{24} & \omega_{24} & \\ \omega_{13} & & \omega_{33} & & & \omega_{36} & \omega_{36} \\ & \omega_{24} & & \omega_{44} & \omega_{44} & & \\ & \omega_{24} & & & & \omega_{66} & \omega_{66} \\ & & \omega_{36} & & & & \omega_{66} \\ & & \omega_{36} & & & & \\ & & & & & & \omega_{66} \end{array} \right) = \Omega : \left. \begin{array}{l} \omega_{11}\omega_{22} - \omega_{12}^2 > 0 \\ \omega_{11}\omega_{33} - \omega_{13}^2 > 0 \\ \omega_{22}\omega_{44} - \omega_{24}^2 > 0 \\ \omega_{33}\omega_{66} - \omega_{36}^2 > 0 \end{array} \right\} \text{ und}$$

$$P(\mathcal{U}_{G_1}) = \left\{ \left( \begin{array}{cccccc} \omega_{11} & \omega_{12} & \omega_{13} & & & \\ \omega_{12} & \omega_{22} & & \omega_{24} & \omega_{36} & \\ \omega_{13} & & \omega_{33} & & \omega_{44} & \omega_{66} \\ & \omega_{24} & & \omega_{36} & \omega_{44} & \\ & & \omega_{36} & & \omega_{66} & \\ & & & & & \omega_{66} \end{array} \right) = \Omega_{G_1} : \left. \begin{array}{l} \omega_{11}\omega_{22} - \omega_{12}^2 > 0 \\ \omega_{11}\omega_{33} - \omega_{13}^2 > 0 \\ \omega_{22}\omega_{44} - \omega_{24}^2 > 0 \\ \omega_{33}\omega_{66} - \omega_{36}^2 > 0 \end{array} \right\}.$$

Die Bijektivität der beiden Parameterräume ist nun offensichtlich. Beachte, dass es sich bei der zugrunde liegenden Bijektion gerade um die Reparametrisierungsabbildung aus Gleichung (5.19) handelt, denn mit der obigen Notation ist  $\Omega_{G_1} = r(\Omega)$ .

Ersetzt man  $G_1$  durch  $G_2$ , so ergeben sich die Parameterräume

$$P(\mathcal{U}, G_2) = \left\{ \left( \begin{array}{cccccc} \omega_{11} & \omega_{12} & \omega_{12} & & & \\ \omega_{12} & \omega_{22} & & \omega_{24} & \omega_{25} & \omega_{24} & \omega_{25} \\ \omega_{12} & & \omega_{22} & & & \omega_{44} & \omega_{44} \\ & \omega_{24} & & \omega_{44} & \omega_{55} & & \\ & \omega_{25} & & & \omega_{55} & \omega_{44} & \omega_{55} \\ & & \omega_{24} & & & & \omega_{55} \\ & & \omega_{25} & & & & \omega_{55} \end{array} \right) : \left. \begin{array}{l} \omega_{11}\omega_{22} - \omega_{12}^2 > 0 \\ \omega_{22}\omega_{44} - \omega_{24}^2 > 0 \\ \omega_{22}\omega_{55} - \omega_{25}^2 > 0 \end{array} \right\} \text{ und}$$

$$P(\mathcal{U}_{G_2}) = \left\{ \left( \begin{array}{cccccc} \omega_{11} & \omega_{12} & & & & \\ \omega_{12} & \omega_{22} & \omega_{24} & \omega_{25} & & \\ & \omega_{24} & \omega_{44} & & & \\ & \omega_{25} & & \omega_{55} & & \\ & & & \omega_{55} & & \end{array} \right) : \left. \begin{array}{l} \omega_{11}\omega_{22} - \omega_{12}^2 > 0 \\ \omega_{22}\omega_{44} - \omega_{24}^2 > 0 \\ \omega_{22}\omega_{55} - \omega_{25}^2 > 0 \end{array} \right\}.$$

Auch hier ist die Bijektivität offensichtlich.

Abschließend soll die Verteilung des Maximum-Likelihood-Schätzers  $\hat{\Omega} = \psi(W) \in P(\mathcal{U}, G)$  beschrieben werden. Betrachte dazu die reparametrisierte Zufallsgröße  $r(\psi(W))$  mit  $r$  aus Gleichung (5.19). Es handelt sich hierbei per Konstruktion um eine  $P(\mathcal{U}_G)$ -wertige Zufallsgröße. Wie weiter oben bereits erwähnt wurde, besteht die Boxenmenge von  $\mathcal{V}_G$  gerade aus den Repräsentantenboxen von  $\mathcal{V}$ , das heißt  $\mathcal{B}(\mathcal{V}_G) = \mathcal{B}(\mathcal{V})^\sim$ . Ebenfalls per

Konstruktion ergeben sich für  $\Omega \in P(\mathcal{U}, G)$  die Cholesky-Parameter der reparametrisierten Matrix  $r(\Omega)$  zu

$$r(\Omega)_{[B]\bullet} = \Omega_{[B]\bullet}, \quad r(\Omega)_{[B]\bullet} = \Omega_{[B]\bullet} \quad \text{für alle } [B] \in \mathcal{B}(\mathcal{V}_G), \quad (5.20)$$

siehe die Definition der Abbildung  $r$  im Beweis von Lemma 5.10. Es sei nochmals hervorgehoben, dass bei der Interpretation der Indizes  $[B]\bullet$  und  $[B]\bullet$  Vorsicht geboten ist. Bei der Matrix  $\Omega$  sind damit a priori die Teilmatrizen gemeint, die zur Box  $[B]$  und deren Eltern im ursprünglichen Graphen  $\mathcal{V}$  korrespondieren. In der Reparametrisierung  $r(\Omega)$  dagegen werden die Box  $[B]$  und ihre Eltern im Graphen  $\mathcal{V}_G$  betrachtet. Da die Repräsentantenmenge jedoch wie in Lemma 5.8 gewählt wurde, macht dies keinen Unterschied.

Mit Hilfe von Gleichung (5.20) lassen sich die Verteilungen und Abhängigkeiten der Cholesky-Parameter von  $r(\psi(W))$  aus den Theoremen 5.5 und 5.6 herleiten. Das folgende Theorem beschreibt, wie diese zu einer gemeinsamen verallgemeinerten Riesz-Verteilung zusammengefasst werden können.

**Theorem 5.12 (Die Verteilung des reparametrisierten ML-Schätzers)**

Sei  $\mathcal{V}$  ein Annahme 2.6 erfüllender azyklisch gemischter Graph mit Gerüst  $\mathcal{U} = \mathcal{V}^\sim$  und sei  $G \subseteq \text{Orth}(\mathcal{V})$  eine endliche Gruppe ohne boxeninterne Symmetrien. Außerdem sei  $\mathcal{V}_G = (V_G, D_G, U_G)$  die  $G$ -reduzierte Version von  $\mathcal{V}$  aus Lemma 5.8 und  $\mathcal{U}_G = (\mathcal{V}_G)^\sim$  deren Gerüst. Sei des Weiteren  $W \sim \mathbb{R}_p(\Omega, \lambda, \mathcal{V})$  eine verallgemeinert Riesz-verteilte Zufallsmatrix mit  $\Omega \in P(\mathcal{U}, G)$  und  $\lambda = (\beta, \dots, \beta)^T \in \mathbb{R}^k$  wie in Gleichung (5.1). Schließlich sei  $r : P(\mathcal{U}, G) \rightarrow P(\mathcal{U}_G)$  die Reparametrisierungsabbildung aus Gleichung (5.19). Dann ist

$$r(\psi(W)) \sim \mathbb{R}_{|\mathcal{V}_G|}(r(\Omega), \gamma, \mathcal{V}_G),$$

wobei  $\gamma \in \mathbb{R}^{|\mathcal{B}(\mathcal{V})|}$  mit  $\gamma_{[B]} = \beta|\text{Orb}[B]|$  für alle  $[B] \in \mathcal{B}(\mathcal{V})^\sim$  ist.

**Beweis**

Sei  $[B_1], \dots, [B_m]$  eine Nummerierung der Boxen von  $\mathcal{V}_G$ , so dass für  $V_j := [B_1] \dot{\cup} \dots \dot{\cup} [B_j]$  die Box  $[B_j]$  maximal in  $(\mathcal{V}_G)_{V_j}$  ist,  $j = 1, \dots, m$ . Solch eine Nummerierung existiert immer, da  $\mathcal{V}$  und damit auch  $\mathcal{V}_G$  azyklisch ist. Der Beweis läuft induktiv über  $m$ .

*Induktionsanfang:* Für  $m = 1$  gibt es in  $\mathcal{V}$  nur ein Orbit bezüglich  $G$ , das heißt  $\mathcal{B}(\mathcal{V}) = \text{Orb}[B_1]$ . Der Graph  $\mathcal{V}$  besteht also aus  $|\text{Orb}[B_1]|$  disjunkten, vollständigen, ungerichteten Teilgraphen. Folglich handelt es sich bei  $\mathcal{V}_G$  um einen vollständigen, ungerichteten Graphen über den Knoten aus  $[B_1]$ . Per Definition der Reparametrisierungs-

abbildung  $r$  ist also

$$\begin{aligned} r(\psi(W)) &= r(\psi(W))_{[B_1]} = \psi(W)_{[B_1]} = \psi(W)_{[B_1]\bullet} \quad \text{und} \\ r(\Omega) &= r(\Omega)_{[B_1]} = \Omega_{[B_1]} = \Omega_{[B_1]\bullet}, \end{aligned}$$

da  $[B_1]$  sowohl in  $\mathcal{V}$  als auch in  $\mathcal{V}_G$  keine Eltern hat. Zudem ist  $[B_1]$  maximal in  $\mathcal{V}$ , so dass aus Theorem 5.6 die Verteilung

$$r(\psi(W)) = \psi(W)_{[B_1]\bullet} \sim \mathcal{W}_{[B_1]} \left( 2\beta|\text{Orb}[B_1]| - 0, \frac{1}{2\beta|\text{Orb}[B_1]|} \Omega_{[B_1]\bullet} \right)$$

folgt. Wegen Lemma 3.26 auf Seite 54 entspricht diese klassische Wishart-Verteilung der verallgemeinerten Riesz-Verteilung über dem vollständigen Graphen  $\mathcal{V}_{[B_1]} = \mathcal{V}_G$  mit Erwartungswertparameter  $\Omega_{[B_1]\bullet} = r(\Omega)$  und Formparameter  $\beta|\text{Orb}[B_1]| = \gamma_{[B_1]}$ . Dies zeigt den Induktionsanfang.

*Induktionsschritt:* Sei jetzt  $m$  beliebig. Aufgrund der Induktionsannahme ist

$$r(\psi(W))_{V_G \setminus [B_m]} \sim \mathbb{R}_{|V_G \setminus [B_m]|} \left( r(\Omega)_{V_G \setminus [B_m]}, \gamma_{V_G \setminus [B_m]}, (\mathcal{V}_G)_{V_G \setminus [B_m]} \right).$$

Des Weiteren ist  $[B_m]$  eine maximale Box in  $\mathcal{V}_G$  und per Konstruktion ist

$$r(\psi(W))_{[B_m]\bullet} = \psi(W)_{[B_m]\bullet} \quad \text{und} \quad r(\psi(W))_{[B_m]\bullet} = \psi(W)_{[B_m]\bullet},$$

denn Gleichung (5.20) gilt insbesondere für  $\psi(W) \in P(\mathcal{U}, G)$ . Beachte außerdem, dass für

$$V_{\text{Orb}} \equiv V_{\text{Orb}}([B_m]) = V \setminus \left\{ \bigcup_{[A] \in \text{Orb}[B_m]} [A] \right\}$$

die Matrizen  $\psi(W)_{V_{\text{Orb}}}$  und  $r(\psi(W))_{V_G \setminus [B_m]}$  bijektiv aufeinander abbildbar sind. Es ist nämlich  $V_G \setminus [B_m] \subseteq V_{\text{Orb}}$  und die Matrix  $r(\psi(W))_{V_G \setminus [B_m]}$  entsteht aus  $\psi(W)_{V_{\text{Orb}}}$  durch Löschen von mehrfach auftretenden Blöcken, siehe den Beweis von Lemma 5.10. Da  $[B_m]$  in  $\mathcal{V}$  maximal ist, folgen also aus den Theoremen 5.5 und 5.6 und aus Gleichung (5.20) die Verteilungsaussagen

- (i)  $r(\psi(W))_{[B_m]\bullet} \perp \{r(\psi(W))_{[B_m]\bullet}, r(\psi(W))_{V_G \setminus [B_m]}\}$ ,
- (ii)  $\mathcal{L}(r(\psi(W))_{[B_m]\bullet}) = \mathcal{W}_{[B_m]} \left( 2\beta|\text{Orb}[B_m]| - \langle B_m \rangle, \frac{1}{2\beta|\text{Orb}[B_m]|} r(\Omega)_{[B_m]\bullet} \right)$ ,
- (iii)  $\mathcal{L}(r(\psi(W))_{[B_m]\bullet} | r(\psi(W))_{V_G \setminus [B_m]}) =$   
 $= \mathcal{N}_{[B_m] \times \langle B_m \rangle} \left( r(\Omega)_{[B_m]\bullet}, \frac{1}{2\beta|\text{Orb}[B_m]|} r(\Omega)_{[B_m]\bullet} \otimes \psi(W)_{\langle B_m \rangle}^{-1} \right)$ .

Beachte außerdem, dass die Konstruktion der Reparametrisierungsabbildung  $r$  im Beweis von Lemma 5.10 die Gleichheit

$$\psi(W)_{\langle B_m \rangle} = r(\psi(W))_{\langle B_m \rangle}$$

impliziert, da  $\langle B_m \rangle$  in  $\mathcal{V}$  einen vollständigen Graphen induziert, siehe den Beweis von Lemma 2.21 auf Seite 30. Außerdem wird dabei die geschickte Wahl der Repräsentantenmenge aus Lemma 5.8 ausgenutzt, um der Indexmenge  $\langle B_m \rangle$  eine wohldefinierte Bedeutung zu geben.

Unter Beachtung dieser Gleichheit, der Punkte (i) bis (iii) und der Induktionsannahme folgt die Behauptung schließlich direkt aus dem Hauptsatz zur verallgemeinerten Riesz-Verteilung.  $\square$

Theorem 5.12 ermöglicht es schlussendlich, die Verteilung des Maximum-Likelihood-Schätzers des Kovarianzparameters in den untersuchten invarianten graphischen Modellen zu beschreiben.

**Theorem 5.13 (Hauptsatz zur Maximum-Likelihood-Schätzung)**

Sei  $\mathcal{V}$  ein Annahme 2.6 erfüllender azyklisch gemischter Graph und  $\mathcal{U} = \mathcal{V}^\sim$  dessen Gerüst. Außerdem sei  $G \subseteq \text{Orth}(\mathcal{V})$  eine endliche Gruppe ohne boxeninterne Symmetrien. Seien des Weiteren  $X_1, \dots, X_n \sim \mathcal{N}_p(0, \pi^{-1}(\Omega))$  unabhängig für ein  $\Omega \in P(\mathcal{U}, G)$ , wobei  $n \geq [B] + \langle B \rangle$  für alle Boxen  $[B] \in \mathcal{B}(\mathcal{V})$  ist. Definiere  $S := \frac{1}{n} \sum_{i=1}^n X_i X_i^T$  und  $W := \pi(S)$ . In diesem Fall existiert der Maximum-Likelihood-Schätzer  $\hat{\Omega}$  für  $\Omega \in P(\mathcal{U}, G)$  fast sicher und ist gegeben durch

$$\hat{\Omega} = \psi(W).$$

Sei außerdem die  $G$ -reduzierte Version von  $\mathcal{V}$  aus Lemma 5.8 wieder bezeichnet mit  $\mathcal{V}_G = (V_G, D_G, U_G)$  und  $\mathcal{U}_G := (\mathcal{V}_G)^\sim$  sei deren Gerüst. Des Weiteren sei  $\mathcal{B}(\mathcal{V})^\sim$  die zu  $\mathcal{V}_G$  korrespondierende Repräsentantenmenge, das heißt  $\mathcal{B}(\mathcal{V})^\sim = \mathcal{B}(\mathcal{V}_G)$ . Ist  $r : P(\mathcal{U}, G) \rightarrow P(\mathcal{U}_G)$  die Reparametrisierungsabbildung aus Gleichung (5.19), so folgt der reparametrisierte Schätzer  $r(\hat{\Omega})$  im Falle der Existenz der folgenden verallgemeinerten Riesz-Verteilung über dem reduzierten Parameterraum  $P(\mathcal{U}_G)$ ,

$$r(\hat{\Omega}) \sim \mathbb{R}_{|V_G|}(r(\Omega), \gamma, \mathcal{V}_G),$$

wobei  $\gamma \in \mathbb{R}^{|\mathcal{B}(\mathcal{V}_G)|}$  mit  $\gamma_{[B]} = \frac{n}{2} |\text{Orb}[B]|$  für alle Repräsentantenboxen  $[B] \in \mathcal{B}(\mathcal{V})^\sim$ .

**Beweis**

Die Existenzaussage und die Form des Schätzers folgen aus Theorem 4.3 und Bemerkung 4.4 ab Seite 58. Außerdem impliziert Theorem 4.2 auf Seite 57, dass  $W$  einer verallgemeinerten Riesz-Verteilung zur intrinsischen Darstellung  $\mathcal{V}_{\mathcal{U}}$  folgt mit Parametern  $\Omega$  und  $\lambda = (\frac{n}{2}, \dots, \frac{n}{2})$ . Da der Formparameter  $\lambda$  identische Einträge hat, kann anstatt der intrinsischen Darstellung  $\mathcal{V}_{\mathcal{U}}$  auch die vorgegebene Darstellung  $\mathcal{V}$  gewählt werden, ohne dass sich die Verteilung ändert, siehe Theorem 3.23 auf Seite 51. Die Behauptung folgt dann direkt aus Theorem 5.12.  $\square$

Es sei hervorgehoben, dass der reduzierte Schätzer  $r(\hat{\Omega})$  einer Verteilung folgt, deren Träger  $P(\mathcal{U}_G)$  der Parameterraum eines rein graphischen Modells ist. Es handelt sich dabei aber nicht um die Verteilung des Maximum-Likelihood-Schätzers der Kovarianzmatrix in diesem graphischen Modell  $P(\mathcal{U}_G)$ . Deren Formparameter ist nämlich von der Form  $(n/2, \dots, n/2)$ , siehe Theorem 4.2 auf Seite 57. Der Formparameter der Verteilung von  $r(\hat{\Omega})$  dagegen hat im Allgemeinen nicht-konstante Komponenten, da sich die Kardinalitäten der einzelnen Orbits  $Orb[B]$  für  $[B] \in \mathcal{B}(\mathcal{V})_{\sim}$  unterscheiden können.

Bei der Entwicklung von Likelihood-Quotienten-Tests in Kapitel 6 wird der Maximalwert der Likelihood-Funktion über dem invarianten graphischen Modell eine wichtige Rolle spielen.

**Korollar 5.14 (Maximalwert der Likelihood-Funktion)**

*Seien die Voraussetzungen wie in Theorem 5.13. Dann hat die Likelihood-Funktion über dem invarianten graphischen Modell  $PD(\mathcal{U}, G)$  den Maximalwert*

$$L_S(\hat{\Sigma}) = (2\pi e)^{-\frac{np}{2}} \prod_{[B] \in \mathcal{B}(\mathcal{V})_{\sim}} \det(\psi(W)_{[B]\bullet})^{-\frac{n}{2}|Orb[B]|}.$$

**Beweis**

Für eine bekannte Stichproben-Kovarianzmatrix  $S$  ist die Likelihood-Funktion gegeben durch

$$L_S(\Sigma) = (2\pi)^{-\frac{np}{2}} \det(\Sigma)^{-\frac{n}{2}} e^{-\frac{n}{2}\text{tr}(\Sigma^{-1}S)} \quad \text{für alle } \Sigma \in PD(\mathcal{U}, G), \quad (5.21)$$

vergleiche Gleichung (4.1) auf Seite 56. Mit  $\hat{\Omega} = \pi(\hat{\Sigma})$  folgt wegen Theorem 3.13 auf Seite 45 die Identität

$$\det(\hat{\Sigma}) = \det(\pi^{-1}(\hat{\Omega})) = \prod_{[B] \in \mathcal{B}(\mathcal{V})} \det(\hat{\Omega}_{[B]\bullet}) = \prod_{[B] \in \mathcal{B}(\mathcal{V})} \det(\psi(W)_{[B]\bullet}),$$

vergleiche Theorem 4.3. Außerdem liegt  $\psi(W)$  fast sicher in  $P(\mathcal{U}, G)$ . Wegen Theorem

3.14 von Seite 45 und der Konstruktion der  $G$ -reduzierten Version  $\mathcal{V}_G$  folgt daher

$$\det(\hat{\Sigma}) = \prod_{[B] \in \mathcal{B}(\mathcal{V})} \det(\psi(W)_{[B]\bullet}) = \prod_{[B] \in \mathcal{B}(\mathcal{V})_{\sim}} \det(\psi(W)_{[B]\bullet})^{|Orb[B]|}. \quad (5.22)$$

Des Weiteren gilt für alle  $\Sigma \in PD(\mathcal{U}, G)$  die Gleichheit

$$\mathrm{tr}[\Sigma^{-1}\psi(S)] = \frac{1}{|G|} \sum_{\rho \in G} \mathrm{tr}[\Sigma^{-1}\rho S \rho^T] = \frac{1}{|G|} \sum_{\rho \in G} \mathrm{tr}[(\rho^T \Sigma \rho)^{-1} S] = \mathrm{tr}[\Sigma^{-1} S]. \quad (5.23)$$

Hierbei geht insbesondere die  $G$ -Invarianz der Matrix  $\Sigma$  ein, das heißt für alle Gruppenelemente  $\rho \in G$  ist  $\rho \Sigma \rho^T = \Sigma$ . Des Weiteren ergibt sich unter Berücksichtigung von Gleichung (5.5) in Madsen [20] die Identität

$$\begin{aligned} \mathrm{tr}[\hat{\Sigma}^{-1}\psi(S)] &= \\ &= \sum_{[B] \in \mathcal{B}(\mathcal{V})} \mathrm{tr}\left[\hat{\Omega}_{[B]\bullet}^{-1} \left( \psi(S)_{[B]} - \psi(S)_{[B]} \hat{\Omega}_{[B]\bullet}^T - \hat{\Omega}_{[B]\bullet} \psi(S)_{[B]}^T + \hat{\Omega}_{[B]\bullet} \psi(S)_{\langle B \rangle} \hat{\Omega}_{[B]\bullet}^T \right)\right]. \end{aligned}$$

Da für alle Boxen  $[B] \in \mathcal{B}(\mathcal{V})$  die Identitäten  $\hat{\Omega}_{[B]\bullet} = \psi(S)_{[B]\bullet}$  und  $\hat{\Omega}_{\langle B \rangle} = \psi(S)_{\langle B \rangle}$  gelten, siehe den Beweis von Theorem 4.3, vereinfacht sich dies zu

$$\begin{aligned} \mathrm{tr}[\hat{\Sigma}^{-1}\psi(S)] &= \sum_{[B] \in \mathcal{B}(\mathcal{V})} \mathrm{tr}\left[\psi(S)_{[B]\bullet}^{-1} \left( \psi(S)_{[B]} - \psi(S)_{[B]} \psi(S)_{\langle B \rangle} \psi(S)_{[B]}^T \right)\right] = \\ &= \sum_{[B] \in \mathcal{B}(\mathcal{V})} \mathrm{tr}[I_{[B]}] = |V| = p. \end{aligned} \quad (5.24)$$

Einsetzen der Gleichungen (5.22), (5.23) und (5.24) in die Likelihood-Funktion (5.21) liefern dann die Behauptung.  $\square$

Das folgende Korollar wird ebenfalls bei der Entwicklung von Likelihood-Quotienten-Tests in Kapitel 6 von Nutzen sein.

**Korollar 5.15 (Ein spezieller Erwartungswert)**

Seien die Voraussetzungen wie in Theorem 5.13. Für alle  $\delta > \max_{[B] \in \mathcal{B}(\mathcal{V})_{\sim}} \left\{ \frac{[B] + \langle B \rangle - 1}{2|Orb[B]|} - \frac{n}{2} \right\}$  gilt dann die Identität

$$\begin{aligned} \mathbb{E} \left[ \prod_{[B] \in \mathcal{B}(\mathcal{V})_{\sim}} \det(\psi(W)_{[B]\bullet})^{\delta|Orb[B]|} \right] &= \\ &= \frac{2^{\delta p} \det[\pi^{-1}(\Omega)]^{\delta}}{n^{\delta p}} \prod_{[B] \in \mathcal{B}(\mathcal{V})_{\sim}} \left\{ |Orb[B]|^{-\delta|Orb[B]|} \prod_{i=1}^{[B]} \frac{\Gamma\left(\left(\delta + \frac{n}{2}\right)|Orb[B]| - \frac{\langle B \rangle + i - 1}{2}\right)}{\Gamma\left(\frac{n}{2}|Orb[B]| - \frac{\langle B \rangle + i - 1}{2}\right)} \right\}. \end{aligned}$$

**Beweis**

Im Beweis von Theorem 5.12 wurde deutlich, dass die Zufallsgrößen  $\{\psi(W)_{[B]\bullet}\}_{[B]\in\mathcal{B}(\mathcal{V})\sim}$  unabhängig voneinander sind. Außerdem folgt aus Theorem 5.13 für alle Repräsentantenboxen  $[B] \in \mathcal{B}(\mathcal{V})\sim$  die Verteilung

$$\psi(W)_{[B]\bullet} \sim \mathcal{W}_{[B]} \left( n|\text{Orb}[B]| - \langle B \rangle, \frac{1}{n|\text{Orb}[B]|} \Omega_{[B]\bullet} \right).$$

In Theorem 3.9 auf Seite 42 wurden bereits die Momente der Determinante einer Wishartverteilten Zufallsmatrix bestimmt. Im vorliegenden Fall ergibt sich

$$\begin{aligned} & \mathbb{E} \left[ \det(\psi(W)_{[B]\bullet})^{\delta|\text{Orb}[B]|} \right] = \\ & = 2^{\delta[B]|\text{Orb}[B]|} \det \left( \frac{1}{n|\text{Orb}[B]|} \Omega_{[B]\bullet} \right)^{\delta|\text{Orb}[B]|} \prod_{i=1}^{[B]} \frac{\Gamma \left( \delta|\text{Orb}[B]| + \frac{n|\text{Orb}[B]| - \langle B \rangle - i + 1}{2} \right)}{\Gamma \left( \frac{n|\text{Orb}[B]| - \langle B \rangle - i + 1}{2} \right)} \end{aligned}$$

für alle  $\delta > -\left[ \frac{n}{2} - \frac{\langle B \rangle + [B] - 1}{2|\text{Orb}[B]|} \right]$ . Beachte, dass durch die Wahl von  $n$  in Theorem 5.13 insbesondere negative Werte für  $\delta$  zugelassen werden, solange der Absolutbetrag  $|\delta|$  klein genug ist.

Aufgrund der Unabhängigkeit der einzelnen Faktoren im folgenden Erwartungswert ergibt sich schließlich

$$\begin{aligned} & \mathbb{E} \left[ \prod_{[B]\in\mathcal{B}(\mathcal{V})\sim} \det(\psi(W)_{[B]\bullet})^{\delta|\text{Orb}[B]|} \right] = \tag{5.25} \\ & = \prod_{[B]\in\mathcal{B}(\mathcal{V})\sim} \left( \frac{n}{2} |\text{Orb}[B]| \right)^{-\delta[B]|\text{Orb}[B]|} \det(\Omega_{[B]\bullet})^{\delta|\text{Orb}[B]|} \prod_{i=1}^{[B]} \frac{\Gamma \left( (\delta + \frac{n}{2})|\text{Orb}[B]| - \frac{\langle B \rangle + i - 1}{2} \right)}{\Gamma \left( \frac{n}{2} |\text{Orb}[B]| - \frac{\langle B \rangle + i - 1}{2} \right)}. \end{aligned}$$

Diese Gleichheit gilt für alle  $\delta > \max_{[B]\in\mathcal{B}(\mathcal{V})\sim} \left\{ \frac{[B] + \langle B \rangle - 1}{2|\text{Orb}[B]|} - \frac{n}{2} \right\}$ . Beachte schließlich, dass per Definition der Orbits und Repräsentantenboxen die Gleichheit

$$\sum_{[B]\in\mathcal{B}(\mathcal{V})\sim} [B]|\text{Orb}[B]| = \sum_{[B]\in\mathcal{B}(\mathcal{V})} [B] = p \tag{5.26}$$

gilt. Des Weiteren ist der Cholesky-Parameter  $\Omega_{[B]\bullet}$  für alle Boxen  $[B]$  aus demselben Orbit identisch, siehe Theorem 3.14 auf Seite 45.. Deswegen ist

$$\prod_{[B]\in\mathcal{B}(\mathcal{V})\sim} \det(\Omega_{[B]\bullet})^{|\text{Orb}[B]|} = \prod_{[B]\in\mathcal{B}(\mathcal{V})} \det(\Omega_{[B]\bullet}) = \det[\pi^{-1}(\Omega)], \tag{5.27}$$

siehe auch die Definition der verallgemeinerten Cholesky-Zerlegung in Theorem 3.13 auf Seite 45. Die Behauptung folgt dann durch Einsetzen dieser zwei Identitäten in Gleichung (5.25).  $\square$

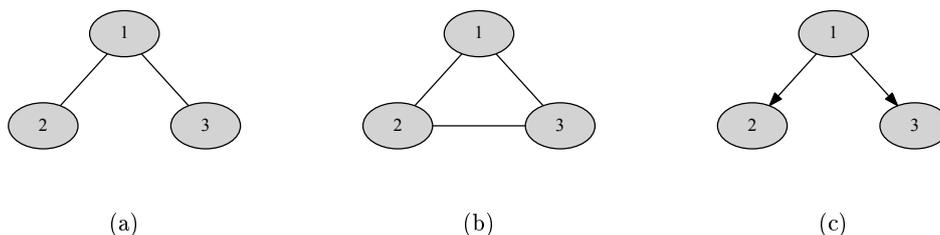
Den Abschluss dieses Kapitels bilden zwei Bemerkungen. Die erste erläutert die Rollen der Graphen  $\mathcal{V}$  und  $\mathcal{U} = \mathcal{V}^\sim$  in invarianten graphischen Modellen. In der zweiten Bemerkung wird kurz zusammengefasst, welche Annahmen an die Modellgraphen und die Invarianzgruppe gestellt wurden.

**Bemerkung 5.16 (Die Rollen von  $\mathcal{U}$  und  $\mathcal{V}$ )**

Das primäre Interesse gilt dem graphischen Modell  $P(\mathcal{U})$  bezüglich eines zerlegbaren Graphen  $\mathcal{U}$ , das heißt der ungerichtete Graph  $\mathcal{U}$  ist vorgegeben. Dieser induziert nun aber einen eindeutig festgelegten azyklisch gemischten Graphen, die intrinsische Darstellung  $\mathcal{V}_{\mathcal{U}}$ . Diese kommt schon bei der Maximum-Likelihood-Schätzung im rein graphischen Modell  $P(\mathcal{U})$  ins Spiel, da dort der Schätzer einer verallgemeinerten Riesz-Verteilung zur Darstellung  $\mathcal{V}_{\mathcal{U}}$  folgt. Da der Formparameter dieser Verteilung identische Einträge hat, folgt sogar noch mehr, siehe Theorem 3.23 auf Seite 51. Dieses besagt, dass statt der intrinsischen Darstellung  $\mathcal{V}_{\mathcal{U}}$  jede weitere Darstellung  $\mathcal{V}$  von  $\mathcal{U}$  gewählt werden kann, ohne dass sich die Verteilung ändert. Damit kann insbesondere ein azyklisch gemischter Graph  $\mathcal{V}$  verwendet werden, der Annahme 2.6 erfüllt.

Unter Umständen kann die Dimension des Parameterraums  $P(\mathcal{U})$  nun aber zu groß sein, um praktische Berechnungen durchzuführen. Will man diesem Problem mit Hilfe von Symmetrien begegnen, so spielt  $\mathcal{V}$  erneut eine wichtige Rolle. Die Gruppe der Invarianzen  $G$  muss nämlich die Struktur der Darstellung  $\mathcal{V}$  berücksichtigen, das heißt  $G \subseteq \text{Orth}(\mathcal{V})$ . Ist solch eine zulässige Gruppe  $G$  gefunden, so liefert sie zusammen mit dem ursprünglichen Graphen  $\mathcal{U}$  das invariante graphische Modell  $P(\mathcal{U}, G)$ . In diesem Schritt hat der azyklisch gemischte Graph  $\mathcal{V}$  folglich einen unmittelbaren Einfluss auf die Modellierung. Oft stehen dem Modellierer nämlich verschiedene Kandidaten für  $\mathcal{V}$  zur Verfügung, die sich insbesondere in der kausalen Abhängigkeitsstruktur unterscheiden können. Da die Gruppe  $G$  die Struktur von  $\mathcal{V}$  berücksichtigen muss, bekommt der Benutzer hierdurch die Möglichkeit, Wissen über Kausalitäten zwischen den beobachteten Variablen ins Modell aufzunehmen. Stünde nur der ungerichtete Graph  $\mathcal{U}$  zur Verfügung, so wäre dies nicht möglich.

Schließlich wird die Darstellung  $\mathcal{V}$  bei der Bestimmung der Verteilung des Maximum-Likelihood-Schätzers im invarianten graphischen Modell  $P(\mathcal{U}, G)$  benötigt. Zusammen mit der Gruppe  $G$  induziert  $\mathcal{V}$  nämlich den reduzierten Graphen  $\mathcal{V}_G$  und insbesondere dessen Gerüst  $\mathcal{U}_G = (\mathcal{V}_G)^\sim$ . Der Maximum-Likelihood-Schätzer im invarianten graphischen



**Abbildung 7:** Ein azyklisch gemischter Graph  $\mathcal{V}$ , dessen Box  $[B] = \{1, 2, 3\}$  unvollständig ist (links). Durch Einfügen einer weiteren Kante kann sie vollständig gemacht werden (Mitte). Dies lässt sich auch durch Umwandeln von ungerichteten in gerichtete Kanten erreichen (rechts).

Modell  $P(\mathcal{U}, G)$  lässt sich dann zu einer  $P(\mathcal{U}_G)$ -wertigen Zufallsgröße reparametrisieren und diese folgt einer verallgemeinerten Riesz-Verteilung zur Darstellung  $\mathcal{V}_G$ .

**Bemerkung 5.17 (Getroffene Annahmen und ihre Auswirkungen)**

Zum Abschluss des Kapitels soll kurz zusammengefasst werden, welche Annahmen getroffen werden mussten, um Theorem 5.13 zu beweisen. Außerdem soll erläutert werden, warum diese Annahmen getroffen wurden und ob sie erhebliche Einschränkungen der Allgemeinheit darstellen.

Betrachte zunächst die Annahme, dass ein untersuchter azyklisch gemischter Graph  $\mathcal{V}$  vollständige Boxen haben muss. Dies stellt keine große Einschränkung dar, da sich dies meist schon durch leichte Modifikation eines gegebenen azyklisch gemischten Graphen erreichen lässt. Betrachte beispielsweise den Graph  $\mathcal{V}$  in Abbildung 7a. Dessen Box  $[B] = \{1, 2, 3\}$  ist zwar nicht vollständig, lässt sich aber durch Einfügen der Kante 2 – 3 vervollständigen, siehe Abbildung 7b. Man kann die Vollständigkeit der Boxen von  $\mathcal{V}$  sogar ohne Hinzufügen von Kanten erreichen, siehe Abbildung 7c. Hier wurden die ungerichteten Kanten 1 – 2 und 1 – 3 zu gerichteten Kanten umgewandelt. Dies führt dazu, dass  $\mathcal{V}$  mehr Boxen bekommt, welche dafür kleiner und im vorliegenden Fall auch vollständig sind. Gerade in praktischen Anwendungen, in denen die Modellgraphen oft sehr viele Knoten enthalten, stellen solche kleinen Modifikationen des Graphen  $\mathcal{V}$  kaum einen Eingriff in die Modellierung dar.

Eine weitere Forderung, die bereits in Annahme 2.1 auf Seite 9 getroffen wurde, ist, dass  $\mathcal{V}$  keine Immoralitäten enthalten darf. Der Grund hierfür ist die sogenannte Markov-Äquivalenz von Graphen. Zwei azyklisch gemischte Graphen  $\mathcal{V}_1$  und  $\mathcal{V}_2$  heißen *markov-äquivalent*, falls die Parameterräume der induzierten graphischen Modelle identisch sind. Andersson, Madigan und Perlman zeigen in ihrem Aufsatz [7], dass ein azyklisch ge-

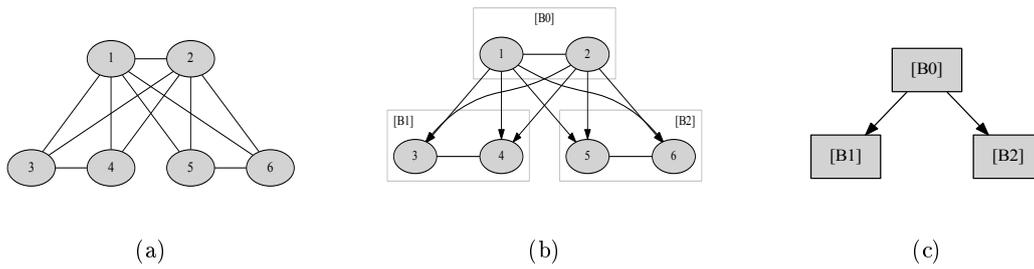
mischer Graph  $\mathcal{V}$  mit Immoralitäten zu keinem zerlegbaren Graphen markov-äquivalent ist. Graphische Modelle über nicht-zerlegbaren Graphen sind jedoch um ein Vielfaches schwieriger zu analysieren als zerlegbare Modelle. Beispielsweise gibt es keine expliziten Formeln für den Maximum-Likelihood-Schätzer der Kovarianzmatrix, stattdessen muss dieser iterativ berechnet werden, siehe Lauritzen [18, Seite 134ff]. Daher hat sich die Forschung rund um Verteilungen, die im Zusammenhang mit graphischen Modellen auftreten, bisher weitestgehend auf zerlegbare Modellgraphen beschränkt, siehe beispielsweise Andersson und Klein [3], Letac und Massam [19], Dawid und Lauritzen [9] und weitere.

Betrachte als nächstes Forderung (A3) aus Annahme 2.6 von Seite 17. Sie besagt, dass für jede Box  $[B]$  die korrespondierende Elternmenge  $\langle B \rangle$  aus einer Vereinigung von Boxen besteht und alle Knoten aus der Elternmenge  $\langle B \rangle$  mit allen Knoten aus der Box  $[B]$  durch gerichtete Kanten verbunden sind. Werden zusätzlich vollständige Boxen unterstellt, siehe Annahme (A1), so handelt es sich gerade um die Klasse von graphischen Modellen, die Madsen in seinem Aufsatz [20] untersucht. Eines der Hauptziele der vorliegenden Arbeit war gerade, die Verteilung der von Madsen bestimmten Schätzer zu ermitteln, daher diese Einschränkung. Sie lässt sich aber auch aus praktischer Sicht motivieren. Seien dazu  $[A]$  und  $[B]$  zwei unterschiedliche Boxen, so dass  $[A]$  unter anderem Eltern von Knoten in  $[B]$  enthält. Gerade in großen Graphen ist es nun kaum begründbar, warum nur ein Teil der Knoten von  $[A]$  einen kausalen Einfluss auf die Knoten in  $[B]$  hat, da  $[A]$  insbesondere vollständig ist. Daher stellt auch die Annahme (A3) keine allzu große Einschränkung dar.

Schließlich wurde auch an die Invarianzgruppe  $G \subseteq \text{Orth}(\mathcal{V})$  eine Forderung gestellt. Genauer gesagt wurde in Annahme 5.1 auf Seite 62 festgelegt, dass  $G$  keine boxeninternen Symmetrien enthalten soll. Den Grund hierfür liefert ein Blick auf den Beweis von Theorem 5.5 ab Seite 66. Dort wurde ausgenutzt, dass sich  $\psi(W)_{[M]}$  als eine Summe von unabhängigen Teilmatrizen von  $W$  darstellen lässt, siehe Lemma 5.3 auf Seite 63. Enthielte  $G$  boxeninterne Symmetrien, so bekäme man nur

$$\psi(W)_{[M]} = \frac{1}{|G|} \sum_{\rho \in G} \rho_{[M, \tau_\rho^{-1}(M)]} W_{[\tau_\rho^{-1}(M)]} \left( \rho_{\langle M, \tau_\rho^{-1}(M) \rangle} \right)^T.$$

Daher können beispielsweise unterschiedliche orthogonale Transformationen von  $W_{[M]}$  in der Summe auftreten, so dass die Summanden im Allgemeinen nicht mehr stochastisch unabhängig sind. Folglich wären die im Beweis der Theoreme 5.5 und 5.6 verwendeten Techniken nicht mehr anwendbar. Da es, wie in der Einleitung bereits erwähnt wurde, viele verschiedene Ansätze für eine Verallgemeinerung der klassischen Wishart-Verteilung gibt, ist es zwar durchaus denkbar, dass durch die Verwendung anderer Techniken



**Abbildung 8:** Ein zerlegbarer Graph  $\mathcal{U}$ , der aus zwei Cliques mit identischer Knotenzahl besteht (links). Bei seiner intrinsischen Darstellung  $\mathcal{V}_{\mathcal{U}}$  handelt es sich um einen azyklisch gemischten Graphen mit drei Boxen (Mitte). Diese erfüllt Annahme 2.6, so dass die kausale Struktur schon durch den Graph der Boxen wiedergegeben wird (rechts).

die Verteilung des Maximum-Likelihood-Schätzers trotzdem noch beschrieben werden könnte. Untersuchungen dieser Art böten jedoch wohl ausreichend Stoff für weitere eigenständige Arbeiten, so dass dieser Ansatz in der vorliegenden Dissertation nicht weiter verfolgt wird.

Ein weiterer Grund dafür ist, dass auch invariante graphische Modelle ohne boxeninterne Symmetrien eine äußerst interessante Modellklasse darstellen. Um dies einzusehen, sei nochmal hervorgehoben, dass die Hauptmotivation für die Einbindung von Symmetrien in ein graphisches Modell die Reduktion der Dimension des Parameterraums ist. Beispiel 2.25 von Seite 35 zeigt schon, dass auch Gruppen ohne boxeninterne Symmetrien die Dimensionalität des Problems erheblich reduzieren können. Außerdem gibt es oft Gruppen  $G \subseteq Orth(\mathcal{V})$  mit boxeninternen Symmetrien, die diese durch reines Umnummerieren der Knoten verlieren, vergleiche Beispiel 5.2 auf Seite 62. Alles in allem verkleinert die Annahme von fehlenden boxeninternen Symmetrien die Klasse der zulässigen Modelle zwar deutlich, trotzdem enthält die verbleibende Klasse sehr viele interessante invariante graphische Modelle.

### 5.3 Beispiel zur Maximum-Likelihood-Schätzung

Zum Abschluss dieses Kapitels sollen die bisher entwickelten Ergebnisse anhand eines Beispiels veranschaulicht werden. Sei dazu  $\mathcal{U} = (V, \emptyset, U')$  ein zerlegbarer Graph, der aus genau zwei Cliques  $C_1$  und  $C_2$  mit identischer Knotenzahl besteht,  $|C_1| = |C_2|$ . Ein Beispiel hierfür ist der Graph in Abbildung 8a mit den Cliques  $C_1 = \{1, 2, 3, 4\}$  und  $C_2 = \{1, 2, 5, 6\}$ . Setze dann  $[B_0] := C_1 \cap C_2$ , sowie  $[B_1] := C_1 \setminus [B_0]$  und  $[B_2] := C_2 \setminus [B_0]$ . Die intrinsische Darstellung  $\mathcal{V} := \mathcal{V}_{\mathcal{U}}$  von  $\mathcal{U}$  ergibt sich dann, indem alle ungerichteten

Kanten zwischen  $[B_0]$  und  $[B_1]$  sowie zwischen  $[B_0]$  und  $[B_2]$  durch gerichtete Kanten ersetzt werden, siehe Algorithmus 3.24 auf Seite 53. Für das obige Beispiel ergibt sich der Graph in Abbildung 8b. Mit anderen Worten ist  $\mathcal{V} = (V, D, U)$  ein azyklisch gemischter Graph, der Annahme 2.6 erfüllt und dessen Graph der Boxen die Form aus Abbildung 8c annimmt, wobei die Boxen  $[B_1]$  und  $[B_2]$  gleich viele Knoten enthalten.

Sei außerdem  $a := |[B_0]|$  und  $b := |[B_1]| = |[B_2]|$ . Die Gesamtzahl der Knoten sei wieder mit  $p$  bezeichnet, das heißt  $p = a + 2b$ . Dann lassen sich die Knoten gemäß  $[B_0] = \{v_1, \dots, v_a\}$ ,  $[B_1] = \{v_{a+1}, \dots, v_{a+b}\}$  und  $[B_2] = \{v_{a+b+1}, \dots, v_p\}$  nummerieren. Sei dann  $\sigma \in \text{Sym}(p)$  diejenige Permutation vom Grad  $p$ , die die Boxen  $[B_1]$  und  $[B_2]$  vertauscht, das heißt

$$\sigma(i) = i, \quad \sigma(a+j) = a+b+j, \quad \sigma(a+b+j) = a+j$$

für alle  $i \in \{1, \dots, a\}$  und alle  $j \in \{1, \dots, b\}$ . Die korrespondierende Permutationsmatrix stellt sich in Blockmatrixform gemäß

$$\rho := \varrho_p(\sigma) = \begin{pmatrix} I_a & 0 & 0 \\ 0 & 0 & I_b \\ 0 & I_b & 0 \end{pmatrix} \in \mathbb{R}^{p \times p}$$

dar. Beachte, dass die Matrix  $\rho$  in der Gruppe der von  $\mathcal{V}$  zugelassenen Transformationen  $\text{Orth}(\mathcal{V})$  liegt. Wähle daher die Gruppe  $G = \{I_p, \rho\}$ , so dass das invariante graphische Modell  $P(\mathcal{U}, G)$  wohldefiniert ist. Dieses Rahmengerüst stellt eine Verallgemeinerung von Beispiel 17.3 in Andersson und Klein [3] dar, wo der Spezialfall  $a = 1 = b$  untersucht wird. Es soll nun die Verteilung des Maximum-Likelihood-Schätzers des Kovarianzparameters  $\Omega \in P(\mathcal{U}, G)$  bestimmt werden.

### Der Maximum-Likelihood-Schätzer

Seien also  $X_1, \dots, X_n$  unabhängig und identisch verteilte Zufallsvektoren, die einer zentrierten Normalverteilung mit Streuungsmatrix  $\Sigma = \pi^{-1}(\Omega)$  für ein

$$\Omega = \begin{pmatrix} \Omega_{[B_0]} & \Omega_{[B_0, B_1]} & \Omega_{[B_0, B_1]} \\ \Omega_{[B_1, B_0]} & \Omega_{[B_1]} & 0 \\ \Omega_{[B_1, B_0]} & 0 & \Omega_{[B_1]} \end{pmatrix} \in P(\mathcal{U}, G)$$

folgen. Insbesondere ist also  $\Omega_{[B_1] \bullet} = \Omega_{[B_2] \bullet}$  und  $\Omega_{(B_1) \bullet} = \Omega_{(B_2) \bullet}$ . Außerdem sei  $n \geq a + b$ , so dass die hinreichende Bedingung für die Existenz des Maximum-Likelihood-Schätzers

aus Theorem 5.13 erfüllt ist. Setze wieder

$$W := \pi \left( \frac{1}{n} \sum_{i=1}^n X_i X_i^T \right) = \begin{pmatrix} W_{[B_0]} & W_{[B_0, B_1]} & W_{[B_0, B_2]} \\ W_{[B_1, B_0]} & W_{[B_1]} & 0 \\ W_{[B_2, B_0]} & 0 & W_{[B_2]} \end{pmatrix}.$$

Wegen den Theoremen 4.1 und 4.2 ab Seite 56 ist  $W$  der Maximum-Likelihood-Schätzer des Kovarianzparameters  $\Omega$  im graphischen Modell  $P(\mathcal{U})$  und folgt einer verallgemeinerten Riesz-Verteilung,

$$W \sim \mathbb{R}_p(\Omega, \lambda, \mathcal{V}) \quad \text{mit} \quad \lambda = \left( \frac{n}{2}, \frac{n}{2}, \frac{n}{2} \right)^T \in \mathbb{R}^3. \quad (5.28)$$

Außerdem besteht wegen Theorem 4.3 auf Seite 58 der Schätzer im invarianten graphischen Modell  $P(\mathcal{U}, G)$  aus der geglätteten Matrix  $\psi(W)$ . Um diese Matrix näher zu beschreiben sei zunächst hervorgehoben, dass sich die von  $G$  induzierten Orbits gemäß

$$\text{Orb}[B_0] = \{[B_0]\} \quad \text{und} \quad \text{Orb}[B_1] = \{[B_1], [B_2]\} = \text{Orb}[B_2]$$

schreiben lassen. Wegen Lemma 5.3 auf Seite 63 lässt sich die geglättete Matrix  $\psi(W)$  also beschreiben durch

$$\begin{aligned} \psi(W)_{[B_0]} &= W_{[B_0]\bullet} = W_{[B_0]}, & \psi(W)_{[B_1]} &= \frac{1}{2} (W_{[B_1]\bullet} + W_{[B_2]\bullet}) W_{[B_0]}, \\ \psi(W)_{[B_1]} &= \frac{1}{2} (W_{[B_1]\bullet} + W_{[B_2]\bullet} + W_{[B_1]\bullet} W_{[B_0]} W_{[B_1]\bullet}^T + W_{[B_2]\bullet} W_{[B_0]} W_{[B_2]\bullet}^T). \end{aligned}$$

Die Cholesky-Parameter von  $\psi(W)$  ergeben sich also zu

$$\begin{aligned} \psi(W)_{[B_0]\bullet} &= W_{[B_0]}, & \psi(W)_{[B_1]\bullet} &= \frac{1}{2} (W_{[B_1]\bullet} + W_{[B_2]\bullet}), \\ \psi(W)_{[B_1]\bullet} &= \frac{1}{2} (W_{[B_1]\bullet} + W_{[B_2]\bullet}) + \frac{1}{4} (W_{[B_1]\bullet} - W_{[B_2]\bullet}) W_{[B_0]} (W_{[B_1]\bullet} - W_{[B_2]\bullet})^T. \end{aligned} \quad (5.29)$$

Um die Verteilung von  $\psi(W)$  zu beschreiben, müssen zunächst die Verteilungen und die Abhängigkeiten der Cholesky-Parameter bestimmt werden.

### Die Verteilungen der Cholesky-Parameter

Die Verteilung von  $\psi(W)_{[B_0]}$  ergibt sich als eine Randverteilung der verallgemeinerten Riesz-Verteilung in Gleichung (5.28) zu

$$\mathcal{L}(\psi(W)_{[B_0]}) = \mathcal{L}(W_{[B_0]}) = \mathbb{R}_a(\Omega_{[B_0]}, (n/2), \mathcal{V}_{[B_0]}), \quad (5.30)$$

siehe Theorem 3.19 auf Seite 49. Des Weiteren folgen aus Gleichung (5.28) und dem Hauptsatz zur verallgemeinerten Riesz-Verteilung, siehe Theorem 3.17 auf Seite 48, die Aussagen

$$W_{[B_1]\bullet} \perp \{W_{[B_1]\bullet}, W_{V \setminus [B_1]}\}, \quad \text{und} \quad W_{[B_2]\bullet} \perp \{W_{[B_2]\bullet}, W_{V \setminus [B_2]}\}, \quad (5.31)$$

$$\mathcal{L}(W_{[B_1]\bullet}) = \mathcal{W}_b \left( n - a, \frac{1}{n} \Omega_{[B_1]\bullet} \right) = \mathcal{L}(W_{[B_2]\bullet}), \quad (5.32)$$

$$\mathcal{L}(W_{[B_1]\bullet} | W_{V \setminus [B_1]}) = \mathcal{N}_{b \times a} \left( \Omega_{[B_1]\bullet}, \frac{1}{n} \Omega_{[B_1]\bullet} \otimes W_{[B_0]}^{-1} \right) = \mathcal{L}(W_{[B_2]\bullet} | W_{V \setminus [B_2]}). \quad (5.33)$$

Da sich beispielsweise  $W_{[B_2]\bullet}$  aus Teilmatrizen von  $W_{V \setminus [B_1]}$  bestimmen lässt, folgt aus Gleichung (5.31) direkt die Unabhängigkeitsbeziehung

$$\{W_{[B_1]\bullet}, W_{[B_2]\bullet}\} \perp \{W_{[B_1]\bullet}, W_{[B_2]\bullet}, W_{[B_0]}\}. \quad (5.34)$$

Des Weiteren erkennt man, dass die Verteilungen in Gleichung (5.33) von der bedingenden Variable jeweils nur durch  $W_{[B_0]}$  abhängen. Damit ist aufgrund von Lemma 2.5(iii) auch

$$W_{[B_1]\bullet} \perp W_{[B_2]\bullet} | W_{[B_0]}. \quad (5.35)$$

Unter Beachtung der Darstellung von  $\psi(W)_{[B_1]\bullet}$  in Gleichung (5.29) ergeben die Gleichungen (5.33) und (5.35) zusammen die Verteilung

$$\begin{aligned} \mathcal{L}(\psi(W)_{[B_1]\bullet} | \psi(W)_{(B_1)}) &= \mathcal{L} \left( \frac{1}{2} (W_{[B_1]\bullet} + W_{[B_2]\bullet}) \middle| W_{[B_0]} \right) \\ &= \mathcal{N}_{b \times a} \left( \Omega_{[B_1]\bullet}, \frac{1}{2n} \Omega_{[B_1]\bullet} \otimes W_{[B_0]}^{-1} \right), \end{aligned} \quad (5.36)$$

vergleiche Theorem 5.5 auf Seite 66. Bei der letzten Gleichheit gehen insbesondere die Theoreme zur Faltung und Skalierung von normalverteilten Zufallsmatrizen ein, siehe Kapitel 3.1.

Des Weiteren sind  $W_{[B_1]\bullet}$  und  $W_{[B_2]\bullet}$  unabhängig, siehe Gleichung (5.31). Zusammen mit Gleichung (5.32) folgt aus den Theoremen zur Faltung und Skalierung von Wishartverteilten Zufallsmatrizen in Kapitel 3.2 die Verteilung

$$\mathcal{L} \left[ \frac{1}{2} (W_{[B_1]\bullet} + W_{[B_2]\bullet}) \right] = \mathcal{W}_b \left( 2(n - a), \frac{1}{2n} \Omega_{[B_1]\bullet} \right).$$

Ebenso folgt unter Beachtung von Gleichung (5.35) und den Theoremen 3.1 und 3.4 die

Beziehung

$$\mathcal{L} \left[ \frac{1}{2} (W_{[B_1]\bullet} - W_{[B_2]\bullet}) \sqrt{W_{[B_0]}} \middle| W_{[B_0]} \right] = \mathcal{N}_{b \times a} \left( 0, \frac{1}{2n} \Omega_{[B_1]\bullet} \otimes I_{[B_0]} \right), \quad (5.37)$$

wobei  $W_{[B_0]} = \sqrt{W_{[B_0]}} \sqrt{W_{[B_0]}}^T$  wieder die klassische Cholesky-Zerlegung meint. Beachte, dass die Verteilung auf der rechten Seite nicht mehr von der bedingenden Variablen  $W_{[B_0]}$  abhängt, so dass auch die unbedingte Verteilung durch die rechte Seite gegeben ist. Per Definition der klassischen Wishart-Verteilung ist also

$$\mathcal{L} \left[ \frac{1}{4} (W_{[B_1]\bullet} - W_{[B_2]\bullet}) W_{[B_0]} (W_{[B_1]\bullet} - W_{[B_2]\bullet})^T \right] = \mathcal{W}_b \left( a, \frac{1}{2n} \Omega_{[B_1]\bullet} \right). \quad (5.38)$$

Unter Beachtung der Gleichungen (5.29) und (5.34), sowie Theorem 3.6 folgt also

$$\mathcal{L} [\psi(W)_{[B_1]\bullet}] = \mathcal{W}_b \left( 2n - a, \frac{1}{2n} \Omega_{[B_1]\bullet} \right), \quad (5.39)$$

vergleiche Theorem 5.6 auf Seite 68. Außerdem folgt aus den Gleichungen (5.34) und (5.37), dass sowohl  $W_{[B_1]\bullet} + W_{[B_2]\bullet}$  als auch  $(W_{[B_1]\bullet} - W_{[B_2]\bullet}) \sqrt{W_{[B_0]}}$  unabhängig von  $W_{[B_0]} = \psi(W)_{[B_0]}$  sind. Da sich  $\psi(W)_{[B_1]\bullet}$  aus diesen beiden Zufallsgrößen zusammensetzt ist folglich

$$\psi(W)_{[B_1]\bullet} \perp \psi(W)_{[B_0]}. \quad (5.40)$$

Des Weiteren implizieren die Gleichungen (5.33) und (5.35) die Verteilungsaussage

$$\mathcal{L} \left[ \begin{pmatrix} W_{[B_1]\bullet} \\ W_{[B_2]\bullet} \end{pmatrix} \middle| W_{[B_0]} \right] = \mathcal{N}_{2b \times a} \left[ \begin{pmatrix} \Omega_{[B_1]\bullet} \\ \Omega_{[B_1]\bullet} \end{pmatrix}, I_2 \otimes \left( \frac{1}{n} \Omega_{[B_1]\bullet} \otimes W_{[B_0]}^{-1} \right) \right], \quad (5.41)$$

vergleiche auch den Beweis von Lemma 3.5 auf Seite 40. Definiere dann die Matrix  $Q := \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \otimes I_b$ . Insbesondere ist dann aufgrund der Rechenregeln des Kronecker-Produkts

$$Q \left( I_2 \otimes \frac{1}{n} \Omega_{[B_1]\bullet} \right) Q^T = \left[ \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}^T \right] \otimes \frac{1}{n} \Omega_{[B_1]\bullet} = I_2 \otimes \frac{2}{n} \Omega_{[B_1]\bullet}$$

Wegen Theorem 3.1 auf Seite 38 folgt damit aus Gleichung (5.41) die Verteilung

$$\begin{aligned} \mathcal{L} \left[ \begin{pmatrix} W_{[B_1]\bullet} + W_{[B_2]\bullet} \\ W_{[B_1]\bullet} - W_{[B_2]\bullet} \end{pmatrix} \middle| W_{[B_0]} \right] &= \mathcal{L} \left[ Q \begin{pmatrix} W_{[B_1]\bullet} \\ W_{[B_2]\bullet} \end{pmatrix} \middle| W_{[B_0]} \right] = \\ &= \mathcal{N}_{2b \times a} \left[ Q \cdot 0, Q \left( I_2 \otimes \frac{1}{n} \Omega_{[B_1]\bullet} \right) Q^T \otimes W_{[B_0]}^{-1} \right] = \mathcal{N}_{2b \times a} \left[ 0, I_2 \otimes \left( \frac{2}{n} \Omega_{[B_1]\bullet} \otimes W_{[B_0]}^{-1} \right) \right]. \end{aligned}$$

Insbesondere sind also  $W_{[B_1]\bullet} + W_{[B_2]\bullet}$  und  $W_{[B_1]\bullet} - W_{[B_2]\bullet}$  unabhängig gegeben  $W_{[B_0]} = \psi(W)_{[B_0]}$ , siehe Korollar 2.2.4.2 in Kollo und von Rosen [17]. Zusammen mit den Gleichungen (5.29) und (5.34) folgt damit die bedingte Unabhängigkeitsbeziehung

$$\psi(W)_{[B_1]\bullet} \perp\!\!\!\perp \psi(W)_{[B_1]\bullet} \mid \psi(W)_{[B_0]}.$$

Unter Beachtung von Gleichung (5.40) und Lemma 2.5(ii) auf Seite 14 folgt schließlich

$$\psi(W)_{[B_1]\bullet} \perp\!\!\!\perp \{ \psi(W)_{[B_1]\bullet}, \psi(W)_{[B_0]} \}, \quad (5.42)$$

siehe Theorem 5.6 auf Seite 68. Abschließend sei hervorgehoben, dass per Konstruktion der Glättungs-Funktion  $\psi$  die Identitäten

$$\psi(W)_{[B_1]\bullet} = \psi(W)_{[B_2]\bullet} \quad \text{und} \quad \psi(W)_{[B_1]\bullet} = \psi(W)_{[B_2]\bullet} \quad \text{für alle } W \in P(\mathcal{U})$$

gelten. Die zur Box  $[B_2]$  korrespondierenden geschätzten Cholesky-Parameter sind somit in einem gewissen Sinne redundant. Dies muss bei der Beschreibung der Verteilung des Maximum-Likelihood-Schätzers  $\hat{\Omega} \in P(\mathcal{U}, G)$  berücksichtigt werden.

### Die Verteilung des Maximum-Likelihood-Schätzers

Aus diesem Grund soll zunächst eine  $G$ -reduzierte Version des Graphen  $\mathcal{V} = (V, D, U)$  konstruiert werden. Setze dazu

$$\mathcal{B}(\mathcal{V})_{\sim} = \{ [B_0], [B_1] \} \quad \text{und} \quad V_G = [B_0] \dot{\cup} [B_1],$$

vergleiche Kapitel 5.2. Außerdem werden sowohl die gerichteten als auch die ungerichteten Kanten einfach aus  $\mathcal{V}$  übernommen,  $D_G = D \cap ([B_0] \times [B_1])$  und  $U_G = U \cap (V_G \times V_G)$ . Dann ist  $\mathcal{V}_G = (V_G, D_G, U_G)$  eine  $G$ -reduzierte Version von  $\mathcal{V}$ , deren Gerüst mit  $\mathcal{U}_G = (\mathcal{V}_G)_{\sim}$  bezeichnet sei.

Des Weiteren sei  $r : P(\mathcal{U}, G) \rightarrow P(\mathcal{U}_G)$  die Reparametrisierungsabbildung aus Gleichung (5.19) von Seite 78, das heißt

$$r \left[ \begin{pmatrix} \Omega_{[B_0]} & \Omega_{[B_0, B_1]} & \Omega_{[B_0, B_1]} \\ \Omega_{[B_1, B_0]} & \Omega_{[B_1]} & 0 \\ \Omega_{[B_1, B_0]} & 0 & \Omega_{[B_1]} \end{pmatrix} \right] = \begin{pmatrix} \Omega_{[B_0]} & \Omega_{[B_0, B_1]} \\ \Omega_{[B_1, B_0]} & \Omega_{[B_1]} \end{pmatrix} \quad \text{für alle } \Omega \in P(\mathcal{U}, G).$$

Insbesondere ist also

$$r[\psi(W)]_{[B_0]} = \psi(W)_{[B_0]}, \quad r[\psi(W)]_{[B_1]^\bullet} = \psi(W)_{[B_1]^\bullet}, \quad r[\psi(W)]_{[B_1]^\bullet} = \psi(W)_{[B_1]^\bullet},$$

wobei sich die Indizes des ursprünglichen Schätzers  $\psi(W)$  auf den Graphen  $\mathcal{V}$  beziehen, während die Indizes des reparametrisierten Schätzers  $r[\psi(W)]$  bezüglich des Graphen  $\mathcal{V}_G$  definiert sind.

Ein Blick zurück auf die Gleichungen (5.30), (5.36), (5.39) und (5.42) zeigt dann, dass für den reparametrisierten Schätzer die folgenden vier Aussagen gelten,

- (i)  $r[\psi(W)]_{[B_1]^\bullet} \perp \{r[\psi(W)]_{[B_1]^\bullet}, r[\psi(W)]_{[B_0]}\}$ ,
- (ii)  $\mathcal{L}[r[\psi(W)]_{[B_1]^\bullet}] = \mathcal{W}_b(2n - a, \frac{1}{2n}r(\Omega)_{[B_1]^\bullet})$ ,
- (iii)  $\mathcal{L}(r[\psi(W)]_{[B_1]^\bullet} | r[\psi(W)]_{[B_0]}) = \mathcal{N}_{b \times a}(r(\Omega)_{[B_1]^\bullet}, \frac{1}{2n}r(\Omega)_{[B_1]^\bullet} \otimes r[\psi(W)]_{[B_0]}^{-1})$ ,
- (iv)  $\mathcal{L}(r[\psi(W)]_{[B_0]}) = \mathbb{R}_a(r(\Omega)_{[B_0]}, (n/2), (\mathcal{V}_G)_{[B_0]})$ .

Da  $V_G \setminus [B_1] = [B_0]$  gilt und  $[B_1]$  eine maximale Box in  $\mathcal{V}_G$  ist, folgt direkt aus Theorem 3.17 von Seite 48, dass  $r[\psi(W)]$  einer verallgemeinerten Riesz-Verteilung folgt. Genauer gesagt ist

$$r[\psi(W)] \sim \mathbb{R}_{a+b}(r(\Omega), \gamma, \mathcal{V}_G) \quad \text{mit} \quad \gamma = \left( \frac{n}{2}, n \right)^T \in \mathbb{R}^2.$$

Dies reproduziert gerade das Ergebnis des Hauptsatzes zur Maximum-Likelihood-Schätzung in invarianten graphischen Modellen, siehe Theorem 5.13 auf Seite 82.

## 6 Likelihood-Quotienten-Tests

Mit Hilfe der Ergebnisse aus den Kapiteln 4 und 5 lässt sich Maximum-Likelihood-Schätzung in einem vorgegebenen invarianten graphischen Modell betreiben. In praktischen Anwendungen steht jedoch oft die Modellauswahl im Vordergrund. Damit ist die Frage gemeint, welches denkbare Modell einen gegebenen Datensatz möglichst gut beschreibt, ohne zu viele Parameter zu verwenden. Ein wichtiges Hilfsmittel zur Beantwortung dieser Frage sind *Likelihood-Quotienten-Tests*. Sie ermöglichen es, bei zwei vorgegebenen, geschachtelten Verteilungsfamilien  $\mathcal{M}_0$  und  $\mathcal{M}$  mit  $\mathcal{M}_0 \subseteq \mathcal{M}$  zu entscheiden, ob die gegebenen Daten schon von dem kleineren Modell  $\mathcal{M}_0$  ausreichend gut beschrieben werden, oder ob dazu das kompliziertere Modell  $\mathcal{M}$  nötig ist. Eine allgemeine Einführung zum Thema Likelihood-Quotienten-Tests bietet beispielsweise Anderson [1, Kapitel 4]. Likelihood-Quotienten-Tests in invarianten graphischen Modellen sind dagegen bisher kaum untersucht worden, nur in Madsen [20, Kapitel 8] finden sich einige kurze Bemerkungen dazu.

### 6.1 Testprobleme in invarianten graphischen Modellen

Die essentielle Voraussetzung für die Durchführung eines Likelihood-Quotienten-Tests ist also, dass die beiden untersuchten Modelle geschachtelt sein müssen. Um im Fall von invarianten graphischen Modellen ein wohldefiniertes Testproblem zu bekommen, müssen folglich zunächst Überlegungen zur folgenden Frage angestellt werden: Wie müssen die zerlegbaren Graphen  $\mathcal{U}_0, \mathcal{U}$  und die endlichen Invarianzgruppen  $G_0, G$  aussehen, damit für die korrespondierenden invarianten graphischen Modelle die Schachtelung  $PD(\mathcal{U}_0, G_0) \subseteq PD(\mathcal{U}, G)$  gilt?

Im Falle von rein graphischen Modellen  $PD(\mathcal{U}_0)$  und  $PD(\mathcal{U})$  ist diese Frage einfach zu beantworten. Da in diesen Modellen eine fehlende Kante im Graphen zu einer (bedingten) Unabhängigkeits-Restriktion in den zulässigen Verteilungen führt, gilt die Inklusion  $PD(\mathcal{U}_0) \subseteq PD(\mathcal{U})$  genau dann, wenn  $\mathcal{U}_0$  durch Streichen von Kanten in  $\mathcal{U}$  entsteht, vergleiche auch Gleichung (2.1) auf Seite 12. Wie Likelihood-Quotienten-Tests in rein graphischen Modellen konkret durchgeführt werden können, wird beispielsweise in Lauritzen [18, Kapitel 5.2.2 und 5.3.3] beschrieben. Werden dagegen auch Invarianzen in das Modell eingebunden, so wird die Charakterisierung wohldefinierter Testprobleme schwieriger.

Sei dazu zunächst ein zerlegbarer Graph  $\mathcal{U}$  gegeben. Konstruiere daraus einen azyklisch gemischten Graphen  $\mathcal{V}$ , der Annahme 2.6 erfüllt und dessen Gerüst mit  $\mathcal{U}$  übereinstimmt,  $\mathcal{V}^\sim = \mathcal{U}$ . Damit wiederum lässt sich eine endliche Gruppe  $G \subseteq Orth(\mathcal{V})$  ohne boxeninterne

Symmetrien auswählen, so dass das invariante graphische Modell  $PD(\mathcal{U}, G)$  wohldefiniert ist. Nun sollen  $\mathcal{U}_0$ ,  $\mathcal{V}_0$  und  $G_0$  so gewählt werden, dass das resultierende Modell  $PD(\mathcal{U}_0, G_0)$  eine Teilmenge von  $PD(\mathcal{U}, G)$  ist.

Aus den oben beschriebenen Gründen kann dies nur der Fall sein, wenn  $\mathcal{U}_0$  aus  $\mathcal{U}$  durch Streichen von Kanten entsteht. Damit lässt sich wiederum ein azyklisch gemischter Graph  $\mathcal{V}_0$  auswählen, der Annahme 2.6 erfüllt und dessen Gerüst gerade  $\mathcal{U}_0$  entspricht. Dies ermöglicht schließlich die Wahl einer endlichen Gruppe  $G_0 \subseteq Orth(\mathcal{V}_0)$  ohne boxeninterne Symmetrien. Zusätzlich muss jedoch darauf geachtet werden, dass die ursprüngliche Gruppe  $G$  eine Untergruppe von  $G_0$  ist,  $G \subseteq G_0$ . Damit legt  $G_0$  mehr Symmetriestriktionen als  $G$  fest und aus Gleichung (2.8) von Seite 33 folgt die Inklusion der Parameterräume  $PD(\mathcal{U}_0, G_0) \subseteq PD(\mathcal{U}, G)$ . Folglich ist das Testproblem

$$\Sigma \in PD(\mathcal{U}_0, G_0) \quad \text{vs.} \quad \Sigma \in PD(\mathcal{U}, G)$$

wohldefiniert, wobei  $\Sigma$  die Streuungsmatrix der beobachteten normalverteilten Daten meint.

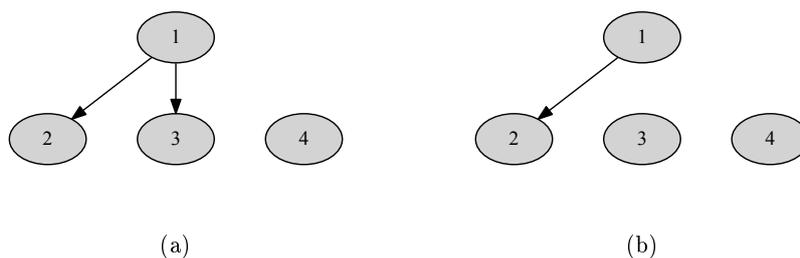
Beachte insbesondere, dass durch diese Konstruktion die Ungleichung

$$\max_{[B] \in \mathcal{B}(\mathcal{V})} \{[B] + \langle B \rangle\} \geq \max_{[A] \in \mathcal{B}(\mathcal{V}_0)} \{[A] + \langle A \rangle\}$$

gilt. Wegen Gleichung (2.3) auf Seite 17 entspricht die linke Seite der Ungleichung nämlich der maximalen Größe einer Clique in  $\mathcal{U}$ , während die rechte Seite die größte Knotenzahl einer Clique in  $\mathcal{U}_0$  angibt. Da  $\mathcal{U}_0$  aus  $\mathcal{U}$  durch Streichen von Kanten entsteht, kann in  $\mathcal{U}_0$  keine größere Clique als in  $\mathcal{U}$  auftauchen und die Ungleichung gilt.

**Bemerkung 6.1** Die beiden Gruppen  $G_0$  und  $G$  müssen insgesamt drei Bedingungen erfüllen. Erstens muss  $G_0 \subseteq Orth(\mathcal{V}_0)$  die Struktur von  $\mathcal{V}_0$  beachten und zweitens muss  $G \subseteq Orth(\mathcal{V})$  die Struktur von  $\mathcal{V}$  respektieren. Drittens muss  $G$  eine Untergruppe von  $G_0$  sein. Diese Liste von Forderungen lässt sich nicht ohne Weiteres vereinfachen, da im Allgemeinen weder  $Orth(\mathcal{V}_0) \subseteq Orth(\mathcal{V})$  noch  $Orth(\mathcal{V}) \subseteq Orth(\mathcal{V}_0)$  gilt.

Um dies einzusehen, betrachte den azyklisch gemischten Graphen  $\mathcal{V}$  in Abbildung 9a. Dessen zulässige Gruppe  $Orth(\mathcal{V})$  besteht aus den Permutationsmatrizen, die zu den Permutationen (1234) und (1324) korrespondieren. Nun werde in  $\mathcal{U} = \mathcal{V}^\sim$  die Kante 1–3 gestrichen und der resultierende Graph sei bezeichnet mit  $\mathcal{U}_0$ . Als korrespondierender azyklisch gemischter Graph werde  $\mathcal{V}_0$  aus Abbildung 9b gewählt. Damit liegt die zuvor zulässige Symmetrie (1324) nicht in  $Orth(\mathcal{V}_0)$ , da in  $\mathcal{V}_0$  der Knoten 2 nicht mit dem Knoten 3 vertauscht werden darf. Im Gegenzug ist aber die für  $\mathcal{V}$  unzulässige Permu-



**Abbildung 9:** Ein azyklisch gemischter Graph  $\mathcal{V}$ , der die Knotenpermutation (1324) zulässt, während (1243) nicht in  $Orth(\mathcal{V})$  liegt (links). Durch Streichen der Kante  $1 \rightarrow 3$  in  $\mathcal{V}$  entsteht der Graph  $\mathcal{V}_0$  (rechts). Dieser erlaubt die Permutation (1324) nicht, im Gegenzug liegt aber (1243) in  $Orth(\mathcal{V}_0)$ .

tation (1243) ein Element von  $Orth(\mathcal{V}_0)$ . Mit anderen Worten ist weder  $Orth(\mathcal{V}_0)$  eine Teilmenge von  $Orth(\mathcal{V})$  noch  $Orth(\mathcal{V})$  eine Teilmenge von  $Orth(\mathcal{V}_0)$ .

Es muss daher im Einzelfall entschieden werden, ob die gewählten Graphen  $\mathcal{U}_0, \mathcal{V}_0, \mathcal{U}$  und  $\mathcal{V}$  ein Testproblem mit angemessenen Invarianzgruppen  $G_0$  und  $G$  zulassen. Sollte dies nicht der Fall sein, so gibt es zwei Möglichkeiten. Entweder modifiziert man die Graphen, um die Wahl von anderen Gruppen  $G_0$  und  $G$  zu ermöglichen. Oder man wählt  $G_0 = \{I_p\} = G$  und testet somit nur die beiden graphischen Modelle  $PD(\mathcal{U}_0)$  und  $PD(\mathcal{U})$  gegeneinander.

**Bemerkung 6.2 (Spezielle Testprobleme)**

Ein interessanter Spezialfall, in dem die gerade beschriebene Problematik vermieden wird, ist die Wahl  $\mathcal{U}_0 = \mathcal{U}$  und  $\mathcal{V}_0 = \mathcal{V}$ . Dann ist in beiden Modellen die Abhängigkeitsstruktur der Variablen identisch und somit werden nur die vorgegebenen Symmetrien getestet. An die Invarianzgruppen muss in diesem Fall nur die Forderung  $G \subseteq G_0 \subseteq Orth(\mathcal{V})$  gestellt werden.

Der dazu duale Fall, das heißt die Abhängigkeitsstrukturen der beiden Modelle werden getestet, während identische Invarianzrestriktionen angenommen werden, gestaltet sich leider nicht so einfach. Wird nämlich  $G_0 = G \subseteq Orth(\mathcal{V})$  fixiert, so ist man bei der Wahl von  $\mathcal{U}_0$  und  $\mathcal{V}_0$  unter Umständen erheblich eingeschränkt. Insbesondere sind nur solche Graphen  $\mathcal{V}_0$  zulässig, für die  $G \subseteq Orth(\mathcal{V}_0)$  ist.

Seien daher im Folgenden  $\mathcal{U}_0 = (V, \emptyset, U_0)$  und  $\mathcal{U} = (V, \emptyset, U)$  zwei zerlegbare Graphen mit derselben Knotenmenge  $V$ , wobei  $U_0 \subseteq U$  ist. Des Weiteren seien  $\mathcal{V}_0$  und  $\mathcal{V}$  zwei azyklisch gemischte Graphen mit Gerüsten  $(\mathcal{V}_0)^\sim = \mathcal{U}_0$  und  $\mathcal{V}^\sim = \mathcal{U}$ , die Annahme 2.6 er-

füllen. Schließlich seien  $G_0$  und  $G$  zwei endliche Gruppen ohne boxeninterne Symmetrien, die  $G_0 \subseteq Orth(\mathcal{V}_0)$ , sowie  $G \subseteq Orth(\mathcal{V})$  und  $G \subseteq G_0$  erfüllen. Die Glättungsfunktionen bezüglich der beiden Gruppen seien bezeichnet mit  $\psi_0 \equiv \psi_{G_0}$  und  $\psi \equiv \psi_G$ , vergleiche Gleichung (4.2) auf Seite 58. Außerdem seien die Maximum-Likelihood-Schätzer der Kovarianzmatrizen in den invarianten graphischen Modellen  $PD(\mathcal{U}_0, G_0)$  und  $PD(\mathcal{U}, G)$  bezeichnet mit  $\hat{\Sigma}_0$  und  $\hat{\Sigma}$ , vergleiche Theorem 4.3 auf Seite 58. Die dazu korrespondierenden umparametrisierten Schätzer seien bezeichnet mit  $\hat{\Omega}_0 = \pi(\hat{\Sigma}_0) \in P(\mathcal{U}_0, G_0)$  und  $\hat{\Omega} = \pi(\hat{\Sigma}) \in P(\mathcal{U}, G)$ , vergleiche Lemma 2.4 auf Seite 13. Schließlich sei die  $G$ -reduzierte Version von  $\mathcal{V}$  aus Lemma 5.8 von Seite 75 bezeichnet mit  $\mathcal{V}_G$  und die  $G_0$ -reduzierte Version von  $\mathcal{V}_0$  aus demselben Lemma heiße  $\mathcal{V}_{G_0}$ . Die korrespondierenden Repräsentantenmengen seien  $\mathcal{B}(\mathcal{V})_{\sim} = \mathcal{B}(\mathcal{V}_G)$  und  $\mathcal{B}(\mathcal{V}_0)_{\sim} = \mathcal{B}(\mathcal{V}_{G_0})$ .

Die Likelihood-Quotienten-Teststatistik ergibt sich nun, indem die beiden Maximum-Likelihood-Schätzer  $\hat{\Sigma}_0$  und  $\hat{\Sigma}$  in die Likelihood-Funktion eingesetzt werden und daraus der Quotient gebildet wird. Wegen Korollar 5.14 ergibt sich

$$Q := \frac{L_S(\hat{\Sigma}_0)}{L_S(\hat{\Sigma})} = \left( \frac{\prod_{[B] \in \mathcal{B}(\mathcal{V})_{\sim}} \det[\psi(W)_{[B]\bullet}]^{|Orb[B]|}}{\prod_{[A] \in \mathcal{B}(\mathcal{V}_0)_{\sim}} \det[\psi_0(W)_{[A]\bullet}]^{|Orb[A]|}} \right)^{\frac{n}{2}}, \quad (6.1)$$

wobei  $W = \pi(S)$  wieder die Projektion der Stichproben-Kovarianzmatrix  $S$  ist. Es sei hervorgehoben, dass bei der Interpretation der hier auftretenden Cholesky-Parameter Vorsicht geboten ist. Da sich die kausalen Strukturen von  $\mathcal{V}_0$  und  $\mathcal{V}$  unterscheiden können, muss darauf geachtet werden, bezüglich welches Graphen die Cholesky-Parameter gemeint sind. Der Deutlichkeit halber werden hier und im Folgenden die Boxen des Graphen  $\mathcal{V}_0$  mit  $[A]$  bezeichnet. Die Indizierung „ $[A]\bullet$ “ bezieht sich dann auf den Graphen  $\mathcal{V}_0$ . Boxen des Graphen  $\mathcal{V}$  werden dagegen mit  $[B]$  gekennzeichnet und der Ausdruck „ $[B]\bullet$ “ ist dementsprechend bezüglich  $\mathcal{V}$  definiert.

Da  $PD(\mathcal{U}_0, G_0)$  ein Teilmodell von  $PD(\mathcal{U}, G)$  ist, impliziert die Existenz des Schätzers  $\hat{\Sigma}$  schon die Existenz von  $\hat{\Sigma}_0$ . Wegen Bemerkung 4.4 auf Seite 59 ist damit ein hinreichendes Kriterium für die fast sichere Existenz der Likelihood-Quotienten-Statistik gegeben durch

$$n > [B] + \langle B \rangle - 1 \quad \text{für alle} \quad [B] \in \mathcal{B}(\mathcal{V})_{\sim}.$$

Später wird die Größe auf der rechten Seite der Ungleichung sowohl für alle Boxen in  $\mathcal{V}$  als auch für alle Boxen in  $\mathcal{V}_0$  von Interesse sein. Um die Notation zu vereinfachen, sei

deswegen im Folgenden

$$\mathcal{B}(\mathcal{V}_0, \mathcal{V})_{\sim} := \mathcal{B}(\mathcal{V}_0)_{\sim} \cup \mathcal{B}(\mathcal{V})_{\sim}. \quad (6.2)$$

Um mit Hilfe der Statistik  $Q$  einen konkreten Likelihood-Quotienten-Test durchführen zu können, soll im nächsten Abschnitt die Verteilung von  $Q$  näher beschrieben werden.

## 6.2 Die Verteilung der Likelihood-Quotienten-Teststatistik

Seien  $\mathcal{U}_0, \mathcal{U}, \mathcal{V}_0, \mathcal{V}, G_0$  und  $G$  wie im letzten Kapitel, so dass  $PD(\mathcal{U}_0, G_0)$  ein Teilmodell von  $PD(\mathcal{U}, G)$  darstellt. Des Weiteren existiere der Maximum-Likelihood-Schätzer  $\hat{\Sigma} \in PD(\mathcal{U}, G)$ , so dass insbesondere auch die Likelihood-Quotienten-Statistik  $Q$  existiert. Das Ziel dieses Abschnitts wird sein, die Verteilung von  $Q$  zu bestimmen. Die Erfahrung zeigt jedoch, dass bei vielen Modellen in der multivariaten Statistik die exakte Verteilung der Likelihood-Quotienten-Testgröße  $Q$  nicht greifbar ist. So lässt sie sich beispielsweise schon für rein graphische Modelle im Allgemeinen nicht bestimmen, vergleiche Kapitel 5.2.2 in Lauritzen [18].

In solchen Fällen muss daher auf Approximationsmethoden zurückgegriffen werden. Besonders akkurate Näherungsverfahren wurden für Verteilungen entwickelt, die in der sogenannten „Box-Familie“ liegen. Diese große Familie von Verteilungen wurde von Box in seinem Aufsatz [8] eingeführt. Sei dazu  $Y$  eine  $(0, \infty)$ -wertige Zufallsvariable. Dann folgt  $Y$  genau dann einer Verteilung aus der Box-Familie, wenn ihre momenterzeugende Funktion für alle  $t$  in einer Umgebung von Null die Identität

$$\mathbb{E} [e^{tY}] = e^{2tz} \prod_{i=1}^a \frac{\Gamma[u_i(1-2t) + \xi_i]}{\Gamma[u_i + \xi_i]} \prod_{j=1}^b \frac{\Gamma[v_j + \eta_j]}{\Gamma[v_j(1-2t) + \eta_j]} \quad (6.3)$$

erfüllt. Dabei müssen die reellen Konstanten  $z, u_i, \xi_i, v_j$  und  $\eta_j$  für alle  $i$  und alle  $j$  wie folgt aussehen: Die Größen  $u_i$  und  $v_j$  müssen positiv sein. Außerdem müssen  $\xi_i$  und  $\eta_j$  so gewählt werden, dass auch  $u_i + \xi_i$  und  $v_j + \eta_j$  positiv sind. Schließlich müssen die beiden Identitäten

$$\sum_{i=1}^a u_i = \sum_{j=1}^b v_j \quad \text{und} \quad z = \sum_{i=1}^a u_i \log(u_i) - \sum_{j=1}^b v_j \log(v_j)$$

erfüllt sein. In diesem Fall gilt die Darstellung (6.3) für alle  $t \in \mathbb{R}$ , für die die Argumente der auftretenden Gammafunktionen positiv sind. Dies ist sowohl für alle negativen  $t$  als auch für alle ausreichend kleinen positiven  $t$  erfüllt, da  $u_i + \xi_i$  und  $v_j + \eta_j$  für alle  $i$  und  $j$  positiv sein müssen.

Um nachzuweisen, dass eine Zufallsvariable einer Verteilung dieses Typs folgt, müssen damit insbesondere höhere Momente bestimmt werden. Im folgenden Theorem wird dies für die Likelihood-Quotienten-Statistik  $Q$  durchgeführt.

**Theorem 6.3 (Momente der Likelihood-Quotienten-Statistik)**

Seien  $\mathcal{U}_0, \mathcal{U}, \mathcal{V}_0, \mathcal{V}, G_0$  und  $G$  wie in Kapitel 6.1, so dass insbesondere  $PD(\mathcal{U}_0, G_0) \subseteq PD(\mathcal{U}, G)$  ist. Seien außerdem  $X_1, \dots, X_n \sim \mathcal{N}_p(0, \Sigma)$  unabhängige Zufallsvektoren mit Streuungsmatrix  $\Sigma \in PD(\mathcal{U}, G)$ . Wie bisher seien  $S := \frac{1}{n} \sum_{i=1}^n X_i X_i^T$  die Stichproben-Kovarianzmatrix und  $W := \pi(S)$  deren Projektion. Des Weiteren sei  $n \geq [B] + \langle B \rangle$  für alle Boxen  $[B] \in \mathcal{B}(\mathcal{V}_0, \mathcal{V})_{\sim}$ , so dass die Likelihood-Quotienten-Statistik  $Q$  aus Gleichung (6.1) fast sicher existiert. Unter der Nullhypothese, das heißt, wenn  $\Sigma \in PD(\mathcal{U}_0, G_0)$  ist, gilt dann die Identität

$$\mathbb{E}[Q^t] = \frac{\prod_{[B] \in \mathcal{B}(\mathcal{V})_{\sim}} \left\{ |Orb[B]|^{-\frac{n}{2}t} |Orb[B]| \prod_{i=1}^{[B]} \frac{\Gamma\left(\frac{n}{2}|Orb[B]|(t+1) - \frac{\langle B \rangle + i - 1}{2}\right)}{\Gamma\left(\frac{n}{2}|Orb[B]| - \frac{\langle B \rangle + i - 1}{2}\right)} \right\}}{\prod_{[A] \in \mathcal{B}(\mathcal{V}_0)_{\sim}} \left\{ |Orb[A]|^{-\frac{n}{2}t} |Orb[A]| \prod_{j=1}^{[A]} \frac{\Gamma\left(\frac{n}{2}|Orb[A]|(t+1) - \frac{\langle A \rangle + j - 1}{2}\right)}{\Gamma\left(\frac{n}{2}|Orb[A]| - \frac{\langle A \rangle + j - 1}{2}\right)} \right\}}$$

für alle  $t > \max \left\{ \frac{[B] + \langle B \rangle - 1}{n|Orb[B]|} - 1 \mid [B] \in \mathcal{B}(\mathcal{V}_0, \mathcal{V})_{\sim} \right\}$ .

**Beweis**

Per Definition der Likelihood-Quotienten-Statistik gilt die Identität

$$Q^{\frac{2\delta}{n}} \cdot \prod_{[A] \in \mathcal{B}(\mathcal{V}_0)_{\sim}} \det(\psi_0(W)_{[A]\bullet})^{\delta|Orb[A]|} = \prod_{[B] \in \mathcal{B}(\mathcal{V})_{\sim}} \det(\psi(W)_{[B]\bullet})^{\delta|Orb[B]|} \quad (6.4)$$

für alle  $\delta \in \mathbb{R}$ , siehe Gleichung (6.1). Des Weiteren sind die Likelihood-Quotienten-Testgröße  $Q$  und die Cholesky-Parameter des Maximum-Likelihood-Schätzers  $\hat{\Sigma}_0$  unter der Nullhypothese unabhängig,

$$Q \perp \left\{ \psi_0(W)_{[A]\bullet} \right\}_{[A] \in \mathcal{B}(\mathcal{V}_0)_{\sim}},$$

vergleiche Madsen [20, Kapitel 8]. Damit folgt aus Gleichung (6.4) die Identität

$$\mathbb{E}\left[Q^{\frac{2\delta}{n}}\right] = \frac{\mathbb{E}\left[\prod_{[B] \in \mathcal{B}(\mathcal{V})_{\sim}} \det(\psi(W)_{[B]\bullet})^{\delta|Orb[B]|}\right]}{\mathbb{E}\left[\prod_{[A] \in \mathcal{B}(\mathcal{V}_0)_{\sim}} \det(\psi_0(W)_{[A]\bullet})^{\delta|Orb[A]|}\right]}. \quad (6.5)$$

Die Erwartungswerte im Zähler und im Nenner lassen sich mit Hilfe von Korollar 5.15

auf Seite 84 bestimmen. Es ergibt sich

$$\mathbb{E} \left[ Q^{\frac{2\delta}{n}} \right] = \frac{\prod_{[B] \in \mathcal{B}(\mathcal{V})_{\sim}} \left\{ |Orb[B]|^{-\delta [B] |Orb[B]} \prod_{i=1}^{[B]} \frac{\Gamma\left(\left(\delta + \frac{n}{2}\right) |Orb[B]| - \frac{\langle B \rangle + i - 1}{2}\right)}{\Gamma\left(\frac{n}{2} |Orb[B]| - \frac{\langle B \rangle + i - 1}{2}\right)} \right\}}{\prod_{[A] \in \mathcal{B}(\mathcal{V}_0)_{\sim}} \left\{ |Orb[A]|^{-\delta [A] |Orb[A]} \prod_{j=1}^{[A]} \frac{\Gamma\left(\left(\delta + \frac{n}{2}\right) |Orb[A]| - \frac{\langle A \rangle + j - 1}{2}\right)}{\Gamma\left(\frac{n}{2} |Orb[A]| - \frac{\langle A \rangle + j - 1}{2}\right)} \right\}}$$

für alle  $\delta > \max_{[B] \in \mathcal{B}(\mathcal{V}_0, \mathcal{V})_{\sim}} \left\{ \frac{[B] + \langle B \rangle - 1}{2|Orb[B]|} - \frac{n}{2} \right\}$ , siehe Gleichung (6.2). Die Behauptung folgt dann durch die Substitution  $t := \frac{2\delta}{n}$ . Beachte, dass dadurch  $t > \frac{[B] + \langle B \rangle - 1}{n|Orb[B]|} - 1$  für alle  $[B] \in \mathcal{B}(\mathcal{V}_0, \mathcal{V})_{\sim}$  sein muss.  $\square$

Mit Hilfe dieses Theorems lässt sich zeigen, dass die Verteilung der *Devianz*  $D := -2 \log(Q)$  in die Familie der Box-Verteilungen fällt.

#### Korollar 6.4 (Box-Eigenschaft der Devianz)

Seien die Voraussetzungen wie in Theorem 6.3. Dann folgt die Devianz  $D = -2 \log(Q)$  unter der Nullhypothese einer Verteilung aus der Box-Familie.

#### Beweis

Wegen Theorem 6.3 ist

$$\begin{aligned} \mathbb{E} \left[ e^{tD} \right] &= \mathbb{E} \left[ e^{-2t \log(Q)} \right] = \mathbb{E} \left[ Q^{-2t} \right] = \\ &= \frac{\prod_{[B] \in \mathcal{B}(\mathcal{V})_{\sim}} |Orb[B]|^{nt [B] |Orb[B]} \prod_{[B] \in \mathcal{B}(\mathcal{V})_{\sim}} \prod_{i=1}^{[B]} \frac{\Gamma\left(\frac{n}{2} |Orb[B]| (1-2t) - \frac{\langle B \rangle + i - 1}{2}\right)}{\Gamma\left(\frac{n}{2} |Orb[B]| - \frac{\langle B \rangle + i - 1}{2}\right)}}{\prod_{[A] \in \mathcal{B}(\mathcal{V}_0)_{\sim}} |Orb[A]|^{nt [A] |Orb[A]} \prod_{[A] \in \mathcal{B}(\mathcal{V}_0)_{\sim}} \prod_{j=1}^{[A]} \frac{\Gamma\left(\frac{n}{2} |Orb[A]| (1-2t) - \frac{\langle A \rangle + j - 1}{2}\right)}{\Gamma\left(\frac{n}{2} |Orb[A]| - \frac{\langle A \rangle + j - 1}{2}\right)}} \end{aligned} \quad (6.6)$$

für alle  $t < \min_{[B] \in \mathcal{B}(\mathcal{V}_0, \mathcal{V})_{\sim}} \left\{ \frac{1}{2} - \frac{[B] + \langle B \rangle - 1}{2n|Orb[B]|} \right\}$ . Definiere nun die Koeffizienten

$$\begin{aligned} u_{[B],i} &:= \frac{n}{2} |Orb[B]|, & \xi_{[B],i} &:= -\frac{\langle B \rangle + i - 1}{2} & \forall [B] \in \mathcal{B}(\mathcal{V})_{\sim}, \forall i \in \{1, \dots, [B]\}, \\ v_{[A],j} &:= \frac{n}{2} |Orb[A]|, & \eta_{[A],j} &:= -\frac{\langle A \rangle + j - 1}{2} & \forall [A] \in \mathcal{B}(\mathcal{V}_0)_{\sim}, \forall j \in \{1, \dots, [A]\}. \end{aligned} \quad (6.7)$$

Damit sind  $u_{[B],i}$  und  $v_{[A],j}$  in jedem Fall positiv. Dies gilt auch für  $u_{[B],i} + \xi_{[B],i}$  und  $v_{[A],j} + \eta_{[A],j}$ , da  $n$  per Voraussetzung für alle  $[B] \in \mathcal{B}(\mathcal{V})_{\sim}$  und alle  $[B] \in \mathcal{B}(\mathcal{V}_0)_{\sim}$  größer

ist als  $[B] + \langle B \rangle - 1$ . Außerdem ist für die obige Wahl

$$\begin{aligned} \sum_{[B] \in \mathcal{B}(\mathcal{V})_{\sim}} \sum_{i=1}^{[B]} u_{[B],i} &= \frac{n}{2} \sum_{[B] \in \mathcal{B}(\mathcal{V})_{\sim}} [B] |Orb[B]| = \frac{np}{2} = \\ &= \frac{n}{2} \sum_{[A] \in \mathcal{B}(\mathcal{V}_0)_{\sim}} [A] |Orb[A]| = \sum_{[A] \in \mathcal{B}(\mathcal{V}_0)_{\sim}} \sum_{j=1}^{[A]} v_{[A],j}, \end{aligned}$$

siehe Gleichung (5.26) und beachte, dass  $\mathcal{V}_0$  und  $\mathcal{V}$  gleich viele Knoten aufweisen. Des Weiteren gilt

$$\begin{aligned} z &:= \sum_{[B] \in \mathcal{B}(\mathcal{V})_{\sim}} \sum_{i=1}^{[B]} u_{[B],i} \log(u_{[B],i}) - \sum_{[A] \in \mathcal{B}(\mathcal{V}_0)_{\sim}} \sum_{j=1}^{[A]} v_{[A],j} \log(v_{[A],j}) = \\ &= \sum_{[B] \in \mathcal{B}(\mathcal{V})_{\sim}} \frac{n}{2} [B] |Orb[B]| \log\left(\frac{n}{2} |Orb[B]|\right) - \sum_{[A] \in \mathcal{B}(\mathcal{V}_0)_{\sim}} \frac{n}{2} [A] |Orb[A]| \log\left(\frac{n}{2} |Orb[A]|\right) \end{aligned}$$

und damit folgt die Identität

$$e^z = \frac{\prod_{[B] \in \mathcal{B}(\mathcal{V})_{\sim}} \left(\frac{n}{2} |Orb[B]|\right)^{\frac{n}{2} [B] |Orb[B]|}}{\prod_{[A] \in \mathcal{B}(\mathcal{V}_0)_{\sim}} \left(\frac{n}{2} |Orb[A]|\right)^{\frac{n}{2} [A] |Orb[A]|}} = \frac{\prod_{[B] \in \mathcal{B}(\mathcal{V})_{\sim}} |Orb[B]|^{\frac{n}{2} [B] |Orb[B]|}}{\prod_{[A] \in \mathcal{B}(\mathcal{V}_0)_{\sim}} |Orb[A]|^{\frac{n}{2} [A] |Orb[A]|}},$$

siehe wieder Gleichung (5.26). Einsetzen dieser Identität in Gleichung (6.6) liefert unter Beachtung der Definitionen in Gleichung (6.7) schließlich

$$\mathbb{E}[e^{tD}] = e^{2tz} \prod_{[B] \in \mathcal{B}(\mathcal{V})_{\sim}} \prod_{i=1}^{[B]} \frac{\Gamma[u_{[B],i}(1-2t) + \xi_{[B],i}]}{\Gamma[u_{[B],i} + \xi_{[B],i}]} \prod_{[A] \in \mathcal{B}(\mathcal{V}_0)_{\sim}} \prod_{j=1}^{[A]} \frac{\Gamma[v_{[A],j} + \eta_{[A],j}]}{\Gamma[v_{[A],j}(1-2t) + \eta_{[A],j}]}.$$

Damit ist die Verteilung der Devianz  $D$  unter der Nullhypothese ein Mitglied der Box-Familie, vergleiche Gleichung (6.3) und die Erläuterungen darunter. Abschließend sei hervorgehoben, dass durch die Voraussetzung an die Stichprobengröße  $n$  in Theorem 6.3 die Abschätzung

$$\frac{1}{2} - \frac{[B] + \langle B \rangle - 1}{2n |Orb[B]|} > \frac{1}{2} - \frac{1}{2 |Orb[B]|} \geq 0$$

für alle Boxen  $[B] \in \mathcal{B}(\mathcal{V}_0, \mathcal{V})_{\sim}$  gilt. Folglich hält Gleichung (6.6) insbesondere für alle  $t$  in einer Umgebung um Null.  $\square$

Korollar 6.4 ist auch aus praktischer Sicht sehr nützlich, da für Verteilungen der Box-Familie sehr akkurate Approximationsverfahren bekannt sind. Die ersten dieser Näherungsmethoden stellte Box in seinem Aufsatz [8] vor, sie wurden jedoch später von Anderson [1] und Jensen [15] jeweils verbessert.

Die meisten Approximationen beruhen darauf, dass eine positive Zufallsvariable  $Y$  mit momenterzeugender Funktion wie in Gleichung (6.3) asymptotisch  $\chi^2$ -verteilt ist mit Freiheitsgrad

$$f = \left( a - 2 \sum_{i=1}^a \xi_i \right) - \left( b - 2 \sum_{j=1}^b \eta_j \right),$$

vergleiche Jensen [15, Kapitel 2]. Eine Verbesserung dieser Approximation wird von Anderson in Kapitel 8.5 seines Buchs [1] vorgeschlagen. Er verwendet eine Bartlett-Korrektur, das heißt eine gewisse Transformation der Zufallsvariablen  $Y$ , um die asymptotische Näherung an die  $\chi^2$ -Verteilung zu verbessern. Auch Jensen schlägt in seinem Aufsatz [15] solch eine Modifikation vor.

In der letztgenannten Arbeit werden jedoch sogar zwei weitere Verbesserungen vorgeschlagen. Die erste versucht, die Verteilung von  $Y$  durch eine Gamma-Verteilung mit den „richtigen“ ersten beiden Momenten zu approximieren. Diese Momente lassen sich mit Hilfe der Darstellung der momenterzeugenden Funktion in Gleichung (6.3) auf einfache Weise berechnen. Die zweite Methode verwendet eine Approximation für exponentielle Verteilungsfamilien, um den Schwanz  $\mathbb{P}[Y \geq y]$  der Verteilung von  $Y$  anzunähern, vergleiche Lauritzen [18, Gleichung (C.13)]. In numerischen Studien zeigt sich, dass dieses Verfahren meist die beste Näherung an die exakte Verteilung liefert, vergleiche Tabelle 1 in Jensen [15].

Um einen Likelihood-Quotienten-Test konkret durchzuführen, kann man also wie folgt vorgehen. Aus der vorgegebenen Stichproben-Kovarianzmatrix  $S$  bestimmt man die konkrete Ausprägung  $q \equiv q(S)$  der Likelihood-Quotienten-Teststatistik  $Q$  gemäß Gleichung (6.1). Aus Korollar 6.4 folgt, dass die Devianz  $-2\log(Q)$  einer Verteilung vom Box-Typ folgt. Insbesondere lässt sich damit die Wahrscheinlichkeit

$$\mathbb{P}[-2\log(Q) > -2\log(q)] \tag{6.8}$$

mit Hilfe der Methode aus Jensen [15] approximieren. Sie beschreibt, wie wahrscheinlich es ist, dass die Devianz  $-2\log(Q)$  einen größeren Wert annimmt als es für die gegebenen Daten  $S$  der Fall ist. In der Literatur wird der Ausdruck (6.8) oft als der *p-Wert* der Devianz bezeichnet. Ist der p-Wert kleiner als ein zuvor festgelegtes Signifikanzniveau

$\alpha \in (0, 1)$ , so wird die Stichprobe als signifikant gewertet und die Nullhypothese „ $\Sigma \in PD(\mathcal{U}_0, G_0)$ “ wird verworfen. Liegt der p-Wert dagegen über dem Signifikanzniveau  $\alpha$ , so kann die Nullhypothese akzeptiert werden und das kleine Modell  $PD(\mathcal{U}_0, G_0)$  ist zur Beschreibung der Daten ausreichend.

## 7 Abschließende Bemerkungen und Ausblick

Graphische Modelle sind ein nützliches Hilfsmittel bei der Modellierung von Abhängigkeitsstrukturen in einem multivariaten Datensatz. In praktischen Anwendungen sind jedoch oft sehr viele Variablen beteiligt, was dazu führt, dass die entsprechenden Modelle sehr komplex werden und nicht mehr effizient analysiert werden können. In solch einem Fall kann die Einbindung von Symmetrien Abhilfe schaffen, da dadurch die Anzahl der Modellparameter erheblich reduziert werden kann. Die resultierenden invarianten graphischen Modelle mit Normalverteilungsannahme stellen das zentrale Thema dieser Dissertation dar.

Ein beliebtes Schätzverfahren für die Kovarianzstruktur in solchen Modellen ist die Berechnung des Maximum-Likelihood-Schätzers. Sie wird ausführlich im Aufsatz von Madsen [20] behandelt, wo unter anderem hinreichende und notwendige Kriterien für die Existenz des Schätzers entwickelt werden. Außerdem wird ein expliziter Ausdruck für den Schätzer angegeben. Soll ein konkreter Datensatz analysiert werden, so ist die Kenntnis des Schätzers allein allerdings noch nicht ausreichend. Um ihn angemessen beurteilen zu können, muss auch dessen Verteilung bekannt sein.

Daher widmet sich Kapitel 5 der vorliegenden Arbeit der Aufgabe, diese Verteilung zu bestimmen. Es zeigt sich, dass es sich um eine verallgemeinerte Riesz-Verteilung handelt, wie sie in Andersson und Klein [3] definiert wird. Insbesondere sei hervorgehoben, dass die einzelnen Formparameter dieser Riesz-Verteilung nicht identisch sind. Dies zeigt nachträglich, dass die Einführung eines multivariaten Formparameters durch Andersson und Klein [3] durchaus einen großen Nutzen mit sich bringt.

Anschließend wird in Kapitel 6 die Teststatistik eines Likelihood-Quotienten-Tests und deren Verteilung unter der Nullhypothese untersucht. Da die exakte Verteilung selbst wohl nicht greifbar ist, werden stattdessen die Momente der Testgröße bestimmt. Dies ermöglicht es nachzuweisen, dass die entsprechende Verteilung in die Box-Familie fällt, so dass sehr akkurate Approximationsverfahren verwendet werden können.

Abschließend sei betont, dass wann immer von einem „graphischen Modell“ die Rede ist, ein Modell gemeint ist, das mit Hilfe eines Graphen bedingte Unabhängigkeitsforderungen an seine Verteilungen stellt. In Normalverteilungsmodellen ist dies äquivalent dazu, dass gewisse Einträge in der Inversen der Kovarianzmatrix, der sogenannten Konzentrationsmatrix, verschwinden. Man könnte also etwas genauer auch von „graphischen Konzentrationsmodellen“ sprechen. In den letzten Jahren wurden allerdings auch verstärkt Modelle untersucht, in denen eine fehlende Kante im Modellgraph zu einer verschwindenden Komponente in der Kovarianzmatrix selbst korrespondiert. Mit anderen

Worten kodiert der Modellgraph dann keine bedingten Unabhängigkeiten mehr, sondern marginale Unabhängigkeiten. Solche „graphischen Kovarianzmodelle“ werden beispielsweise in Drton, Foygel und Sullivant [11] und Drton, Fox und Käuff [10] untersucht. Außerdem liefern die letztgenannten Autoren in ihrem Kommentar [12] einen kleinen Überblick über denkbare Fragestellungen in diesen Modellen.

Da es sich bei solchen graphischen Kovarianzmodellen im Allgemeinen nicht mehr um reguläre, sondern um gekrümmte Exponentialfamilien handelt, sind sie allerdings oft schwieriger zu analysieren als die hier untersuchten graphischen Konzentrationsmodelle. Trotzdem könnte es sich als sinnvoll erweisen, auch in ein Kovarianzmodell Invarianzen einzubinden und so die Dimension des Parameterraums zu reduzieren. Möglicherweise lässt sich auch in diesem Fall die exakte Verteilung des Maximum-Likelihood-Schätzers beschreiben. Einen Schritt in diese Richtung gehen Khare und Rajaratnam [16], die in ihrem Aufsatz eine verallgemeinerte Wishart-Verteilung entwickeln, die auf dem Parameterraum eines graphischen Kovarianzmodells konzentriert ist.

---

**Literatur**

- [1] Theodore Wilbur Anderson. *An Introduction to Multivariate Statistical Analysis*. Wiley, New York, 1984.
- [2] Steen Arne Andersson. Invariant Normal Models. *Annals of Statistics*, 3:132–154, 1975.
- [3] Steen Arne Andersson und Thomas Klein. On Riesz and Wishart Distributions Associated with Decomposable Undirected Graphs. *Journal of Multivariate Analysis*, 101:789–810, 2010.
- [4] Steen Arne Andersson und Jesper Madsen. Symmetry and Lattice Conditional Independence Models in a Multivariate Normal Distribution. *Annals of Statistics*, 26:525–572, 1998.
- [5] Steen Arne Andersson und Michael David Perlman. Normal Linear Regression Models with recursive graphical Markov Structure. *Journal of Multivariate Analysis*, 66:133–187, 1998.
- [6] Steen Arne Andersson und Gregory Gerard Wojnar. Wishart Distributions on Homogeneous Cones. *Journal of Theoretical Probability*, 17:781–818, 2004.
- [7] Steen Arne Andersson, David Madigan, und Michael David Perlman. On the Markov Equivalence of Chain Graphs, Undirected Graphs, and Acyclic Digraphs. *Scandinavian Journal of Statistics*, 24:81–102, 1997.
- [8] George Edward Pelham Box. A General Distribution Theory for a Class of Likelihood Criteria. *Biometrika*, 36:317–346, 1949.
- [9] Philip Dawid und Steffen Lauritzen. Hyper Markov Laws in the Statistical Analysis of Decomposable Graphical Models. *Annals of Statistics*, 21:1272–1317, 1993.
- [10] Mathias Drton, Chris Fox, und Andreas Käüfl. On the Existence of Maximum Likelihood Estimators in Linear Structural Equation Models. In Vorbereitung.
- [11] Mathias Drton, Rina Foygel, und Seth Sullivant. Global identifiability of linear structural equation models. *Annals of Statistics*, 39:865–886, 2011.
- [12] Mathias Drton, Chris Fox, und Andreas Käüfl. Comments on: Sequences of regressions and their independencies. *TEST*, 21:255–261, 2012.
- [13] Morris Leroy Eaton. *Multivariate Statistics – A vector space approach*. Wiley, New York, 1983.
- [14] David Edwards. *Introduction to Graphical Modelling*. Springer, New York, 1995.
- [15] Jens Ledet Jensen. A Large Deviation-Type Approximation for the Box Class of Likelihood Ratio Criteria. *Journal of the American Statistical Association*, 86:437–440, 1991.

- [16] Kshitij Khare und Bala Rajaratnam. Wishart Distributions for Decomposable Covariance Graph Models. *Annals of Statistics*, 39:514–555, 2011.
- [17] Tonu Kollo und Dietrich von Rosen. *Advanced Multivariate Statistics with Matrices*. Springer, Dordrecht, 2005.
- [18] Steffen Lauritzen. *Graphical Models*. Oxford University Press, New York, 1996.
- [19] Gérard Letac und Hélène Massam. Wishart Distributions for Decomposable Graphs. *Annals of Statistics*, 35:1278–1323, 2007.
- [20] Jesper Madsen. Invariant Normal Models With Recursive Graphical Markov Structure. *Annals of Statistics*, 28:1150–1178, 2000.
- [21] John Monahan. *A primer on linear models*. Chapman & Hall/CRC, 2008.
- [22] Michael David Perlman. Comment: Group Symmetry Covariance Models. *Statistical Science*, 2:421–425, 1987.
- [23] Mark R. Sepanski. *Algebra*. American Mathematical Society, 2010.
- [24] Parikshit Shah und Venkat Chandrasekaran. Group Symmetry and Covariance Regularization. *ArXiv e-prints*, 2011.
- [25] Joe Whittaker. *Graphical Models in Applied Multivariate Statistics*. Wiley, Chichester, 1990.

# Andreas Käufl

## Lebenslauf

---

### Studium

- April 2011** Abschluss der integrierten Master-Phase des Promotionsstudiengangs „TopMath“ mit einer Arbeit zum Thema „Die Modellierung von Abhängigkeitsstrukturen mittels diskreter graphischer Modelle“ bei Prof. Dr. Friedrich Pukelsheim; Durchschnittsnote aller Leistungen im Master: 1.0
- September 2009** Abschluss der Bachelor-Phase von „TopMath“ mit einer Arbeit zum Thema „Die Modellierung von Enthaltungen in gewichteten Abstimmungssystemen“ bei Prof. Dr. Friedrich Pukelsheim; Durchschnittsnote aller Leistungen im Bachelor: 1.1
- seit Oktober 2008** Mathematischer Elite-Studiengang „TopMath – Angewandte Mathematik mit Promotion“ (TU München und Universität Augsburg), der in eine Bachelor- und eine kombinierte Master- und Promotions-Phase gegliedert ist
- Oktober 2006** Studiengang „Wirtschaftsmathematik B. Sc.“ an der Universität Augsburg

---

### Schulbildung

- 1997 – 2006** Besuch des Leonhard-Wagner-Gymnasiums Schwabmünchen sowie Erwerb der allgemeinen Hochschulreife mit einer Durchschnittsnote von 1.3
- 1993 – 1997** Besuch der Grundschule Straßberg

---

### Stipendien und Auszeichnungen

- Juli 2011** Aufnahme in die Promotionsförderung der Studienstiftung des deutschen Volkes
- November 2009** Verleihung eines mit 300 Euro dotierten TopMath-Study-Awards für Leistungen während der Bachelor-Phase
- Juli 2008** Aufnahme in die Studienförderung der Studienstiftung des deutschen Volkes
- Juni 2008** Aufnahme in den Elite-Studiengang „TopMath – Angewandte Mathematik mit Promotion“ der TU München und der Universität Augsburg
- Juni 2006** Verleihung eines Buchpreises als Zweitbester des Abiturjahrgangs

---

### Berufspraxis

- August/September 2008** Zweimonatiges Praktikum in der Controlling-Abteilung von Böwe Bell+Howell in Durham, North Carolina (USA).
- seit Oktober 2007** Tätigkeit als Übungsgruppenleiter für verschiedene Vorlesungen: Informatik I, Stochastik für Lehramt, Optimierung I, Wahrscheinlichkeitstheorie, Statistik I

---

### Auslandsaufenthalte

- August – Dezember 2011** Fünfmonatiger Forschungsaufenthalt am Department of Statistics der University of Chicago auf Einladung von Prof. Mathias Drton.
- Sommer 2008** Zweimonatiges Praktikum bei Böwe Bell+Howell in North Carolina (USA).
- Mai/Juni 2004** Aufenthalt in Minneapolis (USA) im Rahmen eines Schüleraustauschprogramms

# Andreas Käufl

## *Veröffentlichungen*

---

2011

Olga Birkmeier, Andreas Käufl, Friedrich Pukelsheim. *Abstentions in the German Bundesrat and ternary decision rules in weighted voting systems*. *Statistics and Decisions*, 28, 1–16.

---

2012

Mathias Drton, Chris Fox, Andreas Käufl. *Comments on: Sequences of regressions and their independencies*. *TEST*, 21, 255–261.

---

In Arbeit

Mathias Drton, Chris Fox, Andreas Käufl. *On the existence of maximum-likelihood-estimators in structural equation models*.