

Bedingte Unabhängigkeit und Graphische Modelle

Bachelorarbeit
am Fachbereich Mathematik und Informatik
der Philipps-Universität Marburg

vorgelegt von
Tobias Zwingmann

April 2013
Betreuer: Prof. Dr. Hajo Holzmann

Inhaltsverzeichnis

1. Vorbetrachtungen	6
1.1. Das Schur-Komplement und die Adjunkte einer Matrix	6
1.2. Elemente der Stochastik	12
1.3. Graphentheorie	16
1.3.1. Ungerichtete Graphen	17
1.3.2. Gerichtete Graphen	21
2. Modelle bedingter Unabhängigkeit	32
2.1. Bedingte Unabhängigkeit	32
2.2. Graphische Modelle	41
2.2.1. Ungerichtete graphische Modelle	41
2.2.2. Gerichtete Graphen ohne Zykel	46
A. Anhang	55

In dieser Arbeit geht es um den Begriff der bedingten Unabhängigkeit von Zufallsvariablen und eine Möglichkeit der Verbindung derselben mit der Graphentheorie.

Die Arbeit gibt einen Überblick über die grundlegenden Begriffe und Eigenschaften graphischer Modelle; besonderer Wert wird auf die Herleitung zugrunde liegender Aussagen und Anschaulichkeit der Probleme gelegt.

Generell basiert diese Ausarbeitung auf [SS09], wobei viele hinführende Aussagen ergänzt und vorhandene ausführlicher besprochen werden.

Die Grundidee bei graphischen Modellen ist eine graphische Faktorisierung von Wahrscheinlichkeitsmodellen, worauf zum Beispiel auch die Bayes'sche Statistik aufbaut. Im Endeffekt stellt dies eine Mischung aus Graphen- und Wahrscheinlichkeitstheorie dar und ist eine mögliche Antwort auf die Frage, wie bestimmte Wahrscheinlichkeitsmodelle kompakt repräsentiert werden können.

Es hat sich herausgestellt, dass diese Herangehensweise einen sehr natürlichen Zugang zum Umgang mit Unsicherheit in vielen Problemfeldern bietet. Die Anwendungen erstrecken sich über viele Gebiete, die a priori nichts mit Mathematik zu tun haben, wie beispielsweise Mustererkennungen (Bioinformatik), Diagnosen in der Medizin bis hin zu Hilfeassistenten in Betriebssystemen.

Von besonderem Interesse bei den Verteilungen, die den graphischen Modellen zugrunde gelegt werden, sind der diskrete Fall und der Fall, in dem eine Normalverteilung betrachtet wird.

Diese Fälle kommen häufig zur Anwendung und bedürfen daher eines tiefgreifenden Verständnisses, weswegen wir immer wieder darauf zurückkommen. Praktisch daran ist, dass die bedingten Unabhängigkeitsbedingungen in beiden Fällen in äquivalente Bedingungen „übersetzt“ werden können: Einerseits in quadratische Polynomgleichungen und andererseits in eine Bedingung an die Kovarianzmatrix. Diese Bedingungen sind algebraischer Natur, wofür eine eigene Theorie zur Verfügung steht.

Es existieren noch weitere Zusammenhänge zwischen der Algebra und dem hier behandelten Thema, insbesondere in Bezug auf eine mögliche Parametrisierung der Wahrscheinlichkeitsmodelle. Hierauf wollen wir in dieser Arbeit aber nicht weiter eingehen, da eine fundierte Einführung den Rahmen sprengen würde und verweisen daher nur auf Kapitel 3.3 aus [SS09].

Kapitel 1 beschäftigt sich mit – für unsere Zwecke – nützlichen Elementen der linearen Algebra, der Stochastik und der Graphentheorie. Unser Hauptaugenmerk liegt hierbei im Schaffen einer Basis, mit der spätere Fragen fundiert beantwortet werden können.

Einen Teil davon widmen wir der Betrachtung des Schur-Komplements einer Matrix, welches uns bei normalverteilten Zufallsvektoren – erstmals bei bedingter Normalverteilung, später erneut – begegnen wird.

In [SS09] werden die benötigten Aussagen über das Schur-Komplement als bekannt vorausgesetzt, da dieses Thema aber von besonderem Interesse ist, werden wir sie detailliert beweisen. Vor allem geht es dabei um die Bestimmung des Rangs und der Determinante einer (gegebenenfalls quadratischen) Matrix mithilfe des Schur-Komplements. Die Berechnung der Determinante folgt unmittelbar aus der Herleitung des Schur-Komplements, wohingegen die Berechnung des Rangs mehr Theorie erfordert. Die Überlegungen führen auf die „Rank Additivity Formula“, welche den Rang der betrachteten Matrix in zwei Summanden aufteilt: Den Rang des Schur-Komplements und den Rang einer gewissen Untermatrix.

Nach einem Blick auf die Adjunkte einer Matrix, von der nur kurz eine Verbindung zu der Inversen hergestellt werden soll, wenden wir uns ausgewählten Elementen der Stochastik zu. Wir wiederholen dort die Definition eines normalverteilten Zufallsvektors und was wir unter stochastischer Unabhängigkeit sowie bedingten Verteilungen verstehen.

Weiter wird ein Ausblick auf den Zusammenhang zwischen bedingter Unabhängigkeit und Polynomgleichungen gegeben, indem für diskret verteilte Zufallsvektoren ein algebraisches Kriterium für stochastische Unabhängigkeit bewiesen wird. Wie sich herausstellen wird, bilden die oben erwähnten Polynomgleichungen für bedingte Unabhängigkeit eine Erweiterung der hier betrachteten Gleichungen.

Der nächste Teil der Vorbetrachtungen behandelt die Graphentheorie, wo sowohl ungerichtete als auch gerichtete Graphen eingeführt werden.

Von besonderem Interesse sind hierbei Zusammenhangseigenschaften der Graphen, welche wir mit den bedingten Unabhängigkeitsbedingungen der korrespondierenden Wahrscheinlichkeitsverteilungen verknüpfen.

Diese sind im ungerichteten Fall anschaulich zu handhaben, in gerichteten Graphen aber weitaus komplizierter. Daher verwenden wir einen großen Teil auf die Einführung der d -Separation in gerichteten Graphen, insbesondere wie diese Eigenschaft „abgelesen“ werden kann. Unser Ziel ist dabei, eine Verbindung zu den leichteren Zusammenhangseigenschaften in ungerichteten Graphen herzustellen. Die Überlegungen dazu gipfeln im Beweis von „Pearl’s Separation Criterion“, welches wir im Hauptteil maßgeblich nutzen werden.

Der Hauptteil der Arbeit startet mit einer allgemeinen Einführung der bedingten Unabhängigkeit und der Herleitung erster Eigenschaften derselben, den Graphoid-Axiomen. Hier werden wir die algebraischen Bedingungen für bedingte Unabhängigkeit an normal- und diskret verteilte Zufallsvektoren ausführlich beweisen.

Im Abschnitt über graphische Modelle werden wir unsere Überlegungen dann ausschließlich auf den Graphoid-Axiomen aufbauen.

Zunächst werden wir dabei ungerichtete graphische Modelle, die Markov-Felder, genauer verstehen. Die Hauptaussage in diesem Teil wird die Äquivalenz zwischen den ungerichteten, paarweisen und den ungerichteten, globalen Markov-Eigenschaften sein.

Wir kommen danach zu einer intensiveren Betrachtung der gerichteten graphischen Modelle, den Bayes'schen Netzen. Hier werden wir die d-Separation nutzen, um mehr Informationen aus einem gerichteten Graphen ablesen zu können. Dazu besprechen wir ein ausführliches Beispiel, welches die Intuition hinter der Verbindung der d-Separation mit bedingter Unabhängigkeit erklären soll.

Die zweite Hauptaussage dieses Teils der Arbeit wird die Äquivalenz der gerichteten, lokalen und der gerichteten, globalen Markov-Eigenschaft eines Graphen darstellen. Diese Beziehung ist meiner Meinung nach erstaunlich, da die genannten Eigenschaften auf den ersten Blick wenig gemeinsam haben.

Als Abschluss des Kapitels und dieser Arbeit gehen wir kurz auf eine spezifische Eigenschaft gerichteter Graphen ein: Die Markov-Äquivalenz behandelt die Frage, ob zwei gerichtete Graphen die gleichen Markov-Eigenschaften besitzen können.

Kapitel 1.

Vorbetrachtungen

In diesem Kapitel wollen wir einige Definitionen, Lemmata und Sätze vorwegnehmen, um spätere Beweise kürzer gestalten zu können. Notationen, die hier eingeführt werden, werden wir im kompletten Text weiter nutzen.

Wir beginnen mit Ergebnissen der linearen Algebra in Abschnitt 1.1, welche uns in den nachfolgenden Abschnitten beim Arbeiten mit Varianz-Matrizen helfen werden.

Daraufhin wenden wir uns der Stochastik zu: In Abschnitt 1.2 wiederholen wir einige Aussagen und Definitionen der elementaren Wahrscheinlichkeitstheorie. Hier gehen wir besonders auf normalverteilte Zufallsvektoren und die Unabhängigkeit von Zufallsvariablen ein.

Als letzten Teil der Vorbetrachtungen geben wir in Abschnitt 1.3 einen Überblick über Graphentheorie, eingeteilt in ungerichtete und gerichtete Graphen (Abschnitt 1.3.1 beziehungsweise Abschnitt 1.3.2).

1.1. Das Schur-Komplement und die Adjunkte einer Matrix

Das Schur-Komplement und die Adjunkte zu einer gegebenen Matrix sind Ergebnisse der linearen Algebra; diese befasst sich insbesondere mit der Theorie der Vektorräume und linearen Abbildungen zwischen diesen. Wir machen uns hier die Ergebnisse im Hinblick auf die Theorie der Matrizen zunutze.

Dazu bezeichne in der gesamten Arbeit $\mathcal{M}_n(\mathbb{R})$ die $n \times n$ -Matrizen mit Einträgen in \mathbb{R} , also $\mathcal{M}_n(\mathbb{R}) := \mathbb{R}^{n \times n}$, und $\mathcal{GL}_n(\mathbb{R}) \subset \mathcal{M}_n(\mathbb{R})$ die invertierbaren Matrizen in $\mathcal{M}_n(\mathbb{R})$.

Einige der folgenden Aussagen sind auf Grundlage von [NN11] entstanden; da dort eine strengere Definition des Schur-Komplements genutzt wird, mussten die Aussagen verallgemeinert und die Beweise geringfügig angepasst werden.

Betrachten wir nun eine nicht notwendigerweise quadratische Matrix M , die sich wie folgt partitionieren lässt:

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix},$$

mit $A \in \mathbb{R}^{p \times k}$, $D \in \mathbb{R}^{q \times l}$, $B \in \mathbb{R}^{p \times l}$ und $C \in \mathbb{R}^{q \times k}$, $p, k, q, l \in \mathbb{N}$.

Für spätere Fragestellungen wird es hilfreich sein, Aussagen über die Determinante – falls M quadratisch ist – und den Rang einer solchen Matrix zu treffen. Wir benötigen dazu allerdings eine Diagonalgestalt von M und nehmen daher eine Faktorisierung der Matrix vor:

Ist D invertierbar (insbesondere $l = q$), so können wir M schreiben als

$$M = \begin{pmatrix} E_p & BD^{-1} \\ 0 & E_q \end{pmatrix} \begin{pmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{pmatrix} \begin{pmatrix} E_k & 0 \\ D^{-1}C & E_l \end{pmatrix} =: Q'TQ. \quad (1.1)$$

Falls A invertierbar ist (insbesondere $k = p$), können wir analog diese Faktorisierung vornehmen:

$$M = \begin{pmatrix} E_p & 0 \\ CA^{-1} & E_q \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{pmatrix} \begin{pmatrix} E_p & A^{-1}B \\ 0 & E_l \end{pmatrix}. \quad (1.2)$$

Das gibt Anlass zu folgender

Definition 1.1.1. Gegeben sei eine Matrix $M \in \mathbb{R}^{m \times n}$, die sich wie oben partitionieren lässt (m, n entsprechend gewählt).

Ist $l = q$ und D invertierbar, so ist das *Schur-Komplement von D in M* die $p \times k$ -Matrix

$$M|D = A - BD^{-1}C.$$

Wenn $k = p$ gilt und A invertierbar ist, heißt die $q \times l$ -Matrix

$$M|A = D - CA^{-1}B$$

das *Schur-Komplement von A in M* .

Die nachfolgenden Ausführungen über das Schur-Komplement beziehen sich jeweils auf $M|A$ und $M|D$, falls A^{-1} beziehungsweise D^{-1} existiert.

Außerdem sind die Dimensionen von Vektorräumen, soweit nicht anders angegeben, über \mathbb{R} zu betrachten.

Die in (1.1) und (1.2) vorgestellte Faktorisierung ist die sogenannte *AITKEN-Block-Diagonalisierung*; aus dieser ergibt sich:

Folgerung 1.1.2. Sei M quadratisch und partitioniert wie oben.

Dann gilt

$$\det(M) = \det(M|D) \det(D) = \det(A) \det(M|A).$$

Beweis. Ausnutzen der Multiplikativität der Determinante und Anwendung des Kästchensatzes (siehe Anhang) in der AITKEN-Block-Diagonalisierung liefert das Ergebnis. \square

Aus dieser Gleichheit schließen wir, dass das Schur-Komplement einer quadratischen Matrix genau dann invertierbar ist, wenn M und D oder M und A invertierbar sind.

Neben der Determinante wollen wir noch etwas über den Rang der Matrix M aussagen, unser Ziel ist folgendes Lemma, manchmal bezeichnet als „Rank Additivity Formula“.

Lemma 1.1.3. *Für M partitioniert wie oben, nicht notwendigerweise quadratisch, ist*

$$\text{rank}(M) = q + \text{rank}(M|D) = p + \text{rank}(M|A).$$

Dazu beachten wir die Tatsache, dass wir eine Matrix $M \in \mathbb{R}^{m \times n}$ als lineare Abbildung $f_M : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $f_M(x) = Mx$ betrachten können.

In diesem Zusammenhang kann der Rang einer Matrix als Dimension des Bildes der zugehörigen linearen Abbildung aufgefasst werden, also

$$\text{rank}(M) = \dim(\text{bild}(f_M)).$$

Eine invertierbare Matrix hat dann nach dem *Dimensionsatz* (siehe Anhang) vollen Rang.

Sind $M \in \mathbb{R}^{m \times n}$, $N \in \mathbb{R}^{n \times r}$, so gilt, wie leicht einzusehen ist, für deren zugehörige lineare Abbildungen $f_{MN} = f_M \circ f_N$ und die Beziehung $\text{bild}(f_{MN}) = \text{bild}(f_M|_{\text{bild}(f_N)})$.

Damit können wir eine der *Sylvester'schen Ungleichungen* beweisen:

Lemma 1.1.4. *Für zwei Matrizen $M \in \mathbb{R}^{m \times n}$, $N \in \mathbb{R}^{n \times r}$ gilt*

$$\text{rank}(MN) \leq \min(\text{rank}(M), \text{rank}(N)).$$

Sind insbesondere $G \in \mathcal{GL}_n(\mathbb{R})$ und $G' \in \mathcal{GL}_m(\mathbb{R})$ invertierbar, so ist

$$\text{rank}(MG) = \text{rank}(G'M) = \text{rank}(M). \tag{1.3}$$

Beweis. Es ist offensichtlich $\text{bild}(f_{MN}) \subset \text{bild}(f_M)$, also $\text{rank}(MN) \leq \text{rank}(M)$. Wenn wir die Dimensionsformel auf $f_M|_{\text{bild}(f_N)}$ anwenden, so folgt nach Umstellen

$$\begin{aligned} \dim(\text{bild}(f_M|_{\text{bild}(f_N)})) &= \dim(\text{bild}(f_N)) - \dim(\ker(f_M|_{\text{bild}(f_N)})) \\ &\leq \dim(\text{bild}(f_N)) \\ &= \text{rank}(N). \end{aligned}$$

Wie oben bemerkt, ist aber $\dim(\text{bild}(f_M|_{\text{bild}(f_N)})) = \dim(\text{bild}(f_{MN})) = \text{rank}(MN)$, also $\text{rank}(MN) \leq \text{rank}(N)$.

Zusammenfassen der beiden Ungleichungen beweist die Behauptung.

Sei G nun invertierbar. Wir wissen, dass $\text{rank}(M) \leq \min(m, n)$ gelten muss (wegen der Gleichheit von Zeilen- und Spaltenrang); das werden wir nun zusätzlich ausnutzen, denn damit und der eben bewiesenen Aussage folgt:

$$\begin{aligned}
 \text{rank}(MG) &\leq \min(\text{rank}(M), n) \\
 &= \text{rank}(M) \\
 &= \text{rank}((MG)G^{-1}) \\
 &\leq \min(\text{rank}(MG), n) \\
 &= \text{rank}(MG),
 \end{aligned}$$

in dem Fall gilt also Gleichheit.

Die andere Gleichheit folgt analog. \square

Jetzt haben wir das Werkzeug um Lemma 1.1.3 zu beweisen:

Beweis von Lemma 1.1.3. Sei $M = Q'TQ$ mit Q', T, Q wie in der Faktorisierung (1.1) auf Seite 7 angegeben. Der Beweis für Faktorisierung (1.2) funktioniert analog.

Q' und Q sind invertierbar, da $\det(Q') = \det(Q) = 1$ nach Anwendung des Kästchensatzes gilt.

Dann ist mit zweimaliger Anwendung des vorigen Lemmas

$$\text{rank}(M) = \text{rank}((Q'T)Q) = \text{rank}(Q'T) = \text{rank}(T).$$

Da T neben den quadratischen Matrizen auf der Hauptdiagonalen nur Nullen als Einträge besitzt, ist $\text{rank}(T) = \text{rank}(M|D) + \text{rank}(D)$, beziehungsweise in der anderen Faktorisierung $\text{rank}(T) = \text{rank}(M|A) + \text{rank}(A)$, was insgesamt die Behauptung zeigt. \square

Wir wissen mittlerweile, wie das Schur-Komplement mit der Invertierbarkeit einer Matrix zusammenhängt. Es stellt sich die Frage, ob die Inverse einer Matrix mit Hilfe des Schur-Komplements angegeben werden kann. Wir kommen auf nachstehendes Ergebnis.

Lemma 1.1.5. *Sei M invertierbar und partitioniert wie oben, dann ist*

$$\begin{aligned}
 M^{-1} &= \begin{pmatrix} A^{-1} + A^{-1}B(M_A)^{-1}CA^{-1} & -A^{-1}B(M_A)^{-1} \\ -(M_A)^{-1}CA^{-1} & (M_A)^{-1} \end{pmatrix} \\
 &= \begin{pmatrix} (M_D)^{-1} & -(M_D)^{-1}BD^{-1} \\ -D^{-1}C(M_D)^{-1} & D^{-1} + D^{-1}C(M_D)^{-1}BD^{-1} \end{pmatrix}.
 \end{aligned}$$

Hierbei seien $M_D := M|D$ und $M_A := M|A$.

Beweis.

Mit der AITKEN-Block-Diagonalisierung (1.2) folgt

$$M^{-1} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \left(\begin{pmatrix} E_p & 0 \\ CA^{-1} & E_q \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & M_A \end{pmatrix} \begin{pmatrix} E_p & A^{-1}B \\ 0 & E_l \end{pmatrix} \right)^{-1}$$

$$\begin{aligned}
 &= \begin{pmatrix} E_p & A^{-1}B \\ 0 & E_l \end{pmatrix}^{-1} \begin{pmatrix} A & 0 \\ 0 & M_A \end{pmatrix}^{-1} \begin{pmatrix} E_p & 0 \\ CA^{-1} & E_q \end{pmatrix}^{-1} \\
 &= \begin{pmatrix} E_p & -A^{-1}B \\ 0 & E_l \end{pmatrix} \begin{pmatrix} A^{-1} & 0 \\ 0 & (M_A)^{-1} \end{pmatrix} \begin{pmatrix} E_p & 0 \\ -CA^{-1} & E_q \end{pmatrix} \\
 &= \begin{pmatrix} A^{-1} & -A^{-1}B(M_A)^{-1} \\ 0 & (M_A)^{-1} \end{pmatrix} \begin{pmatrix} E_p & 0 \\ -CA^{-1} & E_q \end{pmatrix} \\
 &= \begin{pmatrix} A^{-1} + A^{-1}B(M_A)^{-1}CA^{-1} & -A^{-1}B(M_A)^{-1} \\ -(M_A)^{-1}CA^{-1} & (M_A)^{-1} \end{pmatrix}.
 \end{aligned}$$

Die zweite Identität folgt analog aus der anderen Diagonalisierung (1.1). □

Das soll unsere Einführungen bezüglich des Schur-Komplements beenden.

Kommen wir jetzt noch zu der sogenannten *adjunkten* oder *komplementären Matrix* einer Matrix G . Die Adjunkte von G existiert nur, wenn G quadratisch ist; sie beinhaltet die sogenannten Kofaktoren von G .

Mithilfe der Adjunkten lässt sich das Inverse einer invertierbaren Matrix berechnen, worauf wir auch letztendlich hinaus wollen.

Definition 1.1.6. Sei $G \in \mathcal{M}_n(\mathbb{R})$. Definiere $G_{ij} \in \mathcal{M}_{n-1}(\mathbb{R})$ für $i, j \in \{1, \dots, n\}$ als die Matrix, die aus G durch Streichen der i -ten Zeile und j -ten Spalte entsteht. Darauf aufbauend seien

$$\tilde{g}_{ij} := (-1)^{i+j} \det(G_{ji}).$$

die *Kofaktoren* von G .

Die *komplementäre* oder *adjunkte Matrix* oder einfach *Adjunkte* von G ist die Matrix $\tilde{G} := (\tilde{g}_{ij})_{i,j=1}^{n,n}$.

Die Adjunkte zu einer gegebenen Matrix fasst also die (vozeichenbehafteten) Minoren der Matrix zusammen; sie wird manchmal auch mit $\text{adj}(G)$ bezeichnet.

Wir vereinbaren hier noch Folgendes:

Falls wir auf einzelne Einträge einer Matrix verweisen wollen, werden wir Kleinbuchstaben benutzen, um nicht für zwei verschiedene Dinge die gleiche Schreibweise zu nutzen. Genauer, ist $M \in \mathbb{R}^{m \times n}$ eine Matrix, so bezeichnet m_{ij} den Eintrag von M in der i -ten Zeile und j -ten Spalte; M_{ij} nutzen wir als die Matrix, die durch das eben beschriebene Streichen von Zeile i und Spalte j entsteht.

Befinden wir uns in der Situation, in der Teilmengen $A \subset \{1, \dots, m\}$ und $B \subset \{1, \dots, n\}$ gegeben sind, wollen wir mit $M_{A,B}$ die Matrix

$$M_{A,B} := (m_{ij})_{i \in A, j \in B}$$

bezeichnen. In Worten ist $M_{A,B}$ die Matrix, die die Zeilen und Spalten enthält, die von A und B indiziert werden.

Betrachten wir wieder die adjunkte Matrix einer Matrix G . Die Struktur der Spalten von \tilde{G} lässt einen Zusammenhang zwischen ihr und der Determinante von G vermuten; tatsächlich gilt

Lemma 1.1.7. *Für eine Matrix $G \in \mathcal{M}_n(\mathbb{R})$ gilt*

$$\tilde{G}G = \det(G)E_n.$$

Wenn G invertierbar ist, ergibt sich insbesondere die Inverse von G als

$$G^{-1} = \frac{1}{\det(G)}\tilde{G}.$$

Beweis. Für $i, j \in \{1, \dots, n\}$ muss $\sum_{k=1}^n \tilde{g}_{ik}g_{kj} = \det(G)\delta_{ij}$ gelten (δ_{ij} bezeichnet das Kronecker-Delta¹).

Fall I) $i = j$

$$\sum_{k=1}^n \tilde{g}_{ik}g_{ki} = \sum_{k=1}^n (-1)^{i+k} \det(G_{ki})g_{ki} = \det(G)$$

nach dem Laplace'schen Entwicklungssatz.

Fall II) $i \neq j$

Sei H die Matrix, die aus G durch Ersetzen der i -ten durch die j -te Spalte entsteht. Da die Determinante alternierend ist, gilt

$$\begin{aligned} 0 = \det(H) &= \sum_{k=1}^n h_{ki}(-1)^{k+i} \det(H_{ki}) \\ &= \sum_{k=1}^n g_{kj}(-1)^{k+i} \det(G_{ki}) \\ &= \sum_{k=1}^n \tilde{g}_{ik}g_{kj}, \end{aligned}$$

was die Behauptung zeigt, wobei die letzte Gleichheit aus der Definition der Kofaktoren folgt. □

¹ $\delta_{ij} = \begin{cases} 1 & \text{falls } i = j \\ 0 & \text{falls } i \neq j \end{cases}$

1.2. Elemente der Stochastik

Die Stochastik beschäftigt sich mit der Beschreibung und Analyse von Zufallsprozessen. Sie stellt die Methoden bereit, die benötigt werden, um vom Zufall beeinflusste Vorgänge und Zusammenhänge vertieft zu verstehen. Wichtige Anwendungen finden sich beispielsweise in der Wirtschaft bei der Berechnung von Optionspreisen und der Medizin beim Test neuer Medikamente.

Zu den bekanntesten Problemen der Stochastik zählen das „Ziegenproblem“ und das „Teilungsproblem“, welches oft als Ursprung der Theorie angeführt wird.

Wir wollen in diesem Abschnitt kurz auf die Definition eines normalverteilten Zufallsvektors eingehen und einige grundlegende Aussagen aus der elementaren Stochastik angeben. Außerdem legen wir wieder Notationen für spätere Nutzung fest.

Es seien stets $n \in \mathbb{N}$, $X = (X_1, \dots, X_n)$ ein n -dimensionaler Zufallsvektor mit Werten in $\mathcal{X} = \prod_{j=1}^n \mathcal{X}_j$, wobei die $(\mathcal{X}_i, \Sigma_i)$ messbare Räume sind (jede Zufallsvariable X_i nimmt also Werte in \mathcal{X}_i an; die Σ_i sind σ -Algebren).

Zur Verkürzung der Schreibweise definieren wir $[n] := \{1, \dots, n\}$ und für jedes $A \subset [n]$ den Vektor X_A als den von $A \in [n]$ indizierten Teilvektor von X – also $X_A := (X_\alpha)_{\alpha \in A}$.

Weiter sei $f(x) = f(x_1, \dots, x_n) : \mathcal{X} \rightarrow [0, \infty)$ die Dichtefunktion von X bezogen auf ein Produktmaß ν auf \mathcal{X} . Dabei ist ein Produktmaß ein Maß $(\nu_1, \dots, \nu_n) = \nu : \mathcal{A}_\sigma(\Sigma_1 \times \dots \times \Sigma_n) \rightarrow [0, \infty]$, das für alle messbaren $S_1 \times \dots \times S_n \in \mathcal{X}$ die Gleichheit

$$\nu(S_1 \times \dots \times S_n) = \nu_1(S_1) \cdot \dots \cdot \nu_n(S_n)$$

erfüllt ($\mathcal{A}_\sigma(\Sigma_1 \times \dots \times \Sigma_n)$ ist die kleinste σ -Algebra, die $\Sigma_1 \times \dots \times \Sigma_n$ enthält). Ein solches Produktmaß existiert nach der Maß- und Integrationstheorie für messbare Räume immer und ist sogar eindeutig bestimmt (siehe [Agr12]).

Zusätzlich fordern wir die Stetigkeit von f auf \mathcal{X} .

Beispiel 1.2.1. 1. Sei $\mathcal{X} = \mathbb{R}^n$, dann ist die geforderte Stetigkeit die aus der Analysis bekannte Stetigkeit und $\nu = \mu_L$ das Lebesgue-Maß.

2. Ist \mathcal{X} endlich, ist die Stetigkeitsbedingung immer erfüllt (da jede Funktion in einem isolierten Punkt stetig ist).

(Mit diesen beiden Fällen werden wir uns auch im Hauptteil intensiv beschäftigen.)

Wie gewohnt werden wir mit f_A die marginale Dichte von X_A bezeichnen; diese ist für $x_A \in \mathcal{X}_A$ über

$$f_A(x_A) := \int_{\mathcal{X}_{[n] \setminus A}} f(x_A, x_{[n] \setminus A}) d\nu_{[n] \setminus A}(x_{[n] \setminus A})$$

definiert (nach Umsortieren). Für $[n] \setminus A = \{i_1, \dots, i_k\}$, $k \leq n$, lässt sich mit Anwendung des Produktmaßes und nach dem Satz von Fubini der Ausdruck umformen zu

$$\int_{\mathcal{X}_{i_k}} \cdots \int_{\mathcal{X}_{i_1}} f(x_1, \dots, x_n) d\nu_{i_1}(x_{i_1}) \cdots d\nu_{i_k}(x_{i_k}).$$

Ebenfalls wie gewohnt sei für disjunkte $A, B \subset [n]$ die bedingte Dichte von X_A gegeben X_B mit $f_{A|B}$ bezeichnet, nach Definition also

$$f_{A|B}(x_A|x_B) := \frac{f_{A \cup B}(x_A, x_B)}{f_B(x_B)}.$$

Beachte, dass $f_{A|B}$ offensichtlich nur für $x_A \in \mathcal{X}_A$, $x_B \in \mathcal{X}_B$ mit $f_B(x_B) > 0$ wohldefiniert ist.

Des Weiteren: Sind X_1, \dots, X_n Zufallsvariablen mit Werten in $[r_1], \dots, [r_n]$, $r_j \in \mathbb{N}$, und P als gemeinsame Verteilung, definieren wir $p_{i_1, \dots, i_n} := P(X_1 = i_1, \dots, X_n = i_n)$, wobei $i_j \in [r_j]$. In diesem Fall ist die marginale Dichte von $X_1, \dots, X_{k-1}, X_{k+1}, \dots, X_n$, $k \in [n]$, eine endliche Summe. Wir definieren

$$p_{i_1, \dots, i_{k-1}, +, i_{k+1}, \dots, i_n} := \sum_{j=1}^{r_k} p_{i_1, \dots, i_{k-1}, j, i_{k+1}, \dots, i_n}.$$

Sei wieder $A \subset [n]$, $[n] \setminus A = \{i_1, \dots, i_k\}$ für ein $k \leq n$. Die marginale Verteilung von X_A notieren wir mit

$$p_{x_A, +} := P(X_A = x_A) = \sum_{j_1=1}^{r_{i_1}} \cdots \sum_{j_k=1}^{r_{i_k}} P(X_A = x_A, X_{i_1} = j_1, \dots, X_{i_k} = j_k)$$

für $x_A \in \prod_{l \in A} [r_l]$. Diese Schreibweise deutet demnach an, dass über alle Indizes, die nicht in A enthalten sind, summiert wird.

Hier halten wir ein kurzes Ergebnis über Unabhängigkeit von diskreten Zufallsvariablen fest, welches wir im weiteren Verlauf noch verallgemeinern werden. Als Schreibweise für Unabhängigkeit von Zufallsvariablen nutzen wir „ \perp “.

Proposition 1.2.2. *Seien X, Y Zufallsvariablen mit Werten in $[r]$ bzw. $[c]$ für $r, c \in \mathbb{N}$. Definiere $p_{ij} := P(X = i, Y = j)$ für alle $i \in [r]$, $j \in [c]$.*

Dann sind X und Y genau dann unabhängig, wenn die Matrix

$$M := (p_{ij})_{i,j=1}^{r,c} \in \mathbb{R}^{r \times c}$$

den Rang 1 hat.

Beweis. „ \Rightarrow “:

Seien X und Y unabhängig. Nach Definition der Unabhängigkeit ist für alle $i \in [r], j \in [c]$ $P(X = i, Y = j) = P(X = i)P(Y = j)$. Die Matrix P lässt sich also schreiben als

$$M = (P(X = 1), \dots, P(X = r))^T (P(Y = 1), \dots, P(Y = c)),$$

solche Matrizen haben immer Rang 1 (eine kurze Begründung dazu befindet sich im Anhang).

„ \Leftarrow “:

Gelte $\text{rank}(M) = 1$, dann ist $M = a^T b$ für zwei Zeilenvektoren $a \in \mathbb{R}^r, b \in \mathbb{R}^c$. Da M nur positive Einträge besitzt, können o.B.d.A. a und b ebenfalls mit nur positiven Einträgen gewählt werden. Seien a_+, b_+ die Summe der Einträge in a beziehungsweise b .

Es gilt $P(X = i) = a_i b_+, P(Y = j) = a_+ b_j$, was sich aus der Matrix ablesen lässt, und $a_+ b_+ = \sum_{k=1}^r \sum_{l=1}^c a_k b_l$ nach dem allgemeinen Distributivgesetz und $\sum_{k=1}^r \sum_{l=1}^c a_k b_l = 1$ nach der Definition von M .

Dann aber ist $P(X = i, Y = j) = m_{ij} = a_i b_j = (a_i b_+)(a_+ b_j) = P(X = i)P(Y = j)$, also sind X und Y unabhängig. \square

Die Aussage ist äquivalent dazu, dass alle 2-Minoren der Matrix M Determinante 0 haben, was uns zu einer algebraischen Bedingung (in Form von Polynomgleichungen) für Unabhängigkeit führt:

$$X, Y \text{ unabhängig} \Leftrightarrow m_{i_1 j_1} m_{i_2 j_2} - m_{i_1 j_2} m_{i_2 j_1} = 0$$

für alle $i_1, i_2 \in [r], j_1, j_2 \in [c]$ mit $i_1 \neq i_2, j_1 \neq j_2$.

Eine in der Stochastik grundlegende Definition ist die eines *normalverteilten Zufallsvektors*. Die Normalverteilung spielt unter anderem eine große Rolle für den *zentralen Grenzwertsatz*. Die Normalverteilung wird bei uns im Weiteren ebenfalls besondere Aufmerksamkeit erhalten, dazu die folgende

Definition 1.2.3. Seien $\mu \in \mathbb{R}^n, \Sigma \in \mathcal{M}_n(\mathbb{R})$ symmetrisch und positiv definit. Ein \mathbb{R}^n -wertiger Zufallsvektor X heißt *normalverteilt mit den Parametern Σ und μ* , wenn die Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$,

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

die zugehörige Dichte ist.

Für einen normalverteilten Zufallsvektor X schreiben wir wie üblich $X \sim \mathcal{N}(\mu, \Sigma)$.

Wie der Titel der Arbeit nahelegt, werden wir meist mit bedingten Verteilungen arbeiten. Im nächsten Satz sehen wir auch direkt ein Beispiel für die Anwendung des besprochenen Schur-Komplements.

Satz 1.2.4. Seien $Y \sim \mathcal{N}(\mu_1, \Sigma_1)$, $Z \sim \mathcal{N}(\mu_2, \Sigma_2)$ und

$$X = \begin{pmatrix} Y \\ Z \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_1 & \Sigma_{12} \\ (\Sigma_{12})^T & \Sigma_2 \end{pmatrix} \right) =: \mathcal{N}(\mu, \Sigma),$$

mit $\mu_1 \in \mathbb{R}^r$, $\mu_2 \in \mathbb{R}^{n-r}$, $\Sigma_1 \in \mathbb{R}^{r \times r}$, $\Sigma_2 \in \mathbb{R}^{(n-r) \times (n-r)}$ und $\Sigma_{12} \in \mathbb{R}^{r \times (n-r)}$.

Dann ist

$$(Y|Z=z) \sim \mathcal{N}(\mu_{Y|Z=z}, \Sigma|\Sigma_2),$$

wobei $\mu_{Y|Z=z} = \mu_1 + \Sigma_{12}\Sigma_2^{-1}(z - \mu_2)$.

Der Beweis ist aus [Hol11] entnommen, wobei unsere bisherigen Ergebnisse und Notationen genutzt werden.

Beweis. Nach Definition der bedingten Unabhängigkeit ist mit $x = (y^T, z^T)^T$

$$\begin{aligned} f_{Y|Z}(y|z) &= \frac{\frac{1}{\sqrt{\det \Sigma (2\pi)^n}} \exp(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu))}{\frac{1}{\sqrt{\det \Sigma_2 (2\pi)^{n-r}}} \exp(-\frac{1}{2}(z - \mu_2)^T \Sigma_2^{-1}(z - \mu_2))} \\ &= \frac{\exp(-\frac{1}{2}((x - \mu)^T \Sigma^{-1}(x - \mu) - (z - \mu_2)^T \Sigma_2^{-1}(z - \mu_2)))}{\sqrt{\frac{\det \Sigma}{\det \Sigma_2} (2\pi)^r}}. \end{aligned}$$

Über das Schur-Komplement können wir Σ^{-1} berechnen, da Σ_2 positiv definit und somit invertierbar ist. Es bezeichne $\Sigma|_2$ das Schur-Komplement von Σ_2 in Σ .

Mit Folgerung 1.1.2 ist dann $\det(\Sigma) = \det(\Sigma_2) \det(\Sigma|_2)$ und mit Lemma 1.1.5 gilt

$$\Sigma^{-1} = \begin{pmatrix} \Sigma|_2^{-1} & -\Sigma|_2^{-1}\Sigma_{12}\Sigma_2^{-1} \\ -\Sigma_2^{-1}(\Sigma_{12})^T \Sigma|_2^{-1} & \Sigma_2^{-1} + \Sigma_2^{-1}(\Sigma_{12})^T \Sigma|_2^{-1}\Sigma_{12}\Sigma_2^{-1} \end{pmatrix}.$$

Betrachten wir den Ausdruck in dem Exponenten der exp-Funktion, gilt aufgrund dieser Gleichheit

$$\begin{aligned} &\left(-\frac{1}{2}((x - \mu)^T \Sigma^{-1}(x - \mu) - (z - \mu_2)^T \Sigma_2^{-1}(z - \mu_2)) \right) \\ &= \begin{pmatrix} y - \mu_1 \\ z - \mu_2 \end{pmatrix}^T \begin{pmatrix} \Sigma|_2^{-1} & -\Sigma|_2^{-1}\Sigma_{12}\Sigma_2^{-1} \\ -\Sigma_2^{-1}(\Sigma_{12})^T \Sigma|_2^{-1} & \Sigma_2^{-1} + \Sigma_2^{-1}(\Sigma_{12})^T \Sigma|_2^{-1}\Sigma_{12}\Sigma_2^{-1} \end{pmatrix} \begin{pmatrix} y - \mu_1 \\ z - \mu_2 \end{pmatrix} \\ &\quad - (z - \mu_2)^T \Sigma_2^{-1}(z - \mu_2), \end{aligned}$$

was nach Ausmultiplizieren den Ausdruck

$$\begin{aligned} [\dots] &= (y - \mu_1)^T \Sigma|_2^{-1}(y - \mu_1) - (z - \mu_2)^T \Sigma_2^{-1}(\Sigma_{12})^T \Sigma|_2^{-1}(y - \mu_1) \\ &\quad - (y - \mu_1)^T \Sigma|_2^{-1}\Sigma_{12}\Sigma_2^{-1}(z - \mu_2) + (z - \mu_2)^T (\Sigma_2^{-1} + \Sigma_2^{-1}(\Sigma_{12})^T \Sigma|_2^{-1}\Sigma_{12}\Sigma_2^{-1})(z - \mu_2) \\ &\quad - (z - \mu_2)^T \Sigma_2^{-1}(z - \mu_2) \end{aligned}$$

ergibt; zweimaliges Anwenden des Distributivgesetzes liefert uns dann letztendlich

$$\begin{aligned}
 [\dots] &= (y - \mu_1)^T \Sigma|_2^{-1} (y - \mu_1 - \Sigma_{12} \Sigma_2^{-1} (z - \mu_2)) \\
 &\quad - (z - \mu_2)^T \Sigma_2^{-1} (\Sigma_{12})^T \Sigma|_2^{-1} ((y - \mu_1) - \Sigma_{12} \Sigma_2^{-1} (z - \mu_2)) \\
 &= (y - \mu_1 - \Sigma_{12} \Sigma_2^{-1} (z - \mu_2))^T \Sigma|_2^{-1} (y - \mu_1 - \Sigma_{12} \Sigma_2^{-1} (z - \mu_2)) \\
 &= (x - \mu_{Y|Z=z})^T \Sigma|_2^{-1} (x - \mu_{Y|Z=z}).
 \end{aligned}$$

Einsetzen der beiden Terme liefert die Behauptung. \square

1.3. Graphentheorie

Nachfolgend sind einige grundlegende Definitionen aus der Graphentheorie aufgeführt, die wir in Kapitel 2.2 brauchen werden.

Allgemein beschäftigt sich die Graphentheorie mit der Darstellung und Untersuchung von Graphen (grob gesagt einer Struktur aus Knoten und Verbindungen zwischen diesen) und deren Beziehungen zueinander; eine wichtige Anwendung findet sich beispielsweise in der Informatik und der Netzwerktheorie.

Bekannte graphentheoretische Probleme sind zum Beispiel das des „Traveling Salesman“ - ein Handlungsreisender sucht eine möglichst optimale Route durch gewisse Punkte (Städte) - und das „Königsberger Brückenproblem“ - gibt es eine Route oder gar einen Rundweg, so dass eine Person jede Brücke über den Pregel in Königsberg genau einmal nutzt?

Wir werden Graphen zur Darstellung von Abhängigkeiten zwischen Zufallsvariablen heranziehen und die teilweise sehr anschaulichen Beweistechniken dieser Theorie nutzen. Sehr hilfreich ist hierbei die Tatsache, dass die von uns behandelten Graphen immer als Bild bestehend aus Punkten und Verbindungslinien zwischen diesen dargestellt werden können.

„Unsere“ Graphen können in *gerichtete* und *ungerichtete Graphen* eingeteilt werden; den beiden Typen widmen wir jeweils einen kurzen Abschnitt².

²Allgemein gibt es neben den gerichteten und ungerichteten Graphen noch mehr Typen von Graphen, mit denen wir uns nicht weiter beschäftigen wollen.

1.3.1. Ungerichtete Graphen

Definition 1.3.1. Ein *einfacher ungerichteter Graph* G ist ein Paar (V, E) mit einer beliebigen Menge V , genannt *Knotenmenge*, und einer *Kantenmenge* E mit

$$E \subset \{\{i, j\} | i, j \in V\}.$$

Sei $S \subset V$. Der Graph $G_S := (S, E_S)$, mit $E_S := \{\{i, j\} \in E | i, j \in S\}$ heißt von S *induzierter Teil- oder Subgraph*.

Eine Kante in einem ungerichteten Graphen heißt *Schleife*, falls $i = j$.

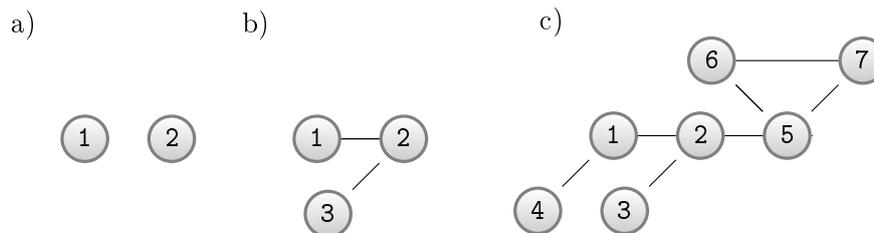
Per Definition sind die Kanten demnach im Fall einer Schleife ein-, sonst zweielementige Mengen.

Die Kantenmenge kann auch als Teilmenge der direkten Summe $V \times V$ betrachtet werden – dann muss für jede Kante $(i, j) \in E$ auch $(j, i) \in E$ gelten. Im nächsten Abschnitt werden wir Tupel der Form $(i, j) \in V \times V$ *gerichtete Kanten* nennen. Dabei ist zu beachten, dass eine ungerichtete Kante in der Darstellung eines Graphen meist ohne Pfeilspitzen angegeben wird, was aber bei *gerichteten Kanten* – siehe Beispiel 1.3.9 – eigentlich der Fall ist. Nach dieser Richtlinie wäre eine ungerichtete Kante dann ein Doppelpfeil; diese Notation wird allerdings selten genutzt, außer bei *gemischten Graphen*, in denen sowohl gerichtete als auch ungerichtete Kanten erlaubt sind.

Der induzierte Teilgraph entsteht aus einem Graphen durch Reduktion auf die Knoten in $S \subset V$ und Weglassen der Kanten, die nicht beide Endpunkte in S haben.

Dazu zunächst einige Beispiele.

Beispiel 1.3.2. Folgende Abbildungen sind Darstellungen von ungerichteten Graphen:



Dabei kann, wie leicht zu sehen ist, der Graph in b) als Teilgraph von dem in c) mit $S = \{1, 2, 3\}$ betrachtet werden.

In ungerichteten Graphen besteht die Möglichkeit, mehrere Kanten zwischen denselben Knoten zuzulassen und Kanten mit einem *Gewicht* zu versehen. Wir wollen später ungerichtete, ungewichtete Graphen mit endlich vielen Knoten ohne Schleifen und ohne Mehrfachkanten betrachten, daher wird darauf hier nicht näher eingegangen.

Bis zum Ende dieses Abschnitts sei $G = (V, E)$ stets ein ungerichteter Graph.

Definition 1.3.3. Zwei Knoten $i, j \in V$ heißen *adjazent* oder *benachbart*, in Zeichen $i \sim j$, wenn sie durch eine Kante verbunden sind, also $\{i, j\} \in E$.

Ein *ungerichteter Pfad* vom Knoten i zum Knoten j ist eine endliche, nicht-leere Folge $\pi = \{i, i_2, \dots, i_{k-1}, j\} \subset V$ von Knoten mit $i = i_1 \sim i_2 \sim \dots \sim i_k = j$.

Die Knoten i und j heißen *Endknoten vom Pfad* π , die Knoten i_2, \dots, i_{k-1} heißen *Zwischenpunkte von* π und $|\pi| := k$ heißt *Länge* des Pfades.

Aufbauend auf dem Pfadbegriff definieren wir noch die nachfolgenden Begriffe, die unter anderem die Notation erleichtern.

Definition 1.3.4. Sei $\pi = \{i, i_2, \dots, i_{k-1}, j\} \subset V$ ein ungerichteter Pfad.

Mit $\pi^{-1} := (j, i_{k-1}, \dots, i_2, i)$ bezeichnen wir den zu π *inversen Pfad*.

Der Pfad $\pi(i_l, i_m) := (i_l, i_{l+1}, \dots, i_{m-1}, i_m)$ für $1 < l < m < k$ heißt *Teilpfad von* π *zwischen* i_l *und* i_m .

Der Pfad π heißt *minimal*, wenn kein anderer Pfad zwischen i und j existiert, der weniger Knoten enthält, also kürzer ist.

In obigem Beispiel gibt es in Fall a) offensichtlich keinen Pfad vom Knoten 1 zum Knoten 2. Das Beispiel verallgemeinern wir und gelangen zu einem intuitiven graphentheoretischen Zusammenhangsbegriff:

Definition 1.3.5. G heißt *zusammenhängend*, falls für je zwei Knoten $i, j \in V$ ein Pfad von i nach j existiert. Ist dies nicht der Fall, heißt der Graph *nicht-zusammenhängend*.

Im weiteren Verlauf der Arbeit brauchen wir einen Trennungsbegriff in Bezug auf Teilmengen der Kanten für Graphen. Dieser kann im ungerichteten Fall auf dem Zusammenhang aufgebaut werden – im gerichteten Fall werden wir mehr Hilfsmittel brauchen.

Definition 1.3.6. Seien $A, B, C \subset V$ paarweise disjunkt, $A, B \neq \emptyset$.

Falls der Graph $G_{V \setminus C}$ nicht-zusammenhängend ist, so heißen die Mengen A und B *getrennt durch* C .

Mit anderen Worten: Enthält jeder Pfad von einem Knoten $i \in A$ zu einem Knoten $j \in B$ einen Knoten der Menge C , so trennt C die Menge A von der Menge B .

Im vorigen Beispiel sind die Graphen in b) und c) zusammenhängend, der in a) nicht.

Als Abschluss der ungerichteten Graphen beweisen wir noch ein eher technisches Lemma. Die Notwendigkeit dieser Aussage ist nicht sofort ersichtlich, wird aber im Hauptteil unbedingt benötigt. Thematisch passt es hier gut in den Kontext und wir sehen eine erste Anwendung des Trennungsbegriffs.

Lemma 1.3.7. Seien $A, B, C \subset V$ paarweise disjunkt, $A, B \neq \emptyset$ und C trenne die Mengen A und B . Weiter existiere von jedem Knoten $v \in V \setminus (A \cup B \cup C)$ ein ungerichteter Pfad zu einem $u \in A \cup B \cup C$.

Dann existieren Mengen $A^* \supset A$ und $B^* \supset B$, die von C getrennt werden und $A^* \cup B^* \cup C = V$ erfüllen.

Beweis. Wegen der Existenz eines ungerichteten Pfades von $v \in V$ zu einem $u \in A \cup B \cup C$ und der Trennungseigenschaft von C , kann die globale Struktur von G wie in Abbildung 1.1 schematisch dargestellt werden.

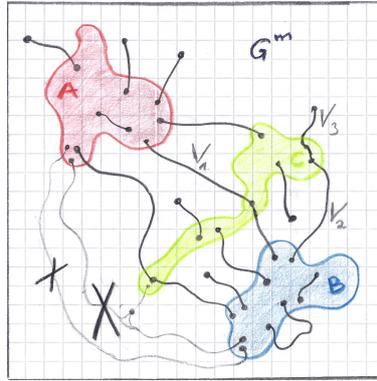


Abbildung 1.1.: Schematische Darstellung von G im Fall von Lemma 1.3.7

Wie in Abbildung 1.1 angedeutet, wollen wir die Knoten, die nicht in $A \cup B \cup C$ enthalten sind, in die drei Mengen V_1, V_2, V_3 aufteilen. V_1 und V_3 werden zum Schluss zu A und V_2 zu B hinzugefügt.

Dazu unterscheiden wir zwei Fälle für $v \in V \setminus (A \cup B \cup C)$, von denen mindestens einer eintreten muss:

Fall I) Es existiert ein ungerichteter Pfad $\mu_{vw} = \{v, v_2, \dots, v_{k-1}, w\}$ zu einem Knoten $w \in A \cup B$, wobei $v_i \notin C$ für alle $i \in \{2, \dots, k-1\}$.

Fall II) Es existiert ein ungerichteter Pfad $\nu_{vu} = \{v, u_2, \dots, u_{l-1}, u\}$ zu einem $u \in C$.

In beiden Fällen wählen wir w und u so, dass μ_{vw} beziehungsweise ν_{vu} minimal sind.

Angenommen es gilt Fall I). Wir bemerken, dass für zwei verschiedene $v, v^* \notin A \cup B \cup C$ die Pfade μ_{vw_1} und $\mu_{v^*w_2}$ keinen Knoten gemeinsam haben, wenn $w_1 \in A$ und $w_2 \in B$ oder umgekehrt gilt, denn: Es ließe sich sonst ein ungerichteter Pfad konstruieren, der A und B verbindet und dabei C umgeht.

Die μ_{vw} können nach ihren Endpunkten aufgeteilt werden, damit konstruieren wir V_1 und V_2 :

Ist $w \in A$, fügen wir alle Punkte v, v_1, \dots, v_{k-1} zu V_1 hinzu; falls $w \in B$ fügen wir v, v_1, \dots, v_{k-1} in V_2 ein.

Wegen der Bemerkung oben, ist diese Zuordnung eindeutig.

Gilt Fall I) nicht, muss Fall II) gelten. Dann fügen wir die Knoten v, u_1, \dots, u_{l-1} zu V_3 hinzu.

Die Mengen $A^* := A \cup V_1 \cup V_3$ und $B^* := B \cup V_2$ werden in G noch immer durch C getrennt, denn: Angenommen es existiert ein ungerichteter Pfad π von einem $v^* \in A^*$ zu einem $w^* \in B^*$, der keine Knoten von C enthält. Wir wählen π minimal, dann enthält π nur zwei Knoten $v \in A^*$ und $w \in B^*$.

Wegen der Konstruktion von A^* und B^* existieren auch ungerichtete Pfade $\mu_{vv'}$ und $\mu_{ww'}$ von v zu einem $v' \in A$ beziehungsweise von w zu einem $w' \in B$, die keine Knoten aus C enthalten. Durch Konkatenation von $\mu_{vv'}$, π und $\mu_{ww'}$ erhalten wir einen Pfad, der von v nach w verläuft, ohne C zu schneiden, im Widerspruch zur Trennung von A und B durch C .

Außerdem gilt $A^* \cup B^* \cup C = V$ nach Konstruktion (da einer der Fälle eintreten muss), womit die Behauptung gezeigt ist. \square

Die Voraussetzung, dass für jedes $v \in V \setminus (A \cup B \cup C)$ ein ungerichteter Pfad zu einem $u \in A \cup B \cup C$ existiert, wird in dem Beweis nicht unbedingt benötigt. Die Aussage stimmt auch ohne diese zusätzliche Annahme (dann existieren „isolierte“ Knoten, diese können beliebig auf A^* und B^* verteilt werden; das letzte Argument über die Trennung von A^* und B^* durch C muss geringfügig angepasst werden).

Wir führen die Bedingung aber hier mit auf, weil in der späteren Anwendung dieses Lemmas genau dieser Fall betrachtet wird.

1.3.2. Gerichtete Graphen

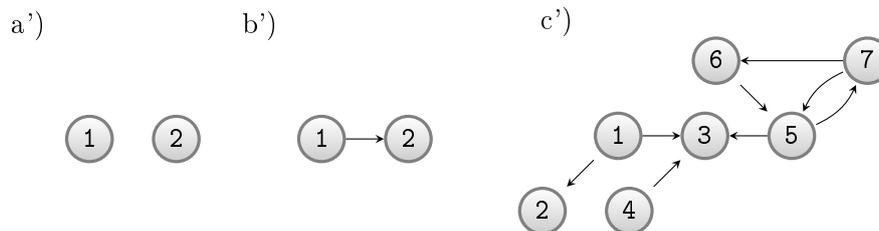
Nehmen wir an, unser „Traveling Salesman“ muss auf seiner Route einige Einbahnstraßen durchqueren. Wie kann dies in einem Graphen verdeutlicht werden?

Definition 1.3.8. Ein Tupel $G = (V, E)$ mit einer beliebigen Knotenmenge V und Kantenmenge E heißt *gerichteter Graph*, falls die Kantenmenge geordnete Tupel $(i, j) \in V \times V$ anstatt ungeordneter Mengen enthält.

Sei $S \subset V$ eine Teilmenge. Der Graph $G_S := (S, E_S)$ mit $E_S := \{(i, j) \in E \mid i, j \in S\}$ heißt *induzierter Teil- oder Subgraph*.

Beziehungswise wenn die alternative Definition von ungerichteten Kanten als Tupel in $V \times V$ betrachtet wird, wird die Eigenschaft vergessen, dass mit jeder Kante auch die umgekehrte Kante in der Kantenmenge enthalten sein muss. Wie oben erwähnt, wird in der Darstellung eines gerichteten Graphen die Richtung/Orientierung einer Kante mit einer Pfeilspitze angedeutet.

Beispiel 1.3.9. Durch Hinzufügen einer Orientierung der Kanten werden aus den ungerichteten Graphen von oben gerichtete Graphen:



Dabei ist anzumerken, dass die Darstellung der Kanten zwischen den Knoten 5 und 7 in c') nicht zu „ \leftrightarrow “ zusammengefasst werden sollte. Eine solche Darstellung von Kanten wird wie oben erwähnt manchmal bei gemischten Graphen genutzt, wie in [Ric99].

Zwar ist bei der Darstellung gemischter Graphen unklar, ob eine ungerichtete Kante zwischen zwei Knoten existiert oder eine Kante hin- und eine zurückführt, wenn „ \leftrightarrow “ als Darstellung genutzt wird, dennoch hat die Nutzung von „ \leftrightarrow “ durchaus Vorteile, auf diese gehen wir hier aber nicht weiter ein.

In diesem Abschnitt sei $G = (V, E)$ nun ein gerichteter Graph, falls nicht anders vorausgesetzt.

Wir hatten einen Zusammenhangsbegriff für ungerichteten Graphen kennengelernt. Wie sieht eine solche Beziehung für Teilmengen der Knotenmenge nun im Fall von gerichteten Graphen aus? Um diese Frage in der erforderlichen Tiefe beantworten zu können, benötigen wir einige Vorbereitungen.

Definition 1.3.10. Zwei Knoten $i, j \in V$ heißen *adjazent*, wenn sie durch eine Kante verbunden sind, in Formeln $(i, j) \in E$ oder $(j, i) \in E$.

Ein Knoten $i \in V$ und eine Kante $(j, k) \in E$ heißen *inzident*, falls $j = i$ oder $k = i$ gilt.

Analog zum ungerichteten Fall benötigen wir einen Pfadbegriff.

Definition 1.3.11. Ein *gerichteter Pfad* von einem Knoten $i \in V$ zu einem Knoten $j \in V$ ist ein endliches, nicht-leeres Tupel $\pi = (i = i_1, i_2, \dots, i_{k-1}, i_k = j)$ mit $(i_1, i_2), (i_2, i_3), \dots, (i_{k-1}, i_k) \in E$.

Ignoriert man die Orientierung der Kanten, heißt π wieder *ungerichteter Pfad*; dann muss also $(i_{l-1}, i_l) \in E \vee (i_l, i_{l-1}) \in E$ für alle $l \in \{2, \dots, k\}$ in π gelten.

Betrachten wir wieder obiges Beispiel, existiert in Beispiel c') kein gerichteter Pfad von Knoten 1 zu Knoten 7, da $(3, 5) \notin V$, aber es existiert ein ungerichteter Pfad zwischen diesen Knoten. Des Weiteren existieren unter anderem gerichtete Pfade von den Knoten 1 und 7 zu Knoten 3.

Elemente von V mit gewissen Eigenschaften fassen wir zu Mengen zusammen:

Definition 1.3.12. Sei $i \in V$.

Die Menge

$$\text{de}(i) := \{j \in V \mid \text{es existiert ein gerichteter Pfad von } i \text{ nach } j\}$$

beinhaltet die *Nachfolger* von i und

$$\text{nd}(i) := V \setminus (\{i\} \cup \text{de}(i))$$

die *Nicht-Nachfolger* von i . Elemente der Menge

$$\text{pa}(i) := \{j \in V \mid (j, i) \in E\}$$

heißen *Eltern* oder *direkte Vorgänger* von i . Für $C \subset V$ sei weiter

$$\text{an}(C) := \{w \in V \mid \exists v \in C : v \in \text{de}(w)\};$$

diese Menge enthält also die Knoten $w \in V$, die Vorgänger eines $v \in C$ sind. Des Weiteren benötigen wir die Menge

$$\text{AN}(C) := C \cup \text{an}(C).$$

Für die Teilmenge C definieren wir zusätzlich

$$\text{de}(C) := \bigcup_{i \in C} \text{de}(i)$$

und analog die Mengen $\text{nd}(C)$ sowie $\text{pa}(C)$.

Dabei kommen die Bezeichnungen jeweils von den englischen Übersetzungen, wenn der Graph als Stammbaum einer Familie gesehen wird: „Eltern“ (*parents*), „Nachfolger“ (*descendants*), „Nicht-Nachfolger“ (*non-descendants*) und „Vorgänger“ (*ancestors*).

Ist $(v, w) \in E$, so heißt w auch manchmal *Kind* von v .

Da wir mit nur mit endlichen Graphen arbeiten, können wir hier die folgende Aussage beweisen.

Lemma 1.3.13. Sei G ein Graph, in dem für alle $v \in V$ keine gerichteten Pfade $(v, v_1, \dots, v_{k-1}, v)$ existieren.

Dann existiert ein Knoten v_0 mit $\text{de}(v_0) = \emptyset$.

Beweis. Angenommen die Aussage stimmt nicht, es existiert also kein Knoten $v_0 \in V$ mit $\text{de}(v_0) = \emptyset$.

Wir wählen ein beliebiges $v \in V$. Weiter wählen wir ein $v_1 \in \text{de}(v)$ ($\neq \emptyset$ nach Voraussetzung) und definieren den gerichteten Pfad $\pi_1 := (v, v_1)$.

Iterativ wählen wir ein $v_{i+1} \in \text{de}(v_i)$ und fügen den Knoten zu π_{i-1} hinzu, sodass der gerichtete Pfad $\pi_i = (v, v_1, \dots, v_i)$ entsteht. Die Menge der Knoten, die auf π_i liegen, wird mit jedem Schritt größer und das Verfahren endet nicht, da jeder Knoten Nachfolger besitzt.

Da $|V| < \infty$, muss nach genügend vielen Schritten ein Knoten v_m existieren, sodass $(v_m, v_{m+1}, \dots, v_n, v_m)$ ein Teilpfad von einem π_i ist, also insbesondere selbst gerichtet. Ein solcher Teilpfad steht aber im Widerspruch zu den Voraussetzungen. \square

Nach diesem kurzen Zwischenspiel kommen wir zurück zu den Definitionen, die wir aus der Graphentheorie benötigen.

Definition 1.3.14. Sei $\pi = (i_1, \dots, i_k)$ ein ungerichteter oder gerichteter Pfad in G .

Die Knoten i_2, \dots, i_{k-1} heißen *Zwischenpunkte* von π .

Sei π nun ein ungerichteter Pfad. Ein i_l heißt *Kollisionspunkt* (*collider*) von π , wenn die zu i_l inzidenten Kanten von der Form



sind (in Formeln: $(i_{l-1}, i_l), (i_{l+1}, i_l) \in E$).

Wir nähern uns einer genaueren Beschreibung des „Zusammenhangs“ in gerichteten Graphen. Bevor wir zur formalen Definition dieses Begriffs kommen, entwickeln wir drei Regeln darüber, was wir unter Zusammenhang verstehen wollen und formalisieren diese Regeln dann zu einer endgültigen Definition. Die Regeln selbst können schon als eigene Definitionen betrachtet werden. Wir lassen die Begriffe vorerst so stehen und konkretisieren nur die benötigten Begriffe später in formalen Definitionen.

Regel I) Zwei Knoten $i, j \in V$ heißen *d-zusammenhängend*³, falls ein ungerichteter Pfad ohne Kollisionspunkte zwischen i und j existiert.

Im folgenden Beispiel sind die Knoten 1, 2, 3, 4 paarweise d-zusammenhängend, ebenso die Knoten 4, 5, 6, 7. Knoten 4 ist ein Kollisionspunkt auf jedem weiteren Pfad, also ist kein Knoten der Menge $\{1, 2, 3\}$ d-zusammenhängend mit einem Knoten der Menge $\{5, 6, 7\}$.



Einen ungerichteten Pfad ohne Kollisionspunkte nennen wir einen *nicht-blockierten Pfad*.

Hier geben wir eine kurze Aussage an, die wir später in einem unserer wichtigsten Ergebnisse nutzen werden.

³Das „d“ kommt von „directional“

Lemma 1.3.15. *Existiere ein ungerichteter Pfad $\pi = (i, i_2, \dots, i_{k-1}, j)$ ohne Kollisionspunkte zwischen i und j .*

Dann ist entweder π selbst oder π^{-1} ein gerichteter Pfad oder es existiert ein $k_0 \in \{2, \dots, k-1\}$, sodass $\pi^{-1}(i_{k_0}, i)$ und $\pi(i_{k_0}, j)$ gerichtete Pfade sind.

Es existiert mit anderen Worten ein k_0 , sodass die Kantenfolge des Pfades, der den d -Zusammenhang garantiert, folgende Gestalt hat:



Beweis. Angenommen die Aussage stimmt nicht, also weder π noch π^{-1} noch $\pi^{-1}(i_{k_0}, i)$ und $\pi(i_{k_0}, j)$ sind gerichtete Pfade für beliebiges $k_0 \in \{2, \dots, k-1\}$.

Wir starten in i und gehen den ungerichteten Pfad entlang.

Zeigt die erste Kante von i weg, muss ein Knoten existieren, zu dem zwei Kanten von π zeigen, da π sonst auch ein gerichteter Pfad ist. Dieser Knoten ist nach Definition ein Kollisionspunkt, was einen Widerspruch zum d -Zusammenhang von i und j liefert.

Daher zeigt die erste Kante von π auf i . Es existiert ein $i_{k_1} \in \{i_2, \dots, i_{k-2}\}$, von dem beide inzidenten Kanten auf π wegführen – sonst wäre π^{-1} ein gerichteter Pfad.

O.B.d.A. wählen wir k_1 mit dieser Eigenschaft so, dass $\pi(i, i_{k_1})$ minimal ist⁴. Dann kann $\pi(i_{k_1}, j)$ nach Voraussetzung kein gerichteter Pfad sein.

Da außerdem die erste Kante von $\pi(i_{k_1}, j)$ von i_{k_1} wegführt, existiert auf $\pi(i_{k_1}, j)$ ein Kollisionspunkt und damit auch auf π selbst. Das ist ein Widerspruch zum d -Zusammenhang von i und j . \square

Kommen wir zur zweiten Regel.

Regel II) Zwei Knoten $i, j \in V$ heißen *d-zusammenhängend gegeben $C \subset V \setminus \{i, j\}$* , falls ein nicht blockierter Pfad zwischen i und j existiert, der kein Element von C enthält. Existiert kein solcher Pfad, heißen i und j *d-separiert gegeben C* .

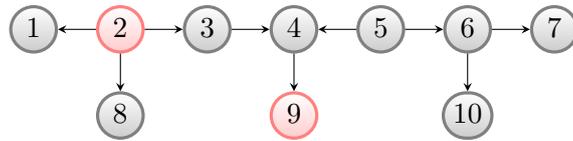


Wie angedeutet sei $C = \{2, 6\}$. Dann sind die Knoten 4, 5 d -separiert von Knoten 7 und die Knoten 3, 4 von Knoten 1 gegeben C , da jeder ungerichtete Pfad über einen der roten Punkte läuft. Die übrigen Beziehungen bleiben erhalten.

Wir definieren als letztes für die Menge C eine „befreiende“ Eigenschaft:

Regel III) Falls C einen Kollisionspunkt i_l oder einen Nachfolger von diesem enthält, werden alle Pfade, die i_l als Kollisionspunkt enthalten, als nicht blockierte Pfade betrachtet.

⁴Wenn wir $\pi(i, i_{k_1})$ nicht minimal wählen, existiert auf $\pi(i, i_{k_1})$ ein Kollisionspunkt, ein Widerspruch zum d -Zusammenhang von i und j .



Hier wird nun die blockierende Eigenschaft von Knoten 4 aufgehoben, sodass Knoten 3 nun mit jedem der Knoten 5, 6, 7, 9, 10 d-zusammenhängend ist. Die übrigen Beziehungen bleiben wieder bestehen.

Diese drei Regeln eignen sich durchaus als Definition des d-Zusammenhangs von zwei Knoten; wir geben noch eine Definition an, die obige Entwicklung charakterisiert und aus der sich die Regeln ableiten lassen.

Definition 1.3.16. Zwei Knoten $i, j \in V$ sind *d-zusammenhängend gegeben* $C \subset V \setminus \{i, j\}$, falls ein ungerichteter Pfad π von i nach j existiert, sodass

- I) alle Kollisionspunkte von π in $C \cup \text{an}(C)$ [= $\text{AN}(C)$] sind und
- II) kein Nicht-Kollisionspunkt von π in C ist.

Ein solcher Pfad heißt *aktiv gegeben* C oder *nicht-blockiert von* C .

Seien $A, B, C \subset V$ paarweise disjunkt, $A, B \neq \emptyset$.

C *d-separiert* die Mengen A und B , falls keine Knoten $i \in A$ und $j \in B$ existieren, sodass i und j d-zusammenhängend gegeben C sind – also kein aktiver Pfad von einem Knoten aus A zu einem Knoten aus B gegeben C existiert.

Die Definition der d-Separation ist optimal auf unsere Bedürfnisse zugeschnitten. Sie ist beim ersten Lesen unübersichtlich, selbst mit obigen Regeln im Hinterkopf, es gibt aber neben diesen Regeln noch eine schöne Charakterisierung der d-Separation.

Dazu definieren wir:

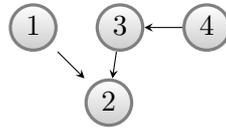
Definition 1.3.17. Der *Moral-Graph* von G ist der ungerichtete Graph $G^m := (V, E^m)$ mit $\{i, j\} \in E^m$, $i \neq j$, falls

1. $(i, j) \in E$ oder
2. $(j, i) \in E$ oder
3. i und j besitzen ein gemeinsames Kind.

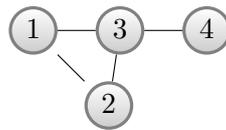
Die Kanten, die im Moral-Graph aus dem zweiten Fall entstehen, nennen wir *hinzugefügte Kanten*.

Wie der Name also andeutet, ist dieser Graph ein, in der allgemein bekannten Wortbedeutung, sehr moralischer Graph: Er entsteht aus G durch „vergessen“ der Orientierung und „verheiraten“ von Eltern, die ein gemeinsames Kind haben und trotzdem (!) noch nicht „vermählt“ sind.

Beispiel 1.3.18. Sei G folgender gerichteter Graph:



Der zugehörige Moral-Graph ist dann



Das nachstehende Lemma, das sogenannte „Pearl’s Separation Criterion“, führt die Frage der d-Separation auf ungerichtete Graphen zurück. Es ist ebenfalls ein Ergebnis, was wir für eine unserer Hauptaussagen maßgeblich nutzen werden. Das Kriterium wird oft zitiert und benutzt, dennoch paradoxerweise selten bewiesen.

Meistens wird auf Lemma 3.25 aus [Lau96] mit dem dort angegebenen, vergleichsweise kurzen Beweis verwiesen. Wie der Autor aber nach Veröffentlichung selbst richtig stellt, ist dieser wohl falsch, da er einen Fall unberücksichtigt lässt. Dennoch wird das Zitat weiter genutzt und nicht weiter vertieft.

Inwieweit dieser fehlende Fall zum Tragen kommt, habe ich nicht herausfinden können. Auf eine entsprechende Frage an den Autor erhielt ich sinngemäß die folgende Antwort: Bei Existenz von ungerichteten Pfaden $A--B$ und $B--C$ kann es bei deren Zusammenfügen zu dem ungerichteten Pfad $A--B--C$ vorkommen, dass dieser zusammengefügte Pfad einen Selbstschnitt enthält. Das sei aber nicht erlaubt.

Warum diese Selbstschnitte ein Problem darstellen sollen, konnte ich bisher nicht endgültig klären; eine erste Betrachtung der dabei auftretenden Fälle scheint keinen Widerspruch zu liefern. Auch den Bezug zu dem Fehler in seinem Beweis habe ich nicht herstellen können.

Selbst nach weiterer intensiver Suche in mehreren Artikeln, die ein ähnliches Thema behandeln und das genannte Kriterium nutzen, habe ich keinen Text gefunden, in dem ein „richtiger“ Beweis angegeben wurde, bis auf das Paper [Ric99]. Auch einem Paper von Sadeghi und Lauritzen, auf das Professor Lauritzen in seiner Antwort auf obige Frage explizit verweist, konnte ich keine Details entnehmen.

Daher bauen wir unseren Beweis auf [Ric99] auf. Dort wird ein allgemeinerer Fall von Graphen (die erwähnten gemischten Graphen) betrachtet, unter denen die gerichteten, azyklischen Graphen einen Spezialfall darstellen. Der Zusammenhangsbegriff, der dort genutzt wird (die sogenannte *m-Separation*), wird bei der Einschränkung auf gerichtete, azyklische Graphen zu der von uns benötigten d-Separation.

Lemma 1.3.19 (Pearl's Separation Criterion). *Seien A, B, C disjunkte Teilmengen von V , A, B nicht-leer. Dann gilt:
 A und B werden genau dann in G von C d -separiert, wenn A und B in $(G_{AN(A \cup B \cup C)})^m$ von C getrennt werden.*

Beweis. Wir zeigen hier jeweils die Negation.

„ \Leftarrow “:

Angenommen A und B sind d -zusammenhängend gegeben C . Dann existiert für zwei Knoten $i \in A$ und $j \in B$ ein Pfad $\pi = (i, i_2, \dots, i_{k-1}, j)$, dessen Knoten alle in $AN(A \cup B \cup C)$ liegen, denn für einen Knoten $\gamma \in \pi$ gilt:

Ist γ ein Kollisionspunkt, so muss wegen des d -Zusammenhangs $\gamma \in AN(C)$ gelten; offensichtlich ist dann auch $\gamma \in AN(A \cup B \cup C)$.

Falls γ kein Kollisionspunkt ist, muss mindestens eine der inzidenten Kanten von γ weg-führen, diese Kante betrachten wir jetzt. Es können insgesamt vier Fälle auftreten:

a)

b)

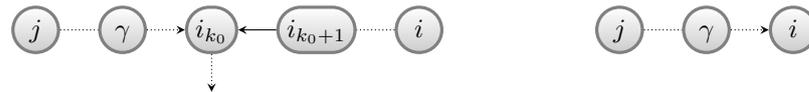


(Die gepunkteten Kanten ohne Orientierung sollen andeuten, dass die Orientierung auf diesem Teilpfad irrelevant für den weiteren Beweis ist; die gepunkteten Kanten mit Ori-entierung deuten an, dass alle Kanten auf diesem Teilpfad die angegebene Orientierung haben.)

Die zwei anderen Fälle sind wie die bisher Dargestellten, wobei i und j jeweils vertauscht werden müssen:

a')

b')



In Fall a) muss erneut wegen des d -Zusammenhangs $i_{k_0} \in AN(C)$ gelten, damit gilt $\gamma \in AN(C)$ und schließlich wieder $\gamma \in AN(A \cup B \cup C)$. Analog im Fall a').

In Fall b) ist $\gamma \in AN(\{j\})$, wegen $j \in B$ also $\gamma \in AN(B)$ und somit $\gamma \in AN(A \cup B \cup C)$. Betrachten wir noch den Fall b'), gilt analog zu voriger Überlegung $\gamma \in AN(\{i\})$. Damit folgt auch hier $\gamma \in AN(A \cup B \cup C)$.

Also ist π tatsächlich ein Pfad, der nur Knoten aus $AN(A \cup B \cup C)$ enthält.

Jeder Kollisionspunkt auf dem Pfad π lässt in dem Moral-Graph eine Kante zwischen den Eltern entstehen (falls noch nicht vorhanden), wodurch in $(G_{AN(A \cup B \cup C)})^m$ ein Pfad ent-steht, der C umgeht: Jede der neu entstandenen Kanten kann keinen Endpunkt in C haben,

da sonst zwei Kollisionspunkte in π direkt aufeinander folgen würden, was in gerichteten, azyklischen Graphen nicht möglich ist.

„ \Rightarrow “:

Nehmen wir umgekehrt an, dass C die Mengen A und B in $(G_{AN(A \cup B \cup C)})^m$ nicht trennt; es existiert also ein ungerichteter Pfad $\pi = (i_1, \dots, i_k)$ von $i_1 \in A$ nach $i_k \in B$, der keine Knoten von C enthält. Wir wählen π minimal (siehe Definition 1.3.4).

Der Pfad π besitzt Kanten, die schon in $G_{AN(A \cup B \cup C)}$ vorkommen, und solche, die aus einer „Vermählung“ entstanden sind – in diesem Fall gibt es in $G_{AN(A \cup B \cup C)}$ einen Kollisionspunkt γ .

Sei ν die Folge von Kanten in $G_{AN(A \cup B \cup C)}$, die schon in π vorkommen, und den Kanten $(i_l, \gamma), (i_{l+1}, \gamma) \in E$ ($\gamma = i_l$ für gewisse $l \in [n]$). Die Kanten, die also nicht in $G_{AN(A \cup B \cup C)}$ liegen, werden „umgangen“.

Jeder Nicht-Kollisionspunkt von ν ist auch in π enthalten und somit nicht in C . Da alle Knoten von ν nach Konstruktion in $AN(A \cup B \cup C)$ liegen, sind alle Kollisionspunkte in $AN(A \cup B \cup C)$. Aus dem nächsten Lemma folgt dann die Behauptung. \square

Das nachstehende Lemma ist technischer Natur, es konstruiert unter bestimmten Voraussetzungen einen Pfad, der den d -Zusammenhang von zwei Punkten garantiert.

Lemma 1.3.20. *Seien A, B, C disjunkte Teilmengen von V , A, B nicht-leer.*

Weiter seien $u \in A$ und $v \in B$ Knoten, die verbunden werden durch einen Pfad π mit den Eigenschaften:

- *kein Nicht-Kollisionspunkt ist in C und*
- *jeder Kollisionspunkt ist in $AN(A \cup B \cup C)$.*

Dann existieren Knoten $u^ \in A$ und $v^* \in B$, die d -zusammenhängend gegeben C sind.*

Beweis. Sei π der vorausgesetzte Pfad. Wir starten in u und „gehen“ diesen Pfad entlang.

Wir wählen d als den letzten Knoten auf π , der in $AN(A) \setminus AN(C)$ liegt, falls dieser existiert, sonst $d = u$.

Analog wählen wir e als den ersten Knoten auf π , der nach d in $AN(B) \setminus AN(C)$ auftaucht, falls dieser existiert, sonst $e = v$.

Seien die Knoten $u^* \in A$ und $v^* \in B$ so, dass $d \in AN(u^*)$ und $e \in AN(v^*)$ (diese existieren nach Definition von d und e).

Wir halten fest:

- a) Alle Zwischenpunkte von $\pi(d, e)$ liegen in $AN(A \cup B \cup C)$.
- b) Nach Definition von u^* gibt es einen gerichteten Pfad ν_{du^*} von d nach u^* .
- c) Nach Definition von v^* gibt es einen gerichteten Pfad ν_{ev^*} von e nach v^* .
- d) Kein Knoten von ν_{du^*} oder ν_{ev^*} liegt in $AN(C)$, insbesondere nicht in C .
- e) Kein Knoten von ν_{ev^*} ist ein Vorgänger von u^* .

Den Fall a) wollen wir genauer herleiten, die übrigen sind offensichtlich. Wir führen dazu Induktion über die Anzahl n der Kollisionspunkte auf $\pi(d, e) =: \mu$

Für $n = 0$ sind d und e per Definition d -zusammenhängend und nach Lemma 1.3.15 tritt einer der drei nachstehenden Fälle auf:

Fall I) μ ist ein gerichteter Pfad.

Dann sind alle Knoten auf μ Elemente in $\text{AN}(x)$ und somit in $\text{AN}(A \cup B \cup C)$, da $x \in \text{AN}(A \cup B \cup C)$.

Fall II) μ^{-1} ist ein gerichteter Pfad.

Hier sind alle Knoten auf μ Elemente in $\text{AN}(d)$ und somit wieder in $\text{AN}(A \cup B \cup C)$, da $d \in \text{AN}(A)$.

Fall III) Es existiert ein Knoten v_0 auf μ , sodass $\mu^{-1}(v_0, d)$ und $\mu(v_0, e)$ gerichtete Pfade sind.

Nun sind alle Knoten auf $\mu^{-1}(v_0, d)$ Elemente in $\text{AN}(d)$ und damit in $\text{AN}(A \cup B \cup C)$. Des Weiteren sind alle Knoten auf $\mu(v_0, e)$ Elemente in $\text{AN}(e)$, also auch in $\text{AN}(A \cup B \cup C)$.

Das zeigt den Induktionsanfang.

Habe μ nun $n+1$ Kollisionspunkte und sei x der erste Kollisionspunkt auf μ nach d . Nach Voraussetzung ist $x \in \text{AN}(A \cup B \cup C)$.

Die Pfade $\mu(d, x)$ und $\mu(x, e)$ haben keinen beziehungsweise $n-1$ Kollisionspunkte.

Für $\mu(d, x)$ kann mit Lemma 1.3.15 analog zum Induktionsanfang gefolgert werden, dass $\mu(d, x)$ in $\text{AN}(A \cup B \cup C)$ verläuft.

Für $\mu(x, e)$ folgt mit dem Induktionsanfang, dass auch hier alle Knoten in $\text{AN}(A \cup B \cup C)$ liegen.

Damit liegen alle Knoten von $\mu = \pi(d, e)$ in $\text{AN}(A \cup B \cup C)$, was in a) behauptet wurde.

Betrachten wir die Aussage a) noch genauer. Sei dazu v ein Zwischenpunkt von $\pi(d, e)$, also ist v insbesondere ein Knoten, der auf π nach d und vor e liegt.

Angenommen $v \notin \text{AN}(C)$. Dann muss $v \in \text{AN}(A \cup B) = \text{AN}(A) \cup \text{AN}(B)$ liegen. Gilt $v \in \text{AN}(A)$, ist also $v \in \text{AN}(A) \setminus \text{AN}(C)$, was einen Widerspruch zur Maximalität der Wahl von d darstellt. Gilt andererseits $v \in \text{AN}(B)$, folgt $v \in \text{AN}(B) \setminus \text{AN}(C)$, in diesem Fall erhalten wir einen Widerspruch zur Minimalität der Wahl von e .

Somit muss für alle Zwischenpunkte v von $\pi(d, e)$ sogar $v \in \text{AN}(C)$ gelten, diese Aussage bezeichnen wir mit a').

Mit 1.3.2 und b) folgt, dass $\pi(d, e)$ und ν_{du^*} nur den Knoten d gemeinsam haben. Analog haben $\pi(d, e)$ und ν_{e, v^*} nur den Knoten e gemeinsam. Aus den Beobachtungen b) und e) folgt, dass ν_{du^*} und ν_{e, v^*} keinen Knoten gemeinsam haben.

Sei μ der Pfad, der durch Konkatenation von ν_{du^*} , $\pi(d, e)$ und ν_{e, v^*} entsteht.

Aus der Konstruktion dieses Pfades folgt nun:

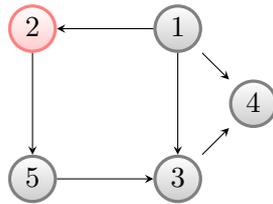
Jeder Zwischenpunkt von ν_{du^*} und ν_{e, v^*} ist ein Nicht-Kollisionspunkt und nicht in C . Jeder Kollisionspunkt auf μ ist demnach auf $\pi(d, e)$ und wegen 1.3.2) somit in $\text{AN}(C)$. Weiter liegt kein Nicht-Kollisionspunkt von μ in C nach Voraussetzung. Außerdem sind d und e Nicht-Kollisionspunkte von μ und nicht in C .

Demnach sind u^* und v^* d -zusammenhängend via μ . □

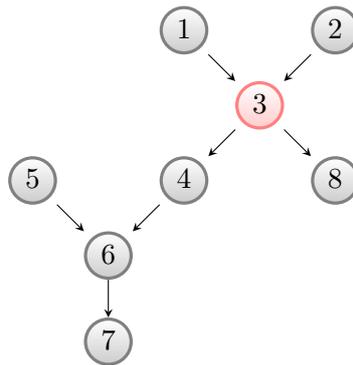
Damit haben wir ein gutes Verständnis des Begriffs „d-Zusammenhang“ erworben. Es soll aber noch bemerkt werden, dass diese Charakterisierung in der Praxis einen Mehraufwand bedeutet (da ein neuer Graph erstellt werden muss) und von daher beispielsweise für eine programmiertechnische Umsetzung eher uninteressant ist.

Wir betrachten noch zwei Beispiele zur Festigung des Begriffs.

Beispiel 1.3.21. Im ersten Beispiel sind die Knoten 1 und 5 d-separiert gegeben Knoten 2, aber durch keine andere Teilmenge von $\{2, 3, 4\}$.



Nachfolgend sind beispielsweise die Knoten 1 und 4, 4 und 5 sowie 2 und 7 d-separiert gegeben Knoten 3; die Knoten 4 und 6 sowie 1 und 2 sind d-zusammenhängend.



Diese Beispiele sollen die Vorbetrachtungen beenden.

Kapitel 2.

Modelle bedingter Unabhängigkeit

Bedingte Unabhängigkeitsbedingungen sind Einschränkungen, die an die gemeinsamen Verteilungen mehrerer Zufallsvariablen gestellt werden. Diese Eigenschaft drückt die Unabhängigkeit zweier Ereignisse beziehungsweise Zufallsvariablen gegeben eines dritten Ereignisses beziehungsweise einer dritten Zufallsvariable aus.

Als Modell bedingter Unabhängigkeit werden wir eine Familie von Verteilungen betrachten, die gewissen bedingten Unabhängigkeitsbedingungen genügen.

2.1. Bedingte Unabhängigkeit

Hier wollen wir den Begriff der bedingten Unabhängigkeit allgemein einführen. Als Motivation betrachten wir folgendes Szenario:

Beispiel 2.1.1. Gegeben seien die Ereignisse B , G und R mit der Wahrscheinlichkeitsverteilung P , die in folgendem Bild dargestellt sind – wobei die blaue Fläche für $P(B)$, die grüne Fläche für $P(G)$ und die rote Fläche für $P(R)$ stehen soll (jeweils im Bezug auf die Gesamtfläche):

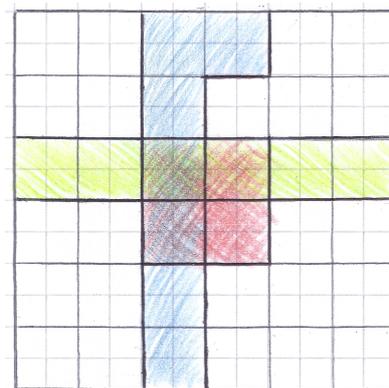


Abbildung 2.1.: Darstellung der Ereignisse B , G und R ; die Ereignisse B und G sind abhängig.

Beispielsweise gilt $P(B) = \frac{7}{36}$ und $P(B|R) = \frac{1}{2}$. Nun betrachten wir die stochastische Abhängigkeit von B und G . Es gilt:

$$P(B)P(G) = \frac{7}{36} \cdot \frac{6}{36} = \frac{7}{216} \neq \frac{1}{36} = P(B \cap G),$$

also sind B und G nicht unabhängig. Aus dem Bild lässt sich aber ablesen, dass

$$P(B|R)P(G|R) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} = P(B \cap G|R).$$

Wenn also R vorausgesetzt ist, sind die Ereignisse B und G unabhängig.

Das wollen wir verallgemeinern zu einer Definition über beliebige Zufallsvariablen.

Definition 2.1.2. Seien $A, B, C \subset [n]$ paarweise disjunkt.

Der Zufallsvektor X_A ist *bedingt unabhängig von X_B gegeben/unter X_C* genau dann, wenn

$$f_{A \cup B|C}(x_A, x_B|x_C) = f_{A|C}(x_A|x_C)f_{B|C}(x_B|x_C)$$

für ν -fast alle x_A, x_B und x_C mit $f_C(x_C) > 0$.

In Zeichen: $X_A \perp\!\!\!\perp X_B|X_C$, manchmal auch zu $A \perp\!\!\!\perp B|C$ abgekürzt.

Wählen wir $C = \emptyset$, ist diese Definition genau die der stochastischen Unabhängigkeit von Zufallsvariablen, wie in Kapitel 1 eingeführt.

Die Definition drückt den Umstand aus, dass, wenn die Wahrscheinlichkeit von X_A gegeben X_C bekannt ist, dies keine Aussage über die Wahrscheinlichkeit von X_B zulässt; beziehungsweise, wenn X_C voraussetzt wird, ist X_B irrelevant, um die Wahrscheinlichkeit von X_A zu bestimmen. Es kann also die folgende äquivalente Definition genutzt werden:

$$X_A \perp\!\!\!\perp X_B|X_C \Leftrightarrow P(X_A|X_B, X_C) = P(X_A|X_C).$$

Geben wir mehr als eine Einschränkung an Zufallsvariablen im Hinblick auf bedingte Unabhängigkeit vor, so sprechen wir von einem Modell für bedingte Unabhängigkeit.

Die Familie von Verteilungen, die mit einem solchen Modell korrespondiert, genügt dann einer Menge von Einschränkungen $\mathcal{C} = \{A_1 \perp\!\!\!\perp B_1|C_1, \dots, A_l \perp\!\!\!\perp B_l|C_l\}$, wobei A_i, B_i und C_i für alle $i \in [l]$ paarweise disjunkt sein müssen.

Im obigen Beispiel hatten wir gesehen, dass zwei abhängige Zufallsvariablen bedingt auf eine dritte Zufallsvariable unabhängig (gegeben der dritten Zufallsvariable) werden können. Nun stellt sich die Frage, ob zwei unabhängige Zufallsvariablen bedingt auf eine dritte unabhängig (gegeben der dritten Zufallsvariable) bleiben. Das nachfolgende Beispiel gibt darauf eine Antwort.

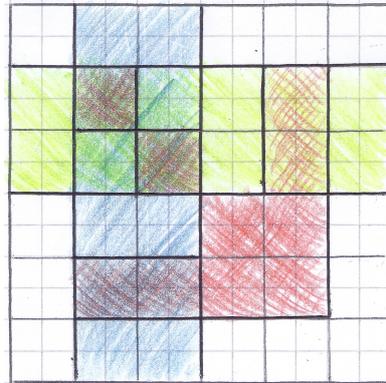


Abbildung 2.2.: Darstellung der Ereignisse B , G und R ; die Ereignisse werden bei Bedingung abhängig

Beispiel 2.1.3. Ähnlich des einführenden Beispiels betrachten wir das Szenario aus Abbildung 2.2:

Es lässt sich nun analog zu Beispiel 2.1.1 nachrechnen, dass $B \perp\!\!\!\perp G$ aber $B \not\perp\!\!\!\perp G|R$ gilt.

Zufallsvariablen können ihr Verhalten in Bezug auf stochastische Unabhängigkeit demnach bei Bedingung auf eine dritte Variable in beide Richtungen verändern. Ähnliche Beispiele lassen sich dafür finden, dass sich die Un-/Abhängigkeit von zwei Zufallsvariablen bei Bedingung auf eine dritte Zufallsvariable *nicht* ändert.

Im weiteren Text seien stets $A, B, C \subset [n]$ paarweise disjunkt, $A, B \neq \emptyset$.

Um ein Gefühl für die Arbeitsweise mit der bedingten Unabhängigkeit zu erhalten, beweisen wir einige – mehr oder weniger – direkte Folgerungen aus der Definition. Sie stellen ein nützliches Hilfsmittel dar.

Proposition 2.1.4. *Für die bedingten Unabhängigkeitsbedingungen gelten folgende Eigenschaften:*

I) *Symmetrie:*

$$X_A \perp\!\!\!\perp X_B|X_C \Rightarrow X_B \perp\!\!\!\perp X_A|X_C$$

II) *Dekomposition:*

$$X_A \perp\!\!\!\perp X_{B \cup D}|X_C \Rightarrow X_A \perp\!\!\!\perp X_B|X_C$$

III) *Schwache Vereinigung:*

$$X_A \perp\!\!\!\perp X_{B \cup D}|X_C \Rightarrow X_A \perp\!\!\!\perp X_B|X_{C \cup D}$$

IV) Kontraktion:

$$X_A \perp\!\!\!\perp X_B | X_{C \cup D} \text{ und } X_A \perp\!\!\!\perp X_D | X_C \Rightarrow X_A \perp\!\!\!\perp X_{B \cup D} | X_C$$

V) Schnittaxiom: Falls $f(x) > 0$ für alle $x \in \mathcal{X}$ erfüllt ist, gilt

$$X_A \perp\!\!\!\perp X_B \perp\!\!\!\perp X_{C \cup D} \text{ und } X_A \perp\!\!\!\perp X_C | X_{B \cup D} \Rightarrow X_A \perp\!\!\!\perp X_{B \cup C} | X_D$$

Beweis. Seien $x_C \in \mathcal{X}_C$ stets so, dass $f_C(x_C) > 0$, und (falls nötig) $x_D \in \mathcal{X}_D$ so, dass $f_{C \cup D}(x_C, x_D) > 0$ gilt.

I) folgt aus der Kommutativität der Multiplikation in \mathbb{R} , denn

$$\begin{aligned} X_A \perp\!\!\!\perp X_B | X_C &\Rightarrow f_{B \cup A | C}(x_B, x_A | x_C) \\ &= f_{A \cup B | C}(x_A, x_B | x_C) \\ &= f_{A | C}(x_A | x_C) f_{B | C}(x_B | X_C) \\ &= f_{B | C}(x_B | X_C) f_{A | C}(x_A | x_C) \\ &\Rightarrow X_B \perp\!\!\!\perp X_A | X_C \end{aligned}$$

II) Gelte $X_A \perp\!\!\!\perp X_{(B \cup D)} | X_C$, also

$$\frac{f_{A \cup B \cup C \cup D}(x_A, x_B, x_C, x_D)}{f_C(x_C)} = \frac{f_{A \cup C}(x_A, x_C)}{f_C(x_C)} \cdot \frac{f_{B \cup D \cup C}(x_B, x_D, x_C)}{f_C(x_C)}.$$

Integration der linken Seite über \mathcal{X}_D liefert

$$\begin{aligned} \frac{1}{f_C(x_C)} \int_{\mathcal{X}_D} f_{A \cup B \cup C \cup D}(x_A, x_B, x_C, x_D) d\nu_D(x_D) \\ = \frac{f_{A \cup B \cup C}(x_A, x_B, x_C)}{f_C(x_C)}. \end{aligned}$$

Integration der rechten Seite über \mathcal{X}_D dann

$$\begin{aligned} \frac{f_{A \cup C}(x_A, x_C)}{(f_C(x_C))^2} \int_{\mathcal{X}_D} f_{B \cup D \cup C}(x_B, x_D, x_C) d\nu_D(x_D) \\ = \frac{f_{A \cup C}(x_A, x_C)}{f_C(x_C)} \cdot \frac{f_{B \cup C}(x_B, x_C)}{f_C(x_C)}. \end{aligned}$$

Insgesamt also $X_A \perp\!\!\!\perp X_B | X_C$.

III) Gelte wieder $X_A \perp\!\!\!\perp X_{(B \cup D)} | X_C$, also (nach Multiplikation mit $\frac{f_C(x_C)}{f_{C \cup D}(x_C, x_D)}$ und umsortieren)

$$\begin{aligned} \frac{f_{A \cup B \cup C \cup D}(x_A, x_B, x_C, x_D)}{f_{C \cup D}(x_C, x_D)} &= \frac{f_{B \cup D \cup C}(x_B, x_D, x_C)}{f_{C \cup D}(x_C, x_D)} \cdot \frac{f_{A \cup C}(x_A, x_C)}{f_C(x_C)} \\ &= f_{B | C \cup D}(x_B | x_C, x_D) \cdot \frac{f_{A \cup C}(x_A, x_C)}{f_C(x_C)} \end{aligned}$$

Es bleibt damit

$$f_{A|CUD}(x_A|x_C, x_D) = f_{A|C}(x_A|x_C)$$

zu zeigen, Einsetzen liefert dann die Behauptung.

Mit II) gilt $X_A \perp\!\!\!\perp X_D | X_C$, das heißt

$$\begin{aligned} \frac{f_{AUDUC}(x_A, x_D, x_C)}{f_C(x_C)} &= \frac{f_{AUC}(x_A, x_C)}{f_C(x_C)} \cdot \frac{f_{DUC}(x_D, x_C)}{f_C(x_C)} \\ \Leftrightarrow \frac{f_{AUDUC}(x_A, x_D, x_C)}{f_{DUC}(x_D, x_C)} &= \frac{f_{AUC}(x_A, x_C)}{f_C(x_C)}, \end{aligned}$$

was zu zeigen war.

- IV) Da $X_A \perp\!\!\!\perp X_B | X_{CUD}$ vorausgesetzt wird, ist nach Multiplizieren mit $\frac{f_{CUD}(x_C, x_D)}{f_C(x_C)}$ die Gleichheit

$$f_{AUBUD|C}(x_A, x_B, x_D|x_C) = f_{AUD|C}(x_A, x_D|x_C) f_{B|CUD}(x_B|x_C, x_D)$$

erfüllt.

Wegen der zweiten Voraussetzung gilt

$$\frac{f_{AUDUC}(x_A, x_D, x_C)}{f_C(x_C)} = \frac{f_{AUC}(x_A, x_C)}{f_C(x_C)} \cdot \frac{f_{DUC}(x_D, x_C)}{f_C(x_C)}.$$

Durch Multiplikation dieser Gleichung mit $f_C(x_C)$ und Einsetzen des Resultats in obige Gleichung folgt dann die Behauptung.

- V) Die beiden Voraussetzungen liefern

a) $f_{AUB|CUD}(x_A, x_B|x_C, x_D) = f_{A|CUD}(x_A|x_C, x_D) f_{B|CUD}(x_B|x_C, x_D)$

b) $f_{AUC|BUD}(x_A, x_C|x_B, x_D) = f_{A|BUD}(x_A|x_B, x_D) f_{C|BUD}(x_C|x_B, x_D)$

Wir multiplizieren a) mit $f_{CUD}(x_C, x_D)$, b) mit $f_{BUD}(x_B, x_D)$ und erhalten

a') $f_{AUBUCUD}(x_A, x_B, x_C, x_D) = f_{A|CUD}(x_A|x_C, x_D) f_{BUCUD}(x_B, x_C, x_D)$

b') $f_{AUBUCUD}(x_A, x_B, x_C, x_D) = f_{A|BUD}(x_A|x_B, x_D) f_{BUCUD}(x_B, x_C, x_D)$

Gleichsetzen der beiden Ausdrücke und Kürzen von $f_{BUCUD}(x_B, x_C, x_D)$ (das ist möglich, da $f(x) > 0$ vorausgesetzt wurde) ergibt dann

$$f_{A|CUD}(x_A|x_C, x_D) = f_{A|BUD}(x_A|x_B, x_D).$$

Da die rechte Seite der Gleichung nicht von x_C abhängt (x_C taucht in dem Ausdruck nicht auf), können wir

$$f_{A|CUD}(x_A|x_C, x_D) = f_{A|D}(x_A|x_D)$$

folgern.

Setzen wir dieses Ergebnis in a') ein und teilen durch $f_D(x_D)$, haben wir die Behauptung gezeigt. \square

In den Voraussetzungen dieser Proposition haben wir für das Schnittaxiom $f(x) > 0$ gefordert, diese Forderung stellte sich im Verlauf des Beweises als zu stark heraus. Schlimmstenfalls muss $f_{B \cup C \cup D}(x_B, x_C, x_D) > 0$ gelten. Die Bedingung kann noch weiter abgeschwächt werden, darauf wollen wir hier aber nicht näher eingehen.

Wie in [HG90] können diese Eigenschaften als Axiome betrachtet werden. Ein Unabhängigkeitsmodell, das den Eigenschaften I) bis IV) genügt, wird dann *Semi-Graphoid* genannt, falls Eigenschaft V) zusätzlich gilt heißt das Modell *Graphoid*.

Ein Ergebnis über sogenannte *diskrete Modelle* halten wir hier fest:

Wir betrachten den Fall, dass X ein diskreter Zufallsvektor ist, der seine Werte wie in den Vorbetrachtungen in $\mathcal{X} = \prod_{j=1}^n [r_j]$ annimmt.

In dieser Situation lassen sich die bedingten Unabhängigkeitsbedingungen wie angekündigt in ein System von quadratischen Polynomgleichungen überführen, welches eine Erweiterung zu Proposition 1.2.2 darstellt.

Proposition 2.1.5. *Es gilt die Eigenschaft $X_A \perp\!\!\!\perp X_B | X_C$ genau dann, wenn*

$$p_{i_A, i_B, i_C, +} p_{j_A, j_B, i_C, +} - p_{j_A, i_B, i_C, +} p_{i_A, j_B, i_C, +} = 0$$

für alle $i_A, j_A \in \mathcal{X}_A, i_B, j_B \in \mathcal{X}_B, i_C \in \mathcal{X}_C$ gilt.

Beweis. „ \Rightarrow “:

Es gelte $X_A \perp\!\!\!\perp X_B | X_C$, also

$$P(X_A = i_A, X_B = i_B | X_C = i_C) = P(X_A = i_A | X_C = i_C) \cdot P(X_B = i_B | X_C = i_C)$$

für alle $i_A \in \mathcal{X}_A, i_B \in \mathcal{X}_B, i_C \in \mathcal{X}_C$.

Nach Definition der bedingten Unabhängigkeit und unter Ausnutzung der verkürzten Notation von diskreten Verteilungen ist diese Gleichung äquivalent zu

$$\frac{p_{i_A, i_B, i_C, +}}{p_{i_C, +}} = \frac{p_{i_A, i_C, +}}{p_{i_C, +}} \cdot \frac{p_{i_B, i_C, +}}{p_{i_C, +}}$$

für alle $i_A \in \mathcal{X}_A, i_B \in \mathcal{X}_B, i_C \in \mathcal{X}_C$.

Für $i_A, j_A \in \mathcal{X}_A, i_B, j_B \in \mathcal{X}_B, i_C \in \mathcal{X}_C$ gilt dann

$$\frac{p_{i_A, i_B, i_C, +}}{p_{i_C, +}} \cdot \frac{p_{j_A, j_B, i_C, +}}{p_{i_C, +}} = \frac{p_{i_A, i_C, +}}{p_{i_C, +}} \cdot \frac{p_{i_B, i_C, +}}{p_{i_C, +}} \cdot \frac{p_{j_A, i_C, +}}{p_{i_C, +}} \cdot \frac{p_{j_B, i_C, +}}{p_{i_C, +}}$$

Auf der rechten Seite können wir wegen der bedingten Unabhängigkeit von A und B den ersten und vierten sowie den zweiten und dritten Term zusammenfassen, sodass

$$\frac{p_{i_A, i_B, i_C, +}}{p_{i_C, +}} \cdot \frac{p_{j_A, j_B, i_C, +}}{p_{i_C, +}} = \frac{p_{i_A, j_B, i_C, +}}{p_{i_C, +}} \cdot \frac{p_{j_A, i_B, i_C, +}}{p_{i_C, +}}$$

gilt; multiplizieren mit $(p_{i_C, +})^2$ und umstellen der resultierenden Gleichung liefert die Behauptung.

„ \Leftarrow “:

Gelte nun

$$p_{i_A, i_B, i_C, +} p_{j_A, j_B, i_C, +} - p_{j_A, i_B, i_C, +} p_{i_A, j_B, i_C, +} = 0$$

für alle $i_A, j_A \in \mathcal{X}_A$, $i_B, j_B \in \mathcal{X}_B$, $i_C \in \mathcal{X}_C$.

Seien $i_A \in \mathcal{X}_A$ und $i_B \in \mathcal{X}_B$.

Des Weiteren seien s und t jeweils die Anzahl von verschiedenen Ausgängen der Zufallsvariable X_A beziehungsweise X_B .

Wir halten zunächst $j_B \in \mathcal{X}_B$ fest und addieren die Terme, die von X_A indiziert werden:

$$p_{i_A, i_B, i_C, +} \cdot \sum_{k=1}^s p_{(j_A)_k, j_B, i_C, +} - \sum_{k=1}^s p_{(j_A)_k, i_B, i_C, +} \cdot p_{i_A, j_B, i_C, +} = 0$$

$$\Leftrightarrow p_{i_A, i_B, i_C, +} \cdot p_{j_B, i_C, +} - p_{i_B, i_C, +} \cdot p_{i_A, j_B, i_C, +} = 0$$

Addieren der Terme, die von X_B indiziert werden, liefert

$$p_{i_A, i_B, i_C, +} \cdot \sum_{l=1}^t p_{(j_B)_l, i_C, +} - p_{i_B, i_C, +} \cdot \sum_{l=1}^t p_{i_A, (j_B)_l, i_C, +} = 0$$

$$\Leftrightarrow p_{i_A, i_B, i_C, +} \cdot p_{i_C, +} - p_{i_B, i_C, +} \cdot p_{i_A, i_C, +} = 0.$$

Wir dividieren letztendlich durch $(p_{i_C, +})^2$, stellen die resultierende Gleichung wieder um und beenden damit den Beweis. \square

Für $C = \emptyset$ sind das genau die Gleichungen, die nach Proposition 1.2.2 aufgeführt werden.

Ein weiterer wichtiger Spezialfall ist der des multivariat normalverteilten Zufallsvektors; auch hier zeigt sich, dass bedingte Unabhängigkeit eine algebraische Bedingung ist, genauer eine Bedingung an die Kovarianzmatrix der Normalverteilung.

Im Fall 1-dimensionaler, unabhängiger Zufallsvektoren ist bekannt, dass deren Korrelationskoeffizient 0 ist, insbesondere auch deren Kovarianz. Diese Aussage wollen wir für die Normalverteilung in höhere Dimensionen verallgemeinern.

Lemma 2.1.6. *Seien $X \in \mathbb{R}^p$ und $Y \in \mathbb{R}^q$ gemeinsam multivariat normalverteilte Zufallsvariablen mit $(X^T, Y^T)^T \sim \mathcal{N}(\mu, \Sigma)$. Folgende Aussagen sind äquivalent:*

I) $X \perp\!\!\!\perp Y$

II) $\Sigma = \begin{pmatrix} \Sigma_X & 0 \\ 0 & \Sigma_Y \end{pmatrix}$

III) $\text{Cov}(X_i, Y_j) = 0$ für alle $i \in [p], j \in [q]$.

Sind die Zufallsvariablen X und Y (nicht gemeinsam) multivariat normalverteilt, so gelten diese Äquivalenzen im Allgemeinen nicht, sondern nur die Implikation

$$X \perp\!\!\!\perp Y \Rightarrow \text{Cov}(X_i, Y_j) = 0 \text{ für alle } i \in [p], j \in [q].$$

Diese gilt aber sogar für beliebige Verteilungen.

Beweis. „II) \Leftrightarrow III)“ ist klar nach Definition der Kovarianzmatrix.

„I) \Rightarrow II)“:

Seien $i \in [p]$, $j \in [q]$ und es gelte $X \perp\!\!\!\perp Y$ – damit gilt insbesondere $X_i \perp\!\!\!\perp Y_j$.

Die Kovarianz kann (nach [Hol11]) über

$$\text{Cov}(X_i, Y_j) = \text{E}(X_i Y_j) - \text{E}(X_i) \text{E}(Y_j)$$

berechnet werden. Die Unabhängigkeit von X_i und Y_j impliziert $\text{E}(X_i Y_j) = \text{E}(X_i) \text{E}(Y_j)$ (genaueres siehe Anhang), woraus $\text{Cov}(X_i, Y_j) = 0$ folgt.

„II) \Rightarrow I)“:

Nach Voraussetzung haben Σ und Σ^{-1} die nachstehenden Formen:

$$\Sigma = \begin{pmatrix} \Sigma_X & 0 \\ 0 & \Sigma_Y \end{pmatrix} \quad \text{und} \quad \Sigma^{-1} = \begin{pmatrix} (\Sigma_X)^{-1} & 0 \\ 0 & (\Sigma_Y)^{-1} \end{pmatrix}.$$

Mit dem Kästchensatz gilt offensichtlich $\det(\Sigma) = \det(\Sigma_X) \det(\Sigma_Y)$.

Nach diesen Vorbereitungen betrachten wir die gemeinsame Dichte f von X und Y ; wir können wie folgt umformen:

$$\begin{aligned} f(x, y) &= \frac{1}{\sqrt{\det(\Sigma)(2\pi)^{p+q}}} \exp \left[-\frac{1}{2} \begin{pmatrix} x^T - \mu_X^T & y^T - \mu_Y^T \end{pmatrix} \begin{pmatrix} \Sigma_X^{-1} & 0 \\ 0 & \Sigma_Y^{-1} \end{pmatrix} \begin{pmatrix} x - \mu_X \\ y - \mu_Y \end{pmatrix} \right] \\ &= \frac{1}{\sqrt{\det(\Sigma_X)(2\pi)^p} \sqrt{\det(\Sigma_Y)(2\pi)^q}} \exp \left[-\frac{1}{2} (x - \mu_X)^T \Sigma_X (x - \mu_X) \right. \\ &\quad \left. -\frac{1}{2} (y - \mu_Y)^T \Sigma_Y (y - \mu_Y) \right] \\ &= \frac{1}{\sqrt{\det(\Sigma_X)(2\pi)^p}} \exp \left[-\frac{1}{2} (x - \mu_X)^T \Sigma_X (x - \mu_X) \right] \cdot \\ &\quad \frac{1}{\sqrt{\det(\Sigma_Y)(2\pi)^q}} \exp \left[-\frac{1}{2} (y - \mu_Y)^T \Sigma_Y (y - \mu_Y) \right] \\ &= f_X(x) \cdot f_Y(y) \end{aligned}$$

Damit sind X und Y stochastisch unabhängig. □

Nach unserer intensiven Beschäftigung mit dem Schur-Komplement und dem vorigen Lemma ist die erwähnte algebraische Bedingung an die Kovarianzmatrix eine direkte Folgerung.

Proposition 2.1.7. *Für ein $X \sim \mathcal{N}(\mu, \Sigma)$ gilt die Bedingung $X_A \perp\!\!\!\perp X_B | X_C$ genau dann, wenn für die Untermatrix $\Sigma_{AUC, BUC}$ von Σ die Gleichheit*

$$\text{rank}(\Sigma_{AUC, BUC}) = |C|$$

erfüllt ist.

Beweis. Die bedingte Verteilung von $X_{A \cup B}$ gegeben $X_C = x_C$ ist nach Lemma 1.2.4 die Verteilung

$$\mathcal{N}(\mu_{A \cup B} + \Sigma_{A \cup B, C} \Sigma_{C, C}^{-1} (x_C - \mu_C), \Sigma | \Sigma_{C, C}).$$

Die Kovarianzmatrix hat also die Form

$$\Sigma | \Sigma_C = \begin{pmatrix} \Sigma_{A, A} - \Sigma_{A, C} \Sigma_{C, C}^{-1} \Sigma_{C, A} & \Sigma_{A, B} - \Sigma_{A, C} \Sigma_{C, C}^{-1} \Sigma_{C, B} \\ \Sigma_{B, A} - \Sigma_{B, C} \Sigma_{C, C}^{-1} \Sigma_{C, A} & \Sigma_{B, B} - \Sigma_{B, C} \Sigma_{C, C}^{-1} \Sigma_{C, B} \end{pmatrix}.$$

Definieren wir die Matrix $S := \Sigma_{A \cup C, B \cup C}$, so sind in $\Sigma | \Sigma_{C, C}$ die Einträge neben der Hauptdiagonalen genau das Schur-Komplement $S | \Sigma_{C, C}$ beziehungsweise $(S | \Sigma_{C, C})^T$.

Für $f_{A \cup B | C}$ folgt aus Lemma 2.1.6

$$X_A \perp\!\!\!\perp X_B | X_C \Leftrightarrow S | \Sigma_{C, C} = 0.$$

In den Vorbetrachtungen hatten wir in Lemma 1.1.3 eine Aussagen über den Rang des Schur-Komplements gemacht; damit gilt

$$S | \Sigma_{C, C} = 0 \Leftrightarrow \text{rank } S = \text{rank } \Sigma_{C, C}.$$

Es ist aber $\text{rank } \Sigma_{C, C} = |C|$, demnach gilt insgesamt die Äquivalenz

$$X_A \perp\!\!\!\perp X_B | X_C \Leftrightarrow \text{rank}(\Sigma_{A \cup C, B \cup C}) = |C|,$$

was die Behauptung war. □

Damit schließen wir die allgemeinen Einführungen über die bedingte Unabhängigkeit ab.

2.2. Graphische Modelle

In diesem Abschnitt betrachten wir einen Zufallsvektor $X = (X_v | v \in V)$ für eine Menge V und einen ungerichteten oder gerichteten Graph $G = (V, E)$; die Knoten von G indizieren also den Vektor X .

Wir verknüpfen X mit dem Graphen G , indem wir die Kanten $\{v, w\} \in E$ beziehungsweise $(v, w) \in E$ von G als Abhängigkeiten der Zufallsvariablen X_v und X_w interpretieren – genauer als gewisse bedingte Unabhängigkeitsbedingungen.

Diese Beziehungen sind als paarweise und globale beziehungsweise lokale und globale Markov-Eigenschaften des Graphen bekannt, je nachdem ob G ungerichtet oder gerichtet ist. Als Hauptergebnis werden wir in ungerichteten und gerichteten Graphen die Äquivalenz der beiden jeweiligen Markov-Eigenschaften beweisen.

Obgleich wir zunächst auf die Normalverteilung eingehen, werden wir allgemein keine spezielle Verteilung für den Zufallsvektor X voraussetzen, sondern auf die „Axiome“ aus Proposition 1.2.2 zurückgreifen.

2.2.1. Ungerichtete graphische Modelle

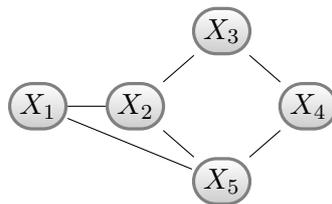
Hier seien $G = (V, E)$ ein ungerichteter Graph und X ein Zufallsvektor, der, wie eben beschrieben, von den Knoten von G indiziert wird.

Definition 2.2.1. Der Zufallsvektor X besitzt die *paarweisen, ungerichteten Markov-Eigenschaften bezüglich G* , falls

$$\{v, w\} \notin E \Rightarrow X_v \perp\!\!\!\perp X_w | X_{V \setminus \{v, w\}}.$$

Dazu zunächst ein Beispiel:

Beispiel 2.2.2. Sei X ein Zufallsvektor, der die ungerichteten, paarweisen Markov-Eigenschaften bezüglich G besitzt, dabei sei G wie folgt:



Die zugehörigen paarweisen Markov-Eigenschaften sind

- $X_1 \perp\!\!\!\perp X_3 | (X_2, X_4, X_5)$

- $X_1 \perp\!\!\!\perp X_4 | (X_2, X_3, X_5)$
- $X_2 \perp\!\!\!\perp X_4 | (X_1, X_3, X_5)$
- $X_3 \perp\!\!\!\perp X_5 | (X_1, X_2, X_4)$

Wegen der Symmetrie der Bedingungen brauchen die übrigen nicht aufgeführt zu werden.

Als erste Aufwärmübung machen wir uns darüber Gedanken, was es heißt, dass ein multivariat normalverteilter Zufallsvektor die aus einem Graphen resultierenden, paarweisen, ungerichteten Markov-Eigenschaften besitzt.

Folgerung 2.2.3. Sei $X \sim \mathcal{N}(\mu, \Sigma)$. Die folgenden Aussagen sind äquivalent:

- I) $X_v \perp\!\!\!\perp X_w | X_{V \setminus \{v, w\}}$
- II) $\det(\Sigma_{wv}) = 0$
- III) $(\sigma^{-1})_{vw} = 0$

Beweis. „II) \Leftrightarrow III)“ folgt aus lemma 1.1.7, denn:

$$\det(\Sigma)(\sigma^{-1})_{vw} = \det(\Sigma_{wv}) \text{ und } \det(\Sigma) \neq 0.$$

Für „I) \Leftrightarrow II)“ benutzen wir Proposition 2.1.7, demnach gilt

$$X_v \perp\!\!\!\perp X_w | X_{V \setminus \{v, w\}} \Leftrightarrow \text{rank}(\Sigma_{wv}) = |V \setminus \{v, w\}| = n - 2. \quad (2.1)$$

„I) \Rightarrow II)“:

Gelte $X_v \perp\!\!\!\perp X_w | X_{V \setminus \{v, w\}}$, also $\text{rank}(\Sigma_{wv}) = n - 2$.

Die Matrix Σ_{wv} hat $n - 1$ Zeilen und Spalten und somit keinen vollen Rang, also gilt $\det(\Sigma_{wv}) = 0$.

„II) \Rightarrow I)“:

Gilt umgekehrt $\det(\Sigma_{wv}) = 0$, kann die Matrix Σ_{wv} keinen vollen Rang besitzen; damit muss $\text{rank} \Sigma_{wv} \leq n - 2$ gelten.

Wir schauen uns die Gestalt von Σ_{wv} genauer an, mit $C := V \setminus \{v, w\}$ ist

$$\Sigma_{wv} = \begin{pmatrix} \Sigma_{\{v\}, \{w\}} & \Sigma_{\{v\}, C} \\ \Sigma_{C, \{w\}} & \Sigma_{C, C} \end{pmatrix}.$$

Die Matrix $\Sigma_{C, C}$ ist positiv definit, also hat $\Sigma_{C, C}$ vollen Rang, nämlich $\text{rank} \Sigma_{C, C} = n - 2$. Weil $\Sigma_{C, C}$ als Teilmatrix in Σ_{wv} auftaucht, hat auch Σ_{wv} selbst mindestens $n - 2$ linear unabhängige Zeilen. Somit gilt die Ungleichung $\text{rank} \Sigma_{wv} \geq n - 2$.

Fassen wir die beiden Ungleichungen zusammen, folgt insgesamt $\text{rank}(\Sigma_{wv}) = n - 2$ und mit (2.1) die Behauptung. □

Wenden wir uns nun dem allgemeinen Fall zu, in dem keine spezielle Verteilung für X vorgegeben ist. Im Weiteren werden wir nur die Semi-Graphoid-Axiome als Voraussetzung nutzen.

Betrachten wir erneut Beispiel 2.2.2. Falls für den Zufallsvektor das Schnittaxiom gilt, können wir aus Proposition 2.1.4 weitere Eigenschaften des zugehörigen Zufallsvektors ablesen; beispielsweise gilt dann $X_1 \perp\!\!\!\perp (X_3, X_4) | (X_2, X_5)$.

Diese Überlegung verallgemeinern wir, indem wir die Markov-Eigenschaften eines Graphen mit dem Zusammenhangsbegriff aus Kapitel 1 in Verbindung bringen. Dazu definieren wir eine Unabhängigkeitsbedingung für die X_v , die auf der globalen Struktur von G aufbaut.

Definition 2.2.4. X besitzt die *ungerichtete, globale Markov-Eigenschaft bezüglich G* , falls die Implikation

$$A \text{ und } B \text{ werden in } G \text{ durch } C \text{ getrennt} \Rightarrow X_A \perp\!\!\!\perp X_B | X_C$$

gilt.

Die Begrifflichkeiten erweitern wir ein wenig:

Besitzt ein Zufallsvektor die ungerichteten, paarweisen oder globalen Markov-Eigenschaften, so sagen wir auch, dass die zugehörige Dichte f diese besitzt.

Ist also ein Graph G gegeben und sollen die damit „verträglichen“ Verteilungen f bestimmt werden, müssen diese Verteilungen allen paarweisen und/oder globalen Markov-Eigenschaften, die der Graph besitzt, genügen.

Alternativ heißt das dadurch repräsentierte bedingte Unabhängigkeitsmodell beziehungsweise eine einzelne Dichte f und X dann auch *konsistent mit G* , ohne genaueren Bezug zu ungerichteter, paarweiser oder ungerichteter, globaler Markov-Eigenschaft.

Diese Sprechweise wird durch unser Hauptergebnis über ungerichtete graphische Modelle gerechtfertigt:

Satz 2.2.5. *Ist X ein Zufallsvektor mit einer Dichte f , die dem Schnittaxiom genügt, dann besitzt f die paarweisen Markov-Eigenschaften bezüglich G genau dann, wenn f die globalen Markov-Eigenschaften bezüglich G besitzt.*

Salopp gesagt bedeutet diese Aussage:

$$\begin{aligned} &\{\text{ungerichtete, paarweise Markov-Eigenschaft}\} \\ &\Leftrightarrow \{\text{ungerichtete, globale Markov-Eigenschaft}\}. \end{aligned}$$

Der Beweis beruht hauptsächlich auf graphentheoretischen Überlegungen.

Beweis. „ \Leftarrow “:

Jedes Paar von nicht-adjazenten Knoten $v, w \in V$ wird durch $V \setminus \{v, w\}$ in G getrennt. Dann gilt $X_v \perp\!\!\!\perp X_w | X_{V \setminus \{v, w\}}$ nach Voraussetzung, was ebenfalls eine der paarweisen, ungerichteten Markov-Eigenschaften darstellt.

„ \Rightarrow “:

C trenne die Mengen A und B in G . Wir führen den Beweis per Induktion über $c := |V \setminus C|$.

Da $A, B \neq \emptyset$, gilt $c \geq 2$; falls Gleichheit gilt, ist $|A| = |B| = 1$. In diesem Fall ist $X_A \perp\!\!\!\perp X_B | X_C$ eine der paarweisen Markov-Eigenschaften von G , was uns den Induktionsanfang liefert.

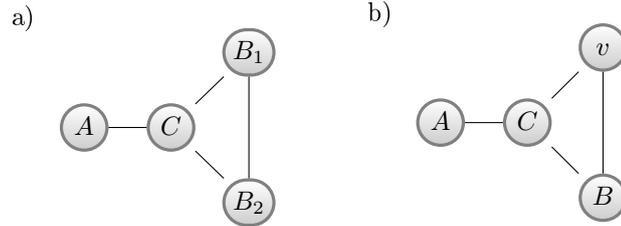
Sei nun $c \geq 3$. Dann ergeben sich die Fälle

- a) $A \cup B \cup C = V$ und
- b) $A \cup B \cup C \subsetneq V$.

In Fall a) muss wegen $c \geq 3$ die Menge A oder B mindestens zwei Elemente enthalten. Gelte dies o.B.d.A für B , dann können disjunkte, nicht-leere $B_1, B_2 \subset B$ gewählt werden mit $B_1 \cup B_2 = B$.

In Fall b) existiert ein $v \notin A \cup B \cup C$.

Die globalen Strukturen der Graphen, die zu diesen Fällen korrespondieren, lassen sich demnach wie folgt darstellen:



Induktionsschritt a): Wähle B_1 und B_2 wie beschrieben.

Dann trennt $C \cup B_1$ die Menge A von B_2 und $C \cup B_2$ die Menge A von B_1 . Es ist $|V \setminus (C \cup B_1)| < c$, somit folgt aus der der Induktionsvoraussetzung

$$X_A \perp\!\!\!\perp X_{B_2} | (X_C, X_{B_1}).$$

Analog folgt

$$X_A \perp\!\!\!\perp X_{B_1} | (X_C, X_{B_2}).$$

Mit dem Schnittaxiom (Proposition 2.1.4) gilt in Fall a) insgesamt die Behauptung.

Induktionsschritt b): Wähle nun $v \in V$, $v \notin A \cup B \cup C$.

Dann werden A und B von $C \cup \{v\}$ getrennt und die Induktionsvoraussetzung liefert

$$X_A \perp\!\!\!\perp X_B | (X_C, X_v).$$

Nach der vorausgesetzten Trennungseigenschaft von C muss jeder Pfad von A nach B Knoten der Menge C enthalten. Also trennt $C \cup A$ den Knoten v von B oder $C \cup B$ trennt

v	X_1	X_2	X_3	X_4
$\text{pa}(v)$	\emptyset	$\{X_1\}$	$\{X_2\}$	$\{X_2, X_3\}$
$\text{nd}(v)$	\emptyset	$\{X_1\}$	$\{X_1, X_2\}$	$\{X_1, X_2, X_3\}$

Also erfüllt die Verteilung von X die gerichteten, lokalen Markov-Eigenschaften $X_3 \perp\!\!\!\perp X_1 | X_2$ und $X_4 \perp\!\!\!\perp X_1 | (X_2, X_3)$.

Auch im Fall gerichteter Graphen wollen wir weitere Bedingungen aus den lokalen Markov-Eigenschaften ableiten. Es wird sich zeigen, dass hier ebenfalls eine Trennungseigenschaft eine Rolle dabei spielt, genauer die d-Separation in einem gerichteten Graphen (hier sei an Definition 1.3.16 erinnert).

Definition 2.2.9. Der Zufallsvektor X besitzt die *gerichteten, globalen Markov-Eigenschaften bezüglich G* , falls die Implikation

$$C \text{ d-separiert } A \text{ und } B \Rightarrow X_A \perp\!\!\!\perp X_B | X_C$$

gilt.

Wie im ungerichteten Fall übertragen wir die gerichtete, lokale und globale Markov-Eigenschaft auch auf Dichten und bedingte Unabhängigkeitsmodelle.

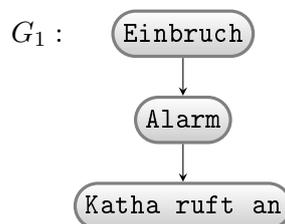
Es stellt sich bei dieser Definition die Frage: Warum wird die d-Separation verwendet? Was ist die Intuition dahinter?

Um dies zu verstehen, interpretieren wir die Kanten in G als Möglichkeit, Informationen in beide Richtungen der Kante weiterzuleiten beziehungsweise die Orientierung der Kante als eine Art Implikation.

Angenommen wir befinden uns in folgender Situation¹:

Wir sind in den Urlaub gefahren und haben vor der Abreise unser Haus mit einem Alarm gegen Einbrüche gesichert. Falls dieser ausgelöst wird, soll unsere Nachbarin Katha uns darüber per Anruf informieren.

Dieses Szenario können wir in den gerichteten Graphen G_1 übersetzen:



¹Das Beispiel basiert auf „Judea Pearl’s Burglary Alarm Example“, siehe [Pea88].

Betrachten wir vorerst nur die Knoten „Alarm“ und „Katha ruft an“:

- Wissen wir, dass der Alarm ausgelöst wurde, steigt für uns die Wahrscheinlichkeit, dass Katha anruft.
- Erhalten wir einen Anruf von Katha, werden wir eher an einen Einbruch denken.

Wir beobachten also: „Alarm“ und „Katha ruft an“ sind stochastisch abhängig, da sich, bei Erhalt von Informationen über das Eintreten des einen Ereignisses, unsere Wahrscheinlichkeit für das Eintreten des anderen verändert.

Jetzt betrachten wir den vollständigen Graphen G_1 .

- Wissen wir, dass ein Einbruch stattfindet, gehen wir davon aus, dass der Alarm ausgelöst wird und Katha uns anruft.
- Erhalten wir einen Anruf von Katha, werden wir denken, dass der Alarm aufgrund eines Einbruchs ausgelöst wurde.

Analog zu oben sind „Einbruch“ und „Katha ruft an“ stochastisch abhängig. Beachte, dass ein gerichteter Pfad von „Einbruch“ zu „Katha ruft an“ existiert!

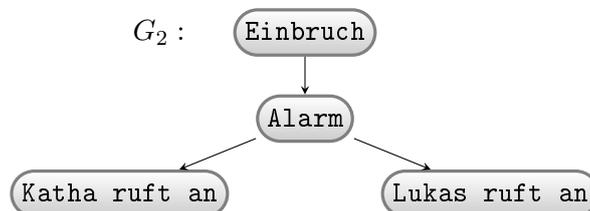
Nehmen wir jetzt an uns wäre bekannt, dass der Alarm ausgelöst wurde; das entspricht der Bedingung auf ein Ereignis. Daraufhin ruft Katha an. Werden wir dem Ereignis eines Einbruchs deswegen eine (noch) höhere Wahrscheinlichkeit zuordnen?

Die Antwort sollte „Nein“ lauten, da uns die Tatsache des Anrufs keine neuen Informationen liefert. „Einbruch“ und „Katha ruft an“ sind demnach bei Bedingung auf „Alarm“ unabhängig.

In allen Fällen taucht natürlicherweise die d-Separation auf:

Beispielsweise werden „Einbruch“ und „Katha ruft an“ von „Alarm“ d-separiert. Nach den globalen, gerichteten Markov-Eigenschaften aus Definition 2.2.9 sind die Ereignisse unabhängig, was wir auch intuitiv gefolgert haben.

Erweitern wir das Szenario. Da Katha schon älter ist und wir nicht sicher sein können, dass sie im Fall eines Einbruchs den Alarm auch hört und uns anruft, haben wir zusätzlich Lukas gebeten, aufzupassen und gegebenenfalls anzurufen. Damit erhalten wir



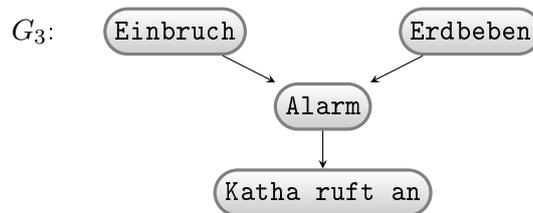
Wir wissen wieder, dass der Alarm ausgelöst wurde oder schließen darauf, weil wir Informationen über einen Einbruch erhalten haben. Daraufhin ruft uns Lukas an. Ist nun für uns die Wahrscheinlichkeit, dass Katha ebenfalls anruft, höher als vor Lukas Anruf?

Auch hier sollte die Antwort „Nein“ lauten, da unsere Erwartung von Kathas Anruf nur von unserem Glauben darüber abhängt, ob sie aufpasst oder nicht. Wir bemerken hier, dass alle ungerichteten Pfade von „Katha ruft an“ zu „Lukas ruft an“ durch „Alarm“ oder „Einbruch“ blockiert werden.

Wieder nach Definition wie auch nach unserer Intuition sind die Ereignisse „Lukas ruft an“ und „Katha ruft an“ bei Bedingung auf „Alarm“ oder „Einbruch“ unabhängig.

Ändern wir die Situation erneut ein wenig, um das Auftreten eines Kollisionspunktes zu verstehen:

Katha hat uns von ihrer Sorgfalt überzeugt, sodass Lukas nicht mehr anrufen muss. Außerdem haben wir vom Hersteller der Alarmanlage die Information bekommen, dass der Alarm auch von einem Erdbeben ausgelöst werden kann. Das führt uns zu



Angenommen wir wissen, dass der Alarm ausgelöst wurde, oder wir nehmen dies an, weil Katha angerufen hat (wir bedingen also auf „Alarm“ oder „Katha ruft an“). Das hat Einfluss auf unseren Glauben an einen Einbruch.

Dann erfahren wir, dass in der Nähe unseres Hauses ein Erdbeben stattgefunden hat. Hat diese neue Information Einfluss auf unsere Wahrscheinlichkeit für einen Einbruch?

In diesem Fall lautet die Antwort „Ja“:

Der Alarm hat einen Einbruch für uns „wahrscheinlicher gemacht“, aber nachdem wir Informationen über das Erdbeben erhalten hatten, können wir den Alarm mit etwas anderem erklären (in der Literatur meist „explain away“ genannt), womit wir einen Einbruch weniger in Betracht ziehen.

Bedingt auf „Alarm“ sind daher, auch nach Definition 2.2.9, „Einbruch“ und „Erdbeben“ abhängig.

Wir sehen also, dass die d-Separation eine intuitive Wahl für die Trennungseigenschaft bei gerichteten Graphen darstellt, zumindest wenn Kanten als Abhängigkeiten der Knoten interpretiert werden.

Für die beiden gerichteten Markov-Eigenschaften wollen wir das Analogon zu der Äquivalenz der beiden ungerichteten Markov-Eigenschaften beweisen.

Dafür benötigen wir noch nachstehende Aussage, die wir mit den bisherigen Betrachtungen leicht einzusehen ist.

Folgerung 2.2.10. *Seien A und B durch C d-separiert und es gelte $\text{AN}(A \cup B \cup C) = V$. Dann können A und B so zu A^* und B^* erweitert werden, dass $A^* \cup B^* \cup C = V$ gilt und A^* und B^* noch immer durch C d-separiert sind.*

Beweis. Mit Lemma 1.3.19 trennt C die Menge A von der Menge B im Moral-Graph $(G_{\text{AN}(A \cup B \cup C)})^m$. Da nach Voraussetzung $\text{AN}(A \cup B \cup C) = V$ erfüllt ist, gilt $(G_{\text{AN}(A \cup B \cup C)})^m = G^m$ und in G^m existiert für jeden Knoten $v \in V$ ein ungerichteter Pfad zu einem $u \in A \cup B \cup C$.

Nun können wir Lemma 1.3.7 nutzen; demnach existieren A^* und B^* , sodass die Gleichheit $A^* \cup B^* \cup C = V$ gilt und C die Mengen A^* und B^* noch immer trennt.

Wieder mit Lemma 1.3.19 folgt, dass A^* und B^* von C in G d-separiert werden, womit die Behauptung gezeigt ist. \square

Kommen wir zur Hauptaussage dieses Kapitels, in der interessanterweise das Schnittaxiom nicht benötigt wird.

Satz 2.2.11. *Eine Dichte f besitzt die gerichtete, lokale Markov-Eigenschaft bezüglich G genau dann, wenn f die gerichtete, globale Markov-Eigenschaft bezüglich G besitzt.*

Drücken wir dies wieder etwas salopp aus, reduziert sich die Aussage auf:

$$\begin{aligned} & \{\text{gerichtete, globale Markov-Eigenschaft}\} \\ & \Leftrightarrow \{\text{gerichtete, lokale Markov-Eigenschaft}\}. \end{aligned}$$

In dem Beweis werden wir aufgrund der Übersichtlichkeit die verkürzte Schreibweise der bedingten Unabhängigkeit nutzen, siehe Definition 2.1.2.

Beweis. „ \Leftarrow “:

Zu zeigen ist, dass $\text{pa}(v)$ die Mengen $\{v\}$ und $\text{nd}(v) \setminus \text{pa}(v)$ d-separiert.

Angenommen, $\{v\}$ und $\text{nd}(v) \setminus \text{pa}(v)$ sind d-zusammenhängend gegeben $\text{pa}(v)$; dann existieren ein $w \in \text{nd}(v) \setminus \text{pa}(v)$ und ein zugehöriger Pfad $\pi = (v, u_1, \dots, u_{k_1}, w)$, der den d-Zusammenhang garantiert.

Es muss $u_1 \notin \text{pa}(v)$ gelten, denn: Falls $u_1 \in \text{pa}(v)$, kann u_1 kein Kollisionspunkt auf π sein, ein Widerspruch zur Definition des d-Zusammenhangs gegeben $\text{pa}(v)$, da $\text{pa}(v)$ keine Nicht-Kollisionspunkte enthalten darf. Demnach gilt $(v, u_1) \in E$, die Kante zwischen u_1 und v ist in Richtung u_1 orientiert und es darf wegen der Azyklizität nicht zusätzlich $(u_1, v) \in E$ gelten. Also ist tatsächlich $u_1 \notin \text{pa}(v)$.

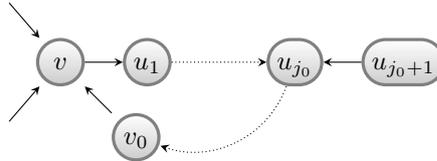
Da $w \in \text{nd}(v)$, existieren ein $l \geq 1$ und $j_1, \dots, j_l \in [k-1]$, sodass u_{j_i} für alle $i \in [l]$ ein Kollisionspunkt auf π ist.

Wir wählen $j_0 \in \{j_1, \dots, j_l\}$ so, dass $\pi(v, u_{j_0})$ minimal ist unter den $\pi(v, u_{j_i})$. Damit ist u_{j_0} der Kollisionspunkt mit kleinstem Abstand zu v , folglich ist $\mu := \pi(v, u_{j_0})$ ein gerichteter Pfad.

Wegen des d-Zusammenhangs muss $u_{j_0} \in \text{AN}(\text{pa}(v)) = \text{pa}(v) \cup \text{an}(\text{pa}(v))$ gelten (u_{j_0} ist ein Kollisionspunkt).

Falls $u_{j_0} \in \text{pa}(v)$ gilt, ist die Folge $(v, u_1, \dots, u_{j_0}, v)$ ein nicht-trivialer Zykel in G , ein Widerspruch zur Voraussetzung.

Falls $u_{j_0} \in \text{an}(\text{pa}(v))$ ist, insbesondere $u_{j_0} \in \text{an}(v_0)$ für ein $v_0 \in \text{pa}(v)$, erhalten wir folgendes Bild:



Auch in diesem Fall erhalten wir einen Widerspruch zur Azyklizität von G . Somit gilt $u_{j_0} \notin \text{pa}(v) \cup \text{an}(\text{pa}(v))$, was ein Widerspruch zum d-Zusammenhang darstellt.

„ \Rightarrow “:

Wir führen den Beweis per Induktion über $|V|$.

Für $|V| \leq 2$ ist nichts zu zeigen.

Angenommen, die Behauptung gilt für Graphen mit n Knoten. Wir betrachten einen Graphen mit $|V| = n + 1$.

Seien A und B von C d-separiert und $V_0 := \text{AN}(A \cup B \cup C)$. Es können global zwei Fälle auftreten, nämlich

Fall I) $V_0 \subsetneq V$ und

Fall II) $V_0 = V$.

In Fall I) gilt insbesondere $A, B, C \subset V_0$ nach Definition; des Weiteren werden A und B ebenfalls in G_{V_0} von C d-separiert. Wir beobachten, dass die lokale, gerichtete Markov-Eigenschaft in G_{V_0} gilt, da sie durch G induziert wird (die Eigenschaft überträgt sich auf G_{V_0} , da die Mengen $\text{pa}(v)$ für alle $v \in V_0$ in G und G_{V_0} übereinstimmen).

Da V_0 weniger Elemente als V enthält, folgt mit der Induktionsvoraussetzung die globale gerichtete Markov-Eigenschaft für G_{V_0} und somit $X_A \perp\!\!\!\perp X_B | X_C$, was zu zeigen war.

In Fall II) können wir wegen der vorigen Folgerung o.B.d.A. annehmen, dass $A \cup B \cup C = V$ gilt.

Nach Voraussetzung werden A und B von C d-separiert, demnach trennt C die Mengen A und B in dem Moral-Graphen $(G_{\text{AN}(A \cup B \cup C)}) = (G_V)^m = G^m$.

Wir wählen $\lambda \in V$ mit $\text{de}(\lambda) = \emptyset$. Ein solcher Knoten existiert nach Lemma 1.3.13.

Wir halten fest:

Aufgrund der d-Separation durch C muss entweder $\text{pa}(\lambda) \subset A \cup C$ oder $\text{pa}(\lambda) \subset B \cup C$ gelten – denn Eltern, die in A und B liegen, wären in G^m adjazent, ein Widerspruch zur Trennung von A und B durch C .

Wegen $\lambda \in V$ erhalten wir die drei Fälle

a) $\lambda \in A$,

b) $\lambda \in B$ und

c) $\lambda \in C$.

Im Fall a) muss nach obiger Beobachtung $\text{pa}(\lambda) \subset (A \setminus \{\lambda\}) \cup C$ gelten, die lokale gerichtete Markov-Eigenschaft liefert

$$\begin{aligned} & \lambda \perp\!\!\!\perp ((A \cup B \cup C) \setminus \{\lambda\}) \setminus \text{pa}(\lambda) \mid \text{pa}(\lambda) \\ \Leftrightarrow & \lambda \perp\!\!\!\perp [B \setminus \text{pa}(\lambda)] \cup [(A \setminus \{\lambda\}) \cup C] \setminus \text{pa}(\lambda) \mid \text{pa}(\lambda) \\ \Leftrightarrow & \lambda \perp\!\!\!\perp B \cup [(A \setminus \{\lambda\}) \cup C] \setminus \text{pa}(\lambda) \mid \text{pa}(\lambda). \end{aligned}$$

Mit der schwachen Vereinigung aus Proposition 2.1.4 folgt

$$\lambda \perp\!\!\!\perp B \mid [(A \setminus \{\lambda\}) \cup C] \quad (2.2)$$

Im Graph $(G^m)_{V \setminus \{\lambda\}}$ trennt C offensichtlich die Mengen $A \setminus \{\lambda\}$ und B ; daher gilt diese Beziehung auch in $(G_{V \setminus \{\lambda\}})^m$, weil dieser Graph höchstens so viele Kanten wie $(G^m)_{V \setminus \{\lambda\}}$ enthält (es können also keine zusätzlichen Pfade entstehen). Mit Pearl's Separations-Kriterium werden $A \setminus \{\lambda\}$ und B von C in $G_{V \setminus \{\lambda\}}$ d-separiert und, da $|V \setminus \{\lambda\}| = n$ gilt, erhalten wir mit der Induktionsvoraussetzung

$$(A \setminus \{\lambda\}) \perp\!\!\!\perp B \mid C. \quad (2.3)$$

Die Kontraktion aus Proposition 2.1.4 angewandt auf (2.2) und (2.3) impliziert letztendlich $X_A \perp\!\!\!\perp X_B \mid X_C$.

Der Fall b) verläuft analog.

Wenden wir uns Fall c) zu, bemerken wir als Erstes, dass, wegen der Trennung von A und B durch C in G^m und weil λ keine Nachfolger hat, die Menge $C \setminus \{\lambda\}$ die Mengen A und B auch in $(G^m)_{V \setminus \{\lambda\}}$ trennt. Somit gilt diese Beziehung, wie eben auch, in $(G_{V \setminus \{\lambda\}})^m$ und die Induktionsvoraussetzung liefert

$$A \perp\!\!\!\perp B \mid (C \setminus \{\lambda\}), \quad (2.4)$$

da $|V \setminus \{\lambda\}| = n$.

Angenommen $\text{pa}(\lambda) \subset A \cup C$.

Aus den gerichteten, lokalen Markov-Eigenschaften folgt die Beziehung

$$\lambda \perp\!\!\!\perp B \cup [A \cup (C \setminus \{\lambda\})] \setminus \text{pa}(\lambda) \mid \text{pa}(\lambda).$$

Mit der schwachen Vereinigung gilt dann die Eigenschaft

$$B \mid A \cup (C \setminus \{\lambda\}). \quad (2.5)$$

Die Kontraktion, angewandt auf (2.4) und (2.5), liefert die Bedingung $A \cup \{\lambda\} \perp\!\!\!\perp B \mid C \setminus \{\lambda\}$ und die schwache Vereinigung abschließend $A \perp\!\!\!\perp B \mid C$.

Der Fall $\text{pa}(\lambda) \subset B \cup C$ verläuft analog. \square

Geiger und Pearl haben allgemein gezeigt, dass das Kriterium der d-Separation in folgendem Sinne nicht verbessert werden kann: Zu gegebenem Graphen lässt sich ein Zufallsvektor X und eine Wahrscheinlichkeitsfunktion P finden, sodass

$$X_A \perp\!\!\!\perp X_B | X_C \Leftrightarrow C \text{ d-separiert } A \text{ von } B.$$

Wir hatten einen konstruktiven Beweis dieser Aussage schon im ungerichteten Fall gesehen, Proposition 2.2.6 gilt aber auch, wenn wir einen gerichteten, azyklischen Graphen statt eines ungerichteten betrachten und „Teilen“ durch „d-Separation“ ersetzen.

Bei Betrachtung der gerichteten, globalen Markov-Eigenschaften stellt sich die Frage, ob zwei verschiedene Graphen die gleichen Markov-Eigenschaften haben können.

Die Antwort darauf lautet „ja“, der einfachste Fall zweier solcher Graphen ist das nachstehende Beispiel.

Beispiel 2.2.12. Gegeben seien die folgenden Graphen:



In allen drei Fällen ist die gerichtete, globale Markov-Eigenschaft die Beziehung $X_1 \perp\!\!\!\perp X_3 | X_2$.

Um einen Begriff für solche Graphen zu bekommen, definieren wir:

Definition 2.2.13. Seien $G_1 = (V_1, E_1)$ und $G_2 = (V_2, E_2)$ gerichtete, azyklische Graphen.

Weiter seien X und Y Zufallsvektoren, die die gerichteten, globalen Markov-Eigenschaften bezüglich G_1 beziehungsweise G_2 besitzen.

Die Graphen G_1 und G_2 heißen *Markov-äquivalent*, falls $|V_1| = |V_2| =: n$ und

$$X_A \perp\!\!\!\perp X_B | X_C \Leftrightarrow Y_A \perp\!\!\!\perp Y_B | Y_C$$

für alle $A, B, C \subset [n]$ gilt.

Definition 2.2.14. Sei $G = (V, E)$ ein gerichteter, azyklischer Graph. Wir betrachten die möglichen Konstellationen für $a, b, c \in V$ mit $a \neq b$.

1. $a \rightarrow c \leftarrow b$ heißt *Senken-orientierte V-Konfiguration*,
2. $a \leftarrow b \rightarrow c$ heißt *Quellen-orientierte V-Konfiguration* und
3. $a \rightarrow b \rightarrow c$ heißt *Übergangs-orientierte V-Konfiguration*.

Wie lässt sich nun obige Frage allgemein beantworten? Obiges Beispiel legt eine Beziehung zu den Moral-Graphen nahe; etwa, dass Gleichheit der Moral-Graphen äquivalent ist zu Markov-Äquivalenz. Leider ist die Gleichheit der Moral-Graphen nur notwendig für die Markov-Äquivalenz von zwei Graphen.

Die Aussage, die die Frage beantwortet, geben wir ohne Beweis an, vor allem wegen seiner Länge.

Satz 2.2.15. *Zwei Graphen $G_1 = (V, E_1)$ und $G_2 = (V, E_2)$ sind Markov-äquivalent genau dann, wenn ihre Skelette und ihre Senken-orientierten V -Konfigurationen übereinstimmen.*

Dabei haben zwei Graphen $G_1 = (V, E_1)$ und $G_2 = (V, E_2)$ dasselbe Skelett, falls

I) $(v, w) \in E_1 \setminus E_2 \Rightarrow (w, v) \in E_2$ und

II) $(v, w) \in E_2 \setminus E_1 \Rightarrow (w, v) \in E_1$.

Betrachten wir also G_1 und G_2 als ungerichtete Graphen, müssen diese die selben Kanten und Knoten besitzen, um die gleichen Skelette zu besitzen.

Anhang A.

Hier sollen kurz die oben angesprochenen aber nicht weiter ausgeführten Aussagen aufgegriffen und zumindest skizzenhaft bewiesen werden.

In den Vorbetrachtungen haben wir den Dimensionssatz für lineare Abbildungen zwischen Vektorräumen genutzt:

Satz A.0.16. *Seien K ein Körper, V, W endlich-dimensionale K -Vektorräume und $f : V \rightarrow W$ eine K -lineare Abbildung. Dann gilt*

$$\dim(V) = \dim(\ker(f)) + \dim(\text{bild}(f))$$

Beweis. $\ker(f)$ und $\text{bild}(f)$ sind Untervektorräume von V beziehungsweise W .

Es gibt einen Untervektorraum $U \subset V$ mit $V = \ker(f) \oplus U$, wir definieren $g : U \rightarrow \text{bild}(f)$ mit $g = f|_U$.

Dann gilt:

g ist ein Monomorphismus, denn:

Sei $u \in U$ mit $g(u) = 0$. Nach Definition von g dann auch $f(u) = 0$ und damit $u \in \ker(f) \cap U$. Daraus folgt $u = 0$, da $U \cap \ker(f) = \{0\}$ nach Definition der direkten Summe. Somit gilt $\ker(g) = \{0\}$, was zur Injektivität von g äquivalent ist.

g ist ein Epimorphismus, denn:

Seien $w \in \text{bild}(f)$ und $v \in V$ mit $f(v) = w$. Da $V = \ker(f) \oplus U$ existieren $x \in \ker(f)$ und $u \in U$ mit $v = x + u$. Dann gilt $g(u) = f(u) = f(v - x) = f(v) - f(x) = f(v)$, also $w \in \text{bild}(g)$, was zu zeigen war.

Damit ist g ein Isomorphismus und es ist $\dim(U) = \dim(\text{bild}(f))$. Außerdem

$$\begin{aligned} \dim(V) &= \dim(\ker(f) \oplus U) \\ &= \dim(\ker(f)) + \dim(U) \\ &= \dim(\ker(f)) + \dim(\text{bild}(f)) \end{aligned}$$

□

Des Weiteren hatten wir uns in der linearen Algebra Eigenschaften von Rang-1-Matrizen zunutze gemacht.

Lemma A.0.17. Seien K ein Körper, $u \in K^m$, $v \in K^n$ für $m, n \in \mathbb{N}$. Die Matrix

$$u \cdot v^T = \begin{pmatrix} u_1 v_1 & u_1 v_2 & \cdots & u_1 v_n \\ u_2 v_1 & u_2 v_2 & \cdots & u_2 v_n \\ \vdots & \vdots & \ddots & \vdots \\ u_m v_1 & u_m v_2 & \cdots & u_m v_n \end{pmatrix} \in K^{m \times n}$$

hat Rang 1, falls $u, v \neq 0$, sonst ist es die Nullmatrix.

Beweis. Falls $u = 0$ oder $v = 0$, sind offensichtlich alle Einträge 0.

Seien also $u, v \neq 0$ und o.B.d.A. $v_1 \neq 0$. Addieren wir das $-\frac{v_i}{v_1}$ -fache der ersten Spalte zur i -ten Spalte für $i \in \{2, \dots, n\}$, entstehen stets Nullspalten. Die resultierende Matrix hat trivialerweise Rang 1, also auch die Matrix $u \cdot v^T$, da elementare Spaltenoperationen den Spaltenrang nicht verändern. \square

Bei der Berechnung einiger Determinanten hatten wir den Kästchensatz genutzt. Der hier angegebene Beweis basiert auf einem Ergebnis der linearen Algebra über multilineare Abbildungen, welches wir hier ohne Beweis zitieren:

Satz A.0.18. Seien R ein Ring, $D : \mathcal{M}_n(R) \rightarrow R$ eine alternierende, multilineare Abbildung.

Dann gilt für alle Matrizen $A \in \mathcal{M}_n(R)$ die Gleichheit $D(A) = \det(A) \cdot d(E_n)$.

Damit kann nun der Kästchensatz bewiesen werden, formal lautet der Satz wie folgt:

Satz A.0.19 (Kästchensatz). Seien R ein Ring, $A \in \mathcal{M}_n(R)$ und A_1, \dots, A_m ($m \leq n$) quadratische Matrizen mit

$$A = \begin{pmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_m \end{pmatrix}.$$

Dann gilt

$$\det(A) = \prod_{i=1}^m \det(A_i).$$

Beweis. Es reicht, diese Behauptung für $m = 2$ zu zeigen, der Rest folgt induktiv.

Seien $n_1, n_2 \in \mathbb{N}$ mit $n_1 + n_2 = n$, $A_1 \in M_{n_1}(\mathbb{R})$, $A_2 \in M_{n_2}(\mathbb{R})$ und

$$A = \begin{pmatrix} A_1 & B \\ 0 & A_2 \end{pmatrix}.$$

Für alle $A_1 \in \mathcal{M}_{n_1}(\mathbb{R})$ sei $D_{A_2 B}(A_1) := \det \begin{pmatrix} A_1 & B \\ 0 & A_2 \end{pmatrix}$. Die Abbildung $D_{A_2 B}$ ist offensichtlich multilinear und alternierend.

Mit dem vorigen Satz folgt, dass $D_{A_2, B}(A_1) = \det(A_1) \cdot D_{A_2, B}(E_{n_1})$ gelten muss. Entwicklung nach den ersten n_1 Spalten ergibt $D_{A_2, B}(E_{n_1}) = \det(A_2)$, also

$$\begin{aligned} \det \begin{pmatrix} A_1 & B \\ 0 & A_2 \end{pmatrix} &= D_{A_2, B}(A_1) \\ &= \det(A_1) \cdot D_{A_2, B}(E_{n_1}) \\ &= \det(A_1) \cdot \det(A_2). \end{aligned}$$

□

Als letztes erwähnten wir eine Aussage über den Erwartungswert von zwei unabhängigen Zufallsvariablen, die eine einfachere Berechnung desselben ermöglicht:

Satz A.0.20. *Seien $X, Y : \mathcal{X} \rightarrow [0, \infty)$ unabhängige (1-dimensionale) Zufallsvariablen mit gemeinsamer Dichte f , welche stetig bezüglich eines Produktmaßes ν auf \mathcal{X} ist. Dann gilt für den Erwartungswert*

$$E(XY) = E(X) E(Y).$$

Beweis. Der Erwartungswert einer eindimensionalen Zufallsvariable Z mit Dichte $g : \mathcal{X} \rightarrow [0, \infty)$, die bezüglich eines Maßes ν stetig ist, wird berechnet über

$$E(Z) = \int_{\mathcal{X}} z \cdot g(z) d\nu(z).$$

Für $E(XY)$ folgt daher

$$\begin{aligned} E(XY) &= \int_{\mathcal{X}} x \cdot y \cdot f(x, y) d\nu(x, y) \\ &= \int_{\mathcal{X}} x \cdot f_X(x) \cdot y \cdot f_Y(y) d\nu(x, y) \\ &= \int_{\mathcal{X}} x \cdot f_X(x) d\nu(x) \int_{\mathcal{X}} y \cdot f_Y(y) d\nu(y) \\ &= E(X) \cdot E(Y) \end{aligned}$$

□

Literaturverzeichnis

- [Agr12] AGRICOLA, Ilka: *Maß- und Integrationstheorie*. Juni 2012
- [HG90] HECKERMANN, David und GEIGER, Dan: *Separable and Transitive Graphoids*. 1990
- [LLD90] LEIMER, H.-G., LAURITZEN, S. L., DAWID, A. P. und LARSEN, B.N.: *Independence Properties of Directed Markov Fields*. 1990
- [Hol11] HOLZMANN, Hajo: *Elementare Stochastik*. 2011
- [PG88] PEARL, Judea und GEIGER, Dan: *On the logic of influence diagrams*. 1988
- [Lau96] LAURITZEN, Steffen L.: *Graphical Models*. Oxford University Press, 1996
- [PMA01] PERLMAN, Michael D., MADIGAN, David und ANDERSSON, Steen A.: *A Characterization of Markov Equivalence Classes for Acyclic Digraphs*. 2001
- [NN11] UNBEKANNT: *Partitioned Matrices and the Schur Complement*. Juli 2011
<http://www.colorado.edu/engineering/cas/courses.d/IFEM.d>
- [Pea88] PEARL, Judea: *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, 1988
- [Ric99] RICHARDSON, Thomas: *Markov Properties for Acyclic Directed Mixed Graphs*. 1999
- [SS09] SULLIVANT, Seth, DRTON, Mathias: *Lectures on Algebraic Statistics*. Birkhäuser Verlag AG, 2009

Erklärung

Hiermit versichere ich, dass ich die vorliegende Bachelorarbeit selbst verfasst habe. Ich habe mich keiner anderen als der von mir ausdrücklich bezeichneten Quellen bedient. Die Arbeit in vorliegender oder ähnlicher Form habe ich noch nie zu Prüfungszwecken vorgelegt.

Marburg, den 23. April 2013