Paradoxien in quantitativen Modellen der Individualdiagnostik

Dissertation

zur Erlangung des Doktorgrades

an der Fakultät Erziehungswissenschaft, Psychologie und Bewegungswissenschaft, Fachbereich Psychologie der Universität Hamburg

vorgelegt von

Pascal Jordan

Hamburg, 2013

Promotionsprüfungsausschuss (Tag der mündlichen Prüfung: 3.6.2013):

Prof. Dr. Gerhard Tutz (Vorsitzender)

Prof. Dr. Martin Spieß (1. Dissertationsgutachter)

Prof. Dr. Volker Franz (2. Dissertationsgutachter)

Prof. Dr. Manfred Amelang (1. Disputationsgutachter)

Prof. Dr. Matthias Burisch (2. Disputationsgutachter)

I like mathematics because it is not human and has nothing particular to do with this planet or with the whole accidental universe - because, like Spinoza's God, it won't love us in return.

Bertrand Russell (Letter to Ottoline Morrell, 1912)

Zusammenfassung

Die vorliegende Arbeit unternimmt den Versuch einer Verallgemeinerung eines von Hooker, Finkelman und Schwartzman (2009) erstmals für eine spezielle Klasse von (mehrdimensionalen) Item-Response-Modellen beschriebenen Paradoxons, welches im Kern darauf beruht, dass eine Person durch das Geben von ein oder mehreren Falschantworten eine vorteilhafte Diagnose hinsichtlich einer zu diagnostizierenden (latenten) Fähigkeit erlangen kann.

Mittels des Begriffs einer „reverse-rule"-Funktion kann - wie in dieser Arbeit gezeigt wird - das Konstruktionsprinzip von Hooker u.a. (2009) sowohl für ordinale Item-Response-Modelle als auch für Modelle, die gegen Regularitätsbedingungen (Logkonkonkavität; zweimal stetig differenzierbare Item-Response-Funktionen) verstoßen, übernommen werden. Als Nebenprodukt erlaubt dieses Konzept zudem eine vereinheitlichende Darstellung des paradoxen Effekts für diskrete und stetige Fähigkeitsdimensionen.

Die Betrachtung des Effekts in speziellen Modellklassen zeigt ferner seine Omnipräsenz sowie seine vielfältigen Erscheinungsformen auf: (1) Im faktorenanalytischen Modell impliziert jede nichtnegative Ladungsmatrix, die von Einfachstruktur abweicht, einen paradoxen Effekt bezüglich (mindestens) einer latenten Dimension. (2) In Item-Response-Modellen mit zusätzlicher Schnelligkeitskomponente kann unter bestimmten Bedingungen jede beliebig schlechte Testperformance durch eine zügige Antwort auf ein Item kompensiert werden. (3) In longitudinalen Item-Response-Modellen kann eine bessere Performance in einem später administrierten Test - je nach Abhängigkeitsmodellierung - eine Zu- oder Abnahme in der geschätzten Fähigkeitsdifferenz bezüglich früherer Zeitpunkte induzieren.

Vor dem Hintergrund, dass keine zufriedenstellende Behebungsmöglichkeit des paradoxen Effekts zu existieren scheint, erhält die Erörterung seiner testtheoretischen Relevanz zusätzliche Bedeutung. Als Kerneigenschaft mehrdimensionaler IRT-Modelle stellt der paradoxe Effekt einen omnipräsenten Antagonismus zwischen statistischer Effizienz und individualdiagnostischer Fairness dar, der sich - wie in der abschließenden Diskussion argumentiert wird - nahtlos in eine Reihe bereits bekannter Kritikpunkte des Latenten-Variablen-Ansatzes (Rotationsproblem, Mehrdeutigkeit der latenten Variable, ...) einfügen lässt.

Inhaltsverzeichnis

Tabelle 1: *Notation*

Symbol	Bedeutung
$\boldsymbol{\theta}$	Vektor latenter Fähigkeiten
$\hat{\boldsymbol{\theta}}$	Vektor der geschätzten latenten Fähigkeiten
τ	Latente „Speedkomponente“ der zu diagnostizierenden Person
κ_i	Klassifikationsrelevanter Schwellenwert der i-ten Fähigkeitsdimension
η	Fähigkeitsdifferenz $\eta := \theta_2 - \theta_1$
p	Länge des Vektors $\boldsymbol{\theta}$
u_i	Erzielter Score bezüglich des i-ten Items
t_i	Responsezeit bezüglich des i-ten Items
k	Anzahl der Items
$\boldsymbol{a}_i$	Itemdiskriminationsvektor des i-ten Items
A	Ladungsmatrix mit i-ter Zeile $\boldsymbol{a}_i$
ξ_i	Inverse Messfehlervarianz des i-ten Items
$\boldsymbol{\gamma}$	Vektor der Itemparameter
l_{i,u_i}	Loglikelihoodbeitrag des i-ten, mit Score u_i beantworteten Items
l_u	Vom Antwortvektor $\boldsymbol{u}$ induzierte Loglikelihood
γ	Logarithmierte Prioridichte
σ^{ij}	Eintrag (i, j) einer inversen Kovarianzmatrix
x_i	Die i-te Komponente des Vektors $\boldsymbol{x}$
$\boldsymbol{x}_{-i}$	Der Vektor $\boldsymbol{x}$ ohne dessen i-te Komponente
$\boldsymbol{e}_i$	i-ter Einheitsvektor
I_p	$p \times p$-Einheitsmatrix

Tabelle 2: *Notation*

Symbol	Bedeutung
A_{-i}	Verkürzung der Matrix A um ihre i-te Zeile
$A_{-(i)}$	Verkürzung der Matrix A um ihre i-te Spalte
δ_{ij}	Kroneckers Delta
$\overline{\mathbb{R}}$	Die um die Werte $\pm\infty$ erweiterten reellen Zahlen
$\mathbb{B}_\epsilon(\boldsymbol{x_0})$	Menge aller $\boldsymbol{x}$ mit Abstand $\lvert\boldsymbol{x} - \boldsymbol{x_0}\rvert < \epsilon$
$0^+(C)$	$\stackrel{\triangle}{=} \{y \,\vert\, C + y \subset C\}$ - *recession cone* der Menge C
C	Nichtnegativer Orthant des $\mathbb{R}^k$
$\mathscr{F}$	σ-Algebra
$L^1(\mu)$	Menge aller (messbaren) reellwertigen Funktionen f mit $\int \lvert f\rvert \mathrm{d}\mu < \infty$
μ	Positives Maß
$s(\cdot)$	Einelementige Lokalisation einer mengenwertigen Lösungsabbildung
$\hat{\boldsymbol{x}}(y)$	Menge aller bedingten Maximumsstellen einer Funktion $f(\boldsymbol{x}, y)$ bezüglich fest gewähltem y-Wert (y-Schnitt der Funktion)
$rg(A)$	Rang der Matrix A
$\nabla_\theta f(\boldsymbol{\gamma}, \boldsymbol{\theta})$	Matrix der Ableitungen der Funktion $g_{\boldsymbol{\gamma}}(\boldsymbol{\theta}) := f(\boldsymbol{\gamma}, \boldsymbol{\theta})$
$\nabla f(\boldsymbol{\gamma}, \boldsymbol{\theta})$	Matrix der Ableitungen der Funktion f.

Bemerkung: Die Notation bezieht sich lediglich auf die wesentlichen Kerngrößen. Darüber hinaus sind in den einzelnen Kapiteln zusätzliche Terme definiert, die nicht zwingend in den oben angegebenen Tabellen aufgeführt sind.
Mitunter mag es vorkommen, dass ein Symbol (z.B. Ω) doppelt besetzt ist. Der Kontext sollte jedoch stets eine eindeutige Identifikation der Bedeutung des entsprechenden Symbols gewährleisten.

Vorwort

In diesem Vorwort wird ein kurzer Leitfaden dargeboten, der dem Leser den Einstieg in die Thematik des paradoxen Effekts erleichtern soll.

Ein kurzer Leitfaden zum Einstieg

Der Kern des Paradoxons kann den ersten beiden Kapiteln entnommen werden. Dort erfolgt sowohl eine informelle (Kapitel 1 und 2.1) als auch eine formelle (Kapitel 2.3) Einführung des paradoxen Effekts. Diese beiden Kapitel sind - mit Ausnahme des Abschnitts „parametrische Abhängigkeit des paradoxen Effekts“, der beim ersten Lesen übersprungen werden kann - essenziell für das Verständnis der darauffolgenden Kapitel. Die Herleitung im Zuge von funktionalbasierten Schätzern (letzter Abschnitt des zweiten Kapitels) wird jedoch an keiner weiteren Stelle der Arbeit mehr aufgegriffen, so dass diese Verallgemeinerung zunächst ausgelassen werden kann.
Die Themen des dritten Kapitels (Spezialformen des paradoxen Effekts) sind hingegen so aufgebaut, dass jedes Unterkapitel unabhängig von den anderen Unterkapiteln gelesen werden kann. Dem Leser, der einen schnellen Überblick über die Thematik gewinnen möchte, sei empfohlen, mit dem Unterkapitel über linear-kompensatorische Modelle zu beginnen und anschließend direkt zu den Abschlusskapiteln (Kapitel 4 und 5) fortzuschreiten. Der Hauptaugenmerk in Kapitel 4 sollte hierbei den Unterkapiteln 4.3 sowie (vor allem) 4.4 gewidmet werden, da diese maßgeblich in die abschließende Diskussion und Bewertung des paradoxen Effekts (Kapitel 5) eingehen.

1 Einleitung

Das Konzept einer latenten Variable ist ein integraler Bestandteil der psychologischen Forschung. Konstrukte wie „Intelligenz“, „Extraversion“ oder „Leistungsmotivation“ stellen Beispiele von für die psychologische Theorie zentralen Begriffen dar, die per Definition nicht direkt beobachtbar sind. Mithilfe einer geeigneten Auswahl von Messinstrumenten (Items) soll es laut vorherrschender Meinung jedoch möglich sein, Rückschlüsse auf diese unbeobachtbaren Konstrukte vorzunehmen. Grundlage hierfür bildet ein stochastisches Modell, das den Ausdruck: „Ein Item misst eine latente Größe“ formalisiert und zugleich eine empirische Überprüfung der Messqualitäten des Messinstruments, d.h. des Fragebogens, ermöglichen soll.
Die postulierte Klasse stochastischer Modelle bestimmt hierbei maßgeblich den Prozess der Skalenkonstruktion. Das der Skalenkonstruktion zugrundeliegende Ziel besteht dabei entweder in der Ergründung von (statistischen) Zusammenhängen zwischen latenten und manifesten Größen in der Population (gruppen- bzw. populationsbasierte Perspektive) oder im Einsatz der Skala zu individualdiagnostischen Zwecken (individuumzentrierte Perspektive).
Im letzteren Fall soll die Fähigkeit einer Person anhand des beobachteten Antwortmusters prädiziert werden. Wenngleich die Vorgehensweise bei diesem Rückschluss (=Schätzung) aus statistisch-methodischer Perspektive zunächst unproblematisch erschien, da für eine reichhaltige Klasse stochastischer Modelle Prinzipien zur Herleitung statistisch adäquater Schätzer zur Verfügung stehen, wies dennoch ein kürzlich erschienener Artikel (Hooker u.a., 2009) auf eine fundamentale Problematik dieser Schätzer bezüglich der Fairness einer auf ihnen basierenden Individualdiagnose hin. Diese Problematik äußert sich in Gestalt eines Paradoxons, dessen Kern darin begründet liegt, dass eine Person durch falsches (richtiges) Antworten eine vorteilhafte (nachteilhafte) Diagnose erlangen kann. Unter gewissen, in den folgenden Kapiteln näher zu erläuternden Umständen wird somit falsches (richtiges) Antworten belohnt (bestraft).
Ziel dieser Arbeit ist es, (1) dieses Paradoxon umfassend im Hinblick auf seine testtheoretischen und individualdiagnostischen Implikationen zu untersuchen; (2) den Gültigkeitsbereich des Paradoxons zu ergünden, sowie (3) die mathematisch-formale Ursache seiner Entstehung darzulegen. Grundlegend für alle drei Aspekte ist dabei Kapitel 2, in dem nach einer kurzen Einführung der formal-statistischen Ausgangssi-

tuation der Individualdiagnostik sowohl die „Konstruktion“ des Paradoxons in einer großen Modellklasse erfolgt, als auch die dem Konstruktionsprozess inhärente mathematische Ursache dargelegt wird.

Während die Formulierung des paradoxen Effekts in Kapitel 2 in sehr allgemeiner Form erfolgt, widmet sich Kapitel 3 der Prävalenz des paradoxen Effekts in konkreten Modellklassen, die sich als Spezialfall des abstrakten Grundmodells aus Kapitel 2 darstellen lassen. Es handelt sich dabei zum Einen um Modelle, die in der Praxis der Skalenkonstruktion eine dominierende Rolle einnehmen (etwa linearkompensatorische Modelle wie z.B. das faktorenanalytische Modell), zum Anderen um Modelle, die zentraler Gegenstand aktueller Forschungsprojekte sind und die im Hinblick auf eine Effizienzsteigerung der Diagnose entwickelt wurden. Unter letzterer Klasse können beispielsweise Modelle, die neben der „üblichen“ Erfassung des Antwortmuster zusätzlich die itemspezifischen Reaktionszeiten der Person nutzen, um einen diagnostisch potentiell relevanten Informationsgewinn zu erlangen, subsumiert werden. Neben der Frage, ob dieser Zugewinn an diagnostischer Aussagekraft mit einer Zunahme an paradoxen Effekten einhergeht, und somit durch eine Verringerung an Fairness erkauft wird, soll auch der für diese Modellklasse charakteristische Einfluss der Antwortzeiten auf die Schätzung thematisiert und in Bezug zum paradoxen Effekt gebracht werden.

Ausgehend von der in den Kapiteln 2 und 3 anhand unterschiedlicher Modellklassen dargelegten Omnipräsenz des paradoxen Effekts befasst sich Kapitel 4 primär mit der Darlegung potentieller Auflösungsmöglichkeiten des paradoxen Effekts. Als Leitgedanke dieses Kapitels fungiert die Frage, ob sich statistisch „vernünftige“ Schätzer für die latenten Fähigkeiten der Person ableiten lassen, die nicht von dem paradoxen Effekt betroffen sind, d.h. deren Einsatz eine zugleich faire Diagnose ermöglicht.

Basierend auf der diesem Kapitel zugrundeliegenden Fragestellung - der Suche nach der Vereinbarkeit zwischen statistischer Effizienz und individuumbezogener Fairness - wird in Kapitel 5 die Frage aufgeworfen, ob der paradoxe Effekt geradezu ein notwendiges Kennzeichen eines statistisch angemessenen Schätzers ist, und ob somit Fairness und statistische Präzision im Rahmen der modellbasierten Diagnostik gegensätzliche Pole bilden. Die Erörterung dieses potentiellen Antagonismus mündet schließlich in einer generellen Diskussion der Bedeutung des Paradoxons für die psychologische Testtheorie.

2 Paradoxe Effekte in mehrdimensionalen Skalen

Nach einer kurzen, informellen Einführung des paradoxen Effekts (Kapitel 2.1) und der Kennzeichnung der statistisch relevanten diagnostischen Ausgangssituation (Kapitel 2.2) folgt in Kapitel 2.3 die detaillierte mathematische Ergündung des paradoxen Effekts. Die entscheidende Bedingung zur Herleitung des paradoxen Effekts wird dabei in Kapitel 2.4 näher betrachtet und zu einem bekannten Assoziationskonzept für Zufallsvariablen umgeformt. Letzteres wird eine stochastische Interpretation der fundamentalen Ursache des Paradoxons ermöglichen.

2.1 Informelle Einführung des paradoxen Effekts

Im Rahmen der Messung von latenten Konstrukten innerhalb der quantitativen Psychologie ist der Prozess der Skalenkonstruktion mit Hilfe von Latenten-Variablen-Modellen ein dominierendes Thema. *Nach* der Konstruktion einer „adäquaten“ Skala erscheint aus statistischer Sicht der Einsatz der Skala zur Schätzung der latenten Fähigkeiten von Personen unproblematisch, da in vielen Fällen auf Standardresultate der Maximum-Likelihood-Theorie (MLT) bzw. Bayes-Theorie (BT) zurückgegriffen werden kann.
Aus dieser Perspektive schien eine nähere Betrachtung der Eigenschaften dieser Fähigkeitsschätzungen unnötig, da die betreffende MLT/BT unter Regularitätsbedingungen (asymptotisch) optimale Schätzungen ermöglicht (siehe z.B. Berger, 1985; Le Cam, 1986; Lehmann und Casella, 1998). Ausgehend von dieser Optimalität lag es fern, den Einsatz der Schätzer anzuzweifeln. Insbesonders schien implizit stets die Annahme zu gelten, dass diese statistisch angemessenen Schätzer generell vernünftige Eigenschaften aufweisen, die nicht mit intuitiven Erwartungshaltungen kollidieren.
Das 2009 im Kontext der Fähigkeitsschätzung, von Hooker, Finkelman und Schwartzman bewiesene Paradoxon beschreibt erstmalig nicht nur einen Bruch mit dieser impliziten Annahme, es legt ferner nahe, dass statistisch optimale Schätzer kontraintuitives Verhalten bezüglich der Testfairness aufweisen können. Letzteres lässt sich am besten in Form des folgenden, fiktiven Beispiels von Hooker u.a. (2009) demonstrieren:

„Jane and Jill are fast friends who are nonetheless intensely competitive. At the end of high school they each take an entrance exam for a prestigious university. After the exam, they compare notes and discover that they gave the same answers for every question but the last. On checking their materials, it is clear that Jane answered this question correctly, but Jill answered incorrectly. They are therefore very surprised, when the test results are published, to find that Jill passed but Jane did not!“ (Hooker u.a., 2009, S.419)

Auch wenn dieses Resultat zunächst befremdlich erscheint, so lässt es sich dennoch mathematisch fundieren. Das wesentliche Prinzip des Paradoxons beruht auf einer kontraintuitiven Klassifikation zweier Personen. Es ist möglich, dass eine Person einen Test (augenscheinlich) besser absolviert als eine andere Person und trotzdem eine schlechtere Klassifikation erhält. Letzteres stellt den *interindividuellen* Aspekt des Paradoxons dar, der auf den Vergleich zweier Personen beruht. Zugleich lässt sich das obige Beispiel jedoch auch *intraindividuell* anhand der Frage, wie Jane's Klassifikation ausgefallen wäre, wenn sie das letzte Item nicht gelöst hätte, interpretieren. Diese auf den ersten Blick redundant anmutende Unterscheidung zwischen intra- und interindividueller Interpretation des Effekts wird jedoch eine zentrale Bedeutung bei der Diskussion von Methoden zur Vermeidung des paradoxen Effekts in Kapitel 4 einnehmen. Beide Formen widersprechen der selten explizit ausgedrückten Forderung an eine Testprozedur, dass „bessere“ Antworten zu einer „besseren“ Klassifikation führen sollten und hinterfragen somit die Sinnhaftigkeit der zugrundeliegenden Schätzmethodik.
Das Beispiel gibt jedoch keine genaue Auskunft über die konkrete Struktur der Testprozedur, die dieses paradoxe Resultat „ermöglicht“. Insbesonders könnte die Vermutung naheliegen, dass das Beispiel einen fiktiv konstruierten, „pathologischen“ Effekt beschreibt, der bei „normal“ konstruierten Skalen praktisch nicht vorkommt.
Zentrales Anliegen der folgenden Kapitel ist es, dieser Vermutung nachzugehen, und dabei den Gültigkeitsbereich des paradoxen Effekts zu präzisieren. Zu diesem Zweck ist es zunächst erforderlich, die statistische Ausgangssituation im Kontext der Individualdiagnostik zu betrachten und grundlegende Notation, die für die mathematische Ergründung des Effekts unabdingbar ist, einzuführen.

2.2 Statistische Basis der Individualdiagnostik

Klassifikation statistischer Schätzmethoden

Schätzmethoden in stochastischen Modellen lassen sich nach dem ihnen zugrundeliegenden Paradigma klassifizieren. Das frequentistische Paradigma basiert dabei im Wesentlichen auf Konzepten wie Suffizienz oder Anzillarität (siehe z.B. Casella und Berger, 2002, S.272 ff.); das Likelihood-Paradigma fußt auf der Likelihood-Funktion (siehe z.B. Casella und Berger, 2002, S.292 ff.), d.h. der Dichte der beobachteten Daten aufgefasst als Funktion des dahinterliegenden, unbekannten Parameters; und das bayesianische Paradigma betrachtet neben den beobachtbaren Daten auch die unbekannten Parameter als Zufallsgrößen in einem „gemeinsamen“ Wahrscheinlichkeitsmodell (Berger, 1985).
Wenngleich die Unterscheidung dieser drei Paradigmen für das Verständnis statistischer Aspekte i.A. sinnvoll ist, so erweist sich bezüglich der Herleitung des paradoxen Effekts eine andere Form der Differenzierung von Schätzmethoden als zweckdienlicher. Die Unterteilung von Schätzmethoden, die sich für die folgenden Abschnitte als hilfreich erweisen wird, unterscheidet ob die Schätzmethodik auf der Maximierung einer (noch genauer zu charakterisierenden) Funktion (Typ 1) oder auf einem Funktional einer (noch zu spezifizierenden) Verteilungsfunktion (Typ 2) beruht.
Nach einer Einführung der für die folgenden Kapitel relevanten formalen Ausgangssituation wird zunächst das Paradoxon für Schätztyp 1 hergeleitet. Ein anschließendes Kapitel (Kapitel 2.4) behandelt dann die notwendige Erweiterung für Schätztyp 2.

Formale Ausgangssituation der Individualdiagnostik

Sei $\boldsymbol{\theta^T} = (\theta_1, \ldots, \theta_p)$ der Vektor der latenten Fähigkeiten der zu diagnostizierenden Person und seien $u_1, u_2, \ldots u_k$ die von der Person auf Item 1 bis Item k erzielten Itemscores, dann repräsentiert die Likelihood-Funktion

$$L_u(\boldsymbol{\theta}) = P(u_1, u_2, \ldots u_k | \, \boldsymbol{\theta})$$

die Wahrscheinlichkeit des beobachteten Antwortmuster $u_1, \ldots u_k$ als Funktion der latenten Fähigkeiten.
Man beachte, dass keine Voraussetzungen an das Skalenniveau der Itemscores ge-

stellt werden. Ebenso wird nicht zwingend postuliert, dass alle Itemscores das gleiche Skalenniveau aufweisen. Analoges gilt für die Unterscheidung zwischen binären und polychotomen Daten sowie für die Abgrenzung zwischen stetigen[1] und diskreten Daten.

Somit umschließt die folgende Darstellung faktorenanalytische Modelle (stetige, intervallskalierte Itemscores), ordinale Item-Response-Modelle (kategoriale, ordinalskalierte Itemscores) als auch binäre Item-Response-Modelle (dichotome Itemscores).

Anhand der konkreten Gestalt der Funktion L_u erfolgt die statistische Schätzung der unbekannten Fähigkeiten $\boldsymbol{\theta}$. Der ML-Schätzer (MLE) ist dabei als der Wert $\hat{\boldsymbol{\theta}}$ definiert, der L_u maximiert - vorausgesetzt ein solches Maximum existiert. Analog ist der Bayesianische-Modal-Schätzer (MAP: *maximum a-posteriori*) als der Wert $\hat{\boldsymbol{\theta}}$ definiert, der die Funktion

$$L_u(\boldsymbol{\theta}) \cdot p(\boldsymbol{\theta}) \tag{2.1}$$

maximiert. $p(\cdot)$ bezeichnet hierbei die Prioriverteilung, die vorhandenes Vorwissen über die Fähigkeiten der Person wiedergeben soll.

Die Form der Funktionen L_u und p ergibt sich dabei unmittelbar aus der abgeschlossenen Phase der Testkonstruktion und Normierung, in der die Charakteristiken der Messinstrumente/Items (z.B. Itemschwierigkeiten) geschätzt und statistische Beziehungen wie z.B. Korrelationen zwischen den latenten Fähigkeiten modellbasiert bestimmt werden. Ein gängiges Beispiel für diese Prozedur ist im Kontext der linearen Faktorenanalyse durch die Festlegung einer orthogonalen Faktorenstruktur (unkorrelierte Fähigkeiten) mit anschließender Schätzung der Ladungsmatrix (Charakteristikum der Items) der Items gegeben.

Für Zwecke der Individualdiagnostik wird üblicherweise angenommen, dass die in der Konstruktions- bzw. Normierungsphase geschätzten Parameter (z.B. Itemdiskriminationsparameter) mit den tatsächlichen Werten dieser Parameter übereinstimmen. Letzteres kann für eine hinreichend große Normierungsstichprobe näherungsweise akzeptabel sein und wird im Folgenden stets unterstellt.

Ausgehend von dieser Annahme stellt $\boldsymbol{\theta}$ den einzig unbekannten Parameter in (2.1) dar. Basierend auf der Schätzung dieses Parameters können Klassifikationen erfolgen. Eine simple Form der Klassifikation (Segall, 2010) verwendet festgelegte Schwellenwerte $\kappa_1, \kappa_2 \ldots, \kappa_p \, (\kappa_i \in \overline{\mathbb{R}} \; \forall i)$ auf den einzelnen Dimensionen, deren Überschreitung

[1] Im Falle stetiger Itemscores ist der Ausdruck L_u jedoch als Dichtefunktion (in $u_1, \ldots u_k$) zu interpretieren.

erforderlich ist, um eine positive Klassifikation zu erhalten. Mit anderen Worten: Das Individuum erzielt bezüglich dieser Klassifikationsregel genau dann eine positive Klassifikation, wenn $\hat{\theta}_i > \kappa_i$ für $i = 1, \ldots p$ gilt. Die Wahl $\kappa_i = -\infty$ ermöglicht es ferner, den Fall einer von der i-ten Fähigkeit unabhängigen Klassifikation darzustellen. Im Extremfall kann die Klassifikation auf lediglich einer Fähigkeit basieren, obwohl p-Dimensionen in die Beantwortung der Items eingehen ($\exists!\ i : \kappa_i > -\infty$). Die übrigen Dimensionen werden dann als Störmdimensionen („nuisance“ dimensions) bezeichnet. Paradoxe Effekte bei der Fähigkeitsschätzungen können bei Verwendung dieses Klassifikationsschemas zu paradoxen Klassifikationen führen. Man nehme beispielsweise an, dass eine Person alle Schwellenwerte überschreitet bis auf den i-ten. Wenn nun in diesem Kontext Item j, welches von der Person richtig beantwortet wurde, einen paradoxen Effekt aufweist, dann hätte diese Person durch falsches Beantworten des j-ten Items eine Erhöhung des Schätzwerts der i-ten Dimension erzielen können. Die Konsequenz kann eine Überschreitung von κ_i und somit eine - wenn übrige klassifikationsrelevante Schwellenwerte nicht unterschritten werden - durch Falschantworten herbeigeführte positive Klassifikation sein. Die Schätzwerte der übrigen Dimensionen können dabei durch diese falsche Antwort durchaus geringer ausfallen. Solange die zugehörigen Schwellenwerte jedoch nicht unterschritten werden, besitzt dies keine negativen Folgen für das Individuum.
Im folgenden, für diese Arbeit zentralen Kapitel wird eine generelle formale Bedingung herausgearbeitet, die Paradoxien bei der Fähigkeitsschätzung in einem p-dimensionalen ($p \geq 2$) Test verursacht.

2.3 Konstruktiver Beweis der Existenz des paradoxen Effekts

Zunächst werden Paradoxien für Schätztyp 1 betrachtet. Der ML-Schätzer der Fähigkeiten ist definiert[2] als

$$\hat{\boldsymbol{\theta}}_k = \arg\max_{\boldsymbol{\theta}} l_u(\boldsymbol{\theta}),$$

wobei l_u den Logarithmus von L_u bezeichnet und der Index k den Umstand wiedergibt, dass in die Berechnung des Schätzers Informationen von k Items einfließen.

[2]Bedingungen, die die Existenz eines ML-Schätzers garantieren, sind in Anhang A näher dargelegt. Im Folgenden wird stets die Existenz des Schätzers unterstellt.

Durch Hinzufügen eines weiteren Items ergibt sich unter der Annahme der bedingten Unabhängigkeit die erweiterte Loglikelihood

$$l_{u,u_{k+1}}(\boldsymbol{\theta}) := l_u(\boldsymbol{\theta}) + \log(f_{u_{k+1}}(\boldsymbol{\theta})),$$

in der $\log(f_{u_{k+1}}(\boldsymbol{\theta}))$ den Beitrag des hinzugefügten Items darstellt. Wenn der neue ML-Schätzer mit $\hat{\boldsymbol{\theta}}_{k,u_{k+1}}$ bezeichnet wird, dann lässt sich die von dem Item induzierte Veränderung des Schätzwerts durch $\hat{\boldsymbol{\theta}}_{k,u_{k+1}} - \hat{\boldsymbol{\theta}}_k$ ausdrücken. Analog lässt sich anhand des Terms $\hat{\boldsymbol{\theta}}_{k,u'_{k+1}} - \hat{\boldsymbol{\theta}}_k$ $(u'_{k+1} > u_{k+1})$ die Veränderung angeben, wenn anstelle des Itemscores u_{k+1} der höhere Itemscore u'_{k+1} erzielt worden wäre.
Die folgenden Überlegungen werden darlegen, unter welchen Bedingungen an l_u und $l_{u,u_{k+1}}$ Veränderungen bezüglich einzelner Dimensionen positiv bzw. negativ ausfallen. Neben einer separaten Herleitung für zwei- und höherdimensionale Fälle wird dabei zunächst zur Vereinfachung angenommen, dass das letzte (hinzugefügte) Item lediglich eine Dimension misst - d.h., dass die Funktion $f_{u_{k+1}}(\boldsymbol{\theta})$ für jede Wahl von u_{k+1} nur von der eindimensionalen Größe θ_p abhängt. Auch wenn diese Zusatzannahme zunächst restriktiv erscheint, so wird sich in späteren Abschnitten zeigen, dass sie für die große Modellklasse der linear-kompensatorischen IRT-Modelle immer via Rotation der Dimensionen erfüllbar ist und dass im generellen Fall Abschwächungen dieser Annahme möglich sind, die die folgenden Ergebnisse qualitativ nicht verändern.

Konstruktionsprinzip in zweidimensionalen Tests

Maßgeblich für die Untersuchung des paradoxen Effekts ist - wie bereits von Hooker u.a. (2009) demonstriert wurde - das Verhalten der folgenden, implizit definierten Abbildung, die die Stelle des bedingten Maximums wiedergibt:

$$\hat{\theta}_1(\theta_2) := \arg\max_{\theta_1} l_u(\theta_1, \theta_2).$$

Diese Abbildung wird im Folgenden mit „bedingte Maximumsabbildung“ bezeichnet. Bei Kenntnis der Ausprägung der zweiten Dimension gibt sie den Wert der unbekannten ersten Dimension $\hat{\theta}_1(\theta_2)$ an, unter dem das Zustandekommen der beobachteten Itemscores die größte Wahrscheinlichkeit besitzt. Die bedingte Maximumsabbildung ist prinzipiell - da i.A. weder die Existenz noch die Eindeutigkeit des Maximums garantiert ist - eine *mengenwertige* Abbildung, d.h. $\hat{\theta}_1(\theta_2)$ gibt die *Menge* aller bedingten Maximumsstellen an. Wenngleich die folgenden Betrachtungen stets unter

der Annahme erfolgen, dass $\hat{\theta}_1(\theta_2)$ eine Funktion[3] ist, so wird an einigen Stellen auch eine Verallgemeinerung für den Fall einer mengenwertigen Abbildung angestrebt.
Die wesentliche Verbindung dieser Funktion zum paradoxen Effekt kann über die folgende, heuristische Argumentation hergestellt werden: Wenn $\hat{\theta}_1(\theta_2)$ streng monoton fallend ist, dann bedeutet dies, dass für zwei Personen mit jeweils bekannten Ausprägungen θ_2', θ_2 $(\theta_2' > \theta_2)$ auf der zweiten Dimension und identischer Testleistung, d.h. identischer Likelihood-Funktion, die unbekannte Ausprägung der ersten Fähigkeit so geschätzt wird, dass $\hat{\theta}_1(\theta_2') < \hat{\theta}_1(\theta_2)$ gilt. Mit anderen Worten: Die beiden Dimensionen stehen in einem Trade-Off, derart, dass die bezüglich der zweiten Dimension fähigere Person einen geringeren Schätzwert auf der ersten Dimension erhält.
Im Fall unbekannter Fähigkeit lässt sich analog zu dieser Vorgehensweise argumentieren. Das nun fehlende Wissen, welche Person die höhere Fähigkeit θ_2 besitzt, wird durch die unterschiedliche Beantwortung des letzten, ausschließlich die zweite Dimension messenden Items kompensiert. Der Person, die dieses, ausschließlich die zweite Dimension messende Item besser beantwortet, wird die höhere Fähigkeit θ_2 zugeschrieben. Da das hinzugefügte Item zudem die bedingte Maximumsfunktion nicht ändert, lässt sich anhand der Beobachtung, dass der ML-Schätzer $\hat{\boldsymbol{\theta}}$ die Beziehung $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1(\hat{\theta}_2), \hat{\theta}_2)$ erfüllt, die obige Argumentation, die von einer bekannten Fähigkeit θ_2 ausging, übertragen.
Ziel der folgenden Betrachtungen ist es, diese heuristische Argumentation in eine rigorose Form zu überführen und zugleich eine Eigenschaft einzuführen, die die postulierte Monotonie der bedingten Maximumsfunktion/Maximumsabbildung impliziert. Vor dem Hintergrund dieser Motivation wird folgende Definition getätigt:

Definition 2.3.1 (SMDD-Bedingung)**.** Eine Funktion $f(x, y)$ besitzt (streng) monoton fallende Differenzen (strictly monotone decreasing differences (S)MDD), wenn $f(x', y) - f(x, y)$ (streng) monoton fallend in y für jedes Paar (x', x) mit der Eigenschaft $x' > x$ ist.

Die (SMDD)-Eigenschaft ist eine symmetrische Eigenschaft im folgenden Sinne:

Lemma 2.3.1. *Wenn $f(x, y)$ (S)MDD ist, dann ist $f(x, y') - f(x, y)$ (streng) monoton fallend in x für jedes Paar (y', y) mit der Eigenschaft $y' > y$.*

[3]Für generelle Bedingungen, die die Existenz dieser Funktion garantieren, sei auf Anhang A verwiesen.

Beweis. MDD impliziert, dass

$$f(x', y) - f(x, y) \geq f(x', y') - f(x, y')$$

gilt, wenn $y' > y$ und $x' > x$ ist.
Direkte Umstellung und Neugruppierung von Termen ergibt dann die Behauptung:

$$f(x, y') - f(x, y) \geq f(x', y') - f(x', y).$$

Der Fall strenger Monotonie folgt analog zu diesen Betrachtungen. □

Operationen, die die SMDD Eigenschaft erhalten, sind Gegenstand der folgenden Hilfssätze.

Lemma 2.3.2. *Sei $f(x, y) := f_1(x, y) + f_2(x, y)$. Wenn f_1 MDD und f_2 (S)MDD ist, dann ist f (S)MDD. Die MDD-Bedingung an f_1 ist insbesonders dann erfüllt, wenn f_1 nur von einer Variable abhängt.*

Beweis. Der erste Teil folgt direkt durch Ausschreiben der Bedingungen, da Summen monotoner Funktionen monoton sind.
Eine Funktion f_1, die nur von einer Variable abhängt, ist ferner stets MDD, da $\forall\, y$

$$f_1(x', y) - f_1(x, y) = 0$$

gilt, wenn f_1 nur von y abhängig ist. □

Lemma 2.3.3. *f (S)MDD $\Leftrightarrow$ cf (S)MDD für jede Konstante $c > 0$.*

Beweis. Trivial. □

Mit dem eingeführten Konzept ist es nun möglich, die besagte Monotonie der bedingten Maximumsabbildung herzuleiten.

Lemma 2.3.4. *Sei $\hat{x}(y)$ die bedingte Maximumsabbildung von $f(x, y)$. Wenn $f(x, y)$ die SMDD-Bedingung erfüllt, und wenn y_1, y_2 zwei geordnete Werte ($y_1 < y_2$) bezeichnen, für die $\hat{x}(y_1) \neq \emptyset \wedge \hat{x}(y_2) \neq \emptyset$ ist, dann gilt $x_2 \leq x_1$ für beliebiges $x_2 \in \hat{x}(y_2)$ und $x_1 \in \hat{x}(y_1)$.*

Beweis. Sei gegensätzlich zur Behauptung $y_1 < y_2$ und $x_1 < x_2$. Es gilt:

$$f(x_2, y_2) - f(x_1, y_2) \geq 0.$$

Nach der SMDD-Eigenschaft und obiger Annahme folgt

$$f(x_2, y_1) - f(x_1, y_1) > f(x_2, y_2) - f(x_1, y_2) \geq 0.$$

Demnach ergibt sich der Widerspruch $x_1 \notin \hat{x}(y_1)$.
Man beachte, dass die Gültigkeit des Lemmas bestehen bleibt, wenn die SMDD-Bedingung durch die schwächere MDD-Bedingung ersetzt wird und zusätzlich die Einelementigkeit von $\hat{x}(y_1)$ und $\hat{x}(y_2)$ postuliert wird. □

Lemma 2.3.4 liefert eine Grundlage, um die Monotonie der bedingten Maximumsfunktion zu gewähren.
Eine Loglikelihoodfunktion $l_u(\theta_1, \theta_2)$ mit der SMDD-Eigenschaft impliziert somit stets, dass von zwei Personen mit identischer Testleistung und jeweils bekannten Fähigkeitsausprägungen θ_2, θ_2', diejenige Person, die eine geringere Ausprägung auf der zweiten Dimension aufweist, eine nicht geringere Fähigkeitszuschreibung auf der übrigen Dimension erhält.
Die Ausdehnung dieser antagonistischen Relation auf den Fall unbekannter Dimensionen ist Gegenstand des folgenden, für die Untersuchung des paradoxen Effekts zentralen Satzes:

Satz 2.3.1. *Sei $g(x, y)$ SMDD und sei $h(y)$ eine streng monoton wachsende Funktion von y. Wenn $(x_f, y_f), (x_g, y_g)$ Stellen eines Maximums von $f(x, y) := g(x, y) + h(y)$ bzw. von $g(x, y)$ bezeichnen, dann gelten im Falle von $y_f \neq y_g$ die Ungleichungen*

$$y_f > y_g, \; x_f \leq x_g.$$

Beweis. Nach Lemma 2.3.2 ist f ebenfalls SMDD. Da $h(\cdot)$ zudem nur von einer Variable abhängt, ist die bedingte Maximumsabbildung $\hat{x}(y)$ für f und g identisch. Ferner gilt offenbar $x_f \in \hat{x}(y_f)$ sowie $x_g \in \hat{x}(y_g)$. Demnach wird die Behauptung unter Rückgriff auf Lemma 2.3.4 folgen, wenn gezeigt werden kann, dass $y_f > y_g$ gilt. Wenn $y_f < y_g$ gelten würde, dann wäre $h(y_f) < h(y_g)$ und folglich: (man beachte, dass h nicht von x abhängt!)

$$f(x_f, y_f) - g(x_f, y_f) < f(x_g, y_g) - g(x_g, y_g).$$

Durch Neugruppierung der Form

$$f(x_f, y_f) + g(x_g, y_g) < f(x_g, y_g) + g(x_f, y_f)$$

folgt ein Widerspruch zur Annahme, dass $(x_f, y_f), (x_g, y_g)$ Stellen eines Maximums von f bzw. g darstellen.

□

Satz 2.3.1 stellt das zentrale Resultat zum paradoxen Effekt für Typ-1-Schätzer dar. Unter der Annahme, dass die bedingte Maximumsabbildung existiert (d.h. an den entsprechenden Stellen nicht leer ist), zeigt er hinreichende Bedingungen für ein paradoxes Resultat auf. Um die volle Allgemeinheit des Satzes herauszustellen, sollen einige Spezialfälle dieses Satzes, die in folgenden Kapiteln noch eingehender zu diskutieren sind, kurz dargestellt werden.

a) Wenn g die SMDD-Loglikelihood resultierend aus den ersten k-Items eines *binären MIRT-Modells* repräsentiert ($g = l_u$), und wenn richtiges Beantworten des letzten Items eine streng monoton wachsende Funktion der zweiten Dimension ist, dann zeigt der Satz, dass der zur ersten Dimension korrespondierende Schätzwert durch das richtige (falsche) Beantworten des letzten Items fällt (steigt)[4].

b) Wenn g die SMDD-Loglikelihood resultierend aus den ersten k-Items eines *ordinalen MIRT-Modells* repräsentiert ($g = l_u$), und wenn die Antwortwahrscheinlichkeiten für die Extremkategorien des letzten Items durch streng monotone Funktionen der zweiten Dimension gegeben sind, dann lässt sich analog wie in a) argumentieren, indem das Erreichen des höchsten Itemscores mit dem Erreichen des niedrigsten Itemscores verglichen wird.

c) Wenn g die Loglikelihood resultierend aus den ersten $k+1$-Items eines *faktorenanalytischen Modells* mit positiver Ladungsmatrix repräsentiert ($g = l_{u,u_{k+1}}$), und wenn h die Differenz $\log(f_{u_{k+1}+\epsilon}(\theta_2)) - \log(f_{u_{k+1}}(\theta_2))(\epsilon > 0)$ darstellt, dann ist das Erzielen eines höheren Itemscores auf dem letzten, ausschließlich die zweite Dimension messenden Items verantwortlich für eine Abnahme des zur ersten Dimension korrespondierenden Schätzwertes.

[4]Zur sprachlichen Vereinfachung wird im Folgenden implizit der Fall gleichbleibender Schätzungen ausgeschlossen.

d) Wenn g' eine *separable Loglikelihood* beschreibt (d.h.: $g' := l_u(\theta_1, \theta_2) = l_1(\theta_1) + l_2(\theta_2)$) und wenn die logarithmierte Prioriverteilung γ SMDD ist, dann lässt sich analog zu a) bzw. b) unter Verwendung von Lemma 2.3.2 ein paradoxer Effekt durch Änderung des Itemscores erzielen ($g = l_u + \gamma$). Dieser, erstmalig von Hooker (2010) untersuchte Fall besitzt - wie in Kapitel 3.2 und 3.3 aufgezeigt werden wird - besondere Relevanz im Rahmen von IRT-Modellen zur Veränderungsmessung sowie für IRT-Modelle, die eine zusätzliche Zeitkomponente zur Erhöhung der Messpräzision beinhalten.

Fall d) beschreibt zugleich eine Verallgemeinerung der bisherigen, auf ML-Schätzern beschränkten Diskussion hin zu Bayes-Schätzern. Insbesonders stellt er heraus, dass separable Loglikelihoods, die im ML-Kontext ein Garant für paradoxiefreie Inferenz darstellen, im bayesianischen Kontext in Abhängigkeit von der Prioriverteilung paradoxe Effekte hervorrufen können.
Fall a) bildete ferner den Ausgangspunkt der Herleitung des paradoxen Effekts bei Hooker u.a. (2009). Um den Unterschied zwischen der Methodik von Hooker u.a. und dem obigen, auf dem SMDD-Konzept beruhenden Zugang hervorzuheben, und damit insbesonders den Zugewinn an Allgemeinheit herauszustellen, wird im Folgenden das Beweisprinzip von Hooker u.a. rekapituliert.
Anstelle der allgemeinen Loglikelihood-Funktion $l_u(\theta_1, \theta_2)$ betrachten Hooker u.a. eine Loglikelihood der Form

$$l_u(\theta_1, \theta_2) = \sum_{i \in M(u)} \log(f_i(\theta_1, \theta_2)) + \sum_{i \notin M(u)} \log(1 - f_i(\theta_1, \theta_2)), \tag{2.2}$$

worin $f_i(\theta_1, \theta_2)$ die Item-Response-Funktion des i-ten Items und $M(u)$ die Indexmenge der richtig gelösten Items bezeichnen. $f_i(\theta_1, \theta_2)$ repräsentiert somit in Abhängigkeit von der Fähigkeit (θ_1, θ_2) die bedingte Wahrscheinlichkeit, das Item zu lösen. Wie anhand dieser veränderten Loglikelihood deutlich wird, erfolgt somit eine Beschränkung auf binäre, lokal unabhängige Items.
Zentral für die Herleitung des paradoxen Resultats von Hooker u.a. (2009) ist die Forderung, dass alle zweiten partiellen Ableitungen[5] eines beliebigen Summanden in (2.2) nichtpositiv ausfallen und das mindestens ein Summand existiert, für den diese zweiten partiellen Ableitungen stets echt negativ sind.

[5]Man beachte an dieser Stelle den impliziten Übergang zu stetigen, auf dem $\mathbb{R}^2$ definierten, latenten Variablen.

Ausgehend von der Dekomposition (2.2) und der Forderung an die partiellen Ableitungen zweiter Ordnung beginnt die Konstruktion des paradoxen Effekts mit der Betrachtung der Nullstelle der „partiellen“ Scorefunktion

$$\frac{\partial l_u}{\partial \theta_1}(\theta_1, \theta_2).$$

Für diese Funktion wird das Verhalten der „bedingten Nullstelle“ für θ_1

$$s(\theta_2) := \left\{ \theta_1 | \, \frac{\partial l_u}{\partial \theta_1}(\theta_1, \theta_2) = 0 \right\}$$

in Abhängigkeit von θ_2 betrachtet. $s(\theta_2)$ gibt somit für jeden Wert θ_2 den Wert der ersten Dimension an, der die Scoregleichung $\frac{\partial l_u}{\partial \theta_1}(\theta_1, \theta_2) = 0$ löst. Die Rolle der Funktion $s(\theta_2)$ in der Konstruktion des paradoxen Resultats ist analog zur Rolle der bedingten Maximumsfunktion $\hat{\theta}_1(\theta_2)$ in der bereits geschilderten, verallgemeinerten Herleitung des Paradoxons. In der Tat, wenn $\hat{\theta}_1(\theta_2)$ die Stelle eines bedingten Maximum von l_u wiedergibt, dann muss $\frac{\partial l_u}{\partial \theta_1}(\hat{\theta}_1(\theta_2), \theta_2) = 0$ gelten, da jedes im Inneren einer offenen Menge liegende (lokale) Maximum einer differenzierbaren Funktion einen Stationaritätspunkt der Funktion darstellt. Insbesonders folgt $\hat{\theta}_1(\theta_2) \subset s(\theta_2)$. Die Umkehrung $s(\theta_2) \subset \hat{\theta}_1(\theta_2)$ gilt zwar i.A. nicht, folgt jedoch aus der von obiger Annahme über die Negativität der zweiten Ableitung $\frac{\partial^2 l_u}{\partial \theta_1^2}$ implizierten Konkavität von $h_{\theta_2}(\theta_1) := l_u(\theta_1, \theta_2)$ (siehe Theorem 2.13 aus Rockafellar und Wets (2009)) unter Rückgriff auf Theorem 2A.6 von Dontchev und Rockafellar (2009). Da $h_{\theta_2}(\theta_1)$ unter den obigen Annahmen sogar strikt konkav ist, folgt ferner, dass die bedingte Maximumsabbildung eine echte Funktion - d.h. $\hat{\theta}_1(\theta_2)$ ist höchstens einelementig für jede Wahl von θ_2 - darstellt.

Unter der Dekomposition (2.2) und den Annahmen über die partiellen Ableitungen zweiter Ordnung gilt somit $s(\theta_2) = \hat{\theta}_1(\theta_2)$, wobei $s(\theta_2)$ und $\hat{\theta}_1(\theta_2)$ Funktionen (mit potentiell „leerem Funktionswert“) darstellen. Aufgrund dieser Identifizierung von $s(\theta_2)$ mit $\hat{\theta}_1(\theta_2)$ kann das zuvor dargestellte Konstruktionsprinzip - beruhend auf einer Anwendung von Satz 2.3.1 - übertragen werden, wenn gezeigt werden kann, dass $s(\theta_2)$ bzw. $\hat{\theta}_1(\theta_2)$ monoton fallend in θ_2 ist.

Unter der zusätzlichen Annahme, dass $s(\theta_2)$ global wohldefiniert ist und dass die Summanden zweimal stetig differenzierbar sind, kann Letzteres mit Hilfe des Satzes über implizite Funktionen (siehe z.B. Theorem 1B.1 in Dontchev und Rockafellar (2009)) hergeleitet werden. Eine Anwendung dieses Satzes auf die durch $s(\theta_2)$ gegebene „Lösungsabbildung“

$$s : \theta_2 \mapsto \left\{ \theta_1 : \frac{\partial l_u}{\partial \theta_1}(\theta_1, \theta_2) = 0 \right\}$$

ergibt

$$\frac{\partial s}{\partial \theta_2}(\theta_2) = -\left(\frac{\partial^2 l_u}{\partial \theta_1^2}(s(\theta_2), \theta_2)\right)^{-1} \frac{\partial^2 l_u}{\partial \theta_1 \partial \theta_2}(s(\theta_2), \theta_2) < 0, \tag{2.3}$$

wobei die strikte Negativität der zweiten partiellen Ableitungen sowohl die für die Anwendung des Satzes über implizite Funktionen erforderliche Invertierbarkeit von $\frac{\partial^2 l_u}{\partial \theta_1^2}(s(\theta_2), \theta_2)$ als auch - via (2.3) - die für die Herleitung des paradoxen Effekts zentrale Monotonie der bedingten Maximumsfunktion sichert.

Von diesem Resultat ausgehend, verläuft die Argumentation analog zu der in Satz 2.3.1 skizzierten Methodik. Der Kern des Konstruktionsvorgangs liegt somit in der Begründung der Monotonie der bedingten Maximumsfunktion. Diese Monotonie wird bei Hooker u.a. über die Forderung von zweimal stetig differenzierbaren Item-Response-Funktionen mit strikt negativen zweiten partiellen Ableitungen etabliert, während für die Generalisierung das SMDD-Konzept eingeführt wurde. Letzteres stellt im Gegensatz zu Ersterem keine restriktiven Forderungen bezüglich der Glattheit der Item-Response-Funktionen. Ferner stellt das SMDD-Konzept, wie das folgende Lemma (siehe Theorem 9.40 in Rudin (1976)) zeigen wird, eine echte Verallgemeinerung der zentralen Bedingung von Hooker u.a. (2009) dar.

Lemma 2.3.5 (Rudin (1976), Theorem 9.40)**.** *Wenn $\frac{\partial f}{\partial x}(x, y)$ und $\frac{\partial^2 f}{\partial y \partial x}(x, y)$ für alle (x, y) in einer offenen Menge O existieren, und wenn (x_0, y_0), (x_1, y_1) gegenüberliegende Ecken eines achsenparallelen, abgeschlossenen, in O liegenden Rechtecks darstellen, dann gibt es einen im Inneren des Rechtecks liegenden Punkt (x, y), so dass*

$$f(x_1, y_1) - f(x_1, y_0) - f(x_0, y_1) + f(x_0, y_0) = (x_1 - x_0)(y_1 - y_0)\frac{\partial^2 f}{\partial y \partial x}(x, y) \tag{2.4}$$

gilt.

Beweis. Der Beweis von Rudin (1976) wird rekapituliert:

Zunächst betrachte man die Funktion $v(x) := f(x, y_1) - f(x, y_0)$. Gemäß den Annahmen ist v im Intervall $[x_0, x_1]$ (o.B.d.A sei $x_1 > x_0$) differenzierbar. Eine Anwendung des Mittelwertsatzes der Differentialrechnung ergibt daher

$$v(x_1) - v(x_0) = (x_1 - x_0)\left(\frac{\partial f}{\partial x}(x, y_1) - \frac{\partial f}{\partial x}(x, y_0)\right)$$

für einen Punkt $x \in (x_0, x_1)$.

Da ferner aufgrund der Annahme bezüglich der gemischten Ableitung die Funktion

$u(y) := \frac{\partial f}{\partial x}(x, y)$ im Intervall $[y_0, y_1]$ differenzierbar ist, folgt mittels erneuter Anwendung des Mittelwertssatzes die Existenz eines Punktes $y \in (y_0, y_1)$, so dass

$$v(x_1) - v(x_0) = (x_1 - x_0)(y_1 - y_0)\frac{\partial^2 f}{\partial y \partial x}(x, y)$$

gilt. Dies ist identisch mit der Behauptung. □

Korollar 2.3.1. *Wenn $f(x, y)$ innerhalb einer offenen Menge O zweimal stetig differenzierbar ist und dort eine negative gemischte zweite partielle Ableitung aufweist, dann gilt die SMDD-Eigenschaft für f bezüglich jedem in O liegenden, achsenparallelem Rechteck.*

Beweis. Folgt direkt aus Lemma 2.3.5, da die rechte Seite von (2.4) aufgrund der Negativität der gemischten Ableitung stets negativ ist. Durch Umstellen lässt sich dann (2.4) direkt zur SMDD-Bedingung umformen.
Es sei ferner bemerkt, dass die volle Stärke der Annahmen des Korollars nicht notwendig ist, um die SMDD-Bedingung herzuleiten. Die Existenz der in Lemma 2.3.5 genannten Ableitungen ist ausreichend. Ferner sei darauf hingewiesen, dass in Lemma 2.3.5 nicht notwendigerweise $\frac{\partial^2 f}{\partial y \partial x}(x, y) = \frac{\partial^2 f}{\partial x \partial y}(x, y)$ gelten muss. Letztere partielle Ableitung muss theoretisch nicht einmal existieren. Sollte Sie dennoch existieren, so setzt das Lemma dennoch *nicht* die Vertauschbarkeit der Differentiationsreihenfolge voraus. □

Somit impliziert folglich die von Hooker u.a. zur Konstruktion des paradoxen Effekts verwendete Bedingung die SMDD-Bedingung. Im Fall binärer IRT-Modelle wurde somit durch Satz 2.3.1 eine Verallgemeinerung des paradoxen Effekts erreicht. Die Item-Response-Funktionen müssen weder glatt (im Sinne von differenzierbar) verlaufen , noch Konkavitätsforderungen genügen. Neben dieser Erweiterung für binäre Daten bietet Satz 2.3.1 zudem die Möglichkeit, den paradoxen Effekt auf beliebiges Skalenniveau auszudehnen. Von besonderem Interesse werden hierbei ordinale sowie metrische Response-Modelle sein (siehe Kapitel 3). Ferner sei an dieser Stelle betont, dass die von Hooker u.a. vorgenommene additive Struktur der Loglikelihood, d.h. die Annahme lokal unabhängiger Items, nicht notwendig ist. Wie die Konstruktion des paradoxen Effekts in Satz 2.3.1 zeigt, ist lediglich die stochastische Unabhängigkeit des hinzugefügten Items von den vorherigen k Items erforderlich. Zudem wird anhand der Neuformulierung über die SMDD-Bedingung deutlich, dass die Konkavität

der Loglikelihood (genauer: die Konkavität bezüglich der ersten Komponente θ_1) entgegen Vermutungen, die bei Betrachtung von (2.3) entstehen könnten, in keinem direkten kausalen Zusammenhang zum paradoxen Effekt steht.

Konstruktionsprinzip in $p > 2$ dimensionalen Tests

Die bisherigen Betrachtungen waren auf den Fall zweidimensionaler Tests beschränkt. Wie die anschließenden Ausführungen jedoch zeigen werden, kann über eine modifizierte Version des zweidimensionalen SMDD-Konzepts eine Ausdehnung des paradoxen Effekts auf $p >$ 2-dimensionale Tests erfolgen. Dies sei im Folgenden für Schätztyp 1 skizziert:
Ausgangspunkt der Betrachtungen ist eine Erweiterung des SMDD-Konzepts.

Definition 2.3.2 (eSMDD-Eigenschaft)**.** Eine Funktion $f(\boldsymbol{x}, y)$ heisst erweitert-(S)MDD in y (kurz: e(S)MDD), wenn $f(\boldsymbol{x'}, y) - f(\boldsymbol{x}, y)$ (streng) monoton fallend in y für jedes Paar $(\boldsymbol{x'}, \boldsymbol{x})$ mit der Eigenschaft $\boldsymbol{x'} > \boldsymbol{x}$ ist.

Die Ungleichung $\boldsymbol{x'} > \boldsymbol{x}$ ist nun im Sinne der partiellen Ordnung im $\mathbb{R}^{p-1}$ zu verstehen, d.h. $\boldsymbol{x'} > \boldsymbol{x} \Leftrightarrow (x'_i \geq x_i \ \forall i) \wedge (\exists j : x'_j > x_j)$.
Die Beziehung zum SMDD-Konzept für zwei Variablen ist im folgenden Lemma festgehalten.

Lemma 2.3.6. *Eine Funktion $f(\boldsymbol{x}, y)$ ist genau dann eSMDD in y, wenn f, betrachtet als Funktion von (x_i, y) bei festen Werten $\boldsymbol{x}_{-i}$, SMDD ist für alle $i = 1 \dots p-1$.*

Beweis. Die Funktionsschreibweise $g_{\boldsymbol{x}_{-i}}(x_i, y)$ wird im Folgenden zur Bezeichnung der zweiargumentigen Funktion verwendet, die entsteht, wenn f bei fixierten Werten $\boldsymbol{x}_{-i}$ lediglich als Funktion der i-ten und letzten Komponente aufgefasst wird.
„⇒:“
Sei f eSMDD in y, sei $i \in \{1, \dots p-1\}$ und sei $\boldsymbol{x}_{-i}$ fixiert. Sei ferner $\boldsymbol{x'} := \boldsymbol{x} + (x'_i - x_i)\boldsymbol{e}_i$ mit $x'_i > x_i$. Dann gilt $\boldsymbol{x'} > \boldsymbol{x}$ und eine direkte Anwendung der eSMDD-Eigenschaft zeigt, dass

$$f(\boldsymbol{x'}, y) - f(\boldsymbol{x}, y) = g_{\boldsymbol{x}_{-i}}(x'_i, y) - g_{\boldsymbol{x}_{-i}}(x_i, y)$$

streng monoton fallend in y ist.

„⇐:“
Sei $\boldsymbol{x'} > \boldsymbol{x}$. Man setze $\boldsymbol{z}_0 := \boldsymbol{x}$ und $\boldsymbol{z}_1 := \boldsymbol{x} + (x'_1 - x_1)\boldsymbol{e_1}$. Wenn $\boldsymbol{z}_{i-1}$ entsprechend gewählt ist, und $\boldsymbol{z}_i = \boldsymbol{z}_{i-1} + (x'_i - x_i)\boldsymbol{e}_i$ definiert wird, dann gilt

$$f(\boldsymbol{x'}, y) - f(\boldsymbol{x}, y) = \sum_{i=1}^{p-1} (f(\boldsymbol{z}_i, y) - f(\boldsymbol{z}_{i-1}, y)), \tag{2.5}$$

wobei ein Term der Form $f(\boldsymbol{z}_i, y) - f(\boldsymbol{z}_{i-1}, y)$ - insofern er von Null verschieden ist - aufgrund der SMDD-Eigenschaft streng monoton fallend in y ist. □

Eine einfache Methode zum Nachweis der eSMDD-Eigenschaft, die zugleich die Brücke zur Herleitung des paradoxen Effekts von Hooker u.a. bildet, ist im folgenden Lemma festgehalten:

Lemma 2.3.7. *Sei $f(\boldsymbol{x}, y)$ zweimal stetig differenzierbar in einer offenen Menge O, und sei*

$$\frac{\partial^2 f}{\partial x_i \partial y}(\boldsymbol{x}, y) < 0, \ \forall (\boldsymbol{x}, y) \in O, \ \forall i = 1, \ldots p-1,$$

dann besitzt f die eSMDD-Eigenschaft in y bezüglich jedem achsenparallelen, abgeschlossenen, mindestens zweidimensionalem Rechteck (mit mindestens einer parallel zur y-Achse verlaufenden Kante) in O.

Beweis. Das Rechteck R sei durch $R := [x_1, x'_1] \times [x_2, x'_2] \times \cdots \times [x_{p-1}, x'_{p-1}] \times [y_1, y_2]$ gegeben. In offensichtlicher Schreibweise bezeichne $\boldsymbol{x'}$ den Vektor bestehend aus den ersten $p-1$ Obergrenzen und analog $\boldsymbol{x}$ den Vektor aus den ersten $p-1$ Untergrenzen. Ferner sei für mindestens ein $i \in \{1, \ldots p-1\}$ die Obergenze x'_i echt größer als die Untergrenze x_i. Dann gilt $\boldsymbol{x'} > \boldsymbol{x}$ (im Sinne der partiellen Ordnung im $\mathbb{R}^{p-1}$). Mit der Notation und den Definitionen des Beweises von Lemma 2.3.6 folgt:

$$f(\boldsymbol{x'}, y) - f(\boldsymbol{x}, y) = \sum_{i=1}^{p-1} (f(\boldsymbol{z}_i, y) - f(\boldsymbol{z}_{i-1}, y)), \tag{2.6}$$

Da die in R liegenden Punkte $(\boldsymbol{z}_i, y)$, $(\boldsymbol{z}_{i-1}, y)$ $(y \in [y_1, y_2])$ sich höchstens in der i-ten Koordinate unterscheiden, lässt sich Korollar 2.3.1 auf die Funktion anwenden, die entsteht, wenn f - bei festen Werten der übrigen Komponenten - lediglich als Funktion der i-ten und letzten Komponente betrachtet wird. Folglich ist jeder Term der Teleskopsumme (2.6) streng monoton fallend in y $(\forall\, y \in [y_1, y_2])$ oder 0. Da jedoch mindestens ein i existiert, für das $\boldsymbol{z}_i \neq \boldsymbol{z}_{i-1}$ gilt, folgt die Behauptung. □

Analog zu Lemma 2.3.4 lässt sich eine Beziehung zwischen dem eSMDD-Konzept und der (vektorwertigen) bedingten Maximumsabbildung $\hat{\boldsymbol{x}}(y) := \arg\max_{\boldsymbol{x}} f(\boldsymbol{x}, y)$ herstellen, mit deren Hilfe anschließend das höherdimensionale Pendant zu Satz 2.3.1 bewiesen werden kann.

Lemma 2.3.8. *Sei $\hat{\boldsymbol{x}}(y)$ die bedingte Maximumsabbildung von $f(\boldsymbol{x}, y)$. Wenn $f(\boldsymbol{x}, y)$ die eSMDD-Bedingung bezüglich y aufweist, und wenn y_1, y_2 zwei geordnete Werte ($y_1 < y_2$) bezeichnen, für die $\hat{\boldsymbol{x}}(y_1) \neq \emptyset \wedge \hat{\boldsymbol{x}}(y_2) \neq \emptyset$ ist, dann gilt für beliebiges $\boldsymbol{x_2} \in \hat{\boldsymbol{x}}(y_2)$ und $\boldsymbol{x_1} \in \hat{\boldsymbol{x}}(y_1)$ genau eine der folgenden Aussagen:*

i) $\boldsymbol{x_2} = \boldsymbol{x_1}$

ii) $\exists\, i : x_{2,i} < x_{1,i}$

Beweis. Sei $\boldsymbol{x_2} \neq \boldsymbol{x_1}$. Angenommen für alle i gelte $x_{2,i} \geq x_{1,i}$, dann wären die Vektoren $\boldsymbol{x_2}$ und $\boldsymbol{x_1}$ partiell geordnet und eine Anwendung der eSMDD Bedingung führte auf

$$0 \leq f(\boldsymbol{x_2}, y_2) - f(\boldsymbol{x_1}, y_2) < f(\boldsymbol{x_2}, y_1) - f(\boldsymbol{x_1}, y_1).$$

Aus letzterer Ungleichung folgt

$$f(\boldsymbol{x_2}, y_1) - f(\boldsymbol{x_1}, y_1) > 0,$$

was einen Widerspruch zur Voraussetzung $\boldsymbol{x_1} \in \hat{\boldsymbol{x}}(y_1)$ darstellt.
Es sei ferner bemerkt, dass die eSMDD-Eigenschaft in Gegenwart einer eindeutigen bedingten Maximumsabbildung (genauer: $\hat{\boldsymbol{x}}(y_1)$ und $\hat{\boldsymbol{x}}(y_2)$ seien einelementig) durch die schwächere eMDD-Annahme ersetzt werden kann, ohne die Gültigkeit des Lemmas zu beeinträchtigen. □

Satz 2.3.2. *Sei $g(\boldsymbol{x}, y)$ eSMDD in y und sei $h(y)$ eine streng monoton wachsende Funktion von y. Wenn $(\boldsymbol{x_f}, y_f), (\boldsymbol{x_g}, y_g)$ Stellen eines Maximums von $f(\boldsymbol{x}, y) := g(\boldsymbol{x}, y) + h(y)$ bzw. von $g(\boldsymbol{x}, y)$ bezeichnen und wenn $y_f \neq y_g$ ist, dann gilt $x_{f,i} < x_{g,i}$ für (mindestens) eine Komponente i, insofern $\boldsymbol{x}_f \neq \boldsymbol{x}_g$ ist.*

Der Beweis dieses Satzes verläuft analog zum Beweis von Satz 2.3.1. Ähnlich wie im bereits skizzierten zweidimensionalen Fall erlaubt das Hinzufügen eines ausschließlich die letzte Dimension θ_p erfassenden Items die Konstruktion des paradoxen Effekts, insofern die Loglikelihood $l_u(\boldsymbol{\theta}_{-p}, \theta_p)$ eSMDD in θ_p ist. Das hinzugefügte Item leistet

im Fall eines binären Antwortformats einen streng monoton wachsenden (fallenden) Beitrag zur Loglikelihood, wenn es richtig (falsch) beantwortet wurde. Im Fall eines ordinalen Antwortformats wird der gleiche Zweck von den Extremkategorien erfüllt. Diese direkte Hinzufügung eines monotonen Beitrags ist jedoch nicht zwingend erforderlich, um den paradoxen Effekt zu garantieren. In Abhängigkeit des spezifisch vorliegenden IRT-Modells lassen sich auch Kategorien vergleichen, die kein extremes Antwortverhalten (richtig vs. falsch, höchster Score vs niedrigster Score) widerspiegeln. Der entscheidende Punkt liegt in der (strikten) Monotonie des Quotienten, d.h. solange die beiden zu vergleichenden Itemscores ($u'_{k+1} > u_{k+1}$) auf dem hinzugefügten Item gewähren, dass $l_{u'_{k+1}}(\theta_p) - l_{u_{k+1}}(\theta_p)$ streng monoton wachsend in θ_p ist, ist die Konstruktion des Paradoxons (unter eSMDD) möglich. Dies lässt sich durch die Wahl von $g := l_u + l_{u_{k+1}}$ und $h := l_{u'_{k+1}} - l_{u_{k+1}}$ in Satz 2.3.2 verdeutlichen.
Somit kann der Kern der vorangegangenen Betrachtungen wie folgt wiedergegeben werden: Wenn die Differenz zweier Loglikelihood-Funktionen $l_{u'}(\boldsymbol{\theta}) - l_u(\boldsymbol{\theta})$ eine streng monoton wachsende Funktion von θ_p ist und wenn l_u eSMDD in θ_p ist, dann gibt es mindestens eine Komponente $i \neq p$, bezüglich derer der ML-Schätzer basierend auf u nicht kleiner ausfällt als der ML-Schätzer basierend auf u'. Des Weiteren zeigt der Beweis von Satz 2.3.1, dass der Schätzer *echt* größer ausfällt, wenn die Zusatzannahme einer injektiven bedingten Maximumsfunktion eingeführt wird.
Die restriktivste der Konstruktion zugrundeliegende Annahme ist die Forderung, dass $l_{u'}(\boldsymbol{\theta}) - l_u(\boldsymbol{\theta})$ lediglich von θ_p abhängt. Im Folgenden soll daher dargelegt werden, wie diese Annahme umgangen bzw. durch eine wesentlich schwächere Annahme ersetzt werden kann. Dabei erfolgt eine Fallunterscheidung je nachdem, ob der Schätzer als Nullstelle der Scorefunktion (wie in der Vorgehensweise bei Hooker u.a., 2009) oder als Stelle eines Maximums der Likelihood (wie in der Satz 2.3.1 zugrundegelegten Konstruktion) betrachtet wird.

Parametrische Abhängigkeit des paradoxen Effekts

Die bisherige „Konstruktion“ des paradoxen Effekts beruhte maßgeblich auf der Existenz eines eindimensionalen Items. Im Falle eines M2PL-Modells[6], dessen Item-

[6]Die (S)MDD-Eigenschaft des M2PL-Modells mit nichtnegativen Diskriminationsparametern wird in Anhang D nachgewiesen.

Response-Funktionen die Gestalt

$$P(U_i = 1|\boldsymbol{\theta}) = \frac{\exp(\boldsymbol{a_i^T\theta} + \beta_i)}{1 + \exp(\boldsymbol{a_i^T\theta} + \beta_i)}$$

aufweisen, impliziert dies $\boldsymbol{a}_{k+1}^{\boldsymbol{T}} = \boldsymbol{e}_p$. Letzteres stellt eine äußerst restriktive Forderung dar. Die folgenden Betrachtungen werden jedoch verdeutlichen, dass der paradoxe Effekt auch dann Bestand hat, wenn anstelle des Einheitsvektors $\boldsymbol{e}_p$ ein beliebiger Diskriminationsvektor $\boldsymbol{a}_{k+1}$ in einer ϵ-Nachbarschaft von $\boldsymbol{e}_p$ vorliegt.
Auch wenn diese „Stetigkeit“ bezüglich der Itemparameter des hinzugefügten Items die primäre Motivation der folgenden Betrachtungen darstellt, so umschließt die nachfolgende Argumentation auch den Fall, in dem beliebige, nicht zwingend auf das letzte Item beschränkte Itemparameter verändert werden. Als Nebenprodukt wird dies die Begründung des paradoxen Effekts außerhalb der Klasse der SMDD-Modelle ermöglichen[7].
In Abweichung zur bisherigen Notation bezeichne im Folgenden $f(\boldsymbol{\gamma}, \boldsymbol{\theta})$ eine generelle Schätzfunktion, d.h. eine Funktion $f : \mathbb{R}^l \times \mathbb{R}^p \mapsto \mathbb{R}^p$, die implizit durch die Forderung $f(\boldsymbol{\gamma}, \boldsymbol{\theta}) = \mathbf{0}$ eine mengenwertige Abbildung $\boldsymbol{\theta}(\boldsymbol{\gamma}) := \{\boldsymbol{\theta}|\, f(\boldsymbol{\gamma}, \boldsymbol{\theta}) = \mathbf{0}\}$ definiert, die an der Stelle $\boldsymbol{\gamma}_0$ den momentanen Schätzwert, d.h. den Schätzwert $\boldsymbol{\theta}_0$, den man beim vorliegenden Wert $\boldsymbol{\gamma}_0$ des Itemparameters erhält, beinhaltet.
Diese Abweichung von der bisherigen Notation wurde zum Einen vorgenommen, um der nun im Fokus stehenden parametrischen Abhängigkeit des Schätzers von Itemkenngrößen $\boldsymbol{\gamma}$ Rechnung zu tragen, zum Anderen, um neben der ML-basierten Scoregleichung auch generelle Schätzgleichungen (GEEs; Zeger und Liang, 1986) in die Betrachtungen einzuschließen.
Wenn für einen bestimmten Wert $\boldsymbol{\gamma}_0$ eine Lösung $\boldsymbol{\theta}_0$ existiert (d.h. $\boldsymbol{\theta}(\boldsymbol{\gamma}_0) \neq \emptyset$), dann stellt sich die Frage, ob für in einer Nachbarschaft von $\boldsymbol{\gamma_0}$ liegende Werte $\boldsymbol{\gamma}$ ebenfalls Lösungen $\boldsymbol{\theta}(\boldsymbol{\gamma})$ existieren, die „beliebig nahe“ an $\boldsymbol{\theta}_0$ liegen. Diese Frage kann positiv beantwortet werden, wenn die implizit definierte Lösungsabbildung $\boldsymbol{\theta}(\boldsymbol{\gamma})$ eine stetige, um $\boldsymbol{\gamma_0}$ definierte, einelementige Lokalisation[8] $\hat{\boldsymbol{\theta}}(\boldsymbol{\gamma})$ besitzt, d.h. wenn eine Auswahl von Schätzwerten stetig von dem Itemparameter abhängt. Wenn dies der Fall ist, und wenn $\theta_{0,i} > v$ gilt, dann gilt dieselbe strikte Ungleichung für die aus einer gewissen Nachbarschaft stammenden Werte $\hat{\theta}_i(\boldsymbol{\gamma})$ ($\boldsymbol{\gamma} \in \mathbb{B}_\epsilon(\boldsymbol{\gamma}_0)$). Überträgt man diese Argumentation auf die zweite Schätzgleichung, d.h. auf die Schätzgleichung, die entsteht, wenn die Antwort auf dem hinzugefügten Item verändert wird, dann zeigt

[7] Ein empirisches Beispiel für einen solchen Fall geben Finkelman, Hooker und Wang (2010).
[8] Zur Bedeutung des Begriffs Lokalisation siehe Anhang A.

dies, dass unter der obigen, informell skizzierten Vorgehensweise der paradoxe Effekt Bestand behält, solange ein Itemparameter $\boldsymbol{\gamma}$ aus einer bestimmten Nachbarschaft von $\boldsymbol{\gamma}_0$ gewählt wird.
Das Ziel dieses Abschnitts ist somit erreicht, wenn Bedingungen an die durch f repräsentierte Schätzgleichung gefunden werden können, die die stetige Abhängigkeit einer Lösungsfunktion von den Itemparametern garantieren. Nachfolgend werden drei Resultate aus dem Bereich impliziter Funktionen wiedergegeben, die diese Stetigkeit - und somit den Erhalt des paradoxen Effekts - garantieren.
Der klassische Satz über implizite Funktionen (Dini, 1877), der im Folgenden notationell in Anlehnung an Dontchev und Rockafellar (2009) wiedergegeben wird, bildet die Grundlage des ersten Resultats über die Existenz einer stetigen Lösungsfunktion[9].

Satz 2.3.3 (Klassischer Satz über implizite Funktionen)**.** *Sei f in einer Nachbarschaft des Punktes $(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)$ stetig differenzierbar und sei $f(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0) = \mathbf{0}$. Wenn die partielle Jacobimatrix $\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)$ invertierbar ist, dann existiert eine stetige einelementige Lokalisation $s(\boldsymbol{\gamma})$ (der zugehörigen Lösungsabbildung $\boldsymbol{\theta}(\boldsymbol{\gamma})$) um den Punkt $\boldsymbol{\gamma}_0$ für den Punkt $\boldsymbol{\theta}_0$, die zudem stetig differenzierbar mit zugehöriger Jacobimatrix*

$$\nabla s(\boldsymbol{\gamma}) = -\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\gamma}, s(\boldsymbol{\gamma}))^{-1} \nabla_{\boldsymbol{\gamma}} f(\boldsymbol{\gamma}, s(\boldsymbol{\gamma})) \tag{2.7}$$

ist.

Bemerkungen: 1) Die Notation $s(\boldsymbol{\gamma})$ soll verdeutlichen, dass die so entstandene Lokalisation nicht zwingend identisch mit der tatsächlich verwendeten Auswahl $\hat{\boldsymbol{\theta}}(\boldsymbol{\gamma})$ ist (siehe hierzu auch die generellen Bemerkungen zum Begriff der Lokalisation aus Anhang A). 2) Die Invertierbarkeit der partiellen Jacobimatrix ist im Falle einer Scoregleichung identisch mit der Invertierbarkeit der beobachteten Fisherinformationsmatrix.
Auch wenn der klassische Satz über implizite Funktionen ausreicht, um den paradoxen Effekt für eine Vielzahl an IRT-Modellen zu verallgemeinern, so können diverse Bestandteile - insbesonders die stetige Differenzierbarkeit von f - unnötige Restriktionen darstellen.
Eine erste weniger restriktive Version bietet der Satz von Goursat (Goursat, 1903;

[9]Im Folgenden gelte stets $(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0) \in \text{int dom} f$, wobei int dom$f$ die Menge der inneren Punkte von dom$f := \{(\boldsymbol{\gamma}, \boldsymbol{\theta}) | f(\boldsymbol{\gamma}, \boldsymbol{\theta}) \neq \emptyset\}$ bezeichnet.

siehe auch Dontchev und Rockafellar, 2009, S.20), der in seinem Bezug zur obigen Notation die folgende Gestalt aufweist:

Satz 2.3.4 (Goursat's Satz über implizite Funktionen)**.** *Sei f in einer Nachbarschaft des Punktes $(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)$, der die Gleichung $f(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0) = \mathbf{0}$ erfüllt, differenzierbar in $\boldsymbol{\theta}$, und seien innerhalb dieser Nachbarschaft sowohl $f(\boldsymbol{\gamma}, \boldsymbol{\theta})$ als auch $\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\gamma}, \boldsymbol{\theta})$ stetig. Wenn die partielle Jacobimatrix $\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)$ invertierbar ist, dann existiert eine im Punkt $\boldsymbol{\gamma}_0$ stetige, einelementige Lokalisation $s(\boldsymbol{\gamma})$ (der zugehörigen Lösungsabbildung $\boldsymbol{\theta}(\boldsymbol{\gamma})$) um den Punkt $\boldsymbol{\gamma}_0$ für den Punkt $\boldsymbol{\theta}_0$.*

Trotz der Abwesenheit der restriktiven Forderung nach stetiger Differenzierbarkeit von f in Satz 2.3.4 ist es noch nicht gelungen, auf entsprechende Glattheits- bzw. Differenzierbarkeitsforderungen in $\boldsymbol{\theta}$ verzichten zu können. Der nachfolgende Satz ermöglicht dies durch die Einführung des Konzepts eines strikten Schätzers (*strict estimator*; siehe Dontchev und Rockafellar, 2009, S.36 f. u. S.45):

Definition 2.3.3 (gleichmäßig strikter Schätzer)**.** Eine Funktion $g(\boldsymbol{\theta})$ ist ein $\boldsymbol{\gamma}$-gleichmäßiger strikter Schätzer der Funktion $f(\boldsymbol{\gamma}, \boldsymbol{\theta})$ an der Stelle $(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)$ (*strict estimator of f with respect to $\boldsymbol{\theta}$ uniformly in $\boldsymbol{\gamma}$*), wenn die folgenden Bedingungen erfüllt sind:

i) $\boldsymbol{\theta}_0 \in$ int dom g

ii) $g(\boldsymbol{\theta}_0) = f(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)$.

iii) $\exists \alpha > 0$, $\exists \delta_1 > 0$, $\exists \delta_2 > 0$:

$$|(f(\boldsymbol{\gamma}, \boldsymbol{\theta'}) - g(\boldsymbol{\theta'})) - (f(\boldsymbol{\gamma}, \boldsymbol{\theta}) - g(\boldsymbol{\theta}))| \leq \alpha |\boldsymbol{\theta'} - \boldsymbol{\theta}| \quad \forall \boldsymbol{\theta}, \boldsymbol{\theta'} \in \mathbb{B}_{\delta_1}(\boldsymbol{\theta}_0),\ \boldsymbol{\gamma} \in \mathbb{B}_{\delta_2}(\boldsymbol{\gamma}_0)$$

Die größte untere Schranke der Menge aller α, für die *iii*) gilt, heisst gleichmäßiger Lipschitzmodulus von $f - g$ an der Stelle $(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)$ und wird mit $\widehat{\mathrm{lip}}_{\boldsymbol{\theta}}(f - g, (\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0))$ bezeichnet.

Mithilfe dieser Terminologie kann nun ein wesentlich mächtigerer Satz zur stetigen Abhängigkeit implizit definierter Funktionen wiedergegeben werden:

Satz 2.3.5 (Theorem 2B.5, Dontchev und Rockafellar, 2009)**.** *Sei $f(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0) = 0$. Unter den Annahmen, dass*

a) $f(\cdot, \boldsymbol{\theta}_0)$ *stetig an der Stelle* $\boldsymbol{\gamma}_0$ *ist;*

b) g *ein* $\boldsymbol{\gamma}$*-gleichmäßiger strikter Schätzer von* f *an der Stelle* $(\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)$ *mit* $\widehat{lip}_{\boldsymbol{\theta}}(f - g, (\boldsymbol{\gamma}_0, \boldsymbol{\theta}_0)) = \mu$ *ist;*

c) g^{-1} *eine Lipschitz-stetige, einelementige Lokalisation* σ *um den Punkt* $\mathbf{0}$ *für* $\boldsymbol{\theta}_0$ *mit Lipschitzmodulus* $lip(\sigma, \mathbf{0}) \leq \kappa$ *für eine Konstante* κ *mit der Eigenschaft* $\kappa \cdot \mu < 1$ *aufweist,*

existiert eine im Punkt $\boldsymbol{\gamma}_0$ *stetige, einelementige Lokalisation* $s(\boldsymbol{\gamma})$ *der zugehörigen Lösungsabbildung* $\boldsymbol{\theta}(\boldsymbol{\gamma}) := \{\boldsymbol{\theta} | f(\boldsymbol{\gamma}, \boldsymbol{\theta}) = \mathbf{0}\}$ *um den Punkt* $\boldsymbol{\gamma}_0$ *für den Punkt* $\boldsymbol{\theta}_0$*.*

Bemerkung: Die Sätze 2.3.4 und 2.3.3 können als Spezialfall des Satzes 2.3.5 angesehen werden, indem die durch die Jacobimatrix gegebene affine Funktion die Rolle von g übernimmt.
Da Satz 2.3.5 im Vergleich zum klassischen Satz über implizite Funktionen auf Differenzierbarkeitsforderungen verzichtet, stellt er ein geeignetes Pendant zu der im vorherigen Abschnitt vorgenommenen Verallgemeinerung des paradoxen Effekts, die im Gegensatz zur Herleitung von Hooker u.a. (2009) keine Forderungen bezüglich Differenzierbarkeit stellt, dar.
Die drei dargelegten Sätze über implizite Funktionen „sichern“ somit unter den jeweils angegebene Bedingungen den Erhalt des paradoxen Effekts, wenn anstelle des Parameters $\boldsymbol{\gamma}_0$ ein Parameterwert aus einer bestimmten Nachbarschaft gewählt wurde. Insbesonders zeigt dies für das einführend erwähnte Beispiel eines M2PL-Modells, dass auf die strikte Forderung eines eindimensionalen Items verzichtet werden kann: Alle Diskriminationsvektoren $\boldsymbol{a}_{k+1}$, die in einer bestimmten ϵ-Nachbarschaft des Einheitsvektors $\boldsymbol{e}_p$ liegen, ermöglichen die Konstruktion des paradoxen Effekts.
Neben dieser auf das letzte Item fokussierten Fragestellung ist es jedoch auch mittels der rezipierten Sätze möglich, den paradoxen Effekt außerhalb der Klasse der SMDD-MIRT-Modelle aufzuzeigen. Ein interessantes Beispiel für Letzteres ist durch ein $M2PL$-Modell mit zusätzlichen Rateparametern, welches bereits als Grundlage einer *empirischen* Analyse des paradoxen Effekts fungierte (Finkelman u.a., 2010), gegeben:

$$P(U_i = 1 | \boldsymbol{\theta}) = \gamma_i + (1 - \gamma_i) \cdot \frac{\exp(\boldsymbol{a_i^T \theta} + \beta_i)}{1 + \exp(\boldsymbol{a_i^T \theta} + \beta_i)}$$

Gemäß den obigen Erörterungen über die Existenz stetiger Lokalisationen ist somit die Konstruktion des paradoxen Effekts innerhalb dieser Modellklasse für $\boldsymbol{\gamma^T} =$

$(\gamma_1, \ldots, \gamma_k) \in \mathbb{B}_\epsilon(\mathbf{0})$ möglich, obwohl die zugehörige Loglikelihood-Funktion nicht global die SMDD Eigenschaft aufweist.

Die bisherige Diskussion der parametrischen Abhängigkeit des paradoxen Effekts fand im Rahmen von Schätzgleichungen statt. Auch wenn dies mit der in der Praxis durchgeführten Berechnung eines Schätzers korrespondiert[10], so ist die ursprüngliche, im Kontext des ML-Schätzers dargelegte Motivation des Schätzers als Stelle eines Maximums verlorengegangen. Die abschließenden Betrachtungen dieses Abschnitts sind daher der Diskussion der parametrischen Abhängigkeit im Kontext von Schätzern, die als Stelle des Maximums einer Funktion betrachtet werden können, gewidmet.

Im Folgenden bezeichne $f_1(\boldsymbol{\gamma}, \boldsymbol{\theta})$ die bezüglich $\boldsymbol{\theta}$ zu maximierende Funktion (mit Parameter $\boldsymbol{\gamma}$) korrespondierend zum ersten Szenario (höherer Itemscore auf dem hinzugefügten Item). Analog sei für das Szenario eines geringeren Itemscores $f_2(\boldsymbol{\gamma}, \boldsymbol{\theta})$ definiert. Im Kern des paradoxen Effekts steht das Resultat, dass für einen spezifischen Wert $\boldsymbol{\gamma}_0$ - derjeniege Wert, der die Eindimensionalität des hinzugefügten Items impliziert - eine Komponente θ_i existiert, deren Schätzwerte „verkehrt“ geordnet sind, d.h. die Beziehung $\hat{\theta}_i^{f_1}(\boldsymbol{\gamma}_0) < \hat{\theta}_i^{f_2}(\boldsymbol{\gamma}_0)$ erfüllen. Die Untersuchung, unter welchen Bedingungen diese strikte Ungleichung für Parameterwerte $\boldsymbol{\gamma} \neq \boldsymbol{\gamma}_0$ Gültigkeit behält, ist Gegenstand der folgenden Betrachtungen.

Lemma 2.3.9. *Seien (in obiger Notation) f_1 sowie f_2 usc[11] Funktionen ($\mathbb{R}^l \times \mathbb{R}^p \mapsto \mathbb{R} \cup \{-\infty\}$) mit jeweils eindeutigem bedingten Maximum $\hat{\boldsymbol{\theta}}^{f_1}(\boldsymbol{\gamma_0})$ bzw. $\hat{\boldsymbol{\theta}}^{f_2}(\boldsymbol{\gamma_0})$ an der Stelle $\boldsymbol{\gamma}_0$. Wenn f_1 und f_2 $\boldsymbol{\gamma}$-gleichmäßig niveaubeschränkt in $\boldsymbol{\theta}$ sind und wenn (für $i = 1, 2$) $g_i(\boldsymbol{\gamma}) := f_i(\boldsymbol{\gamma}, \hat{\boldsymbol{\theta}}^{f_i}(\boldsymbol{\gamma_0}))$ stetig an der Stelle $\boldsymbol{\gamma}_0$ ist, dann gibt es eine ϵ-Umgebung von $\boldsymbol{\gamma}_0$, so dass für beliebiges $\boldsymbol{\gamma} \in \mathbb{B}_\epsilon(\boldsymbol{\gamma}_0)$ und jedes Paar $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ mit $\boldsymbol{\theta}_1 \in \hat{\boldsymbol{\theta}}^{f_1}(\boldsymbol{\gamma})$, $\boldsymbol{\theta}_2 \in \hat{\boldsymbol{\theta}}^{f_2}(\boldsymbol{\gamma})$ die unter $\boldsymbol{\gamma}_0$ bestehende strikte Ungleichung $\hat{\theta}_i^{f_1}(\boldsymbol{\gamma}_0) < \hat{\theta}_i^{f_2}(\boldsymbol{\gamma}_0)$ erhalten bleibt.*

Beweis. Sei $\hat{\theta}_i^{f_1}(\boldsymbol{\gamma}_0) < c < \hat{\theta}_i^{f_2}(\boldsymbol{\gamma}_0)$. Gemäß den Ausführungen des Anhangs A (Abschnitt: Maximum einer Funktion) ergeben die usc-Eigenschaft von f_1 sowie die „lokale“ Stetigkeit von $g_1(\boldsymbol{\gamma})$ zusammen mit der gleichmäßigen Niveaubeschränktheit

[10] Die Berechnung eines Schätzers erfolgt in der Praxis meist mittels Algorithmen, die eine Nullstelle einer Funktion suchen. Dies gilt insbesonders auch für den Fall eines als Stelle eines Maximums definierten Schätzers.

[11] Es wird stillschweigend angenommen, dass für f_i eine offene Menge O_i um den Punkt $(\boldsymbol{\gamma}_0, \hat{\boldsymbol{\theta}}^{f_i}(\boldsymbol{\gamma_0}))$ existiert, so dass f_i auf O_i endliche Werte annimmt.

die folgende Aussage (siehe Theorem 1.17 aus Rockafellar und Wets, 2009):
Für jede gegen $\boldsymbol{\gamma}_0$ konvergierende Folge $(\boldsymbol{\gamma}_k)_{k\in\mathbb{N}}$ liegt jeder Häufungspunkt einer beliebigen Auswahlfolge $(\boldsymbol{\theta}_k \in \hat{\boldsymbol{\theta}}^{f_1}(\boldsymbol{\gamma}_k))_{k\in\mathbb{N}}$ in $\hat{\boldsymbol{\theta}}^{f_1}(\boldsymbol{\gamma}_0)$. Dies impliziert die Existenz einer Umgebung $\mathbb{B}_{\epsilon'}$ von $\boldsymbol{\gamma}_0$, für die $\hat{\theta}_i^{f_1}(\boldsymbol{\gamma}) < c$ ist (unabhängig davon, welches Element aus $\hat{\boldsymbol{\theta}}^{f_1}(\boldsymbol{\gamma})$ gewählt wurde.). Gäbe es nämlich eine solche Umgebung nicht, so wähle man für jedes $\epsilon_n := \frac{1}{n}$ ein $\boldsymbol{\gamma}_n \in \mathbb{B}_{\epsilon_n}(\boldsymbol{\gamma}_0)$ und ein $\boldsymbol{\theta}_n \in \hat{\boldsymbol{\theta}}^{f_1}(\boldsymbol{\gamma}_n)$ mit $\hat{\theta}_{n,i} \geq c$. Aufgrund der Wahl von ϵ_n konvergiert $\boldsymbol{\gamma}_n$ gegen $\boldsymbol{\gamma}_0$. Jeder Häufungspunkt (die Existenz eines solchen ist gemäß Theorem 1.17 b aus Rockafellar und Wets (2009) gesichert.) der Folge $\boldsymbol{\theta}_n$ liegt demnach in $\hat{\boldsymbol{\theta}}^{f_1}(\boldsymbol{\gamma_0})$. Für die entsprechende Teilfolge $\boldsymbol{\theta}_{n_v}$ gilt $\hat{\theta}_{n_v,i} \geq c$ für alle v, so dass sich diese Ungleichung auch auf ihren, in $\hat{\boldsymbol{\theta}}^{f_1}(\boldsymbol{\gamma}_0)$ liegenden Limes (bzw. dessen i-te Komponente) überträgt. Da $\hat{\boldsymbol{\theta}}^{f_1}(\boldsymbol{\gamma_0})$ jedoch nur aus einem Element besteht und da dieses Element gemäß der Voraussetzung des Lemmas $\hat{\theta}_i^{f_1}(\boldsymbol{\gamma}_0) < c$ erfüllt, führt dies auf einen Widerspruch.
Eine analoge Argumentation führt auf die Existenz einer Umgebung $\mathbb{B}_{\epsilon''}(\boldsymbol{\gamma}_0)$ in der die Ungleichung $\hat{\theta}_i^{f_2}(\boldsymbol{\gamma}) > c$ gültig ist. Für $\epsilon := \min(\epsilon', \epsilon'')$ gelten folglich beide Abschätzungen und die Behauptung des Lemmas ist demnach bewiesen. □

Die ursprünglich äußerst restriktiv anmutende Bedingung eines ausschließlich eine Dimension messenden Items kann somit mittels Lemma 2.3.9 deutlich abgeschwächt werden, so dass der Kern der Konstruktion des paradoxen Effekts primär aus der (S)MDD-Bedingung zu bestehen scheint.
Im folgenden Kapitel soll diese entscheidende Bedingung, die bisher lediglich in einem rein formalen Sinne eingeführt und zur Herleitung der entsprechenden Resultate verwendet wurde, näher auf ihren (statistischen) Bedeutungsgehalt untersucht werden. Über eine einfache Beziehung zu einem bekannten Assoziationskonzept für Zufallsvariablen wird dies darüber hinaus ermöglichen, den paradoxen Effekt in der Klasse der „Typ-2-Schätzer“ zu etablieren.

2.4 Statistische Bedeutung der „MDD“-Bedingung

Die der Konstruktion des paradoxen Effekts zugrundeliegende MDD-Bedingung soll in diesem Abschnitt näher im Hinblick auf ihre statistische Bedeutung untersucht werden. Eine einfache, in Lemma 2.4.1 darzulegende Umformung des MDD-Konzepts wird eine Verbindung zu einem zentralen Assoziationskonzept für Zufallsvariablen

herstellen. Zunächst soll jedoch kurz das relevante Assoziationskonzept erörtert werden.

Das zu betrachtende Assoziationskonzept beruht auf den Begriffen MTP_2 (*multivariate totally positive*) und MRR_2 (*multivariate reverse rule*), die jeweils Eigenschaften einer (nichtnegativen) Funktion $f(\boldsymbol{x})$ von mehreren Komponenten ($\boldsymbol{x}$) beschreiben.

Eine nichtnegative Funktion f heisst MRR_2 (siehe z.B. Karlin und Rinott, 1980b), wenn für zwei Vektoren $\boldsymbol{x}, \boldsymbol{x'}$ stets

$$f(\boldsymbol{x} \wedge \boldsymbol{x'}) f(\boldsymbol{x} \vee \boldsymbol{x'}) \leq f(\boldsymbol{x}) f(\boldsymbol{x'}) \tag{2.8}$$

gilt. Hierin beschreibt $\boldsymbol{x} \wedge \boldsymbol{x'}$ den Vektor, dessen i-te Komponente durch $\min(x_i, x_i')$ gegeben ist und $\boldsymbol{x} \vee \boldsymbol{x'}$ den Vektor, dessen i-te Komponente durch $\max(x_i, x_i')$ gegeben ist.[12]

Eine nichtnegative Funktion f heisst MTP_2 (Karlin und Rinott, 1980a), wenn (2.8) mit „$\geq$“ anstelle von „$\leq$“ gilt. Im Spezialfall einer Funktion von zwei Variablen wird MTP_2 mit TP_2 (*totally positive*) bzw. MRR_2 mit RR_2 (*reverse rule*) bezeichnet. Diese Spezialfälle sind, wie das folgende Lemma zeigen wird, für die Untersuchung des paradoxen Effekts von zentralem Interesse.

Lemma 2.4.1. *Die Funktion $f(x, y)$ ist genau dann MDD, wenn die Funktion $g(x, y) := \exp(f(x, y))$ RR_2 ist.*

Beweis. MDD bedeutet, dass für $x' > x$ und $y' > y$

$$f(x', y') - f(x, y') \leq f(x', y) - f(x, y)$$

gilt. Dies ist gleichbedeutend zu

$$\exp(f(x', y')) \exp(f(x, y)) \leq \exp(f(x', y)) \exp(f(x, y')).$$

□

Die zur Konstruktion des paradoxen Effekts entscheidende Forderung einer MDD-Loglikelihood ist somit äquivalent zur Forderung einer RR_2 Likelihood. Letztere

[12] Für die Wohldefiniertheit der MRR_2-Eigenschaft ist es - ähnlich wie im Falle des MDD-Konzepts - nicht zwingend erforderlich, dass die einzelnen Komponenten aus $\mathbb{R}$ stammen. Jede geordnete Menge kann anstelle der reellen Zahlen treten.

Eigenschaft kann in Form eines „Wettszenarios“ wie folgt interpretiert werden:
Bezeichne $L_u(\theta_1, \theta_2)$ eine RR_2-Likelihood eines MIRT-Modells. Seien ferner für jede Dimension zwei geordnete Fähigkeitsausprägungen vorgegeben: $\Theta_1 := \{\theta_1, \theta_1'\}$ $(\theta_1 < \theta_1')$, $\Theta_2 := \{\theta_2, \theta_2'\}$ $(\theta_2 < \theta_2')$. Wenn die Vektoren (θ_1', θ_2'), (θ_1, θ_2) die „fähigste“ bzw. „unfähigste“ Person innerhalb dieser Vorgabe beschreiben, dann ist die Wahrscheinlichkeit, dass beide Personen[13] das gleiche Antwortmuster u aufweisen, niemals größer als bei zwei Personen „mittlerer“ bzw. „balancierter“ $((\theta_1', \theta_2), (\theta_1, \theta_2'))$ Fähigkeiten. Wenn eine Wette darauf abgeschlossen werden soll, welches der beiden Paare („Extrempaar“ vs. „Trade-Off-Paar“) eher in der Lage ist, das gleiche Antwortmuster u hervorzubringen, dann ist unter RR_2 somit das „balancierte“ Paar vorzuziehen.
Neben dieser anschaulicheren Interpretation der MDD-Bedingung ermöglicht das Lemma - und hierin liegt sein Hauptnutzen - eine statistische Interpretation der MDD-Bedingung. Zu diesem Zweck wird der unbekannte, zu schätzende Personenparameter (θ_1, θ_2) als Zufallsvektor unter der konstanten (uneigentlichen) Prioriverteilung $p(\theta_1, \theta_2) = c$ betrachtet. Die RR_2-Eigenschaft der Likelihoodfunktion (bzw. die MDD-Eigenschaft von l_u) überträgt sich, da die RR_2-Eigenschaft bei Produktbildung erhalten bleibt (siehe Lemma 2.3.2), direkt auf die Posteriori-Dichte von (θ_1, θ_2). Generell besitzt jede Posteriori-Dichte, die auf einer RR_2-Likelihood sowie einer RR_2-Priori beruht, selber wiederum die RR_2-Eigenschaft. Demnach kann die statistische Interpretation der MDD-Eigenschaft auf die statistische Bedeutung der RR_2-Eigenschaft einer Posterioriverteilung zurückgeführt werden.
Für die statistische Interpretation der RR_2-Eigenschaft stehen zahlreiche Resultate zur Verfügung (siehe z.B: Lehmann, 1966; Karlin und Rinott, 1980b; Lehmann und Romano, 2005, S.70; Marshall, Olkin und Arnold, 2010, Kap.18). Im Folgenden sollen drei Hauptresultate bezüglich der RR_2-Eigenschaft kurz skizziert werden[14].

Lemma 2.4.2. *Sei $l_u(\theta_1, \theta_2) + \gamma(\theta_1, \theta_2)$ MDD. Dann stehen θ_1 und θ_2 a-posteriori in negativer Regressionsabhängigkeit („negative regression dependence“), d.h.*

$$P(\theta_1 > a | \, \theta_2 = b, \, \boldsymbol{u})$$

ist monoton fallend in b für alle a.

[13] Die stochastische Unabhängigkeit der Antwortmuster beider Personen wird dabei stillschweigend unterstellt.

[14] Die Ausdehnung der Resultate auf die SMDD-Eigenschaft kann den entsprechenden Beweisen des Anhangs B entnommen werden.

Beweis. Siehe Lemma B.0.2 aus Anhang B. □

Korollar 2.4.1. *Sei $l_u(\theta_1, \theta_2)+\gamma(\theta_1, \theta_2)$ MDD. Dann stehen θ_1 und θ_2 a-posteriori in negativer Quadrantenabhängigkeit („negative quadrant dependence"), d.h. bezüglich ihrer gemeinsamen Posterioriverteilung gilt:*

$$P(\theta_1 \leq a,\, \theta_2 \leq b|\, \boldsymbol{u}) \leq P(\theta_1 \leq a|\, \boldsymbol{u})P(\theta_2 \leq b|\, \boldsymbol{u}) \quad \forall a, b.$$

Beweis. Siehe Lehmann (1966) oder Anhang B. □

Korollar 2.4.2. *Wenn $l_u(\theta_1, \theta_2)+\gamma(\theta_1, \theta_2)$ MDD ist, dann ist die a-posteriori Kovarianz zwischen θ_1 und θ_2 nichtpositiv - insofern die entsprechenden zweiten Momente endlich sind. Des Weiteren ist die Kovarianz genau dann Null, wenn θ_1 und θ_2 a-posteriori unabhängig sind.*

Beweis. Eine direkte Implikation des Ergebnisses des vorherigen Korollars (siehe auch Lemma B.0.4 aus Anhang B). □

Vor allem die in Lemma 2.4.2 dargelegte stochastische Ordnung erscheint im Hinblick auf die Untersuchung des paradoxen Effekts von besonderem Interesse zu sein. Gemäß der RR_2-Eigenschaft ist θ_1 stochastisch nach θ_2 geordnet, wobei geringere θ_2-Werte Anlass zu stochastisch größeren θ_1-Verteilungen geben. Die bestehende stochastische Ordnung bleibt ferner durch das Hinzufügen eines Items, welches ausschließlich die zweite Dimension misst, erhalten. Da das hinzugefügte Item (nach dem Konstruktionsschema) einen streng monoton wachsenden Beitrag zur Likelihood bzw. Posteriori liefert, ist davon auszugehen, dass θ_2 durch dieses Hinzufügen stochastisch größer ausfällt. In Kombination mit der unveränderten, „gegenläufigen" stochastischen Ordnung der bedingten Verteilung von θ_1 (gegeben θ_2) ergibt sich die Vermutung, dass die Randverteilung von θ_1 durch Hinzufügung des monotonen Beitrags stochastisch kleiner ausfällt als zuvor.
Diese informellen Überlegungen bilden die Grundlage für die Konstruktion des paradoxen Effekts für Typ-2-Schätzer. Sie sind im nachfolgenden Satz formalisiert.

Satz 2.4.1. *Seien $(\Omega_1, \mathscr{F}_1, \mu_1), (\Omega_2, \mathscr{F}_2, \mu_2)$ σ-finite Maßräume mit zugehörigen geordneten Grundmengen (d.h. auf Ω_1 und Ω_2 ist jeweils eine totale Ordnungsrelation definiert). $f(\theta_1, \theta_2)$ bezeichne eine $\mu_1 \times \mu_2$-meßbare RR_2-Dichte eines zweidimensionalen Zufallsvektor (Θ_1, Θ_2).*

Ist $h(\theta_2) > 0$ monoton wachsend und $h \cdot f \in L^1(\mu_1 \times \mu_2)$, dann fällt Θ_1 unter der von h induzierten Dichte $\tilde{f} \propto f \cdot h$ stochastisch kleiner[15] *als unter f aus.*

Beweis. Für messbare Mengen $A \in \mathscr{F}_1$, $B \in \mathscr{F}_2$ gilt gemäß der Definition einer Dichte:

$$P(\Theta_1 \in A, \Theta_2 \in B) = \int_A \int_B f(\theta_1, \theta_2) \mathrm{d}\mu_2 \mathrm{d}\mu_1$$

Die Wahl von $A := \{\theta_1 > z\}$ und $B = \Omega_2$ führt auf

$$P(\Theta_1 > z) = \int_{\Omega_2} \left[\int_{>z} f(\theta_1|\theta_2) \mathrm{d}\mu_1 \right] f_2(\theta_2) \mathrm{d}\mu_2, \tag{2.9}$$

wobei $f_2(\theta_2)$ die marginale Dichte $\int_{\Omega_1} f(\theta_1, \theta_2) \mathrm{d}\mu_1$ von Θ_2 sowie

$$f(\theta_1|\theta_2) := \begin{cases} 0 & \text{für } (\theta_1, \theta_2) \in N \\ \frac{f(\theta_1, \theta_2)}{f_2(\theta_2)} & \text{sonst} \end{cases}$$

die bedingte Dichte von Θ_1 bezeichnen und die Definition $N := \{(\theta_1, \theta_2) | f_2(\theta_2) = \infty\} \cup \{(\theta_1, \theta_2) | \tilde{f}_2(\theta_2) = \infty\} \cup \{(\theta_1, \theta_2) | f_2(\theta_2) = 0\}$ verwendet wurde.

$$P(\Theta_1 > z) = \int_{\Omega_2} \left[\int_{\Omega_1} I_{\{\theta_1 > z\}} f(\theta_1|\theta_2) \mathrm{d}\mu_1 \right] f_2(\theta_2) \mathrm{d}\mu_2. \tag{2.10}$$

Eine analoge Herleitung für die Dichte $\tilde{f}(\theta_1, \theta_2)$ ergibt (unter Beachtung der Gleichheit der bedingten Dichten in beiden Szenarien) mit $\tilde{f}_2(\theta_2) := \frac{f_2(\theta_2) \cdot h(\theta_2)}{\int_{\Omega_2} f_2(\theta_2) \cdot h(\theta_2) \mathrm{d}\mu_2}$

$$\tilde{P}(\Theta_1 > z) = \int_{\Omega_2} \left[\int_{\Omega_1} I_{\{\theta_1 > z\}} f(\theta_1|\theta_2) \mathrm{d}\mu_1 \right] \tilde{f}_2(\theta_2) \mathrm{d}\mu_2. \tag{2.11}$$

Da die Funktion $\frac{f(\theta_1, \theta_2)}{f_2(\theta_2)}$ (für $\infty > f_2(\theta_2) > 0$) offenbar die RR_2-Eigenschaft von f „erbt“, und da $I_{\{\theta_1 > z\}}$ monoton wachsend in θ_1 ist, kann Lemma B.0.2 aus Anhang B angewendet werden. Demnach ist die Funktion $v(\theta_2) := \int_{\Omega_1} I_{\{\theta_1 > z\}} f(\theta_1|\theta_2) \mathrm{d}\mu_1$ monoton fallend (jedenfalls $f_2 \mathrm{d}\mu_2$-f.ü.) in θ_2.
v und h bilden folglich ein Paar ($f_2 \mathrm{d}\mu_2$-f.ü.) gegensätzlich monoton verlaufender Funktionen auf die Lemma B.0.3 (Anhang B) anwendbar ist (man setze $\mu := f_2 \mathrm{d}\mu_2$). Mit $\mu^* := f_2 \mathrm{d}\mu_2$ folgt[16]

$$\int_{\Omega_2} v(\theta_2) h(\theta_2) \mathrm{d}\mu^* \leq \int_{\Omega_2} v(\theta_2) \mathrm{d}\mu^* \cdot \int_{\Omega_2} h(\theta_2) \mathrm{d}\mu^*.$$

[15] Die $\mathscr{F}_1$-Messbarkeit von Mengen des Typs $\{\theta_1 | \theta_1 > z\}$ wird stillschweigend vorausgesetzt.

[16] Man beachte, dass μ^* ein Wahrscheinlichkeitsmaß auf Ω_2 ist und damit insbesonders die Voraussetzung von Lemma B.0.3 erfüllt.

Letzteres impliziert direkt die Behauptung

$$P(\Theta_1 > z) \geq \tilde{P}(\Theta_1 > z).$$

□

Satz 2.4.1 stellt das Pendant zu Satz 2.3.1 für Typ-2-Schätzer dar. Er impliziert - wie nachfolgend demonstriert wird - für eine große Klasse bayesianisch motivierter Schätzer die Existenz eines paradoxen Effekts und bildet somit das Kernresultat für funktionalbasierte Schätzer. Bevor diese Implikationen hervorgehoben werden, erscheint jedoch ein kurzer Rückblick auf die postulierten Annahmen, die den paradoxen Effekt ermöglichen, lohnenswert. Im Kern steht die Annahme einer RR_2-Likelihood bzw. Posterioridichte sowie die Existenz eines ausschließlich die zweite Dimension messenden Items, welches einen monotonen Beitrag generiert. Berücksichtigt man die in Lemma 2.4.1 dargelegte Beziehung zum MDD-Konzept, so wird eine unmittelbare Analogie zur Herleitung im Fall von Typ-1-Schätzern deutlich. Es sei ferner darauf hingewiesen, dass (ähnlich wie im Rahmen von Typ-1-Schätzern diskutiert wurde) durch geringfügige Variation der Annahmen (z.B. strikte RR_2-Eigenschaft anstelle der RR_2-Eigenschaft) entsprechend stärkere Resultate (beispielsweise strikte Ungleichung in Satz 2.4.1) hergeleitet werden können. Die Grundlage für diese Erweiterungen können den entsprechenden Beweisen aus Anhang B (vorrangig Lemma B.0.2 sowie Lemma B.0.3) entnommen werden.
Nach diesen kurzen Zwischenbemerkungen seien nun zwei praxisrelevante Beispiele bayesianisch motivierter Schätzer gegeben, die unter den Gültigkeitsbereich des Satzes fallen:

- Der sogenannte EAP-Schätzer (*expected a posteriori* estimator) ist definiert als Erwartungswert der Posterioriverteilung. Da stochastisch größere Verteilungen offenbar größere Erwartungswerte aufweisen, folgt aus Satz 2.4.1 unmittelbar die Existenz eines paradoxen Effekts für EAP-Schätzer.

- Der zum q-Quantil der Posterioriverteilung korrespondierende Bayes-Schätzer ist definiert als

$$\hat{\theta}_1^q := \inf M_q, \quad M_q := \{z | P(\Theta_1 \leq z | \boldsymbol{u}) \geq q\}.$$

Nach Satz 2.4.1 ist jeder Wert $z \in M_q$ aufgrund der stochastischen Ordnungsrelation zugleich auch Element der Menge

$$\tilde{M}_q := \{z | \tilde{P}(\Theta_1 \leq z | \boldsymbol{u}, u_{k+1}) \geq q\}$$

Hieraus folgt unmittelbar die entsprechende, „paradoxe“ Anordnung der zugehörigen größten unteren Schranken.

Abschließende Bemerkung: Ähnlich wie für Typ-1-Schätzer kann auch für den EAP bzw. für quantilbasierte Schätzer eine parametrische Abhängigkeitsanalyse erstellt werden. Diese Analyse arbeitet Bedingungen heraus, unter denen der paradoxe Effekt auch für den Fall erhalten bleibt, in dem das hinzugefügte Item mehrere Dimensionen misst.
Für den EAP-Schätzer kann hierbei - da dieser als Integral ($f_{\boldsymbol{\gamma}}$ bezeichne eine parametrisierte Posterioridichte)

$$\hat{\theta}_1^{EAP}(\boldsymbol{\gamma}) = \int f_{\boldsymbol{\gamma}} \mathrm{d}\mu,$$

definiert ist - auf „klassische“ Resultate der abstrakten Integration zurückgegriffen werden (z.B. Satz von Vitali/Satz über dominierte Konvergenz, Rudin, 1999, Kap. 1), die es erlauben, Integration und Grenzwertbildung zu vertauschen:

$$\lim_{n\to\infty} \int f_{\boldsymbol{\gamma_n}} \mathrm{d}\mu \stackrel{?}{=} \int \lim_{n\to\infty} f_{\boldsymbol{\gamma_n}} \mathrm{d}\mu = \int f_{\boldsymbol{\gamma_0}} \mathrm{d}\mu. \tag{2.12}$$

Wenn Letzteres möglich ist, und wenn $f_{\boldsymbol{\gamma}}$ stetig an der Stelle $\boldsymbol{\gamma_0}$ ist (was die Gültigkeit des letzten Gleichheitszeichen in (2.12) sichert), dann steht der EAP in stetiger Relation zum Parametervektor $\boldsymbol{\gamma}$ (jedenfalls an der Stelle $\boldsymbol{\gamma_0}$), so dass der paradoxe Effekt, der lediglich and der Stelle $\boldsymbol{\gamma_0}$ explizit hergeleitet wurde, gemäß einer an dem Beweis von Lemma 2.3.9 orientierten Argumentation übertragen werden kann.
Da der zum q-Quantil korrespondierende Quantilschätzer $\hat{\theta}_1^q$ (jedenfalls für eine streng monoton wachsende, stetige Verteilungsfunktion F) als Lösung der Gleichung

$$q - F(\boldsymbol{\gamma}, \theta) = 0$$

darstellbar ist, können die in Kapitel 2.3 skizzierten Sätze über implizite Funktionen erneut verwendet werden, um die stetige Abhängigkeit des paradoxen Effekts von $\boldsymbol{\gamma}$ zu etablieren.

3 Spezialformen des paradoxen Effekts

Im Gegensatz zu Kapitel 2, welches mit einer möglichst generellen Herleitung des paradoxen Effekts befasst war, stehen in diesem Kapitel konkrete MIRT-Modelle, die entweder von großer praktischer Relevanz (linear-kompensatorische Modelle) sind, oder in denen eine zusätzliche diagnostische Facette, die aus den Rahmen der bisherigen Darstellungweise fällt, zum Tragen kommt (Modelle mit zusätzlicher Schnelligkeitskomponente; longitudinale Modelle), im Vordergrund der weiteren Analyse des paradoxen Effekts.

Kapitel 3.1 widmet sich der Klasse der linear-kompensatorischen Modelle, die u.a. das in der Praxis nahezu omnipräsente Modell der (linearen) Faktorenanalyse umfasst. Neben einer allgemeinen, für den Fall binärer Items bereits von Hooker u.a. (2009) angegebenen Bedingung (Satz 3.1.1), die zur Prüfung der Monotonie der bedingten Maximumsfunktion in einem generellen linear-kompensatorischen Modell dienen kann, steht ein auf faktorenanalytische Modelle zugeschnittener Satz, der konkrete Angaben über die Gestalt einer Ladungsmatrix expliziert, unter denen ein paradoxer Effekt vermieden werden kann, im Mittelpunkt der Analyse.

Longitudinale IRT-Modelle, in denen der Fokus auf der zeitlichen Entwicklung einer Fähigkeitskomponente liegt, stellen, ebenso wie IRT-Modelle in denen die Schnelligkeit der Probanden bei der Bearbeitung der Testitems registriert wird, Beispiele von IRT-Modellen dar, die einen zusätzlichen semantischen Aspekt betonen, der mittels der bisherigen, relativ unspezifischen Methodik nicht adäquat herausgearbeitet werden konnte und dessen Analyse im Fokus der Kapitel 3.2 bzw. 3.3 steht.

Kapitel 3.2 behandelt dabei MIRT-Modelle, die neben der „üblichen“ Erfassung des Antwortmuster auch die Antworts- bzw. Reaktionszeiten der Person miteinbeziehen, um so zu einer präziseren Diagnose zu gelangen. Die naheliegende Frage, wie sich eine Verkürzung bzw. Verlängerung der Antwortszeiten auf die Inferenz der unbekannten Fähigkeit auswirkt, bildet den Fokus dieses Kapitels.

Die Betrachtung des paradoxen Effekts im Rahmen longitudinaler Modelle (Kapitel 3.3), d.h. im Rahmen von Modellen, die an mehreren Zeitpunkten eine Testung der Person vornehmen, um so auf Fähigkeitsveränderungen schließen zu können, stellt schließlich den Abschluss der Analyse des paradoxen Effekts in spezifischen MIRT-Modellen dar.

3.1 Paradoxien in linear-kompensatorischen Modellen

Die in Kapitel 2.3 dargelegte Konstruktion des paradoxen Effekts wurde in einem möglichst generellen Rahmen behandelt. Insbesondere wurde keine parametrische Form des MIRT-Modells angenommen. Die Ergebnisse besitzen dementsprechend breite Gültigkeit.
Nichtdestotrotz gibt es Aspekte des paradoxen Effekts, die nur unter restriktiveren (parametrischen) Submodellen untersucht werden können. Ein solches zu diskutierendes Submodell ist durch die Klasse der linear-kompensatorischen Item-Response-Modelle gegeben. Die zu einem linear-kompensatorischen Modell korrespondierende Loglikelihood besitzt die Form

$$l_u(\boldsymbol{\theta}) = \sum_{i=1}^{k+1} l_i(\boldsymbol{a_i^T}\boldsymbol{\theta}), \quad \mathbf{0} \neq \boldsymbol{a}_i \in \mathbb{R}_{0+}^p \,\forall\, i, \tag{3.13}$$

wobei die Funktionen $l_i : \mathbb{R} \mapsto \mathbb{R}$ im Folgenden stets als zweimal stetig differenzierbar mit der Eigenschaft $\frac{\partial^2 l_i}{\partial x^2}(x) < 0 \;\forall\, x \in \mathbb{R}$ vorausgesetzt werden.
Ein Ladungsvektor $\boldsymbol{a}_i$ bestimmt hierbei, welche Fähigkeitswerte $\boldsymbol{\theta}$ als äquivalent bezüglich des i-ten Items anzusehen sind: Alle Fähigkeitsvektoren $\boldsymbol{\theta}$ mit der Eigenschaft $\boldsymbol{a_i^T}\boldsymbol{\theta} = c$ liefern den gleichen Beitrag $l_i(\boldsymbol{a_i^T}\boldsymbol{\theta})$ zur Loglikelihood. Anders formuliert: Diese, in einer Hyperebene des $\mathbb{R}^p$ liegenden Fähigkeitsvektoren erklären das Antwortmuster auf dem i-ten Item gleich gut. Der Ladungsvektor $\boldsymbol{a}_i$ kennzeichnet dabei die Normale der entsprechenden Hyperebene.
Im Kern der nachfolgenden Betrachtungen steht die Frage, in welcher Beziehung die Ladungsmatrix A, deren i-te Zeile durch den Vektor $\boldsymbol{a_i^T}$ gebildet wird, zum paradoxen Effekt steht. Bevor diese Frage im Rahmen der durch (3.13) definierten Modellklasse untersucht wird, sollen jedoch die drei „gängigsten“ Spezialfälle[17] des Grundmodells (3.13) kurz dargelegt werden. Dies wird zugleich die Omnipräsenz dieser Modellklasse in praktischen Anwendungen verdeutlichen.

a) Das in Kapitel 2.3 bereits skizzierte M2PL-Modell stellt ein binäres IRT-Modell der Form (3.13) dar. Man wähle hierzu $l_i(x) := \log(\frac{\exp(x+\beta_i)}{1+\exp(x+\beta_i)})$ im Falle einer richtigen Antwort oder $l_i(x) := \log\left(1 - \frac{\exp(x+\beta_i)}{1+\exp(x+\beta_i)}\right)$ für den Fall einer falschen Antwort.

[17] Für den Nachweis der Negativität der zweiten Ableitung sei auf Anhang D verwiesen.

b) Wählt man je nach vorliegendem Itemscore $l_i(x) := \log(\Phi(x + \tau_{i,1}))$, $l_i(x) := \log(1 - \Phi(x + \tau_{i,L-1}))$ bzw. $l_i(x) := \log(\Phi(x + \tau_{i,l}) - \Phi(x + \tau_{i,l-1}))(\tau_{i,l} > \tau_{i,l-1})$, so erhält man das MGRM (*multidimensional graded response model*) als Spezialfall eines linear-kompensatorischen Modells (Φ bezeichnet hierbei die Verteilungsfunktion der Standardnormalverteilung).

c) Die Wahl $l_i(x) = -\frac{1}{2}\xi_i(u_i - x)^2$ ermöglicht die Einbettung des linearen faktorenanalytischen Modells in die obige Modellklasse.

Somit können die in der Praxis der Skalenkonstruktion vorherrschenden Modelle in die skizzierte Modellklasse eingebettet werden. Jedes M2PL, MGRM und jedes faktorenanalytische Modell stellt - insofern die korrespondierende Ladungsmatrix ausschließlich nichtnegative Einträge aufweist - einen Spezialfall des linear-kompensatorischen Modells dar.
Im Folgenden soll erörtert werden, unter welchen Bedingungen die Änderung eines Itemscores (o.B.d.A des letzten Items) ein paradoxes Resultat impliziert. Hierzu ist es zunächst - in Anlehnung an die Vorgehensweise von Hooker u.a. (2009) - erforderlich, einige technische Vorüberlegungen bzw. Umformungen darzulegen:
Wenn $\boldsymbol{b} := \boldsymbol{a}_{k+1}$ den Ladungsvektor des letzten Items bezeichnet, dann ist es aufgrund der speziellen Gestalt eines linear-kompensatorischen Modells stets möglich, zu gewährleisten, dass das letzte Item ausschließlich eine Dimension misst. Dies wird erzielt durch eine orthonormale Transformation[18] G, die einen neuen Parametervektor $\boldsymbol{\psi} := G \cdot \boldsymbol{\theta}$ induziert und deren letzte Zeile durch den Vektor $\boldsymbol{b}$ gegeben ist:

$$G := \begin{pmatrix} B \\ \boldsymbol{b^T} \end{pmatrix}, \quad G \cdot G^T = G^T G = \boldsymbol{I}_p, \; B\boldsymbol{b} = \boldsymbol{0}_{p-1}. \tag{3.14}$$

Die umparametrisierte Loglikelihood besitzt demnach die Form

$$l_u(\boldsymbol{\psi}) = \sum_{i=1}^{k} l_i(\boldsymbol{a_i^T} G^T \boldsymbol{\psi}) + l_{k+1}(\psi_p), \quad \boldsymbol{0} \neq \boldsymbol{a}_i \in \mathbb{R}^p_{0+} \, \forall \, i. \tag{3.15}$$

Das folgende Lemma ermöglicht erste Aussagen über die zur Konstruktion des paradoxen Effekts relevante bedingte Maximumsabbildung $\hat{\boldsymbol{\psi}}_{-p}(\psi_p)$.

Lemma 3.1.1. *Sei $l_u(\boldsymbol{\psi})$ die Loglikelihood eines linear-kompensatorischen Modells der Form (3.15) und sei $rg(A_0) = p$, wobei A_0 die Matrix bestehend aus den ers-*

[18] Es gelte o.B.d.A $\boldsymbol{b^T b} = 1$.

*ten k-Zeilen von A bezeichnet. Wenn $l_u(\boldsymbol{\psi})$ ψ_p-gleichmäßig niveaubeschränkt (*level bounded*) in $\boldsymbol{\psi}_{-p}$ ist, dann ist $\hat{\boldsymbol{\psi}}_{\boldsymbol{-p}}(\psi_p)$ eine wohldefinierte Funktion auf $\mathbb{R}$.*

Beweis. Der Beweis ist zweiteilig. Im ersten Teil gilt es $\hat{\boldsymbol{\psi}}_{\boldsymbol{-p}}(\psi_p) \neq \emptyset$ zu zeigen. Im zweiten Teil wird die Eindeutigkeit des bedingten Maximums behandelt.

1. $\hat{\boldsymbol{\psi}}_{\boldsymbol{-p}}(\psi_p) \neq \emptyset$: Dies folgt direkt durch Anwendung von Lemma 1.17 aus Rockafellar und Wets (2009).
2. Für die Eindeutigkeit des bedingten Maximums wird wie folgt argumentiert: Aufgrund der (zweimal stetigen) Differenzierbarkeit von l_u gilt

$$\frac{\partial l_u}{\partial \boldsymbol{\psi}_{-p}}(\boldsymbol{\psi}) = \sum_{i=1}^{k} \frac{\partial l_i}{\partial x}(\boldsymbol{a_i^T} G^T \boldsymbol{\psi}) \boldsymbol{a_i^T} B^T. \tag{3.16}$$

Die partielle Jacobimatrix $\nabla_{\boldsymbol{\psi}_{-p}} \frac{\partial l_u}{\partial \boldsymbol{\psi}_{-p}}$ besitzt ferner die Form

$$\frac{\partial^2 l_u}{\partial \boldsymbol{\psi}_{\boldsymbol{-p}}^{\boldsymbol{T}} \boldsymbol{\psi}_{\boldsymbol{-p}}}(\boldsymbol{\psi}) = \sum_{i=1}^{k} B\boldsymbol{a}_i \frac{\partial^2 l_i}{\partial x^2}(\boldsymbol{a_i^T} G^T \boldsymbol{\psi}) \boldsymbol{a_i^T} B^T = B A_0^T W A_0 B^T, \tag{3.17}$$

wobei W eine Diagonalmatrix mit Diagonalelementen $w_{ii} = \frac{\partial^2 l_i}{\partial x^2}(\boldsymbol{a_i^T} G^T \boldsymbol{\psi})$ darstellt. Aus $\frac{\partial^2 l_i}{\partial x^2} < 0$ sowie $rg(A_0) = p$ folgt, dass $B A_0^T W A_0 B^T$ negativ definit für jede Wahl von $\boldsymbol{\psi}$ ist.
Somit wurde (nach Theorem 2.14 aus Rockafellar und Wets (2009)) gezeigt, dass die Funktion $l_u^*(\boldsymbol{\psi_{-p}}) := l_u(\boldsymbol{\psi}_{-p}, \psi_p)$ strikt konkav in $\boldsymbol{\psi}_{-p}$ für jede Wahl von ψ_p ist. Daraus folgt die Eindeutigkeit des bedingten Maximums. □

Bevor die eigentlich interessierende Frage betreffend den Verlauf einer Komponente $\hat{\theta}_1(\psi_p)$ der bedingten Maximumsfunktion beantwortet werden kann, bedarf es eines Zwischenresultats[19] aus der linearen Algebra.

Lemma 3.1.2. *Sei Z eine invertierbare $p \times p$-Matrix und sei B eine $(p-1) \times p$ Matrix mit orthonormalen Zeilen, die zudem $B\boldsymbol{b} = 0$ für einen bestimmten Vektor $\boldsymbol{b} \in \mathbb{R}^p$ mit $\boldsymbol{b^T b} = 1$ erfüllt, dann gilt unter der Voraussetzung $\boldsymbol{b^T} Z^{-1} \boldsymbol{b} \neq 0$ die folgende Relation:*

$$(BZB^T)^{-1} = BZ^{-1}B^T - \frac{BZ^{-1}\boldsymbol{b}\boldsymbol{b^T}Z^{-1}B^T}{\boldsymbol{b^T}Z^{-1}\boldsymbol{b}}. \tag{3.18}$$

[19] Dieses Zwischenresultat wurde bereits in der Herleitung von Hooker u.a. (2009) verwendet.

Beweis. Sei im Folgenden G die in (3.14) definierte Matrix.
Zunächst gilt, wegen $B^TB = I_p - \boldsymbol{bb^T}$:

$$\begin{aligned} BZB^TBZ^{-1}B^T &= BB^T - BZ\boldsymbol{bb^T}Z^{-1}B^T \\ &= I_{p-1} - BZ\boldsymbol{bb^T}Z^{-1}B^T. \end{aligned} \tag{3.19}$$

Da $\boldsymbol{b^T}Z^{-1}\boldsymbol{b} \neq 0$ ist, lässt sich der Ausdruck

$$BZB^T\frac{BZ^{-1}\boldsymbol{bb^T}Z^{-1}B^T}{\boldsymbol{b^T}Z^{-1}\boldsymbol{b}}$$

betrachten, der sich unter Verwendung der Identität $B^TB = I_p - \boldsymbol{bb^T}$ zu

$$\frac{B\boldsymbol{bb}^TZ^{-1}B^T - BZ\boldsymbol{bb^T}Z^{-1}\boldsymbol{bb^T}Z^{-1}B^T}{\boldsymbol{b^T}Z^{-1}\boldsymbol{b}}$$

bzw. unter weiterer Verwendung von $B\boldsymbol{b} = \boldsymbol{0}_{p-1}$ zu

$$-BZ\boldsymbol{bb^T}Z^{-1}B^T$$

ergibt. Aus der Kombination mit (3.19) folgt die Behauptung.

□

Mit Hilfe dieses Zwischenresultats ist es nun möglich, die für die Untersuchung des paradoxen Effekts zentrale Frage nach der Monotonie der bedingten Maximumsfunktion zu beantworten. Der folgende Satz liefert das Kernresultat. Er stellt eine leichte Abwandlung des Theorems 6.1 von Hooker u.a. (2009) dar.

Satz 3.1.1. *Seien im Rahmen eines linear kompensatorischen MIRT-Modells (3.13) mit korrespondierender, durch (3.15) gegebener Umparametrisierung die Bedingungen von Lemma 3.1.1 erfüllt. Dann gilt für die j-te Komponente $\hat{\theta}_j(\psi_p)$ der bedingten Maximumsfunktion $\hat{\boldsymbol{\theta}}(\psi_p) := G^T(\hat{\boldsymbol{\psi}}_{-p}(\psi_p), \psi_p)$ die folgende Relation:*

$$\frac{\partial\hat{\theta}_j}{\partial\psi_p}(\psi_p) = \frac{\boldsymbol{e_j^T}(A_0^TWA_0)^{-1}\boldsymbol{b}}{\boldsymbol{b^T}(A_0^TWA_0)^{-1}\boldsymbol{b}}, \quad w_{ij} := (W)_{i,j} = \delta_{ij}\frac{\partial^2 l_i}{\partial x^2}(\boldsymbol{a_i^T}B^T\hat{\boldsymbol{\psi}}_{-p}(\psi_p) + \boldsymbol{a_i^T}\boldsymbol{b}\,\psi_p) \tag{3.20}$$

Beweis. Der Beweis von Hooker u.a. (2009) wird in leicht modifizierter Form wiedergegeben.
Gemäß der in Lemma 3.1.1 dargelegten Eindeutigkeit der bedingten Maximumsfunktion, lässt sich das Paar $(\hat{\boldsymbol{\psi}}_{-p}(\psi_p), \psi_p)$ als Nullstelle von (3.16) auffassen. Die zweimal

stetige Differenzierbarkeit von l_u in Kombination mit der Beobachtung, dass (3.17) stets nichtsingulär ist, bietet die Grundlage zur Anwendung des Satzes über implizite Funktionen. Mit den in (3.16) und (3.17) dargelegten Formeln und der „gemischten“ Ableitung

$$\frac{\partial^2 l_u}{\partial \psi_p \partial \boldsymbol{\psi}_{-p}}(\boldsymbol{\psi}) = BA_0^T W A_0 \boldsymbol{b}$$

folgt:

$$\nabla \hat{\boldsymbol{\psi}}_{-p}(\psi_p) = -(BA_0^T W A_0 B^T)^{-1} BA_0^T W A_0 \boldsymbol{b},$$

wobei sämtliche Einträge von W an der Stelle $(\hat{\boldsymbol{\psi}}_{-p}(\psi_p), \psi_p)$ zu evaluieren sind. Aufgrund der Beziehung $\hat{\boldsymbol{\theta}}(\psi_p) = B^T \hat{\boldsymbol{\psi}}_{-p}(\psi_p) + \psi_p \boldsymbol{b}$ folgt für die j-te Komponente von $\hat{\boldsymbol{\theta}}(\psi_p)$:

$$\frac{\partial \hat{\theta}_j}{\partial \psi_p}(\psi_p) = -\boldsymbol{e_j^T} B^T (BA_0^T W A_0 B^T)^{-1} BA_0^T W A_0 \boldsymbol{b} + b_j$$

Unter Verwendung der Relation (3.18) mit $Z := A_0^T W A_0$ kann $\frac{\partial \hat{\theta}_j}{\partial \psi_p}(\psi_p)$ wie folgt dargestellt werden:

$$\begin{aligned}
\frac{\partial \hat{\theta}_j}{\partial \psi_p}(\psi_p) &= -\boldsymbol{e_j^T} B^T \left(BZ^{-1}B^T - \frac{BZ^{-1}\boldsymbol{b}\boldsymbol{b^T}Z^{-1}B^T}{\boldsymbol{b^T}Z^{-1}\boldsymbol{b}} \right) BZ\boldsymbol{b} + b_j \\
&= \underbrace{(-\boldsymbol{e_j^T} B^T BZ^{-1}B^T BZ\boldsymbol{b} + b_j)}_{(1)} + \underbrace{\frac{\boldsymbol{e_j^T} B^T BZ^{-1}\boldsymbol{b}\boldsymbol{b^T}Z^{-1}B^T BZ\boldsymbol{b}}{\boldsymbol{b^T}Z^{-1}\boldsymbol{b}}}_{(2)} \quad (3.21)
\end{aligned}$$

Der Ausdruck (1) lässt sich dabei wie folgt umformen ($\boldsymbol{b}^T\boldsymbol{b} = 1, B^TB = I_p - \boldsymbol{b}\boldsymbol{b^T}$):

$$\begin{aligned}
\text{„(1)“} &= -\boldsymbol{e_j^T}(I_p - \boldsymbol{b}\boldsymbol{b^T})Z^{-1}(I_p - \boldsymbol{b}\boldsymbol{b^T})Z\boldsymbol{b} + b_j \\
&= -\boldsymbol{e_j^T}\boldsymbol{b} + b_j + \boldsymbol{e_j^T}\boldsymbol{b}\boldsymbol{b^T}\boldsymbol{b} + \boldsymbol{e_j^T}Z^{-1}\boldsymbol{b}\boldsymbol{b^T}Z\boldsymbol{b} - \boldsymbol{e_j^T}\boldsymbol{b}\boldsymbol{b^T}Z^{-1}\boldsymbol{b}\boldsymbol{b^T}Z\boldsymbol{b} \\
&= b_j + \boldsymbol{e_j^T}Z^{-1}\boldsymbol{b}\boldsymbol{b^T}Z\boldsymbol{b} - \boldsymbol{e_j^T}\boldsymbol{b}\boldsymbol{b^T}Z^{-1}\boldsymbol{b}\boldsymbol{b^T}Z\boldsymbol{b} \quad (3.22)
\end{aligned}$$

Für den Term (2) erhält man auf analoge Weise:

$$\begin{aligned}
\text{„(2)“} &= \frac{\boldsymbol{e_j^T}(I_p - \boldsymbol{b}\boldsymbol{b^T})Z^{-1}\boldsymbol{b}\boldsymbol{b^T}Z^{-1}(I_p - \boldsymbol{b}\boldsymbol{b^T})Z\boldsymbol{b}}{\boldsymbol{b^T}Z^{-1}\boldsymbol{b}} \\
&= \frac{\boldsymbol{e_j^T}\boldsymbol{Z^{-1}}\boldsymbol{b}}{\boldsymbol{b^T}Z^{-1}\boldsymbol{b}} - \boldsymbol{e_j^T}\boldsymbol{b} - \boldsymbol{e_j^T}Z^{-1}\boldsymbol{b}\boldsymbol{b^T}Z\boldsymbol{b} + \boldsymbol{e_j^T}\boldsymbol{b}\boldsymbol{b^T}Z^{-1}\boldsymbol{b}\boldsymbol{b^T}Z\boldsymbol{b} \quad (3.23)
\end{aligned}$$

Einsetzen von (3.22) und (3.23) in (3.21) ergibt die Behauptung. □

Da der Nenner des Terms $\frac{\boldsymbol{e_j^T}(A_0^T W A_0)^{-1}\boldsymbol{b}}{\boldsymbol{b}(A_0^T W A_0)^{-1}\boldsymbol{b}}$ stets negativ ausfällt, entscheidet allein das Vorzeichen des Ausdrucks $\boldsymbol{e_j^T}(A_0^T W A_0)^{-1}\boldsymbol{b}$ über die Monotonie der bedingten

Maximumsfunktion. Die suggestive Schreibweise $\boldsymbol{e}_j^{\boldsymbol{T}}(A_0^T W A_0)^{-1}\boldsymbol{b}$ sollte jedoch nicht darüber hinwegtäuschen, dass das Vorzeichen in Abhängigkeit von ψ_p, welches über die Matrix $W = W(\psi_p)$ in die Berechnung miteingeht, variieren kann.
Wenn $\boldsymbol{e}_j^{\boldsymbol{T}}(A_0^T W(\psi_p) A_0)^{-1}\boldsymbol{b} > 0 \; \forall \, \psi_p$ gilt, dann liefert unter der Voraussetzung der strikten Monotonie von l_{k+1} eine Anwendung von Satz 2.3.1 - man beachte, dass die Forderung einer SMDD-Loglikelihood durch die Monotonie der bedingten Maximumsfunktion austauschbar ist - ein paradoxes Resultat bezüglich der j-ten Komponente, wenn der Response auf dem letzten Item verändert wird.
Wenn $\boldsymbol{e}_j^{\boldsymbol{T}}(A_0^T W(\psi_p) A_0)^{-1}\boldsymbol{b} > 0$ nur für einen bestimmten Bereich $\psi_p \in (\psi_u, \psi_o)$ gilt, dann muss die Veränderung des Responses des k-ten Items nicht zwangsläufig zu einem paradoxen Resultat führen. Umschließt der Bereich jedoch die ψ_p-Komponente des auf den erstem k Items beruhenden ML-Schätzers, so kann mittels einer modifizierten, „lokalen“ Version des Satzes 2.3.1 gezeigt werden, dass stets ein Item (mit Ladungsvektor $\boldsymbol{b}$) aus einem als hinreichend groß anzunehmenden „Itemuniversum“ existiert, welches einen paradoxen Effekt induziert.
Es sei ferner bemerkt, dass Satz 3.1.1 eine einfache Erweiterung im Hinblick auf linear-kompensatorische Modelle mit einem zusätzlichen Strafterm ($\widehat{=}$ log-Priori-Verteilung) $\gamma(\boldsymbol{\theta})$ besitzt. Unter Beachtung, dass für eine penalisierte Loglikelihood der Form (γ sei eine konkave, zweimal stetig differenzierbare Funktion auf $\mathbb{R}^p$)

$$l_u^p(\boldsymbol{\theta}) := l_u(\boldsymbol{\theta}) + \gamma(\boldsymbol{\theta})$$

die korrespondierende, umparametrisierte penalisierte Loglikelihood die Ableitungen

$$\frac{\partial l_u^p}{\partial \boldsymbol{\psi}_{-p}}(\boldsymbol{\psi}) = \left(\sum_{i=1}^{k} \frac{\partial l_i}{\partial x}(\boldsymbol{a}_i^{\boldsymbol{T}} G^T \boldsymbol{\psi}) \boldsymbol{a}_i^{\boldsymbol{T}} + \nabla\gamma(G^T\boldsymbol{\psi}) \right) B^T. \tag{3.24}$$

sowie

$$\begin{aligned} \frac{\partial^2 l_u^p}{\partial \boldsymbol{\psi}_{-p}^{\boldsymbol{T}} \boldsymbol{\psi}_{-p}}(\boldsymbol{\psi}) &= B \left(\sum_{i=1}^{k} \boldsymbol{a}_i \frac{\partial^2 l_i}{\partial x^2}(\boldsymbol{a}_i^{\boldsymbol{T}} G^T \boldsymbol{\psi}) \boldsymbol{a}_i^{\boldsymbol{T}} + \frac{\partial^2 \gamma}{\partial \boldsymbol{\theta}^{\boldsymbol{T}} \partial \boldsymbol{\theta}}(G^T\boldsymbol{\psi}) \right) B^T \\ &= B(A_0^T W A_0 + K) B^T, \qquad K := \frac{\partial^2 \gamma}{\partial \boldsymbol{\theta}^{\boldsymbol{T}} \partial \boldsymbol{\theta}}(G^T\boldsymbol{\psi}). \end{aligned} \tag{3.25}$$

aufweist, lassen sich zwei zu Lemma 3.1.1 und Satz 3.1.1 analoge Resultate etablieren:

Lemma 3.1.3. *Sei $l_u^p(\boldsymbol{\psi})$ die um eine zweimal stetig differenzierbare, konkave Funktion $\gamma^*(\boldsymbol{\psi}) := \gamma(G^T\boldsymbol{\psi})$ penalisierte Loglikelihood eines linear-kompensatorischen Modells der Form (3.15) und sei $rg(A_0) = p$, wobei A_0 die Matrix bestehend aus den*

ersten k-Zeilen von A bezeichnet.
Wenn $l_u^p(\boldsymbol{\psi})$ ψ_p-gleichmäßig niveaubeschränkt in $\boldsymbol{\psi}_{-p}$ ist, dann ist $\hat{\boldsymbol{\psi}}_{-\boldsymbol{p}}(\psi_p)$ eine wohldefinierte Funktion auf $\mathbb{R}$.
Die Forderung der gleichmäßigen Niveaubeschränktheit von l_u^p ist insbesonders stets erfüllt, wenn γ (oder äquivalent γ^*) global niveaubeschränkt und l_u nach oben beschränkt ist.

Beweis. Der erste Teil folgt analog zum Beweis von Lemma 3.1.1.
Die letzte Behauptung resultiert unmittelbar aus folgenden Beobachtungen:

1. Jede global bezüglich $\boldsymbol{\psi}$ niveaubeschränkte Funktion ist lokal ψ_p-gleichmäßig niveaubeschränkt in $\boldsymbol{\psi}_{-p}$.
2. Die Summe einer global niveaubeschränkten Funktion und einer nach oben beschränkten Funktion ist global niveaubeschränkt.

Zu 1.:
Sei γ^* global niveaubeschränkt, d.h. $\forall \alpha \in \mathbb{R}\ \exists \mathrm{M}\,(\text{beschränkte Menge}) : \{\boldsymbol{\psi} | \gamma^*(\boldsymbol{\psi}) \geq \alpha\} \subset M$, und sei ψ^p fest gewählt. Mit $V := \mathbb{B}_1(\psi_p)$ ist dann offenbar

$$\{\boldsymbol{\psi} | \psi_p \in V \ \wedge \gamma^*(\boldsymbol{\psi}) \geq \alpha\} \subset M$$

und damit ist 1. offensichtlich wahr.
Zu 2.:
Sei $\alpha \in \mathbb{R}$, sei ferner β eine obere Schranke für l_u.
Man setze $\alpha^* := \alpha - \beta$. Nach Voraussetzung existiert zu α^* eine beschränkte Menge M, so dass

$$M \supset \{\boldsymbol{\psi} | \gamma^*(\boldsymbol{\psi}) \geq \alpha^*\} = \{\boldsymbol{\psi} | \gamma^*(\boldsymbol{\psi}) + \beta \geq \alpha\} \supset \{\boldsymbol{\psi} | \gamma^*(\boldsymbol{\psi}) + l_u(\boldsymbol{\psi}) \geq \alpha\}.$$

Daraus folgt die zweite Behauptung.
Man beachte ferner, dass aufgrund der Invertierbarkeit der Matrix G die Bedingung der globalen Niveaubeschränktheit an γ^* äquivalent ist mit der globalen Niveaubeschränktheit von γ (um dies einzusehen, nutze man die Tatsache, dass das Bild einer kompakten Menge unter einer stetigen Abbildung stets kompakt ist.). □

Satz 3.1.2. *Sei γ zweimal stetig differenzierbar und konkav und sei $l^p := l_u + \gamma^*$, wobei l_u die Loglikelihood eines linear-kompensatorischen $MIRT$-Modells (3.13) mit korrespondierender, durch (3.15) gegebener Umparametrisierung darstellt.Wenn die*

Bedingungen von Lemma 3.1.3 erfüllt sind, dann gilt für die j-te Komponente $\hat{\theta}_j(\psi_p)$ *der bedingten Maximumsfunktion* $\hat{\boldsymbol{\theta}}(\psi_p) := G^T(\hat{\boldsymbol{\psi}}_{-p}(\psi_p), \psi_p)$ *die folgende Relation:*

$$\frac{\partial \hat{\theta}_j}{\partial \psi_p}(\psi_p) = \frac{\boldsymbol{e_j^T}(A_0^T W A_0 + K)^{-1}\boldsymbol{b}}{\boldsymbol{b^T}(A_0^T W A_0 + K)^{-1}\boldsymbol{b}}, \tag{3.26}$$

mit

$$(W)_{i,j} = \delta_{ij}\frac{\partial^2 l_i}{\partial x^2}(\boldsymbol{a_i^T} B^T \hat{\boldsymbol{\psi}}_{-p}(\psi_p) + \boldsymbol{a_i^T b}\,\psi_p)$$

und

$$K := \frac{\partial^2 \gamma}{\partial \boldsymbol{\theta^T} \partial \boldsymbol{\theta}}(B^T \hat{\boldsymbol{\psi}}_{-p}(\psi_p) + \boldsymbol{b}\,\psi_p)$$

Beweis. Analog zum Beweis des Satzes 3.1.1. Man beachte lediglich, dass Lemma 3.1.2 in diesem Kontext mit der Gleichsetzung $Z := A_0^T W A_0 + K$ Gültigkeit behält, da die Summe einer negativ definiten und einer negativ (semi)definiten Matrix wiederum eine negativ definite - und damit insbesondere invertierbare - Matrix darstellt. □

Diese „bayesianische“ Variante wird insbesonders in den Kapiteln 3.2 (IRT-Modelle mit Schnelligkeitskomponente) und 3.3 (longitudinale IRT-Modelle) eine wichtige Rolle zur Untersuchung des paradoxen Effekts einnehmen.
Für die Zwecke der folgenden Abschnitte reicht jedoch die einfachere Form, d.h. Satz 3.1.1 aus, um interessante Resultate bezüglich des paradoxen Effekts in der Klasse der linear-kompensatorischen Modelle zu etablieren.
Dieser Satz wird es ermöglichen, starke Aussagen bezüglich der Existenz bzw. Prävalenz des paradoxen Effekts herzuleiten. Bevor dies in einem möglichst allgemeinen Rahmen geschieht, soll jedoch zunächst das für die Anwendung bedeutendste Submodell - das lineare faktorenanalytische Modell - eingehender betrachtet werden.

Existenz des paradoxen Effekts in faktorenanalytischen Modellen mit kompensatorischer Ladungsmatrix

Ein faktorenanalytisches Modell, dessen zugehörige Ladungsmatrix aus nichtnegativen Einträgen besteht, stellt einen Spezialfall eines linear-kompensatorischen Modells dar, wie sich mittels der Wahl von

$$l_i(x) := -\frac{1}{2}\xi_i(u_i - x)^2,$$

wobei $\xi_i \in \mathbb{R}_+$ den Präzisionsparameter (inverse Messfehlervarianz) des i-ten Items bezeichnet, demonstrieren lässt. Da ferner $\frac{\partial^2 l_i}{\partial x^2} < 0$ für alle i gilt, sind sämtliche Forderungen an ein linear-kompensatorisches Modell erfüllt.
Als nächsten Schritt gilt es festzustellen, ob die bedingte Maximumsfunktion wohldefiniert ist. Hierzu wird wie folgt argumentiert:
Die umparametrisierte Loglikelihood (3.15) lässt sich in der Form

$$-\frac{1}{2}(\boldsymbol{u} - V\boldsymbol{\psi})^T \Omega^{-1} (\boldsymbol{u} - V\boldsymbol{\psi})$$

schreiben, wobei V eine $(k+1)\times p$ Matrix (Ladungsmatrix nach Durchführung der isometrischen Transformation) bestehend aus den Zeilenvektoren $\boldsymbol{v}_i = \boldsymbol{a}_i^{\boldsymbol{T}} G^T$ darstellt und Ω^{-1} eine Diagonalmatrix mit Einträgen ξ_i. Da das letzte Item zur Berechnung des bedingten Maximums keinen Beitrag liefert, kann ebenso die alternative Form

$$-\frac{1}{2}(\boldsymbol{u}_0 - V_0\boldsymbol{\psi})^T \Omega_0^{-1} (\boldsymbol{u}_0 - V_0\boldsymbol{\psi}),$$

in der mit „0“ indizierte Größe sich jeweils auf den Test der Länge k anstelle von $k+1$ beziehen, bzw. mit $\boldsymbol{u^*} := \boldsymbol{u}_0 - \psi_p \boldsymbol{v}_{(p)}$ ($\boldsymbol{v}_p$ bezeichnet die p-te Spalte von V_0) die Form

$$-\frac{1}{2}(\boldsymbol{u^*}(\psi_p) - V_{-(p)}\boldsymbol{\psi}_{-p})^T \Omega_0^{-1} (\boldsymbol{u^*}(\psi_p) - V_{-(p)}\boldsymbol{\psi}_{-p}),$$

wobei $V_{-(p)}$ die Matrix bestehend aus den ersten $p-1$ Spalten von V_0 bezeichnet, betrachtet werden.
Durch Neudefinition der Vektoren/Matrizen kann die Maximierung dieser Funktion (bezüglich $\boldsymbol{\psi}_{-p}$) auf ein „gewöhnliches“ Kleinste-Quadrate-Problem reduziert werden, welches bekanntlich (siehe z.B. Lax, 2007, S.86) genau dann eindeutig lösbar ist, wenn die zugehörige Matrix - in diesem Fall die Matrix $\Omega_0^{-\frac{1}{2}} V_{-(p)}$ - vollen Spaltenrang aufweist. Letzteres gilt unter der wenig restriktiven Annahme, dass die Ladungsmatrix A_0 vollen Spaltenrang besitzt. Folglich ist Satz 3.1.1 anwendbar.
Da $\frac{\partial^2 l_i}{\partial x^2}(x) = -\xi_i$ ist, gilt in (3.20) $W = -\Omega_0^{-1}$. Somit weist das faktorenanalytische Modell die Besonderheit auf, dass die Ableitung der bedingten Maximumsfunktion $\frac{\partial \hat{\theta}_j}{\partial \psi_p}(\psi_p)$ eine bezüglich ψ_p konstante Funktion ist.
Des Weiteren ist die zur Konstruktion des paradoxen Effekts relevante Monotonie des Beitrags des hinzugefügten Items gegeben. Betrachtet man etwa eine Erhöhung des Itemscores des letzten Items um ϵ ($\epsilon \in \mathbb{R}_+$), so gilt

$$\begin{aligned} l_{k+1,u_{k+1}+\epsilon}(\psi_p) - l_{k+1,u_{k+1}}(\psi_p) &= -\frac{1}{2}\xi_{k+1}(u_{k+1} + \epsilon - \psi_p)^2 + \frac{1}{2}\xi_{k+1}(u_{k+1} - \psi_p)^2 \\ &= \xi_{k+1}\epsilon\psi_p + r(u_{k+1}, \epsilon), \end{aligned}$$

wobei der Term $r(\cdot)$ unabhängig von ψ_p ist. Damit ist die Monotoniebedingung offenbar gewährt.
Ein paradoxer Effekt bezüglich der j-ten Fähigkeitskomponente tritt bei Erhöhung des Itemscores des letzten Items folglich genau dann auf, wenn

$$\boldsymbol{e}_j^{\boldsymbol{T}}(A_0^T W A_0)^{-1}\boldsymbol{a}_{k+1} > 0$$

bzw.

$$\boldsymbol{e}_j^{\boldsymbol{T}}(A_0^T \Omega_0^{-1} A_0)^{-1}\boldsymbol{a}_{k+1} < 0$$

gilt.
Man beachte, dass die vorliegende Argumentation keineswegs impliziert, dass paradoxe Effekte im faktorenanalytischem Modell auftreten. Um einen derartigen Schluss vornehmen zu können, müsste die strikte Ungleichung $\boldsymbol{e}_j^{\boldsymbol{T}}(A_{-l}^T W(l) A_{-l})^{-1}\boldsymbol{a}_l > 0$ für mindestens ein Paar (j, l) bewiesen werden.
Von Interesse ist daher die Beantwortung der Frage, unter welchen Bedingungen an A bzw. an Ω paradoxe Effekte im Rahmen des faktorenanalytischen Modells „garantiert“ werden können.
Im Folgenden soll dieser aufgeworfenen Frage nachgegangen werden. Zunächst bedarf es hierfür einer zusätzlichen Definition:

Definition 3.1.1. Ein faktorenanalytisches Modell heisst *regulär kompensatorisch*, wenn die folgenden Bedingungen an die Itemparameter erfüllt sind:

i) Die Ladungsmatrix A, mit zugehörigen Zeilen $\boldsymbol{a_j} \neq \mathbf{0}$, besteht ausschließlich aus nichtnegativen Einträgen (kompensatorisch).

ii) Die Diagonaleinträge ω_i (ξ_i^{-1}) (Messfehlervarianzen) der Matrix Ω sind positiv (erste Regularitätsbedingung).

iii) Die Matrix A besitzt vollen Spaltenrang $p > 1$(zweite Regularitätsbedingung; Mehrdimensionalität).

Wenn zudem die Ladungsmatrix keine Einfachstruktur aufweist, so heisst das Modell *regulär strikt kompensatorisch*.

Bevor mit der Betrachtung des paradoxen Effekts innerhalb der so definierten Klasse begonnen wird, sollen die in der Definition dargelegten Bedingungen bezüglich ihrer

Restriktivität diskutiert werden.
Die Regularitätsbedingungen stellen keine maßgeblichen Einschränkungen dar. Die erste Regularitätsbedingung ist offensichtlich in Anwendungen stets erfüllt, da sie lediglich Items mit perfekter Messpräzision ausschließt. Die zweite Regularitätsbedingung - nicht jedoch die Mehrdimensionalitätsbedingung - kann ferner durch Umformulierung des faktorenanalytischen Modells stets gewährleistet werden. Ist nämlich die Ladungsmatrix A nicht von vollem Spaltenrang, so lässt sich eine Spalte (o.B.d.A die letzte Spalte) aus den übrigen Spalten linearkombinieren

$$A = \begin{pmatrix} A_{-(p)} & A_{-(p)}\boldsymbol{c} \end{pmatrix} = A_{-(p)} \begin{pmatrix} I & \boldsymbol{c} \end{pmatrix}$$

und die Definition $\boldsymbol{\theta}^* := \boldsymbol{\theta}_{-p} + \boldsymbol{c}\,\theta_p$ liefert ein um eine Dimension reduziertes Modell mit Ladungsmatrix $A_{-(p)}$. Die Wiederholung dieses Vorgangs führt schließlich zu einem reduzierten Modell mit vollem Spaltenrang. Somit verbleibt lediglich die Forderung nach echter Mehrdimensionalität, d.h. die obige Prozedur zur Erlangung des vollen Spaltenrangs stoppt bei einer Matrix mit mindestens zwei Spalten, in Kombination mit einer nichtnegativen Ladungsmatrix. Die Forderung nach Mehrdimensionalität stellt dabei keine echte Einschränkung dar. Wie zahlreichen Literaturquellen (z.B. Shapiro, 1982; Rosenbaum, 1984) zu entnehmen ist - und wie im abschließenden Abschnitt dieser Arbeit noch eingehender diskutiert werden wird - stellen eindimensionale Tests die Ausnahme und mehrdimensionale Tests die Regel dar. Auch wenn die verbleibende Kompensationsannahme auf den ersten Blick restriktiv erscheinen mag, so sei darauf hingewiesen, dass

a) Annahme i) lediglich die Tatsache wiedergibt, dass jede latente Dimension bei Kontrolle der übrigen Dimensionen einen positiven (nichtnegativen) Beitrag zur Beantwortung des Items liefert. Dies stellt ein in vielen Fällen plausibles Szenario dar[20].

b) im Hinblick auf die Diskussion des paradoxen Effekts Annahme i) nahezu unausweichlich erscheint. Dies liegt darin begründet, dass (nur) unter der Annahme eines positiven Beitrags der latenten Dimensionen die intuitive Erwartungshaltung entsteht, dass ein höherer Itemscore stets mit einer höheren Ausprägung aller latenten Dimensionen einhergeht.

[20]Beispielsweise wäre es in einem Physik-Test, der räumliches Vorstellungsvermögen sowie Mathematikkenntnisse misst, verwunderlich, wenn ein höheres räumliches Vorstellungsvermögen (bei Kontrolle der Mathematikfähigkeit) abträglich zur Beantwortung einer Frage wäre.

Ausgehend von diesen Vorbemerkungen zur Restriktivität der in Definition 3.1.1 dargelegten Annahmen, soll nun die Existenz des paradoxen Effekts innerhalb der Testklasse der regulär kompensatorischen faktorenanalytischen Modelle betrachtet werden.
Der Maximum-Likelihood-Schätzer der latenten Fähigkeiten ist durch den Term

$$\hat{\boldsymbol{\theta}} := (A^T\Omega^{-1}A)^{-1}A^T\Omega^{-1}\boldsymbol{u}$$

gegeben. Eine Erhöhung des Itemscores auf dem i-ten Item verringert genau dann nicht den Schätzwert der j-ten latenten Dimension, wenn

$$\boldsymbol{e_j^T}(A^T\Omega^{-1}A)^{-1}A^T\Omega^{-1}\boldsymbol{e_i} \geq 0.$$

gilt. Die obige Ungleichung muss für jede Kombination (i, j) mit $i \in \{1, \ldots, k+1\}$ und $j \in \{1 \ldots p\}$ erfüllt sein, damit der Test nicht dem paradoxen Effekt unterliegt. Dies ist äquivalent zur Nichtnegativität der Einträge der Matrix $(A^T\Omega^{-1}A)^{-1}$.[21]
Im Folgenden wird zunächst der Fall $\Omega = I_{k+1}$, d.h. der Fall gleicher Messfehlervarianzen, betrachtet. Gemäß den bisherigen Bemerkungen ist in diesem Spezialfall der paradoxe Effekt genau dann vermeidbar, wenn die Matrix $(A^TA)^{-1}$ ausschließlich nichtnegative Einträge beinhaltet.
Das nachfolgende Lemma stellt eine Äquivalenz zwischen dieser formalen Bedingung und einer im Kontext der Testkonstruktion zugänglicheren Bedingung her und impliziert somit - wie in den anschließenden Überlegungen ausgeführt werden wird - eine direkte Beziehung des paradoxen Effekts zum Begriff der Einfachstruktur der Ladungsmatrix.

Lemma 3.1.4. *Bei Gültigkeit der Annahmen i) und iii) sind folgende Aussagen äquivalent:*

(a) Die Matrix $(A^TA)^{-1}$ besitzt ausschließlich nichtnegative Einträge.

(b) Die Kreuzproduktmatrix A^TA ist eine positive Diagonalmatrix.

Beweis. Die Implikation (b) $\Rightarrow$ (a) ist offensichtlich.
Die Implikation (a) $\Rightarrow$ (b) wird mittels Induktion nach der Dimensionalität der Matrix A^TA bewiesen.

[21] An dieser Stelle sei betont, dass der Begriff Nichtnegativität auf die Einträge der Matrix bezogen ist. Streng davon abzugrenzen, ist der Begriff einer nicht negativ definiten quadratischen Form.

Induktionsanfang:
Der Fall $p = 2$ lässt sich wie folgt darstellen:

$$A^T A := \begin{pmatrix} a & b \\ b & d \end{pmatrix} \quad (A^T A)^{-1} = \frac{1}{ad - b^2} \begin{pmatrix} d & -b \\ -b & a \end{pmatrix}.$$

Gemäß Annahme *iii*) ist die Matrix $A^T A$ positiv definit. Folglich gilt $\det(A) = ad - b^2 > 0$, was in Kombination mit Annahme *i*) auf $b = 0$ führt.

Induktionsschritt:
Die Aussage des Lemma gelte nun für alle $p \times p$- Matrizen $A^T A$, die aus den Annahmen *i*) sowie *iii*) resultieren. Eine $(p+1) \times (p+1)$ Kreuzproduktmatrix $A^T A$ lässt sich wie folgt darstellen:

$$(A^T A) = \begin{pmatrix} M_p & \boldsymbol{b} \\ \boldsymbol{b^T} & d \end{pmatrix}.$$

Hierbei ist M_p eine $p \times p$ Teilmatrix, die aus dem Kreuzprodukt $(A_{-(p+1)})^T A_{-(p+1)}$ entsteht; $\boldsymbol{b}$ ein $(p \times 1)$ Vektor und d ein positiver Skalar.
Unter der Voraussetzung, dass die Matrix $Q := M_p - d^{-1}\boldsymbol{b}\boldsymbol{b^T}$ invertierbar ist, lässt sich die Inverse der Kreuzproduktmatrix wie folgt darstellen (Harville, 1997, S.99):

$$(A^T A)^{-1} = \begin{pmatrix} Q^{-1} & -Q^{-1}\boldsymbol{b}d^{-1} \\ -d^{-1}\boldsymbol{b^T}Q^{-1} & d^{-1} + d^{-1}\boldsymbol{b^T}Q^{-1}\boldsymbol{b}d^{-1} \end{pmatrix} \quad Q := M_p - d^{-1}\boldsymbol{b}\boldsymbol{b^T},$$

Gemäß (a) besteht die Matrix Q^{-1} ausschließlich aus nichtnegativen Einträgen. Gleiches gilt gemäß Annahme *i*) für den Vektor $\boldsymbol{b}$. Die Nichtnegativität des Terms $-Q^{-1}\boldsymbol{b}d^{-1}$ lässt sich folglich nur durch die Wahl $\boldsymbol{b} = \boldsymbol{0}$ erzielen. Dies impliziert wiederum $Q = M_p$, so dass die Induktionsvoraussetzung auf die Matrix Q anwendbar ist. Dies führt auf die besagte Diagonalstruktur der Matrix $A^T A$.

Somit verbleibt lediglich der Beweis der Invertierbarkeit der Q-Matrix.
Man nehme gegensätzlich zur Behauptung an, dass $Q\boldsymbol{x} = \boldsymbol{0}$ für ein $\boldsymbol{x} \neq \boldsymbol{0}$ gilt. Dann folgt gemäß der Definition von Q:

$$M_p\boldsymbol{x} = \boldsymbol{b}\boldsymbol{b^T}\boldsymbol{x}d^{-1}.$$

Da offenbar $\boldsymbol{b^T}\boldsymbol{x} \neq 0$ gilt (ansonsten wäre $M_p\boldsymbol{x} = \boldsymbol{0}$, was aufgrund der Invertierbarkeit von $A^T A$ unmöglich ist.), ist der Vektor $\tilde{\boldsymbol{x}} := d \cdot (\boldsymbol{b^T}\boldsymbol{x})^{-1} \cdot \boldsymbol{x}$ wohldefiniert. Wie

die Rechnungen

$$M_p\tilde{\boldsymbol{x}} = \boldsymbol{b} \qquad \boldsymbol{b^T}\tilde{\boldsymbol{x}} = d$$

zeigen, liefern die Einträge des $\tilde{\boldsymbol{x}}$-Vektors die Gewichte für eine Linearkombination der letzten Spalte der Matrix A^TA aus den übrigen Spalten. Folglich besitzt die Matrix A^TA nicht vollen Spaltenrang. Letzteres stellt einen Widerspruch zur Annahme *iii*) dar (alternativ lässt sich mittels Theorem 19 von Marsaglia und Styan (1974) argumentieren.). □

Nach Lemma 3.1.4 ist demnach der Schätzer genau dann fair, wenn die Kreuzproduktmatrix A^TA Diagonalstruktur aufweist. Wie das folgende Lemma zeigen wird, ist Letzteres unter den gegebenen Annahmen äquivalent zur Einfachstruktur der Ladungsmatrix.

Lemma 3.1.5. *Unter Annahme i) sind die folgenden Aussagen äquivalent:*

(a) A^TA *ist eine positive Diagonalmatrix.*

(b) A besitzt Einfachstruktur, d.h. in jeder Zeile der Matrix A befindet sich genau ein von Null verschiedener Eintrag.

Beweis. Die Implikation $(b) \Rightarrow (a)$ ist trivial.
Anstelle der Implikation $(a) \Rightarrow (b)$ wird die äquivalente Implikation $\neg(b) \Rightarrow \neg(a)$ bewiesen.
Sei A eine Matrix, die keine Einfachstruktur aufweist. Dann gibt es eine Zeile mit mindestens zwei von Null verschiedenen Einträgen. Aufgrund Annahme *i*) ist das Skalarprodukt der entsprechenden Spalten - gemeint sind die Spalten, in denen sich die zugehörigen, von Null verschiedenen Einträge befinden - positiv. Demnach weist A^TA keine Diagonalgestalt auf. □

In der Gesamtschau von Lemma 3.1.4 und Lemma 3.1.5 ergibt sich das folgende, für das Verständnis des paradoxen Effekts in der faktorenanalytischen Modellklasse zentrale Resultat:

Satz 3.1.3. *Bei Gültigkeit der Annahmen i) und iii) ist der Kleinste-Quadrate-Schätzer genau dann fair, wenn die Ladungsmatrix Einfachstruktur aufweist.*
Innerhalb der Klasse der regulär kompensatorischen faktorenanalytischen Modelle

induziert jedes regulär strikt kompensatorische Modell bei Verwendung des Kleinste-Quadrate-Schätzers einen paradoxen Effekt.

Im Rahmen des faktorenanalytischen Modells ist folglich jeder Test mit Querladungen von einem paradoxen Effekt betroffen: Es existiert stets ein Item, bei dem eine Erhöhung des Itemscores eine Verringerung des Schätzwerts in mindestens einer latenten Dimension hervorruft.
Der skizzierte Effekt soll anhand eines realen Datensatzes illustriert werden. Der Fragebogen zum Screening psychischer Störungen im Jugendalter (SPS-J) erfasst anhand von $k + 1 = 32$ Items vier Dimensionen (Hampel und Petermann, 2006). Diese vier Dimensionen (Subtests) messen Ängstlichkeit/Depressivität (Faktor 1), aggressives-dissoziales Verhalten (Faktor 2), Selbstwertprobleme (Faktor 3) sowie Ärgerkontrollprobleme (Faktor 4). Die 32×4 Ladungsmatrix, die Tabelle 3 von Hampel und Petermann (2006) entnommen werden kann, besitzt vollen Spaltenrang und ist zudem nichtnegativ[22], d.h. jede Dimension steht in einem positiven (genauer: nichtnegativen) Zusammenhang zu dem jeweils erzielten Itemscore. Es handelt sich um einen regulär strikt kompensatorischen Test, der gemäß obiger Darstellung bei Verwendung des Kleinsten-Quadrate-Schätzers einem Paradoxon unterliegt. Spezifischer besagt Satz 3.1.3, dass mindestens ein Item existiert, bei dem eine Erhöhung (Verringerung) des Itemscores den Schätzwert bezüglich mindestens einer Dimension verringert (erhöht)- obwohl alle Dimensionen positiv zur Beantwortung der Fragen beitragen.
Der vorliegende Fall demonstriert die „konservative“ Ausrichtung des Satzes: Obwohl lediglich die Existenz *eines* „unfairen“ Items garantiert ist, gilt im SPS-J der paradoxe Effekt für *jedes* der $k+1 = 32$ Items. Ein bemerkenswert groteskes Beispiel liefert dabei das erste, nach dem Drogen- bzw. Alkoholkonsum des betreffenden Jugendlichen fragende Item. Ein höherer Score auf diesem Item (ceteris paribus) erhöht zwar den Schätzwert der zweiten Dimension (aggressives-dissoziales Verhalten), verringert aber zugleich die Schätzwerte der übrigen Dimensionen. Letzteres kann absurde Konsequenzen für eine Individualdiagnostik nach sich ziehen: Ein erhöhter Score auf dem ersten Item, d.h. „erhöhter Drogenkonsum“ kann dazu führen, dass der Schätzwert für die Dimension „Selbstwertprobleme“ (aber auch für die Dimension Ärgerkontrollprobleme) fällt. Mit anderen Worten: Eine Person, bei der Selbstwertprobleme

[22]Genau genommen besitzt eines der 32 Items eine (geringfügig) negative Ladung auf einer Dimension. Dies ändert jedoch qualitativ nichts an den folgenden Analysen.

aufgrund ihres Antwortmusters diagnostiziert wurden, hätte dies u.U. durch die Angabe eines höheren Drogenkonsums vermeiden können.
Bevor die Übertragung der bisherigen, im Rahmen des Kleinste-Quadrate-Schätzers behandelten Resultate auf den ML-Schätzer bei potentiell verschiedenen Messfehlervarianzen erfolgt, soll an dieser Stelle ein Eindruck über die Stärke des paradoxen Effekts in diesem Datensatz vermittelt werden.
Zur Kennzeichnung der Stärke des paradoxen Effekts im SPS-J erfolgt für jedes Item eine Betrachtung des maximalen Zuwachs, der durch eine Erhöhung des Itemscores des entsprechenden Items um eine Einheit erzielt werden kann, sowie eine Betrachtung des maximalen paradoxen Effekts (die maximale Verringerung), der durch diese Erhöhung induziert wird, d.h. der Zuwachs in der am meisten „profitierenden“ Dimension wird kontrastiert mit der Abnahme in der Dimension, die am stärksten reduziert wird. Um eine einfache Kenngröße zu erhalten, wird für jedes Item der Quotient aus maximal induzierter Abnahme (Zähler) und maximal induzierter Zunahme (Nenner) als Maß des paradoxen Effekts gebildet. Im vorliegenden Fall des SPS-J liegt der kleinste Quotient - korrespondierend zu dem Item mit dem geringsten (relativen) paradoxen Effekt - bei 8% und der größte Quotient - korrespondierend zu dem Item mit dem stärksten (relativen) paradoxen Effekt - bei 85%. Median und Mittelwert der Quotienten liegen bei 34% respektive 38%. Eine Person wird somit durch eine (um eine Einheit) bessere Antwort bei einem Item in einer Dimension l belohnt, erfährt jedoch gleichzeitig eine Abnahme in einer anderen Dimension $j \neq l$, die im Regelfall etwa ein Drittel der Zunahme in der Dimension l beträgt. Dies sollte einen groben Eindruck über die (relative) Stärke des paradoxen Effekts im SPS-J vermitteln.

Auch wenn die bisherige Diskussion auf den Kleinste-Quadrate-Schätzer beschränkt war, so lässt sich - man beachte, dass Ω Diagonalstruktur besitzt - mittels der Beobachtung, dass $\Omega^{-\frac{1}{2}}A$ genau dann Einfachstruktur aufweist, wenn selbiges für A gilt, obiger Satz auf den ML-Schätzer erweitern:

Satz 3.1.4. *Bei Gültigkeit der Annahmen i), ii) und iii) ist der Gewichtete-Kleinste-Quadrate-Schätzer - respektive der ML-Schätzer unter Normalverteilungsannahme - genau dann fair, wenn die Ladungsmatrix Einfachstruktur aufweist.*
Folglich ist jedem regulär strikt kompensatorischen Modell innerhalb der Klasse der regulär kompensatorischen faktorenanalytischen Modelle der paradoxe Effekt inhärent.

Angewandt auf obige Diskussion des paradoxen Effekts im Rahmen des SPS-J lässt sich nach Satz 3.1.4 festhalten, dass die implizit unterstellete Gleichheit der Messfehlervarianzen im Hinblick auf die Existenz des Paradoxons keine Rolle spielt. Unter jeder beliebig gearteten Konstellation der Messfehlervarianzen weist der Test einen paradoxen Effekt auf. Nichtsdestotrotz kommt den Messfehlervarianzen - wie im Folgenden darzustellen sein wird - eine entscheidende Bedeutung bezüglich der (absoluten) Stärke des paradoxen Effekts zu. Wenngleich die formale Argumentation für diese Behauptung noch aussteht, lässt sich informell festhalten, dass eine höhere Messfehlervarianz mit einer weniger zuverlässigen Messung einhergeht. Folglich sollten Messungen mit verringerter Präzision schwächer in die Gesamtinferenz eingehen, was wiederum mit der Erwartung eines geringeren Effekts bei einer Veränderung des betreffenden Itemscores korrespondiert.

Stärke des paradoxen Effekts in faktorenanalytischen Modellen mit kompensatorischer Ladungsmatrix

Um den skizzierten Sachverhalt formal darzustellen, betrachte man die Funktion (mit: $W_{i,j} = \delta_{ij}\xi_j$)

$$\Delta_l(\boldsymbol{\xi}) := \boldsymbol{e_l^T}(A^T W(\boldsymbol{\xi})A)^{-1}A^T W(\boldsymbol{\xi})\boldsymbol{e_j}$$

Diese Funktion gibt in Abhängigkeit der Messpräzisionen $\boldsymbol{\xi}$ (= inverse Messfehlervarianzen) die Änderung der Fähigkeitsschätzung in der l-ten Komponente $\hat{\theta}_l$ bei einer Erhöhung des Itemscores des j-ten Items um eine Einheit wieder. Die zunächst informell aufgestellte Behauptung bezüglich der Monotonie dieser Funktion wird im nachfolgenden Lemma formalisiert und bewiesen[23].

Lemma 3.1.6. *Die Stärke des paradoxen Effekts bezüglich der l-ten Dimension steht - vorausgesetzt die Matrix $(A_{-j})^T A_{-j}$ (A_{-j} bezeichne die Matrix, die aus A durch Streichen der j-ten Zeile hervorgeht) besitzt vollen Spaltenrang - in einem streng monoton wachsenden Zusammenhang zur Messpräzision ξ_j des „verursachenden“ Items. Es gilt:*

$$\frac{\partial}{\partial \xi_j}\Delta_l(\boldsymbol{\xi}) = \boldsymbol{e_l}^T(A^T W(\boldsymbol{\xi})A)^{-1}\boldsymbol{a_j}(1 - \boldsymbol{a_j}^T(A^T W(\boldsymbol{\xi})A)^{-1}\boldsymbol{a_j}\xi_j),$$

[23] Nachfolgend wird stets ohne weitere explizite Erwähnung von der Gültigkeit der Annahmen $i), ii)$ und $iii)$ - d.h. von der Gegenwart eines regulär kompensatorischen Modells - ausgegangen.

wobei $W = \Omega^{-1}$ *eine Diagonalmatrix mit Einträgen* $w_{ii} = \xi_i$ *(*$i = 1, \ldots, k+1$*) bezeichnet.*

Beweis. Im Folgenden sei $\Delta(\boldsymbol{\xi}) := (A^T W(\boldsymbol{\xi}) A)^{-1} A^T W(\boldsymbol{\xi}) \boldsymbol{e_j}$, so dass die l-te Komponente von $\Delta(\boldsymbol{\xi})$ mit $\Delta_l(\boldsymbol{\xi})$ übereinstimmt. Die behauptete Monotonie wird mittels Berechnung der Ableitung der Funktion $\Delta(\boldsymbol{\xi})$ nach ξ_j bewiesen.
Unter Verwendung der Ableitungsregeln im Falle Matrizen- bzw. Vektor-wertiger Funktionen (siehe z.B. Lax, 2007, Kap. 9) ergibt sich zunächst:

$$\frac{\partial}{\partial \xi_j} \Delta(\boldsymbol{\xi}) = \frac{\partial}{\partial \xi_j} \left[(A^T W(\boldsymbol{\xi}) A)^{-1} \right] A^T W(\boldsymbol{\xi}) \boldsymbol{e_j} + (A^T W(\boldsymbol{\xi}) A)^{-1} \frac{\partial}{\partial \xi_j} \left[A^T W(\boldsymbol{\xi}) \boldsymbol{e_j} \right]. \tag{3.27}$$

Bei Verwendung der Identität $B(\boldsymbol{\xi}) B^{-1}(\boldsymbol{\xi}) = I$ ($B(\boldsymbol{\xi}) := A^T W(\boldsymbol{\xi}) A$) lässt sich der erste Ableitungsterm in (3.27) wie folgt schreiben:

$$\begin{aligned} \frac{\partial}{\partial \xi_j} \left[(A^T W(\boldsymbol{\xi}) A)^{-1} \right] &= -(A^T W(\boldsymbol{\xi}) A)^{-1} \frac{\partial}{\partial \xi_j} \left[A^T W(\boldsymbol{\xi}) A \right] (A^T W(\boldsymbol{\xi}) A)^{-1} \\ &= -(A^T W(\boldsymbol{\xi}) A)^{-1} \boldsymbol{a_j} \boldsymbol{a_j}^T (A^T W(\boldsymbol{\xi}) A)^{-1}. \end{aligned}$$

Der zweite Ableitungsterm in (3.27) besitzt die Struktur

$$\frac{\partial}{\partial \xi_j} \left[A^T W(\boldsymbol{\xi}) \boldsymbol{e_j} \right] = \boldsymbol{a_j}.$$

Einsetzen in (3.27) und Linksmultiplikation mit $\boldsymbol{e_l^T}$ führt auf

$$\begin{aligned} \frac{\partial}{\partial \xi_j} \Delta_l(\boldsymbol{\xi}) &= -\boldsymbol{e_l}^T B^{-1}(\boldsymbol{\xi}) \boldsymbol{a_j} \boldsymbol{a_j}^T B^{-1}(\boldsymbol{\xi}) A^T W(\boldsymbol{\xi}) \boldsymbol{e_j} + \boldsymbol{e_l}^T B^{-1}(\boldsymbol{\xi}) \boldsymbol{a_j} \\ &= \boldsymbol{e_l}^T B^{-1}(\boldsymbol{\xi}) \boldsymbol{a_j} (1 - \boldsymbol{a_j}^T B^{-1}(\boldsymbol{\xi}) A^T W(\boldsymbol{\xi}) \boldsymbol{e_j}) \\ &= \boldsymbol{e_l}^T B^{-1}(\boldsymbol{\xi}) \boldsymbol{a_j} (1 - \boldsymbol{a_j}^T B^{-1}(\boldsymbol{\xi}) \boldsymbol{a_j} \xi_j). \end{aligned} \tag{3.28}$$

Die Matrix $A^T W(\boldsymbol{\xi}) A$ besitzt die Darstellung $A^T W(\boldsymbol{\xi}) A = ((A_{-j})^T W^*(\boldsymbol{\xi}) A_{-j} + \xi_j \boldsymbol{a_j} \boldsymbol{a_j}^T)$. Aufgrund der Annahme $rg(A_{-j}) = p$ ist die Matrix $(A_{-j})^T W^*(\boldsymbol{\xi}) A_{-j}$ invertierbar und es gilt (siehe z.B. Mardia, Kent und Bibby, 1979, S.458):

$$(A^T W(\boldsymbol{\xi}) A)^{-1} = ((A_{-j})^T W^*(\boldsymbol{\xi}) A_{-j})^{-1} - \frac{((A_{-j})^T W^*(\boldsymbol{\xi}) A_{-j})^{-1} \boldsymbol{a_j} \boldsymbol{a_j^T} ((A_{-j})^T W^*(\boldsymbol{\xi}) A_{-j})^{-1}}{\xi_j^{-1} + \boldsymbol{a_j^T} ((A_{-j})^T W^*(\boldsymbol{\xi}) A_{-j})^{-1} \boldsymbol{a_j}}$$

Demnach ist $\boldsymbol{e_l^T} (A^T W(\boldsymbol{\xi}) A)^{-1} \boldsymbol{a_j}$ identisch mit

$$\boldsymbol{e_l^T} ((A_{-j})^T W^*(\boldsymbol{\xi}) A_{-j})^{-1} \boldsymbol{a_j} \left(1 - \frac{\boldsymbol{a_j^T} ((A_{-j})^T W^*(\boldsymbol{\xi}) A_{-j})^{-1} \boldsymbol{a_j}}{\xi_j^{-1} + \boldsymbol{a_j^T} ((A_{-j})^T W^*(\boldsymbol{\xi}) A_{-j})^{-1} \boldsymbol{a_j}} \right).$$

Da der Ausruck in Klammern offenbar stets im offenen Einheitsintervall liegt, folgt

$$sgn(\boldsymbol{e_l}^T B^{-1}(\boldsymbol{\xi})\boldsymbol{a_j}) = sgn(\boldsymbol{e_l}^T((A_{-j})^T W^*(\boldsymbol{\xi})A_{-j})^{-1}\boldsymbol{a_j}).$$

Der von ξ_j unabhängige Ausdruck auf der rechten Seite kennzeichnet nach Satz 3.1.1, ob ein Paradoxon bei Veränderung des Itemscores des j-ten Items bezüglich der l-ten Dimension auftritt.
Der Term in Klammern in (3.28) kann mittels der Substitution $\boldsymbol{d_j} := \boldsymbol{a_j}\xi_j^{\frac{1}{2}}$, $D := W(\boldsymbol{\xi})^{\frac{1}{2}}A$ wie folgt dargestellt werden:

$$\boldsymbol{a_j}^T B^{-1}(\boldsymbol{\xi})\boldsymbol{a_j}\xi_j = \boldsymbol{d_j}^T(D^T D)^{-1}\boldsymbol{d_j}.$$

Als Diagonalterm einer „Hat“-Matrix ist der Ausdruck der rechten Seite stets nicht größer als Eins. Die Behauptung des Lemmas wird somit folgen, wenn der Fall $\boldsymbol{d_j}^T(D^T D)^{-1}\boldsymbol{d_j} = 1$ ausgeschlossen werden kann.
Angenommen es gilt $\boldsymbol{d_j}^T(D^T D)^{-1}\boldsymbol{d_j} = 1$, dann sind die übrigen Einträge der j-ten Zeile aufgrund der Idempotenz und Symmetrie der Hat-Matrix gleich Null. Die Vektoren $\boldsymbol{d_i}$ liegen somit orthogonal zum Vektor $\boldsymbol{z} := (D^T D)^{-1}\boldsymbol{d_j}$ $(\forall\, i \neq j)$. Die Menge aller orthogonal zum Vektor $\boldsymbol{z}$ liegenden Vektoren bildet einen $p-1$ dimensionalen Untervektorraum. Da die lineare Hülle der Vektoren $\{\boldsymbol{d_i}|i \neq j\}$ identisch mit der linearen Hülle der Vektoren $\{\boldsymbol{a_i}|i \neq j\}$ ist, folgt $rg(A_{-j}) < p$. Letzteres steht im Widerspruch zur Annahme $rg(A_{-j}) = p$. Somit ist der Term in Klammern in (3.28) stets echt positiv und die Behauptung folgt. □

Auch wenn gemäß Satz 3.1.1 die Messfehlervarianz des veränderten Items im Hinblick auf das Auftreten des paradoxen Effekts irrelevant ist, so steht sie dennoch - via Regulation der *Stärke* des paradoxen Effekts - in einem engen Zusammenhang zum paradoxen Effekt. Treten in einem Test paradoxe Effekte auf und ist man an den „verursachenden“ Items interessiert, so legt Lemma 3.1.6 nahe, zunächst die am wenigsten mit Messfehler behafteten Items zu betrachten.
Es sei jedoch bemerkt, dass Lemma 3.1.6 von der *absoluten* Stärke des paradoxen Effekts handelt. Besitzt das j-te Item bezüglich der l-ten Dimension einen paradoxen Effekt, so gewinnt dieser mit steigender Messpräzision an (absoluter) Stärke. Weist andererseits dasselbe Item bezüglich der v-ten Dimension keinen paradoxen Effekt auf, so wächst der korrespondierende Schätzwert mit steigendem Itemscore und diese Tendenz fällt wiederum umso stärker aus, je größer die Messpräzision des Items ist. Die relative Stärke setzt - analog zur Vorgehensweise am Beispiel des SPS-J - diese

beiden, auf unterschiedlichen Dimensionen beruhenden Tendenzen in Beziehung. Im Gegensatz zur absoluten Stärke des paradoxen Effekts ist die relative Stärke - wie im Folgenden gezeigt werden wird - unabhängig von der Messpräzision des Items.

Korollar 3.1.1. *Unter den Annahmen von Lemma 3.1.6 verläuft die Funktion*

$$f(\boldsymbol{\xi}) := \frac{\Delta_l(\boldsymbol{\xi})}{\Delta_{l'}(\boldsymbol{\xi})},$$

die die relative Stärke des paradoxen Effekts (bei Veränderung des j-ten Itemscores) in Abhängigkeit der Messpräzisionen wiedergibt, konstant bezüglich ξ_j.

Beweis. Eine einfache Rechnung unter Verwendung der Quotientenregel in Kombination mit dem Resultat aus Lemma 3.1.6 führt auf die Behauptung. □

Dieses Resultat schließt die Betrachtung des paradoxen Effekts in der faktorenanalytischen Modellklasse ab. Der folgende Abschnitt wendet sich - wie bereits vor der Besprechung des faktorenanalytischen Modells angekündigt wurde - der Diskussion des paradoxen Effekts in der allgemeinen Klasse linear-kompensatorischer Modelle zu. Im Vordergrund steht dabei die Frage, ob ein zu Satz 3.1.4 äquivalentes Resultat, welches die Existenz eines paradoxen Effekts garantiert, auch in der größeren Klasse der linear-kompensatorischen Modelle nachgewiesen werden kann.

Existenz des paradoxen Effekts in linear-kompensatorischen Modellen

Gemäß (3.20) tritt im Rahmen eines linear-kompensatorischen Modellen ein paradoxer Effekt bezüglich der ersten Fähigkeitskomponente mit Sicherheit auf, wenn ein Ladungsvektor $\boldsymbol{b}$ existiert, so dass für alle Diagonalmatrizen W mit negativen Einträgen die Beziehung

$$\boldsymbol{e_1^T}(A_0^T W A_0)^{-1}\boldsymbol{b} > 0$$

gilt, bzw. äquivalent dazu

$$\boldsymbol{e_1^T}(A_0^T W A_0)^{-1}\boldsymbol{b} < 0$$

für alle positiven Diagonalmatrizen W.
Man nehme für einen $p = 2$-dimensionalen Test zunächst an, dass sämtliche Ladungsvektoren die Länge Eins besitzen und wähle aus diesen Ladungsvektoren, denjeniegen

Vektor ($\boldsymbol{b}$) aus, dessen erste Komponente den geringsten Wert (relativ zu den übrigen Ladungsvektoren) aufweist. A_0 bezeichne die Ladungsmatrix bestehend aus den verbleibenden Vektoren.
Mit diesen Definitionen betrachte man die erste Komponente des Vektors $\boldsymbol{x}$, der die Gleichung $A_0^T W A_0 \boldsymbol{x} = \boldsymbol{b}$ erfüllt.[24]
Nach Cramers Regel besitzt die erste Komponente (bis auf einen positiven Faktor $c = (\det(A_0^T W A_0))^{-1}$) die Darstellung

$$x_1 = \det \begin{pmatrix} b_1 & \sum_i w_i a_{i1} a_{i2} \\ b_2 & \sum_i w_i a_{i2}^2 \end{pmatrix} = b_1 \sum_i w_i a_{i2}^2 - b_2 \sum_i w_i a_{i1} a_{i2}$$

Wegen $b_1 < a_{i1}$ sowie $b_2 > a_{i2}$ für alle i folgt

$$x_1 = \sum_i b_1 w_i a_{i2}^2 - \sum_i b_2 w_i a_{i1} a_{i2} < \sum_i a_{i1} w_i a_{i2}^2 - \sum_i w_i a_{i1} a_{i2}^2 = 0.$$

Somit gilt $x_1 < 0$ unabhängig von der spezifischen Ausprägung der (positiven) Gewichte w_i.
Die dieser Argumentation zugrundeliegende Annahme gleichlanger Itemvektoren ist ferner stets durch Skalarmultiplikation (mit einem Faktor $\lambda_i > 0$) erfüllbar:

$$l_i(\boldsymbol{a}_i^T \boldsymbol{\theta}) = l_i^*(\boldsymbol{a_i^{*T}} \boldsymbol{\theta}), \quad l_i^*(x) := l_i(\lambda_i^{-1} x), \ \boldsymbol{a_i^*} := \lambda_i \boldsymbol{a}_i.$$

Da die so entstandene Funktionen des Typs l_i^* ebenfalls konform mit der Forderung $(l_i^*)'' < 0$ sind, kann der Beweis der Existenz gemäß dem oben skizzierten Prinzip erfolgen.
Es ist somit gelungen, im Rahmen eines zweidimensionalen linear-kompensatorischen Modells den paradoxen Effekt nachzuweisen. Wählt man das Item mit der geringsten (relativen) Ladung auf der ersten Dimension aus, so geht eine Scoreveränderung mit einem paradoxen Effekt auf der ersten Dimension einher. Das „intuitive“ Prinzip, dasjeniege Item auszuwählen, welches in geringster Abhängigkeit zur ersten Dimension steht, um so einen paradoxen Effekt bezüglich der ersten Dimension zu verursachen, scheint auf den ersten Blick generalisierbar. In der Tat, Korollar 6.1 von Hooker u.a. (2009) behauptet, dass ein analoges Resultat in höherdimensionalen linear-kompensatorischen Modellen Gültigkeit besitzt. Ordnet man die Items gemäß ihrer relativen Ladung auf der ersten Dimension, d.h. nach der Größe $\frac{a_{i1}}{|\boldsymbol{a}_i|}$, an, und bezeichnet $\boldsymbol{b}$ den Ladungsvektor des ersten Items bezüglich dieser Reihenfolge, so

[24] Die Existenz und Eindeutigkeit ist gemäß der „üblichen“ Rangbedingung stets gewährt.

besagt das Korollar, dass eine Veränderung bezüglich dieses Items einen paradoxen Effekt auf der ersten Dimension hervorruft. Der Beweis des entsprechenden Korollars ist jedoch - wie im Folgenden dargelegt wird - fehlerhaft.

Grundlage für den Beweis ist die Behauptung (siehe Zwischenbemerkung in Lemma D.2 von Hooker u.a (2009)), dass für jede Auswahl von $p-1$ linear unabhängigen Ladungsvektoren $\boldsymbol{a_{s(2)}}, \ldots, \boldsymbol{a_{s(p)}}$ die erste Komponente der Projektion von $\boldsymbol{b}$ auf das orthogonale Komplement des Spaltenraums[25], der von diesen Ladungsvektoren aufgespannt wird, negativ ausfällt. Gleichwohl die Behauptung im Zweidimensionalen korrekt ist, kann bereits für $p = 3$ ein Gegenbeispiel konstruiert werden. Man setze hierzu

$$\boldsymbol{b} = \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix}, \ \boldsymbol{a_{s(2)}} = \begin{pmatrix} 5 \\ 3 \\ 1 \end{pmatrix}, \ \boldsymbol{a_{s(3)}} = \begin{pmatrix} 2 \\ 2 \\ 4 \end{pmatrix}.$$

Die erste Komponente der Projektion von $\boldsymbol{b}$ auf den von $\boldsymbol{a_{s(2)}}$ und $\boldsymbol{a_{s(3)}}$ aufgespannten Unterraum beträgt (auf zwei Nachkommastellen gerundet) $b_{s1} = 0.91$, so dass die erste Komponente der Projektion von $\boldsymbol{b}$ auf das orthogonale Komplement $b_1 - b_{s1} = 1 - 0.91 = 0.09 > 0$ ist. Obwohl $\boldsymbol{b}$ die geringste relative Ladung auf der ersten Dimension aufweist, besitzt es somit - entgegen der Behauptung von Hooker u.a. (2009) - eine positive erste Komponente bezüglich des orthogonalen Komplementents.

Des Weiteren muss die im Zuge des Beweises von Korollar 6.1 gegebene Bemerkung, dass $\boldsymbol{e}_1^T(A_0^T W A_0)^{-1}\boldsymbol{b} < 0$ für alle positiven[26] Diagonalmatrizen W gültig ist, insofern $\boldsymbol{b}$ gemäß der obigen Prozedur gewählt wurde, angezweifelt werden. Um auch hierfür ein Gegenbeispiel zu erhalten, ergänze man die oben angegebenen Vektoren durch den Vektor $\boldsymbol{a}_4^T := (1, 1, 1)$. Bezeichnet A_0 die Ladungsmatrix bestehend aus den Vektoren $\boldsymbol{a}_{s(2)}, \boldsymbol{a_{s(3)}}$ sowie $\boldsymbol{a}_4$, so resultiert mit $W := I_3$ (3×3 Einheitsmatrix) $\boldsymbol{e}_1^T(A_0^T W A_0)^{-1}\boldsymbol{b} = 3 > 0$.

Die bisherigen Erörterungen beziehen sich nicht direkt auf die Aussage des Korollars, sondern tangieren dessen Beweis. Betrachtet man jedoch die angegebenen Itemparameter als Ladungsvektoren in einem faktorenanalytischen Modell (Spezialform eines

[25] In der Terminologie von Hooker u.a. (2009) ausgedrückt: $\boldsymbol{b_1} - \boldsymbol{b_{s1}} < 0$.

[26] Man beachte, dass (äquivalent zur obigen Darstellung) in Hooker u.a. (2009) die Gültigkeit von $\boldsymbol{e}_1^T(A_0^T W A_0)^{-1}\boldsymbol{b} > 0$ für alle *negativen* Diagonalmatrizen behauptet wird.

linear-kompensatorischen Modells), d.h. formt man die Ladungsmatrix

$$A := \begin{pmatrix} 1 & 1 & 3 \\ 5 & 3 & 1 \\ 2 & 2 & 4 \\ 1 & 1 & 1 \end{pmatrix}$$

und berechnet darauf basierend die zugehörige „Kleinste-Quadrate-Matrix“

$$(A^T A)^{-1} A^T = \begin{pmatrix} 1.00 & 0.50 & -0.50 & -1.50 \\ -1.83 & -0.50 & 0.83 & 2.68 \\ 0.50 & 0.00 & 0.00 & -0.50 \end{pmatrix},$$

so erkennt man zwar in Übereinstimmung mit Satz 3.1.3, dass mindestens ein Item existiert, welches paradoxe Effekte induziert. Der nach Hooker u.a. (2009) zu antizipierende Effekt bleibt jedoch aus: Eine Erhöhung des Itemscores des ersten Items (= das Item mit geringster relativer Ladung auf der ersten Dimension) um eine Einheit bewirkt eine Erhöhung des Schätzwerts bezüglich der ersten Dimension um eine Einheit ($\widehat{=}$ Eintrag 1.00). Der Effekt einer Erhöhung ist ferner doppelt so hoch wie der korrespondierende Effekt bei einer Erhöhung des Itemscores des zweiten Items, welches „massivst“ auf der ersten Dimension lädt.

Neben der Tatsache, dass das erste Item keinen paradoxen Effekt induziert, liefert das Gegenbeispiel eine verblüffende Einsicht: Eine Erhöhung des Itemscores eines (relativ) gering ladenden Items (hier: das erste Item) kann eine stärkere Erhöhung des zugehörigen Schätzwerts hervorrufen als eine vergleichbare Veränderung auf einem hoch ladenden Items (hier: Item 2)!

Die Veränderung der Schätzwerte einzelner Dimensionen steht - wie obige Beispiele erahnen lassen - in einer komplizierten Relation zu den Itemparametern. Verschiebt man jedoch den Akzent von der dimensionsspezifischen Etablierung des Paradoxons hin zu einer globalen Fragestellung, die analog zu Satz 3.1.3 lediglich die *Existenz* des paradoxen Effekts tangiert, so kann der paradoxe Effekt auch innerhalb der höherdimensionalen linear-kompensatorischen Modelle nachgewiesen werden. Letzteres soll an dieser Stelle abschließend erläutert werden.

Man nehme hierzu die Existenz eines Ladungsvektors $\boldsymbol{a}_e$ ($e \in \{1, 2, \dots k\}$) sowie $p-1$ nichtnegativer Vektoren $\boldsymbol{c}_2, \dots, \boldsymbol{c}_p$ an, die zusammen einen konvexen, p-dimensionalen Kegel aufspannen, der sämtliche Ladungsvektoren $\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_k$ umfasst. Jeder Ladungsvektor $\boldsymbol{a}_i$ lässt sich demnach als eindeutige Linearkombination

mit *nichtnegativen* Gewichten schreiben ($\boldsymbol{c}_1 := \boldsymbol{a}_e$):

$$\boldsymbol{a}_i = \sum_{j=1}^{p} \lambda_{ij} \boldsymbol{c}_j.$$

Mit der Transformation $\psi_v := \boldsymbol{c}_v^T \boldsymbol{\theta}$ (für $v = 1, \ldots, p$) kann die ursprüngliche Loglikelihood

$$l(\boldsymbol{\theta}) = \sum_i l_i(\boldsymbol{a}_i^T \boldsymbol{\theta})$$

in den Ausdruck

$$l^*(\boldsymbol{\psi}) = \sum_{i \neq e} l_i(\boldsymbol{\lambda}_i^T \boldsymbol{\psi}) + l_e(\psi_1), \tag{3.29}$$

in dem $\boldsymbol{\lambda}_i$ den Vektor mit Komponenten λ_{ij} ($j = 1, \ldots, p$) bezeichnet, überführt werden.

Gemäß den Ausführungen des Anhangs D ist obige Loglikelihood paarweise SMDD in ψ_1, wenn für jede Dimension $j \neq 1$ mindestens ein Item existiert, welches auf der ersten und zugleich auf der j-ten Dimension *positiv* lädt.

Ist diese (schwache) Bedingung erfüllt, so ist l^* nach Lemma 2.3.6 eSMDD in ψ_1 und die Existenz eines paradoxen Effekts bezüglich der $\boldsymbol{\psi}$-Parametrisierung folgt gemäß Satz 2.3.2. Mittels der invertierbaren Matrix C, deren i-ten Zeile durch den Vektor $\boldsymbol{c}_i$ gebildet wird (für $i = 1, \ldots, p$), lassen sich die zu maximierenden Funktionen l und l^* in eine 1-1-Beziehung der Form

$$l^*(\boldsymbol{\psi} = C\boldsymbol{\theta}) = l(\boldsymbol{\theta})$$

bringen. Bezeichnet $\hat{\boldsymbol{\psi}}$ die Stelle des Maximums von l^*, so ist $\hat{\boldsymbol{\theta}} = C^{-1}\hat{\boldsymbol{\psi}}$ die Stelle des Maximums von l und umgekehrt. Diese Beziehung gilt zwangsläufig auch für die jeweiligen Szenarien, in denen der Itemscore von Item e verändert wird. Da in der $\boldsymbol{\psi}$-Metrik ein Paradoxon bei Umwandlung des Scores resultiert und da die Matrix C ausschließlich nichtnegative Einträge beinhaltet, muss demnach auch bezüglich der $\boldsymbol{\theta}$-Parametrisierung ein paradoxer Effekt resultieren.

Diese Argumentation zeigt abschließend somit auch die Existenz des Paradoxons in höherdimensionalen Modellen auf.

3.2 Paradoxien in Item-Response-Modellen mit zusätzlicher Schnelligkeitskomponente

Die im vorherigen Abschnitt betrachtete Modellklasse der linear-kompensatorischen Modelle stellt zwar die in der Praxis am häufigsten eingesetzte IRT-Modellklasse dar, sie ist jedoch auf den „klassischen“ Fall eines (Querschnitts-)IRT-Modells, in dem Antwortmuster auf homogene Items anhand des Latenten-Variablen-Paradigmas modelliert werden, beschränkt. „Neuere“, in den folgenden Abschnitten zu betrachtende Formen der Item-Response-Modellierung erweitern diese klassische Struktur um verschiedene Aspekte.

Ein Beispiel für eine solche Erweiterung stellen IRT-Modelle dar, die zusätzlich zur Erhebung des „üblichen“ Antwortmusters $u_1, \dots u_k$ die itemspezifischen Antwortzeiten $t_1, \dots t_k$ der Person miterfassen. Wenn Letztere in einer direkten oder indirekten (stochastischen) Beziehung zur Fähigkeit der Person stehen, dann liefert der Einbezug von Responsezeiten in die Modellierung u.U. einen wichtigen Beitrag im Hinblick auf eine Effizienzsteigerung der Diagnose.

Wenngleich die Menge der potentiellen Response-Responsezeit-Modelle aufgrund der Tatsache, dass prinzipiell jedes „gewöhnliche“ IRT-Modell (z.B. 3PL-Modell) für $u_1 \dots u_k$ mit einem entsprechenden Modell der Antwortzeiten (z.B. exponentialverteilte Antwortzeiten, deren Erwartungswert durch eine Funktion der latenten Variablen parametrisiert wird) unter Nutzung lokaler stochastischer Unabhängigkeit kombinierbar ist, eine sehr heterogene Klasse darstellt, so zeichnen sich dennoch in der aktuellen Forschung gewisse Präferenzen (siehe z.B. Fox, 2010, Kap.8; Klein Entink, Kuhn, Hornke und Fox, 2009; van der Linden, 2007) in der Modellierung ab, denen im Folgenden anhand der Auswahl der zu diskutierenden Modelle Rechnung getragen werden soll.

Als gemeinsame Basisannahme aller nachfolgend zu betrachtenden Modelle fungiert die Grundannahme lokal stochastischer Unabhängigkeit zwischen $u_1, \dots u_k, t_1 \dots, t_k$ gegeben ein i.d.R. mehrdimensionaler, aus latenten Variablen bestehender Vektor $\tilde{\boldsymbol{\theta}} = (\boldsymbol{\theta}, \tau)$. Als Konsequenz dieser Annahme ergibt sich die folgende Zerlegung der Loglikelihood

$$l(\tilde{\boldsymbol{\theta}}) = \sum_i \log(f_{i,u_i}(\tilde{\boldsymbol{\theta}})) + \sum_i \log(g_i(t_i|\tilde{\boldsymbol{\theta}})), \tag{3.30}$$

wobei die Notation symbolisch die i.A. vorherrschende Verschiedenheit im Skalenniveau widerspiegelt. $f_{i,u_i}(\tilde{\boldsymbol{\theta}})$ bezeichnet die Wahrscheinlichkeit, auf dem i-ten Item

Kategorie u_i zu erzielen, während $g_i(t_i|\tilde{\boldsymbol{\theta}})$ die Dichtefunktion (ausgewertet an der Stelle der realisierten Antwortszeit t_i) der stetig modellierten Antwortzeit bezüglich des i-ten Items wiedergibt. Die Partitionierung des Vektors $\tilde{\boldsymbol{\theta}}$ spiegelt den Umstand wider, dass über den „gewöhnlichen“ Vektor latenter Variablen $\boldsymbol{\theta}$ hinaus, eine spezifische, ebenfalls latente Variable τ vorliegt, die i.d.R als Schnelligkeit der Person interpretierbar ist.
Im Folgenden wird stets *zur sprachlichen Vereinfachung* der Begriff „Schnelligkeitskomponente“ synonym für τ sowie der Begriff „Fähigkeitskomponente“ für $\boldsymbol{\theta}$ verwendet. Es sei jedoch einschränkend zu diesem Sprachgebrauch bemerkt, dass gleichwohl Situationen existieren können, in denen es sich um semantische Fehlbezeichnungen handeln kann (z.B. wenn $\boldsymbol{\theta}$ keine Fähigkeit repräsentiert).
Im Gegensatz zu dieser strikten Unterscheidung zwischen Fähigkeit und Schnelligkeit steht die symmetrische Form der Grundgleichung, in der beide Größen sowohl in den üblichen, durch $u_1, \dots u_k$ gegebenen Responseteil als auch in die stochastische Modellierung der Antwortzeiten einfließen. Eine gängige zweite Annahme, die der semantischen Verschiedenheit der Parameter Rechnung trägt, ist daher durch die zusätzliche Forderung der Unabhängigkeit des Ausdrucks $f_{i,u_i}(\tilde{\boldsymbol{\theta}})$ $(\forall i)$ von τ gegeben. Mit anderen Worten: Die Schnelligkeitskomponente bestimmt lediglich die Antwortzeiten und spielt keine Rolle bei der Frage, wie akkurat ein Item bearbeitet wird. Die korrespondierende Loglikelihood besitzt daher die vereinfachte Form:

$$l(\tilde{\boldsymbol{\theta}}) = \sum_i \log(f_{i,u_i}(\boldsymbol{\theta})) + \sum_i \log(g_i(t_i|\tilde{\boldsymbol{\theta}})), \tag{3.31}$$

Die analoge Forderung der Unabhängigkeit des Terms $g_i(t_i|\tilde{\boldsymbol{\theta}})$ von $\boldsymbol{\theta}$ gibt den Sachverhalt wieder, dass bei gegebener Schnelligkeitskomponente Antwortzeiten (in einem probabilistischen Sinn) unabhängig von der Fähigkeit $\boldsymbol{\theta}$ der Person sind und führt somit auf ein Teilmodell von (3.31) der Form

$$l(\tilde{\boldsymbol{\theta}}) = \sum_i \log(f_{i,u_i}(\boldsymbol{\theta})) + \sum_i \log(g_i(t_i|\tau)). \tag{3.32}$$

Während sich (nahezu) jedes in der aktuellen Forschung prävalente Modell in die durch Modell (3.31) repräsentierte Klasse einordnen lässt (siehe z.B. den Überblick von Ranger, 2009) wird die zusätzliche Unabhängigkeitsannahme in (3.32) nicht von allen gängigen Modellen geteilt. Nichtsdestotrotz stellt (3.32) eine Modellklasse dar, die in aktuellen Lehrbüchern (Fox, 2010) sowie aktuellen IRT-bezogenen Forschungsansätzen (van der Linden, 2007; van der Linden und Guo, 2008; van der Linden und

Glas, 2010) vorherrschend ist und die zusätzlich in substanzwissenschaftlichen Anwendungen von Responsezeitmodellen (Klein Entink u.a., 2009) Eingang findet.
Die nachfolgende Fokussierung auf diese Modellklasse erfolgt jedoch nicht primär aus Gründen der Aktualität und Praxisnähe sondern aufgrund der Tatsache, dass mit dieser Modellklasse ein bisher wenig beachteter Aspekt des paradoxen Effekts gekennzeichnet werden kann. Im Gegensatz zu bisherigen Betrachtungen besitzt eine Loglikelihood der Form (vorerst erfolgt eine Beschränkung auf eindimensionales θ)

$$l(\tilde{\boldsymbol{\theta}}) = \sum_i \log(f_{i,u_i}(\theta)) + \sum_i \log(g_i(t_i|\tau)). \tag{3.33}$$

eine separable Struktur, die zwangsläufig bei Verwendung des ML-Schätzers frei von paradoxen Effekten ist. Aufgrund der spezifischen Schätzstruktur[27], die bei der Durchführung der Kalibrierung bzw. Normierung des Tests vorhanden ist, wird jedoch im Rahmen dieser Modellklasse im Regelfall eine bayesianische Schätzung der Parameter propagiert (van der Linden, 2007). Diese Vorgehensweise ist primär dadurch motiviert, dass die Kalibrierungsstichprobe nicht nur dem Zweck der Schätzung der Itemparameter dient, sondern auch Informationen über die Verteilung des latenten Vektors $\tilde{\boldsymbol{\theta}}$ in der Population beinhaltet. Folgt man Standardresultaten der Schätz- bzw. Prädiktionstheorie (z.B. Searle, Casella und McCulloch, 2006, Kap. 7), so lässt sich der Fehler[28] bei der Prädiktion[29] von (θ, τ) durch Einbezug dieses Vorwissens verringern. Je nach gewählter Verlustfunktion ergibt sich dabei entweder der MAP-Schätzer (*maximum a-posteriori*), der EAP-Schätzer (*expected a-posteriori*) oder ein anderer, bayesianisch motivierter Schätzer (siehe z.B. Berger, 1985).
Die folgenden Betrachtungen basieren daher vorrangig auf der Modellstruktur (3.33) in Kombination mit einer logarithmierten Prioriverteilung $\gamma(\theta, \tau)$, die das aus der Kalibrierungsstichprobe vorhandene Vorwissen über die gemeinsame Verteilung der Schnelligkeits- und Fähigkeitskomponente in der Grundgesamtheit reflektiert.
Die logarithmierte Posterioriverteilung besitzt damit die Struktur

$$\gamma^p(\tilde{\boldsymbol{\theta}}) = \sum_i \log(f_{i,u_i}(\theta)) + \sum_i \log(g(t_i|\tau)) + \gamma(\theta, \tau). \tag{3.34}$$

[27] Die Popularität dieses Modellierungszugangs könnte ferner auch in einer entsprechend einfachen Schätzbarkeit der Modellparameter im Rahmen des bayesianischen Paradigmas begründet liegen.

[28] Präziser bedeutet dies, dass basierend auf der Verteilung der latenten Variablen eine Prädiktion mit minimalem zu erwartenden quadratischen Verlust berechnet werden kann (Searle u.a., 2006, S. 261).

[29] Der Term „Prädiktion“ wird anstelle des Ausdrucks „Schätzung“ benutzt, um anzudeuten, dass die unbekannte Größe $\tilde{\boldsymbol{\theta}}$ als Zufallsvariable aufgefasst wird.

Bezüglich dieser Funktion soll nachfolgend der Frage nachgegangen werden, welche Auswirkung eine Veränderung des Itemscores bzw. der Responsezeit auf den MAP-Schätzer besitzt. Es wird dabei o.B.d.A der MAP-Schätzer diskutiert. Im Sinne der Herleitung des paradoxen Effekts im Rahmen der Typ-2-Schätzklasse (basierend auf Funktionalen) ließen sich jedoch die anschließenden Betrachtungen analog am Beispiel des EAP durchführen.
Ein erstes Resultat hinsichtlich der Existenz des paradoxen Effekts kann unmittelbar durch eine Anwendung von Satz 2.3.1 auf die Grundmodellgleichung (3.34) erhalten werden:

Satz 3.2.1. *Es gelte das Modell (3.34). Ferner besitze γ die SMDD-Eigenschaft. Wenn u_i, u_i^* zwei geordnete Antwortkategorien $u_i^* > u_i$ mit der Eigenschaft, dass die Differenz der logarithmierten Antwortwahrscheinlichkeiten $\log(f_{i,u_i^*}(\theta)) - \log(f_{i,u_i}(\theta))$ streng monoton wachsend in θ ist, bezeichnen, dann sind unter Regularitätsbedingungen, die die Existenz der MAP-Werte bzw. der bedingten Maximumsfunktionen garantieren, die zu den beiden (log-)Posterioriverteilungen $\gamma_1^p := \sum_l \log(f_{l,u_l}(\theta)) + \sum_l \log(g_l(t_l|\tau)) + \gamma(\theta,\tau)$, $\gamma_2^p := \gamma_1^p + \log(f_{i,u_i^*}(\theta)) - \log(f_{i,u_i}(\theta))$ korrespondierenden MAP-Schätzer - mit $(\hat{\theta},\hat{\tau})(u_i)$ respektive $(\hat{\theta},\hat{\tau})(u_i^*)$ bezeichnet - konträr geordnet, d.h. es gilt*

$$\hat{\theta}(u_i^*) \geq \hat{\theta}(u_i),\ \hat{\tau}(u_i^*) \leq \hat{\tau}(u_i).$$

*Ist γ hingegen SMID (*strictly monotone increasing differences*), dann sind die Schätzwerte gleichsinnig geordnet, d.h.*

$$\hat{\theta}(u_i^*) \geq \hat{\theta}(u_i),\ \hat{\tau}(u_i^*) \geq \hat{\tau}(u_i)$$

gilt.

Beweis. Satz 2.3.1 ergibt mit der Identifikation $\gamma_1^p = g, \gamma_2^p = f, \log(f_{i,u_i^*}) - \log(f_{i,u_i}) = h$ die Behauptung. Für entsprechende Regularitätsbedingung, die die Existenz der bedingten Maximumsfunktion sowie des MAP sichern, sei auf Anhang A verwiesen. □

Die für die Anwendung des Satzes zentrale Monotoniebedingung kann je nach vorliegendem IRT-Modell auf unterschiedliche Art erfüllt sein.
Liegt beispielsweise ein *binäres* Item vor, dessen Lösungswahrscheinlichkeit streng monoton mit der Fähigkeit θ wächst, so erfüllt $\log(f_{i,u_i=1}(\theta)) - \log(f_{i,u_i=0}(\theta))$ die

Monotoniebedingung des Satzes.
Analog erfüllen die Extremkategorien in einem kumulativen ordinalen IRT-Modell die Bedingung, da $\log(f_{i,u_i=L_i}(\theta))-\log(f_{i,u_i=0}(\theta))$ streng monoton wachsend ist, wenn $f_{i,u_i}(\theta)$ einem *kumulativem* Modell folgt und L_i die höchste Antwortkategorie bezeichnet.
Dass in ordinalen Modellen auch Zwischenkategorien betroffen sein können, zeigt das Beispiel eines *„adjacent category (logit) model“* (Agresti, 2002, S. 286f.). Letzteres modelliert den Logarithmus des Wahrscheinlichkeitsquotienten zweier benachbarter Kategorien als lineare Funktion von θ:

$$\log\left(\frac{f_{i,u_i+1}(\theta)}{f_{i,u_i}(\theta)}\right) = \alpha_{i,u_i}\theta + \beta_{i,u_i}.$$

Wenn $\alpha_{i,u_i} > 0$ gilt, dann ist offenbar die Monotoniebedingung aus Satz 3.2.1 für die beiden benachbarten Antwortkategorien $u_i, u_i + 1$ erfüllt. Wenn dies zudem für alle benachbarte Kategorienpaare gilt, dann genügt *jedes* geordnete Kategorienpaar der Bedingung. Das Gleiche gilt natürlich, wenn die lineare Funktion von θ durch eine streng monoton wachsende Funktion ersetzt wird.
Einen weiteren Sonderfall stellen *sequentielle (Logit)-Modelle* (Tutz, 1991) dar, in denen die Kategorie-Response-Funktionen eines Items mit $L_i + 1$ verschiedenen Kategorien durch den Ausdruck

$$f_{i,u_i}(\theta) = (1 - h_{i,u_i}(\theta)) \prod_{l=0}^{u_i-1} h_{i,l}(\theta) \qquad \text{für } u_i < L_i,$$

bzw. für $u_i = L_i$ durch den Ausdruck

$$f_{i,L_i}(\theta) = \prod_{l=0}^{L_i-1} h_{i,l}(\theta),$$

gegeben sind. Die Terme $h_{i,l}(\theta)$, die anhand eines „gewöhnlichen“ binären IRT-Modells mit streng monoton wachsenden Item-Response-Funktionen modelliert werden, geben hierbei die Wahrscheinlichkeiten wieder, die l-te Kategorie zu überschreiten (gegeben, dass die (l)-te Kategorie erreicht wurde). Der Vergleich einer beliebigen Antwortkategorie $u_i \neq L_i$ mit der „höchsten“ Antwortkategorie L_i führt auf die Beziehung

$$\log(f_{i,L_i}(\theta)) - \log(f_{i,u_i}(\theta)) = \sum_{l>u_i-1} \log(h_{i,l}(\theta)) - \log(1 - h_{i,u_i}(\theta)),$$

die offenbar unter Verwendung der strikten Monotonie der Itemstufenfunktionen h_{i,u_i} die besagte Monotoniebedingung des Satzes 3.2.1 impliziert.
Analoge Resultate gelten im Falle einer Veränderung der Responsezeiten, d.h. wenn für $t_i^* < t_i$ der Ausdruck $\log(g(t_i^*|\tau)) - \log(g(t_i|\tau))$ streng monoton wachsend in τ ist, dann lautet die zu Satz 3.2.1 analoge Ungleichung unter SMDD

$$\hat{\theta}(t_i^*) \leq \hat{\theta}(t_i),\ \hat{\tau}(t_i^*) \geq \hat{\tau}(t_i).$$

bzw. unter SMID

$$\hat{\theta}(t_i^*) \geq \hat{\theta}(t_i),\ \hat{\tau}(t_i^*) \geq \hat{\tau}(t_i).$$

Existiert folglich ein Item mit zwei geordneten Responsekategorien $u_i^* > u_i$ oder zwei Responsezeiten $t_i^* < t_i$, die der Monotoniebedingung genügen, dann tritt unter SMDD ein paradoxer Effekt bei einer Veränderung des Itemscores von u_i zu u_i^* (oder von t_i zu t_i^*) auf. Eine kürzere Antwortzeit t_i^* impliziert dann zwar einen höheren Schätzwert der τ-Komponente, führt jedoch zugleich zu einer Verringerung des Schätzwerts der Fähigkeitskomponente. Im Rahmen der schwellenwertbasierten Klassifikation kann folglich der Fall $\hat{\theta}(t_i) > \kappa_1, \hat{\tau}(t_i) > \kappa_2$; $\hat{\theta}(t_i^*) < \kappa_1, \hat{\tau}(t_i^*) > \kappa_2$ auftreten, in dem eine positive Gesamtklassifikation durch eine längere Antwortzeit - bei sonst gleicher Performance - herbeigeführt wird.
Im Gegensatz zu den vorherigen Betrachtungen des paradoxen Effekts ergeben sich jedoch durch die Asymmetrie im Wertebereich bzw. Skalenniveau neue Fragen. Ist z.B. in „gewöhnlichen“ IRT-Modellen der Einfluß einer Antwortänderung auf einem Item durch die endliche Anzahl an verfügbaren Antwortkategorien stets indirekt begrenzt, so kann von einem analogen Verhalten im Rahmen der stetigen Antwortzeiten nicht zwingend ausgegangen werden, was wiederum die Frage aufwirft, unter welchen Bedingungen an das allgemein formulierte Modell (3.34) der Effekt einer Veränderung der Antwortszeit *beschränkt* ausfällt. Die Untersuchung dieser Thematik für eine Teilmodellklasse von (3.34) liegt im Fokus der nachfolgenden Abschnitte.
Bevor jedoch diese Sensitivität der Schätzer in Bezug auf extreme Responsezeiten eingehend dargelegt wird, seien abschließend stichpunktartig einige Bemerkungen im Zuge einer Generalisierung von Satz 3.2.1 gegeben.

- *Generalisierung für eine nicht-separable Loglikelihoodstruktur*: Eine Verallgemeinerung der Resultate im Hinblick auf Kategorie-Response-Funktionen, die nicht frei von τ, bzw. bezüglich Antwortzeiten, die nicht unabhängig von θ sind, kann unter Berufung auf Satz 2.3.4 bzw Satz 2.3.5 vorgenommen werden.

- *Generalisierung bei mehrdimensionalem Fähigkeitsvektor*: Wenn γ^p die Struktur

$$\gamma^p(\tilde{\boldsymbol{\theta}}) = \sum_i \log(f_{i,u_i}(\boldsymbol{\theta})) + \sum_i \log(g(t_i|\tau)) + \gamma(\boldsymbol{\theta}, \tau)$$

aufweist, und wenn γ eSMDD in τ ist, dann impliziert eine direkte Anwendung von Lemma 2.3.8, dass bei einer Verkürzung der Antwortszeit eines (die Monotoniebedingung erfüllenden) Items stets mindestens eine Fähigkeitskomponente existiert, deren zugehöriger Schätzwert abnimmt.

- *Generalisierung für lokale SMDD bzw. lokale SMID-Bedingungen*: Auch wenn die bisherigen Erörterungen zum Verhalten der Schätzwerte bei einer Veränderung des Itemscores stets auf einer globalen SMDD-Bedingung fußten, so lassen sich - wie bereits in vorherigen Abschnitten angedeutet wurde - analoge Aussagen auch bei Vorlage einer lokalen SMDD-Eigenschaft tätigen. Dies sei an einem kurzen Beispiel erläutert.
  Gegeben sei eine bivariate Cauchy-Prioriverteilung der Form

$$\gamma(\theta, \tau) = -\frac{3}{2}\log(1 + \theta^2 + \tau^2).$$

  Berechnung der gemischten zweiten Ableitung ergibt

$$\frac{\partial^2 \gamma}{\partial\theta\partial\tau}(\theta, \tau) = \frac{6\theta\tau}{(1 + \theta^2 + \tau^2)^2}.$$

  Gemäß Lemma 2.3.5 bzw. Korollar 2.3.1 ist γ somit SMID in den Quadranten $\mathbb{R}^2_+, \mathbb{R}^2_-$ und SMDD in den „gemischten“ Quadranten.
  Eine einfache lokale Modifikation von Satz 2.3.1 führt auf folgende Resultate:

  - Für Personen, deren zugehörige Schätzwerte (vor *und* nach der Veränderung des Itemscores) in dem Quadranten $\mathbb{R}^2_+$ liegen, tritt das Paradoxon nicht auf, d.h. bei Variation eines (die Monotoniebedingung erfüllenden) Itemscores verändern sich die Schätzer $\hat{\theta}, \hat{\tau}$ gleichsinnig. Analoges gilt für den $\mathbb{R}^2_-$-Quadranten.
  - Für Personen, deren zugehörige Schätzwerte (vor *und* nach der Veränderung des Itemscores) in dem Quadranten $\mathbb{R}_+ \times \mathbb{R}_-$ liegen, tritt ein paradoxer Effekt auf, d.h. bei Variation eines (die Monotoniebedingung erfüllenden) Itemscores verändern sich die Schätzer $\hat{\theta}, \hat{\tau}$ gegenläufig. Analoges gilt für den $\mathbb{R}_- \times \mathbb{R}_+$-Quadranten

Dieses Beispiel demonstriert das Auftreten eines paradoxen Effekts innerhalb einer Modellklasse, die keine globale SMDD-Bedingung erfüllt. Wie anhand des Beispiels ersichtlich, können in lokalen SMDD-Modellen für gewisse Teilgruppen paradoxe Effekte entstehen, während die Schätzer sich für andere „Gruppen“ fair verhalten.

Nachdem das Vorgehen in verallgemeinerten Modellen kurz skizziert wurde, soll nun der bereits aufgeworfenen Frage nach der Beschränktheit des Schätzers bei Veränderung einer Antwortzeit nachgegangen werden. Die Untersuchung dieser Frage erfordert zunächst eine Eingrenzung der (sehr allgemeinen) Modellklasse (3.34). Diese Eingrenzung wird anhand der folgenden Zusatzannahmen vorgenommen, die nebenbei auch die Existenz des MAP-Schätzers für jedes beliebige Antworts- bzw. Reaktionszeitmuster garantieren:

a) Die Antwortzeiten sind lognormalverteilt, d.h. $g_i(t_i|\tau)$ besitzt die Form

$$g_i(t_i|\tau) = \frac{\alpha_i}{t_i\sqrt{2\pi}} \exp\left(-\frac{1}{2}(\alpha_i^2(\log(t_i) - \beta_i + \tau)^2)\right) I_{\{t_i>0\}},$$

die äquivalent zu einer Normalverteilung für $\log(t_i)$ mit Erwartungswert $\beta_i - \tau$ und Standardabweichung α_i^{-1} ist.

b) Die Priori entspricht einer nichtsingulären, bivariaten Normalverteilung. Die Elemente der entsprechenden Kovarianzmatrix Σ werden mit σ_{jl} bezeichnet, während die Einträge der Inversen von Σ mit σ^{jl} gekennzeichnet werden.

c) Die Funktionen $\log(f_{i,u_i}(\theta))$ sind (mindestens einmal) stetig differenzierbar mit beschränkten Ableitungen $(\log(f_{i,u_i}(\theta)))'$.

Zusatzforderung *c*) ist von (nahezu) jedem „gängigen“ IRT-Modell erfüllt; Forderung *b*) korrespondiert mit der in der Praxis üblichen Wahl einer Prioriverteilung und Forderung *a*) entspricht einer in der Praxis häufig vorkommenden Wahl einer Verteilungsform für Responsezeiten (siehe z.B. Van der Linden, 2007; Klein Entink u.a., 2009; Fox, 2010).
Die Frage nach der Beschränktheit des MAP-Schätzers in Gegenwart extremer Antwortzeiten kann (innerhalb der so definierten Teilklasse) gemäß des folgenden Resultats negativ beantwortet werden.

Satz 3.2.2. *Sei Modell (3.34) mit den Zusatzannahmen a), b) und c) gegeben. Gilt $\sigma_{12} \neq 0$, dann ist $\hat{\theta}$ eine unbeschränkte Funktion von t_j beim Übergang $t_j \to 0$ bzw. $t_j \to \infty$.*
Ist $\sigma_{12} > 0$, dann kann eine Person durch eine entsprechend kurze Antwortzeit bei Item j - unabhängig von ihrem Antwortmuster $u_1, \ldots u_k$ - stets einen beliebig hohen Schätzwert ihrer Fähigkeit erreichen.
Ist $\sigma_{12} < 0$, dann kann eine Person durch eine entsprechend lange Antwortzeit bei Item j - unabhängig von ihrem Antwortmuster $u_1, \ldots u_k$ - stets einen beliebig hohen Schätzwert ihrer Fähigkeit erreichen.

Beweis. Aufgrund der Differenzierbarkeit des Ausdrucks γ^p erfüllt jede Stelle eines Maximums von γ^p die Gleichung

$$h_1(\theta, \tau) := -\frac{\partial \gamma^p(\theta, \tau)}{\partial \theta} = -\sum_i \frac{f'_{i,u_i}(\theta)}{f_{i,u_i}(\theta)} + \sigma^{11}\theta + \sigma^{12}\tau = 0 \tag{3.35}$$

sowie die Gleichung

$$h_2(\theta, \tau) := -\frac{\partial \gamma^p(\theta, \tau)}{\partial \tau} = \sum_i \alpha_i^2(\log(t_i) - \beta_i + \tau) + \sigma^{22}\tau + \sigma^{12}\theta = 0.$$

Letztere erlaubt das explizite Aufstellen der bedingten Maximumsfunktion $\hat{\tau}(\theta)$ gemäß

$$\hat{\tau}(\theta) = -\frac{\sum_i \alpha_i^2(\log(t_i) - \beta_i) + \sigma^{12}\theta}{\sigma^{22} + \sum_i \alpha_i^2}.$$

Einsetzen in (3.35) führt auf den Ausdruck

$$\sum_i \frac{f'_{i,u_i}(\theta)}{f_{i,u_i}(\theta)} = \sigma^{11}\theta - \sigma^{12}\frac{\sum_i \alpha_i^2(\log(t_i) - \beta_i) + \sigma^{12}\theta}{\sigma^{22} + \sum_i \alpha_i^2}$$

$$\sum_i \frac{f'_{i,u_i}(\theta)}{f_{i,u_i}(\theta)} = \left(\sigma^{11} - \frac{(\sigma^{12})^2}{\sum_i \alpha_i^2 + \sigma^{22}}\right)\theta - \sigma^{12}\frac{\sum_i \alpha_i^2(\log(t_i) - \beta_i)}{\sigma^{22} + \sum_i \alpha_i^2}.$$

Nach Annahme *c*) ist die linke Seite eine in θ beschränkte Funktion. Die rechte Seite ist eine affine Funktion von θ mit positiver Steigung[30].
Sei $\sigma^{12} < 0$ (oder äquivalent $\sigma_{12} > 0$) und sei $\theta < \theta_0$ (θ_0 eine beliebige reelle Zahl). Die folgende Überlegung wird beweisen, dass θ als Stelle eines Maximums ausgeschlossen

[30]Dies folgt aufgrund der Tatsache, dass die Determinante $\sigma^{11}\sigma^{22} - (\sigma^{12})^2$ der Inversen einer positiv definiten Matrix Σ positiv ist.

werden kann, indem bei einem beliebigen Item j der Grenzübergang $t_j \to 0$ erfolgt. Wenn eine Zahl $\theta < \theta_0$ tatsächlich ein MAP-Schätzer wäre, dann müsste die Ungleichung (M bezeichne eine nach c) existierende untere Schranke für den Ausdruck der linken Seite)

$$\begin{aligned} M^* : &= M - \left(\sigma^{11} - \frac{(\sigma^{12})^2}{\sum_i \alpha_i^2 + \sigma^{22}}\right)\theta_0 \\ &\leq \sum_i \frac{f'_{i,u_i}(\theta)}{f_{i,u_i}(\theta)} - \left(\sigma^{11} - \frac{(\sigma^{12})^2}{\sum_i \alpha_i^2 + \sigma^{22}}\right)\theta \\ &= -\sigma^{12}\frac{\sum_i \alpha_i^2(\log(t_i) - \beta_i)}{\sigma^{22} + \sum_i \alpha_i^2} \end{aligned} \tag{3.36}$$

gelten. Diese Ungleichung kann jedoch stets durch eine entsprechend kurz gewählte Responsezeit t_j verletzt werden, da der Term (3.36) für $t_j \to 0$ beliebig klein wird. Analog lässt sich im Fall $\sigma^{12} > 0$ und $t_j \to \infty$ argumentieren. □

Eine Ausdehnung auf den Fall mehrdimensionaler Fähigkeiten ist leicht möglich. Die Unterscheidung bezüglich der Richtung der Veränderung stellt jedoch keine einfache Funktion der Kovarianzmatrix mehr dar.

Satz 3.2.3. *Gegeben sei das Modell (3.32) mit der Zusatzannahme a) und einer multivariaten, nichtsingulären Normalverteilung als Prioriverteilung. Ferner seien die Terme* $(\log(f_{i,u_i}(\boldsymbol{\theta})))_{i=1,\ldots k}$ *stetig differenzierbar.*
Wenn τ *nicht a-priori unabhängig von* $\boldsymbol{\theta}$ *ist, dann ist der MAP-Schätzer* $\hat{\boldsymbol{\theta}}$ *unbeschränkt beim Übergang* $t_j \to 0/\infty$.

Beweis. Die Inverse Σ^{-1} der Kovarianzmatrix Σ sei partitioniert gemäß

$$A := \Sigma^{-1} = \begin{pmatrix} A_1 \\ \boldsymbol{a_1^T} \end{pmatrix},$$

wobei der Vektor $\boldsymbol{a_1}$ die letzte Zeile der Matrix A kennzeichnet.
Die zu minimierende Zielfunktion lautet:

$$h(\tilde{\boldsymbol{\theta}}) = -\sum \log(f_{i,u_i}(\boldsymbol{\theta})) + \frac{1}{2}\sum \alpha_i^2(\log(t_i) - \beta_i + \tau)^2 + \frac{1}{2}\tilde{\boldsymbol{\theta}}^T A\tilde{\boldsymbol{\theta}}$$

Das korrespondierende Gleichungssystem der partiellen Ableitungen lautet (in „Spaltenform“):

$$-\begin{pmatrix} \sum_i \frac{\partial \log(f_{i,u_i}(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta^T}} \\ 0 \end{pmatrix} + \begin{pmatrix} \mathbf{0} \\ \sum_i \alpha_i^2(\log(t_i) - \beta_i + \tau) \end{pmatrix} + \begin{pmatrix} A_1\tilde{\boldsymbol{\theta}} \\ \boldsymbol{a_1^T}\tilde{\boldsymbol{\theta}} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ 0 \end{pmatrix}$$

Die letzte Gleichung lässt sich explizit auflösen. Die resultierende Funktion $\hat{\tau}(\boldsymbol{\theta})$ ist affin und besitzt die Gestalt $\hat{\tau}(\boldsymbol{\theta}) = \boldsymbol{b^T\theta} + c(t_1, \dots t_k)$, wobei die "Konstante" $c(t_1, \dots t_k)$ streng monoton fallend in t_j für alle j und zudem, wenn als Funktion einer einzelnen Variable t_j betrachtet, unbeschränkt in beiden Richtungen $t_j \to \infty, t_j \to 0$ ist.
Einsetzen von $\hat{\tau}(\boldsymbol{\theta})$ in die verbliebenen Gleichungen führt auf:

$$\sum_i \frac{\partial \log(f_{i,u_i}(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta^T}} = A_1 \begin{pmatrix} \boldsymbol{I} \\ \boldsymbol{b^T} \end{pmatrix} \boldsymbol{\theta} + A_1 \begin{pmatrix} \mathbf{0} \\ c(t_1, \dots t_k) \end{pmatrix}$$

Wenn die Schätzwerte beim Übergang $t_j \to 0$ beschränkt wären, dann lägen sie in einer kompakten Menge K. Innerhalb einer solchen kompakten Menge ist jede stetige Funktion beschränkt. Folglich ist jede Komponente der linken Gleichungsseite von

$$\sum_i \frac{\partial \log(f_{i,u_i}(\boldsymbol{\theta}))}{\partial \boldsymbol{\theta^T}} - A_1 \begin{pmatrix} \boldsymbol{I} \\ \boldsymbol{b^T} \end{pmatrix} \boldsymbol{\theta} = A_1 \begin{pmatrix} \mathbf{0} \\ c(t_1, \dots t_k) \end{pmatrix}$$

eine beschränkte Funktion innerhalb der kompakten, die Schätzwerte umfassenden Menge K. Dies führt zu einem Widerspruch, da die rechte Seite der Gleichung unbeschränkt beim Grenzübergang $t_j \to 0$ ist - insofern die letzte Spalte der Matrix A_1 mindestens einen von Null verschiedenen Eintrag aufweist. Letztere Bedingung ist wiederum genau dann nicht gegeben, wenn τ a-priori unabhängig von $\boldsymbol{\theta}$ ist.

□

Dieses Resultat deutet auf generelle Probleme bei der Verwendung von „Vorinformationen" in einem Lognormal-Normal-Modell zur simultanen Modellierung von Antwortmustern und Reaktionszeiten hin. Personen können durch extreme Antwortzeiten prinzipiell beliebig hohe Schätzwerte ihrer Fähigkeit erreichen. Wie die Herleitung des Resultats zudem demonstriert, gilt dieser Effekt unabhängig von dem Antwortmuster $u_1, \dots u_k$ und somit insbesonders für die schlechtmöglichste Performance im Test.
Eine Vermeidung dieses Effekts ist jedoch durch Veränderung der Klasse der Prioriverteilungen möglich. Die folgenden, modifizierten Annahmen garantieren - wie im nachfolgenden Satz 3.2.4 gezeigt werden wird - innerhalb der durch (3.34) gegebenen Modellklasse die Beschränktheit des MAP-Schätzers beim Übergang zu extremen Responsezeiten:

a) Die Antwortzeiten sind lognormalverteilt, d.h. $g_i(t_i|\tau)$ besitzt die Form

$$g_i(t_i|\tau) = \frac{\alpha_i}{t_i\sqrt{2\pi}} \exp\left(-\frac{1}{2}(\alpha_i^2(\log(t_i) - \beta_i + \tau)^2)\right) I_{\{t_i>0\}}.$$

b') $\gamma(\theta, \tau)$ ist eine niveaubeschränkte, stetig differenzierbare logarithmierte Priori mit beschränkter partieller Ableitung $\frac{\partial\gamma}{\partial\tau}(\theta, \tau)$ und der Eigenschaft, dass $\lim_{n\to\infty} \frac{\partial\gamma}{\partial\theta}(\theta_n, \tau_n) = 0$ für jede Folge $(\theta_n, \tau_n)_{n\in\mathbb{N}}$, die $\theta_n \to \pm\infty$ und $\tau_n \to \infty$ erfüllt, gilt.

c) Die Funktionen $\log(f_{i,u_i}(\theta))$ sind stetig differenzierbar mit beschränkten Ableitungen.

d) Es existiert keine gegen unendlich konvergierende Folge θ_n mit der Eigenschaft, dass $\sum_i \frac{\partial \log(f_{i,u_i}(\theta_n))}{\partial\theta}$ gegen Null konvergiert.

Auf den ersten Blick stellt *d)* eine sehr technische Forderung dar. Dennoch kann mittels des folgenden Lemmas eine einfache Beziehung zwischen Annahme *d)* und der Existenz eines endlichen Maximum-Likelihood-Schätzers in dem durch $\sum_i \log f_{i,u_i}$ gegebenen, „gewöhnlichem“ IRT-Part hergestellt werden.

Lemma 3.2.1. *Sei $f: \mathbb{R} \to \mathbb{R}$ eine differenzierbare, strikt konkave Funktion. Wenn es einen Wert θ_0 gibt, der die Gleichung $f'(\theta_0) = 0$ erfüllt, dann existiert keine gegen unendlich konvergierende Folge $(\theta_n)_{n\in\mathbb{N}}$ mit der Eigenschaft, dass $f'(\theta_n)$ gegen Null konvergiert.*

Beweis. Sei θ^* ein beliebiger Wert rechts von θ_0. Die strikte Konkavität ist in Kombination mit der Differenzierbakeit äquivalent zur Aussage, dass $f'(\theta)$ streng monoton fallend in θ ist. Man setze $c := f'(\theta^*) < 0$ und nehme entgegen der Behauptung die Existenz einer gegen $+\infty$ konvergierenden Folge $(\theta_n)_{n\in\mathbb{N}}$ mit $f'(\theta_n) \to 0$ an. Die Wahl $\epsilon := -c > 0$ führt zu einem Index $N(\epsilon)$ mit der Eigenschaft $\theta_n > \theta^*$ und $0 > f'(\theta_n) > f'(\theta^*)$ für alle $n \geq N$. Letzteres stellt einen Widerspruch zur Monotonie der Ableitung dar.
Im Fall $(\theta_n)_{n\in\mathbb{N}} \to -\infty$ kann analog argumentiert werden. □

Wenn der „gewöhnliche“ IRT-Part einen Maximum-Likelihood-Schätzer besitzt - was für eindimensionale binäre IRT-Modelle i.d.R. der Fall ist, sobald mindestens eine

richtige und mindestens eine falsche Antwort gegeben wurde - und wenn die korrespondierende Loglikelihood $\sum \log(f_{i,u_i}(\theta))$ strikt konkav ist[31], dann folgt gemäß Lemma 3.2.1 die Gültigkeit der Annahme d).
Nach diesen Zwischenbemerkungen zur Modellannahme d) steht nun die eigentlich interessierende Frage nach der Robustheit des modellbasierten MAP-Schätzers im Vordergrund. Das diesbezügliche Kernresultat liefert der folgende Satz.

Satz 3.2.4. *Unter dem Modell (3.34) mit den Zusatzannahmen $a), b'), c)$ und $d)$ verhält sich der MAP-Schätzer robust gegenüber extrem kurzen Antwortzeiten, d.h. beim Übergang $t_j \to 0$ bleibt der MAP-Schätzer beschränkt.*

Beweis. Unter den obigen Modellannahmen besitzt die zu minimerende Funktion die Gestalt

$$-\sum_i \log(f_{i,u_i}(\theta)) + \frac{1}{2}\sum_i \alpha_i^2(\log(t_i) - \beta_i + \tau)^2 - \gamma(\theta, \tau).$$

Jede Stelle eines Minimums muss, aufgrund der von den Annahmen $a) - c)$ implizierten Differenzierbarkeit der obigen Funktion, das Gleichungssystem[32]

$$-\sum_i (\log(f_{i,u_i}(\theta)))' - \frac{\partial \gamma}{\partial \theta}(\theta, \tau) = 0 \tag{3.37}$$

$$\sum \alpha_i^2(\log(t_i) - \beta_i + \tau) - \frac{\partial \gamma}{\partial \tau}(\theta, \tau) = 0 \tag{3.38}$$

erfüllen.
Sei nun $(t_{j_n})_{n\in\mathbb{N}}$ eine gegen Null konvergierende Folge von Responsezeiten. Ferner bezeichne $(\theta_n, \tau_n)_{n\in\mathbb{N}}$ eine zugehörige Folge an Lösungen des obigen Gleichungssystems (man beachte, dass die Eindeutigkeit der Lösungen für die folgende Argumentation irrelevant ist).
Zunächst betrachte man den Ausdruck (3.38), dessen zweiter Term $-\frac{\partial \gamma}{\partial \tau}(\theta, \tau)$ gemäß Annahme b') eine beschränkte Funktion darstellt, während der erste Teil des ersten Terms beim Übergang $(t_{j_n})_n \to 0$ gegen $-\infty$ konvergiert. Hieraus folgt unmittelbar $\tau_n \to \infty$.
Man nehme nun gegensätzlich zur Behauptung des Satzes an, dass die Folge $(\theta_n)_{n\in\mathbb{N}}$ unbeschränkt ist. Der Übergang zu einer geeigneten Teilfolge ermöglicht o.B.d.A von

[31] Rasch-Modelle, 2PL-Modelle sowie Probitmodelle stellen Beispiele von Modellen dar, die eine strikt konkave Loglikelihood implizieren.

[32] Die Existenz des Minimums ist - wie in Anhang A dargelegt - durch den ersten Teil der Annahme b') in Kombination mit der Beschränktheit (nach oben) der übrigen Terme gewährleistet.

der Tatsache $(\theta_{n_v})_{v\in\mathbb{N}} \to \infty$ (oder $(\theta_{n_v})_{v\in\mathbb{N}} \to -\infty$) auszugehen.
Nach Annahme b') wird der zweite Term aus (3.37) ($\frac{\partial\gamma}{\partial\theta}(\theta_{n_v}, \tau_{n_v})$) gegen Null konvergieren, wenn v gegen unendlich anwächst. Es existiert folglich für jedes $\epsilon > 0$ eine natürliche Zahl V, so dass für alle $v \geq V$ die Beziehung $|\frac{\partial\gamma}{\partial\theta}(\theta_{n_v}, \tau_{n_v})| < \epsilon$ gültig ist. Aber da das Paar $(\theta_{n_v}, \tau_{n_v})$ eine Lösung der Gleichung (3.37) darstellt, gilt $|\sum_i(\log(f_{i,u_i}(\theta_{n_v})))'| < \epsilon$ für die gleiche Menge $v \geq V$. Letzteres steht im Widerspruch zur Modellannahme d).

□

Auch wenn Satz 3.2.4 eine robuste Alternative zur „üblichen“ Normalverteilungsspezifikation darstellt, ist dennoch unklar, welche Verteilungsfamilien der Annahme b') genügen. Das folgende Lemma gibt eine konkrete, mit Annahme b') konforme Verteilungsklasse an.

Lemma 3.2.2. *Die Klasse der (bivariaten) t-Verteilungen erfüllt Annahme b').*

Der Beweis des Lemmas kann mittels des folgenden Hilfssatzes erbracht werden:

Korollar 3.2.1. *Zu jedem (reellen) Polynom zweiten Grades $f(\boldsymbol{x}) := \sum a_i x_i^2 + \sum \alpha_{ij} x_i x_j + c$, welches durch eine positiv definite quadratische Form gegeben ist, existieren Polynome zweiten Grades der Form $g_1(\boldsymbol{x}) = \sum_i \delta_{1i} x_i^2 + c_1$, $g_2(\boldsymbol{x}) = \sum \delta_{2i} x_i^2 + c_2$ mit strikt positiven Koeffizienten, so dass*

$$g_1(\boldsymbol{x}) \leq f(\boldsymbol{x}) \leq g_2(\boldsymbol{x})$$

gilt.

Beweis. Aus $(x_i + x_j)^2 = x_i^2 + x_j^2 + 2x_i x_j$ sowie $(x_i - x_j)^2 = x_i^2 + x_j^2 - 2x_i x_j$ folgt

$$|x_i x_j| \leq \frac{x_i^2 + x_j^2}{2}$$

Deshalb gilt

$$\sum_i a_i x_i^2 + \sum \alpha_{ij} x_i x_j + c \leq \sum_i a_i x_i^2 + \sum |\alpha_{ij}| \left(\frac{x_i^2 + x_j^2}{2}\right) + c,$$

wobei die Koeffizienten der rechten Seite offenbar positiv sind. Demnach folgt der erste Teil der Behauptung.

Wenn A die zum Polynom f gehörende Matrix der quadratischen Form darstellt, d.h. $f(\boldsymbol{x}) = \boldsymbol{x}^{\boldsymbol{T}} A\boldsymbol{x} + c$, dann ist A gemäß Annahme positiv definit und es existiert ein Radius $r > 0$, so dass jede symmetrische Matrix N mit der Eigenschaft

$$|A - N| < r,$$

ebenfalls positiv definit ist (Lax, 2007, S.144 ff.).
Die Wahl $N := A - \epsilon I$ ($\epsilon := \frac{r}{2}$) liefert unmittelbar: $|A - N| = |\epsilon \boldsymbol{I}| = \epsilon < r$. Folglich ist die Matrix $A - \epsilon I$ positiv definit. Es folgt

$$f(\boldsymbol{x}) = \boldsymbol{x}^{\boldsymbol{T}} A\boldsymbol{x} + c = \boldsymbol{x}^{\boldsymbol{T}}(A - N)\boldsymbol{x} + \boldsymbol{x}^{\boldsymbol{T}} N\boldsymbol{x} + c \geq \epsilon \boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{x} + c.$$

Damit ist auch der zweite Teil der Behauptung bewiesen. □

Beweis von Lemma 3.2.2. O.B.d.A sei der Median der bivariaten t-Verteilung gleich Null. Bis auf von (θ, τ) unabhängige Konstanten besitzt die logarithmierte Priori die Form

$$\gamma(\theta, \tau) = -\frac{(v+2)}{2} \log\left(1 + \frac{1}{v}(\sigma^{11}\theta^2 + 2\sigma^{12}\theta\tau + \sigma^{22}\tau^2)\right).$$

Da dieser Ausdruck offenbar stetig differenzierbar ist, verbleibt der Beweis von drei Teilbehauptungen.
1. Es gilt $\frac{\partial\gamma}{\partial\theta}(\theta_n, \tau_n) \to 0$ für eine beliebige Folge, die der in b') aufgeführten Bedingung genügt:
Die partielle Ableitung nach θ ist durch den Ausdruck

$$\frac{\partial\gamma}{\partial\theta}(\theta, \tau) = -\frac{(v+2)}{2v} \cdot \frac{2\sigma^{11}\theta + 2\sigma^{12}\tau}{1 + v^{-1}(\sigma^{11}\theta^2 + 2\sigma^{12}\theta\tau + \sigma^{22}\tau^2)}$$

gegeben.
Nach Korollar 3.2.1 existieren zwei Polynome zweiten Grades mit strikt positiven Koeffizienten $\delta_{zw} > 0 (z = 1, 2; w = 1, 2)$, so dass

$$\frac{(v+2)}{2v} \cdot \frac{|2\sigma^{11}\theta + 2\sigma^{12}\tau|}{1 + v^{-1}(\delta_{21}\theta^2 + \delta_{22}\tau^2)} \leq |\frac{\partial\gamma}{\partial\theta}(\theta, \tau)| \leq \frac{(v+2)}{2v} \cdot \frac{|2\sigma^{11}\theta + 2\sigma^{12}\tau|}{1 + v^{-1}(\delta_{11}\theta^2 + \delta_{12}\tau^2)}$$

Für eine beliebige Folge (θ_n, τ_n), die $\theta_n \to \infty$ (oder $-\infty$) sowie $\tau_n \to \infty$ erfüllt, konvergieren offenbar sowohl die linke als auch die rechte Seite gegen Null. Demnach folgt $\frac{\partial\gamma}{\partial\theta}(\theta_n, \tau_n) \to 0$ und der dritte Teil aus b') ist bewiesen.
2. Die Funktion $\frac{\partial\gamma}{\partial\tau}(\theta, \tau)$ ist beschränkt:
Die partielle Ableitung nach τ ist durch den Ausdruck

$$\frac{\partial\gamma}{\partial\tau}(\theta, \tau) = -\frac{(v+2)}{2v} \cdot \frac{2\sigma^{12}\theta + 2\sigma^{22}\tau}{1 + v^{-1}(\sigma^{11}\theta^2 + 2\sigma^{12}\theta\tau + \sigma^{22}\tau^2)}$$

gegeben. Da jede stetige Funktion auf einer kompakten Menge beschränkt ist, genügt es folglich zu argumentieren, dass eine Zahl M existiert, derart, dass außerhalb einer kompakten Menge $|\frac{\partial\gamma}{\partial\tau}(\theta,\tau)| \leq M$ gilt. Die Argumentation folgt wiederum anhand obiger Polynomialabschätzung angewandt auf $\frac{\partial\gamma}{\partial\tau}(\theta,\tau)$ anstelle von $\frac{\partial\gamma}{\partial\theta}(\theta,\tau)$.
3. Die Funktion $\gamma(\theta,\tau)$ ist niveaubeschränkt:
Offenbar ist $\gamma(\theta,\tau)$ genau dann niveaubeschränkt, wenn die Funktion $f(\theta,\tau) = -(\sigma^{11}\theta^2 + 2\sigma^{12}\theta\tau + \sigma^{22}\tau^2)$ niveaubeschränkt ist. Die Niveaubeschränktheit dieser quadratischen Form folgt wiederum aus direkter Kombination von Theorem 3.26 mit Korollar 3.27 aus Rockafellar und Wets (2009) □

Es sei bemerkt, dass die Forderung der Niveaubeschränktheit „lediglich“ zur Sicherung der Existenz des MAP-Schätzer dient und darüber hinaus für die Herleitung der Beschränktheit des MAP-Schätzers keine Rolle spielt.
Somit konnte anhand Satz 3.2.4 sowie Lemma 3.2.2 skizziert werden, wie eine gegenüber extremen Antwortzeiten robustere Schätzmethode erfolgen kann. Die entsprechende Robustheit dürfte jedoch aus praktischer Sichtweise nur dann tatsächlich von Nutzen sein, wenn der Freiheitsgrad der t-Verteilung entsprechend gering ausfällt. Dies folgt unmittelbar aus dem Ergebnis der Unbeschränktheit des Schätzers im Falle einer Normalverteilungspriori und der Beobachtung, dass die durch Freiheitsgrade v indizierte t-Verteilungfamilie für $v \to \infty$ gegen die Normalverteilung konvergiert.
Zusammenfassend konnte somit gezeigt werden, dass

a) extreme Responsezeiten in der Klasse der Lognormal-Normal-Modelle (mit relativ frei wählbarem Modell für $f_{i,u_i}(\theta)$) zu beliebig hohen Schätzwerten führen können (Satz 3.2.2). Für den Fall negativer (positiver) (Priori-)Korrelation bedeutet dies, dass durch entsprechend langsames (schnelles) Antworten ein hoher Schätzwert der Fähigkeit θ erzielt werden kann. Wenngleich dies zunächst streng genommen nur im Fall negativer Korrelation auf ein paradoxes Resultat hindeutet, so sei dennoch für den Fall positiver Korrelation hervorgehoben, dass die Inkaufnahme einer falschen Antwort auf einem Item durch eine entsprechend schnelle Antwortzeit kompensiert werden kann und somit gegebenenfalls zu einer höher geschätzten Fähigkeit führt als eine korrekte Beantwortung des Items. Dies stellt zwar keinen paradoxen Effekt im eigentlichen Sinne dar, wirft aber gleichwohl erhebliche Zweifel an der Angemessenheit des Schätzverfahrens auf.

b) der Einfluß extremer Responsezeiten durch Veränderung der zugrundeliegenden Prioriverteilungsklasse begrenzt werden kann (Satz 3.2.4).

Diese Effekte können als eine spezielle Eigenschaft von IRT-Modellen mit zusätzlicher Schnelligkeitskomponente angesehen werden. Durch die Asymmetrie in der funktionalen Form zwischen dem Responsezeitmodell (stetige Zeitvariable) und dem „gewöhnlichen“ IRT-Modell resultiert ein spezieller, asymmetrischer Effekt: Extreme Responsezeiten dominieren die Likelihood und führen somit zu beliebig hohen Schätzwerten des Fähigkeitsparameters, während der Einfluß des „üblichen“ Antwortmusters aufgrund der inhärenten Diskretheit stets begrenzt bleibt.

3.3 Paradoxien in longitudinalen Item-Response-Modellen

Dieser Abschnitt ist der Analyse des paradoxen Effekts in Längsschnittmodellen gewidmet. Wenngleich das generelle Konstruktionsprinzip aus Kapitel 2.3 prinzipiell nicht zwischen Quer- und Längsschnittmodellen unterscheidet, scheint dennoch eine separate Diskussion longitudinaler Modelle in diesem Abschnitt aus mehreren Gründen sinnvoll und gerechtfertigt. Zum Einen kann ein grundlegender, vereinheitlichender Kern der in der Praxis verwendeten longitudinalen IRT-Modelle identifiziert werden, der es ermöglicht, von einer gemeinsamen Modellstruktur zu sprechen. Zum Anderen ist - wie nach einer kurzen Darlegung der generellen Modellstruktur zu erörtern sein wird - die semantische Bedeutung des paradoxen Effekts im Falle longitudinaler Messungen deutlich facettenreicher als die bisher verwendete Deutung als unfaires Testresultat.

Bevor diese Bemerkungen konkretisiert werden können, ist es jedoch erforderlich, die im Folgenden zu betrachtende Modellklasse formal einzuführen. Zu diesem Zweck bezeichne u_{t,j_t} den zum t-ten Zeitpunkt ($t = 1, \dots T$) auf dem j_t-ten Item ($j_t = 1, \dots k_t$) des t-ten Tests erzielten Itemscore einer Person[33]. Ferner sei die Person durch einen T-dimensionalen Vektor latenter Fähigkeiten $\boldsymbol{\theta}$ derart gekennzeichnet, dass die Items lokal stochastisch unabhängig sind und dass ferner eine Bearbeitung eines Items zum t-ten Zeitpunkt nur von der t-ten Komponente von $\boldsymbol{\theta}$ abhängt. Die Loglikelihood eines solchen Modells besitzt folglich die Dekomposition

$$l(\boldsymbol{\theta}) = \sum_{t=1}^{T} \sum_{j_t=1}^{k_t} l_{u_{t,j_t}}(\theta_t) \tag{3.39}$$

und ist demnach separabel, so dass sämtliche Paradoxien lediglich durch Verwendung einer entsprechenden Klasse von Prioriverteilungen „erzeugt“ werden.

Die dem paradoxen Effekt zugrundeliegende Fragestellung kann im longitudinalen Kontext wie folgt formuliert werden: Wie wirkt sich eine Erhöhung des Itemscores eines zum t-ten Zeitpunkt bearbeiteten Items ($u_{t,j_t} \nearrow$) auf die Schätzung der Fähigkeit zum Zeitpunkt $\theta_{t'} (t' \neq t)$ aus und welche Konsequenzen besitzt dies im Hinblick auf die Diagnose einer Fähigkeitsveränderung $\theta_{t'} - \theta_{t''}$?

Da die Antworten auf diese Fragen offenbar abhängig von der gegebenen Prioriverteilung ausfallen, sollen zunächst - in Anlehnung an Andrade und Tavares (2005) -

[33] Man beachte, dass diese Formulierung nicht impliziert, dass das j-te Item mehrfach appliziert wird. Das j-te Item im t-ten Test kann verschieden vom j-ten Item des t'-ten Tests sein.

die im Kontext longitudinaler Modelle prävalentesten Prioriverteilungen dargestellt werden, die im Kern alle aus der Familie der multivariaten Normalverteilung stammen und sich ausschließlich in der Struktur der Kovarianzmatrix Σ unterscheiden. Es erfolgt ferner für die anschließende Diskussion eine Beschränkung auf drei Messzeitpunkte ($T = 3$).

1. *Diagonalstruktur:* Σ und Σ^{-1} sind von der Form

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{pmatrix} \qquad \Sigma^{-1} = \begin{pmatrix} (\sigma_1^2)^{-1} & 0 & 0 \\ 0 & (\sigma_2^2)^{-1} & 0 \\ 0 & 0 & (\sigma_3^2)^{-1} \end{pmatrix}$$

   Diese Struktur korrespondiert zu der restriktiven Forderung der Unkorreliertheit (respektive Unabhängigkeit) der latenten Größen zweier verschiedener Zeitpunkte. Ferner ist sie aus offensichtlichen Gründen (separable log-Posteriori) für die weitere Untersuchung des paradoxen Effekts nicht von Belang.

2. *Äquikovarianzmatrix:* Σ und Σ^{-1} sind von der Form

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix} \qquad \Sigma^{-1} = \frac{1}{(1-\rho)\sigma^2} \begin{pmatrix} 1 - \frac{\rho}{1+2\rho} & -\frac{\rho}{1+2\rho} & -\frac{\rho}{1+2\rho} \\ -\frac{\rho}{1+2\rho} & 1 - \frac{\rho}{1+2\rho} & -\frac{\rho}{1+2\rho} \\ -\frac{\rho}{1+2\rho} & -\frac{\rho}{1+2\rho} & 1 - \frac{\rho}{1+2\rho} \end{pmatrix}$$

   Diese Struktur impliziert neben der Varianzgleichheit die Forderung, dass die zu zwei beliebigen Messzeitpunkten korrespondierenden latenten Größen stets die gleiche Kovarianz bzw. Korrelation aufweisen.

3. *Moving-Average:* Σ und Σ^{-1} sind von der Form

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & 0 \\ \rho & 1 & \rho \\ 0 & \rho & 1 \end{pmatrix} \qquad \Sigma^{-1} = \frac{1}{\sigma^2(1-2\rho^2)} \begin{pmatrix} 1-\rho^2 & -\rho & \rho^2 \\ -\rho & 1 & -\rho \\ \rho^2 & -\rho & 1-\rho^2 \end{pmatrix}$$

   Diese Struktur impliziert identische Korrelationen bezüglich zweier benachbarter Zeitpunkte sowie Nullkorrelationen zwischen zeitlich weiter entfernt liegenden latenten Größen.

4. *Autoregressiv:* Σ und Σ^{-1} sind von der Form

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{pmatrix} \qquad \Sigma^{-1} = \frac{1}{\sigma^2(1-\rho^2)} \begin{pmatrix} 1 & -\rho & 0 \\ -\rho & 1+\rho^2 & -\rho \\ 0 & -\rho & 1 \end{pmatrix}$$

Diese Struktur impliziert identische Korrelationen zwischen latenten Variablen, die äquidistant erhoben wurden. Ferner gibt sie den inhaltlich häufig plausiblen Fall wieder, dass Korrelationen (zwischen latenten Größen) umso geringer ausfallen, je weiter die Messzeitpunkte voneinander entfernt liegen.

5. *Unrestringierte Kovarianzmatrix:* Kein Eintrag der Kovarianzmatrix wird fixiert. Es liegt keine parametrische Struktur vor. Einzig die Forderung nach positiven Eigenwerten (in Kombination mit Symmetrie) beschränkt die Beliebigkeit der Einträge in Σ.

Bemerkung: Es sei darauf hingewiesen, dass die den Fällen 2-4 zugrundeliegende Forderung nach Varianzgleichheit aufgrund der „vagen“ Struktur eines Latenten-Variablen-Modells stets durch Transformation (mit anschließender Umdefinition der $l_{u_{t,j_t}}$-Terme) erzielt werden kann.
Im Folgenden soll für die oben aufgelisteten Fälle jeweils gekennzeichnet werden, unter welchen Bedingungen an die freien Parameter die resultierende Log-Posteriori einer eSMDD-Bedingung oder einer eSMID-Bedingung genügt. Die Überprüfung dieser Bedingungen kann dabei gemäß Lemma 2.3.7 anhand einer Betrachtung des Vorzeichens der zweiten partiellen Ableitungen erfolgen. Letztere können für multivariatnormale Prioriverteilungen offenbar direkt anhand der entsprechenden Nebendiagonaleinträge der Matrix $-\Sigma^{-1}$ abgelesen werden.

1. *Diagonalstruktur:* Im Falle einer Diagonalstruktur ist die resultierende Log-Posteriori offenbar separabel, so dass die Änderung eines mit dem Zeitpunkt t verknüpften Itemscores ausschließlich Auswirkungen auf die Schätzung des zugehörigen Fähigkeitsparameters θ_t besitzt.

2. *Äquikovarianzmatrix:* Für $\rho > 0$ sind sämtliche Nebendiagonaleinträge von $-\Sigma^{-1}$ positiv, so dass die resultierende Log-Posteriori die eSMID-Bedingung bezüglich jeder Variablen erfüllt. Eine Erhöhung des Itemscores zu einem beliebigem Zeitpunkt wirkt sich somit nie verringernd bezüglich der Fähigkeitsdiagnose zu anderen Zeitpunkten aus. Im Fall $\rho < 0$ folgt aus der implizierten eSMDD-Eigenschaft, dass die Erhöhung bezüglich mindestens eines anderen Messzeitpunktes einen negativen (d.h. verringernden) Effekt aufweist.

3. *Moving-Average:* Man beachte, dass für diese Struktur stets $|\rho| < \frac{1}{\sqrt{2}}$ gelten muss, um die positive Definitheit der Kovarianzmatrix zu garantieren. Der Vorfaktor $(1 - 2\rho^2)$ ist folglich positiv, womit das Vorzeichen zweier Nebendiagonaleinträge, die sich in der ersten (dritten) Zeile von Σ^{-1} befinden, im Fall $\rho > 0$ stets verschieden ausfällt. Die Log-Posteriori ist ergo weder eSMDD noch eSMID in θ_1 (θ_3). Für $\rho > 0$ und θ_2 resultiert hingegen eine eSMID-Bedingung. Eine Erhöhung des Itemscores zum zweiten Zeitpunkt besitzt somit bezüglich mindestens einer anderen Dimension eine nicht verringernde Wirkung.
   Ist $\rho < 0$, dann sind hingegen alle Nebendiagonaleinträge von $-\Sigma^{-1}$ negativ, so dass eine globale $(S)MRR_2$-Bedingung resultiert, in der von einer Erhöhung eines Itemscores zum Zeitpunkt t ($t = 1, 2, 3$) mindestens ein anderer Zeitpunkt $t' \neq t$ negativ tangiert wird.

4. *Autoregressiv:* Die Log-Posteriori ist eMDD (eMID) bezüglich jeder Variable und somit ist die Posteriori global MRR_2 (MTP_2), wenn $\rho < 0$ (> 0) gilt. Im MTP_2-Fall besitzt einer Erhöhung eines Itemscores zum Zeitpunkt t niemals einen verringernden Effekt auf eine andere Dimension. Im Fall MRR_2 hingegen existiert stets mindestens eine Dimension, deren zugehöriger Schätzwert fällt.

5. *Unrestringierte Kovarianzmatrix:* Das Vorzeichen eines Nebendiagonaleintrages von $-\Sigma^{-1}$ ist identisch mit dem Vorzeichen der Partialkorrelation. Es gilt $\sigma^{i,j} < (>) 0$ genau dann, wenn die Partialkorrelation zwischen θ_i und θ_j bei Kontrolle der übrigen latenten Variable positiv (negativ) ausfällt. Aus dieser Beziehung folgt, dass die resultierende Log-Posteriori genau dann eSMDD bezüglich θ_3 ist, wenn sowohl $\rho_{1,3|2}$ als auch $\rho_{2,3|1}$ negative Werte annehmen[34]. Ist dies erfüllt, dann impliziert eine Erhöhung des Itemscores eines zum dritten Messzeitpunkt applizierten Items (unter der „üblichen“ Monotoniebedingung) eine Verringerung des Fähigkeitsschätzwertes eines früheren Zeitpunkts. Analoge Resultate gelten bei Variation eines Itemscores zum ersten/zweiten Messzeitpunkt.

Der Effekt einer Veränderung eines Itemscores fällt somit je nach Prioriklasse unterschiedlich aus. Im Fall $\rho > 0$ des Moving-Average-Prozesses konnte ferner für θ_1

[34] $\rho_{i,j|l}$ bezeichnet hierbei die Partialkorrelation zwischen der i-ten und j-ten Variable bei Kontrolle der l-ten Variable.

(bzw. θ_3) scheinbar kein einfach interpretierbares, konsistentes Vorzeichenpattern beschrieben werden. Es sei jedoch darauf hingewiesen, dass durch das Hinzufügen der Annahme $\frac{\partial^2 l_{u_t,j_t}}{\partial x^2} < 0$ in (3.39) auf Resultate des Abschnitts 3.1 (insbesondere Satz 3.1.2) zurückgegriffen werden kann, die auch für diesen Fall eine konkrete Angabe des Effekts einer Itemscoreänderung ermöglichen. Für Details sei auf Hooker (2010) verwiesen.

Die Tatsache, dass die Diagnose einer Person zu einem bestimmten Zeitpunkt in Abhängigkeit ihrer Performance bei späteren Zeitpunkten steht, kann bereits unabhängig von der konkret vorliegenden Wirkungsrichtung (eSMDD vs eSMID) als bemerkenswertes Phänomen gelten. Sind darüber hinaus Diagnosen von *Fähigkeitsveränderungen* betroffen, die etwa an der Fragestellung orientiert sein können, ob ein nach dem ersten Zeitpunkt verabreichtes Treatment einen positiven Effekt bezogen auf den zweiten Zeitpunkt aufweist, so kann dies je nach anwendungsbezogenem Kontext den eigentlichen statistischen Zusatznutzen einer Priori („*borrowing strength*"; akkurate Prädiktion im statistischen Sinne) auf der personenbezogenen Ebene konterkarieren.

Die folgenden Analysen sollen diesen zusätzlichen Aspekt - d.h. die Betrachtung der Wirkung der Priori auf die Diagnose einer *Fähigkeitsveränderung* zwischen dem ersten Zeitpunkt und dem zweiten Zeitpunkt, wenn der Response zu einem späteren Zeitpunkt (Zeitpunkt 3) verändert wird - beleuchten.

Zur Vereinfachung der Notation werden die drei Indexmengen I_1, I_2, I_3 eingeführt, wobei I_j genau die Indizes der Items beinhaltet, die zum j-ten Testzeitpunkt dargeboten werden (die Items sind demnach wie „gewöhnlich" von 1 bis k durchnummeriert). Ferner werden die drei zu den Indexmengen korrespondierenden Laufindizes i,j,r eingeführt. Insofern nicht anders bemerkt, gibt die Verwendung des entsprechenden Laufindizes immer die Summation über alle Items der zugehörigen Indexmenge wieder, d.h. $\sum_i l_i(\theta_1)$ ist die Kurzschreibweise für $\sum_{n \in I_1} l_n(\theta_1)$.

Ziel der nachfolgenden Betrachtungen ist die Untersuchung des Verlaufs der bedingten Maximumsfunktion $\hat{\eta}(\theta_3)$, die sich ergibt, wenn die penalisierte Loglikelihood über (θ_1, θ_2) bei bekanntem θ_3 maximiert, und anschließend die Fähigkeitsdifferenz $\hat{\eta}(\theta_3) := \hat{\theta}_2(\theta_3) - \hat{\theta}_1(\theta_3)$ berechnet wird. Wenn diese Funktion streng monoton wachsend (fallend) ist, dann zeigen die Ergebnisse aus Kapitel 2.3, dass (unter der „üblichen" Monotonievoraussetzung) mit einer Erhöhung des Itemscores zum dritten Zeitpunkt eine Erhöhung (Verringerung) in der geschätzten Fähigkeitsdifferenz einher-

geht.
Die zu maximierende, um eine logarithmierte Priori γ erweiterte Loglikelihood besitzt die Form

$$f(\boldsymbol{\theta}) = \sum_i l_i(\theta_1) + \sum_j l_j(\theta_2) + \sum_r l_r(\theta_3) + \gamma(\boldsymbol{\theta}), \tag{3.40}$$

wobei jede der auftretenden l-Terme als zweimal stetig differenzierbar mit negativer zweiter Ableitung angenommen wird und γ zudem eine negativ definite zweite Ableitung über den gesamten Parameterbereich aufweise.[35]
Zur Einführung des interessierenden Parameters erfolgt eine Umparametrisierung:

$$\begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix} \quad \Rightarrow \quad \begin{pmatrix} \mu & := & \theta_1 \\ \eta & := & \theta_2 - \theta_1 \\ \nu & := & \theta_3 \end{pmatrix}$$

Die umparametrisierte Form von f ist durch

$$g(\mu, \eta, \nu) = \sum_i l_i(\mu) + \sum_j l_j(\eta + \mu) + \sum_r l_r(\nu) + \gamma(\mu, \mu + \eta, \nu)$$

gegeben.
Aufgrund der Modellannahmen können die bedingten Maximumsfunktionen $\hat{\mu}(\nu), \hat{\eta}(\nu)$ mit den Lösungsfunktionen der zugehörigen partiellen Ableitungen ($\widehat{=}$ Lösungen der partiellen Scoregleichungen) von g identifiziert werden. Letztere sind von der Form

$$\begin{aligned} \frac{\partial g}{\partial \mu}(\mu, \eta, \nu) &= \sum_i l_i'(\mu) + \sum_j l_j'(\mu + \eta) + \frac{\partial \gamma}{\partial \theta_1}(\mu, \mu + \eta, \nu) + \frac{\partial \gamma}{\partial \theta_2}(\mu, \mu + \eta, \nu) \\ \frac{\partial g}{\partial \eta}(\mu, \eta, \nu) &= \sum_j l_j'(\mu + \eta) + \frac{\partial \gamma}{\partial \theta_2}(\mu, \mu + \eta, \nu). \end{aligned}$$

Die Lösungen des Gleichungssystems $\nabla_{\mu,\eta} g = \mathbf{0}$ sind implizite Funktionen von ν, deren parametrische Abhängigkeit von ν mit Hilfe des Satzes über implizite Funktionen (siehe Satz 2.3.3) untersucht werden kann.
Die Anwendung dieses Satzes erfordert zunächst die Berechnung der zweiten Ableitungen (alle auftretenden Terme sind - insofern nicht anders gekennzeichnet - an der

[35] Diese Annahmen implizieren strikte Konkavität und ermöglichen die Übertragung der in Lemma 3.1.3 aufgeführten Ergebnisse zur Existenz/Eindeutigkeit des (bedingten) Maximums.

Stelle $(\mu, \mu+\eta, \nu)$ zu evaluieren)

$$\begin{aligned}
\frac{\partial^2 g}{\partial \mu^2}(\mu,\eta,\nu) &= \sum_i l_i'' + \sum_j l_j'' + \frac{\partial^2 \gamma}{\partial \theta_1^2} + 2\frac{\partial^2 \gamma}{\partial \theta_1 \partial \theta_2} + \frac{\partial^2 \gamma}{\partial \theta_2^2} \\
\frac{\partial^2 g}{\partial \mu \partial \eta}(\mu,\eta,\nu) &= \sum_j l_j'' + \frac{\partial^2 \gamma}{\partial \theta_1 \partial \theta_2} + \frac{\partial^2 \gamma}{\partial \theta_2^2} \\
\frac{\partial^2 g}{\partial \eta^2}(\mu,\eta,\nu) &= \sum_j l_j'' + \frac{\partial^2 \gamma}{\partial \theta_2^2}
\end{aligned}$$

sowie die Berechnung der Ableitung nach dem Parameter ν:

$$\frac{\partial^2 g}{\partial \mu \partial \nu}(\mu,\eta,\nu) = \frac{\partial^2 \gamma}{\partial \theta_1 \partial \theta_3} + \frac{\partial^2 \gamma}{\partial \theta_2 \partial \theta_3}; \quad \frac{\partial^2 g}{\partial \eta \partial \nu}(\mu,\eta,\nu) = \frac{\partial^2 \gamma}{\partial \theta_2 \partial \theta_3}.$$

Die parametrische Abhängigkeit der (vektorwertigen) bedingten Maximumsfunktion $s(\nu) := (\hat{\mu}(\nu), \hat{\eta}(\nu))$ lässt sich dann mittels der Ableitung

$$\nabla s(\nu) = -(A(\nu))^{-1} \boldsymbol{b}(\nu)$$

beschreiben, wobei $A(\nu)$ die Matrix der zweiten Ableitungen (nach μ und η) und $\boldsymbol{b}(\nu)$ den Vektor der gemischten Ableitungen (erst nach μ/η, dann nach ν) - jeweils ausgewertet an der Stelle $(\hat{\mu}(\nu), \hat{\eta}(\nu), \nu)$ - darstellen.
Zur Untersuchung der Abhängigkeit der Fähigkeitsdifferenz ist lediglich die zweite Zeile der Matrix A^{-1} von Interesse. Da ferner nur das Vorzeichen der Ableitung zur Untersuchung der Monotonie relevant ist, kann die Berechnung der Determinante von A unter Beachtung, dass diese aufgrund der Modellannahmen (negativ definite zweite Ableitungen) stets ein Produkt von zwei negativen Eigenwerten ist, außen vor bleiben.
Es folgt demnach[36]:

$$\operatorname{sgn}(\hat{\eta}'(\nu)) = h_1(\nu) - h_2(\nu)$$

mit

$$h_1(\nu) = \left(\sum_j l_j'' + \frac{\partial^2 \gamma}{\partial \theta_2 \partial \theta_1} + \frac{\partial^2 \gamma}{\partial \theta_2^2}\right)\left(\frac{\partial^2 \gamma}{\partial \theta_3 \partial \theta_1} + \frac{\partial^2 \gamma}{\partial \theta_2 \partial \theta_3}\right) \tag{3.41}$$

$$h_2(\nu) = \left(\sum_i l_i'' + \sum_j l_j'' + \frac{\partial^2 \gamma}{\partial \theta_1^2} + 2 \cdot \frac{\partial^2 \gamma}{\partial \theta_2 \partial \theta_1} + \frac{\partial^2 \gamma}{\partial \theta_2^2}\right)\frac{\partial^2 \gamma}{\partial \theta_2 \partial \theta_3} \tag{3.42}$$

[36]Zur Vereinfachung der Notation wird der Ausdruck „sgn“ im Folgenden unterdrückt. Eine Gleichung der Form „$\operatorname{sgn}(f(x)) = g(x)$“ ist somit zu lesen als: „Das Vorzeichen von $f(x)$ ist identisch mit dem Vorzeichen von $g(x)$“.

Der Ausdruck $\operatorname{sgn}(\hat{\eta}'(\nu))$ wird nun für die in den Fällen 2-4 genannten Kovarianzmatrizen näher betrachtet:
Zunächst sei der Fall einer autoregressiven Korrelationsstruktur gegeben, in dem

$$\gamma(\boldsymbol{\theta}) = -\frac{1}{2\sigma^2(1-\rho^2)}\left(\theta_1^2 - 2\rho\theta_1\theta_2 + (1+\rho^2)\theta_2^2 - 2\rho\theta_2\theta_3 + \theta_3^2\right)$$

ist.
Unter Verwendung der Abkürzung $c := \frac{1}{2\sigma^2(1-\rho^2)} > 0$ folgt

$$\begin{aligned}\frac{\partial\gamma}{\partial\theta_1} &= -c(2\theta_1 - 2\rho\theta_2)\\ \frac{\partial\gamma}{\partial\theta_2} &= -c(-2\rho\theta_1 + 2(1+\rho^2)\theta_2 - 2\rho\theta_3),\end{aligned}$$

so dass die zur Berechnung des Ausdrucks $\operatorname{sgn}(\hat{\eta}'(\nu))$ erforderlichen zweiten partiellen Ableitungen die Form

$$\frac{\partial^2\gamma}{\partial\theta_1^2} = -2c, \quad \frac{\partial^2\gamma}{\partial\theta_2\partial\theta_1} = 2c\rho, \quad \frac{\partial^2\gamma}{\partial\theta_2^2} = -2c(1+\rho^2), \quad \frac{\partial^2\gamma}{\partial\theta_3\partial\theta_1} = 0, \quad \frac{\partial^2\gamma}{\partial\theta_3\partial\theta_2} = 2c\rho$$

besitzen.
Einsetzen in (3.41) bzw. (3.42) führt auf

$$\operatorname{sgn}(\hat{\eta}'(\nu)) = 2c\rho\left(\sum_j l_j'' + 2c\rho - 2c(1+\rho^2) - \sum_i l_i'' - \sum_j l_j'' + 2c - 4c\rho + 2c(1+\rho^2)\right),$$

was nach kurzer Umformung in den Ausdruck

$$\operatorname{sgn}(\hat{\eta}'(\nu)) = -2c\rho\left(\sum_i l_i''(\hat{\mu}(\nu)) - \sigma^{-2}\frac{1}{1+\rho}\right)$$

übergeht.
Unter Beachtung, dass $\frac{1}{1+\rho} > 0\ \forall\,\rho \in (-1,1)$ sowie $l_i''(\mu) < 0$ unabhängig von dem Argument μ der Funktion gilt, folgt

Satz 3.3.1. *Gegeben sei eine penalisierte Loglikelihood der Form (3.40), deren zweimal stetig differenzierbare Terme negative zweite Ableitungen über den gesamten Parameterbereich aufweisen.*
Unter einer autoregressiven Priori führt eine Erhöhung des Itemscores zum dritten Zeitpunkt (bei vorausgesetzter Monotonie des Beitrags) zu einem Anstieg (Rückgang) in der geschätzten Fähigkeitsdifferenz $\hat{\eta} := \hat{\theta}_2 - \hat{\theta}_1$, wenn $\rho > 0$ ($\rho < 0$) gilt.

Mit den Abkürzungen $b := \frac{\rho}{1+2\rho}, c := \frac{1}{2\sigma^2(1-\rho)}$ ergibt sich im Falle der Äquikovarianzstruktur zunächst

$$\frac{\partial^2\gamma}{\partial\theta_1^2} = \frac{\partial^2\gamma}{\partial\theta_2^2} = -2c(1-b), \quad \frac{\partial^2\gamma}{\partial\theta_2\partial\theta_1} = \frac{\partial^2\gamma}{\partial\theta_3\partial\theta_1} = \frac{\partial^2\gamma}{\partial\theta_3\partial\theta_2} = 2bc.$$

Einsetzen in (3.41) bzw. (3.42) führt auf

$$\begin{aligned} \operatorname{sgn}(\hat{\eta}'(\nu)) &= 2bc\left(2\sum_j l_j'' + 4bc - 4c(1-b) - \sum_i l_i'' - \sum_j l_j'' + 4c(1-b) - 4bc\right) \\ &= 2bc\left(\sum_j l_j'' - \sum_i l_i''\right) \end{aligned} \tag{3.43}$$

Man beachte, dass $bc > 0$ für $\rho > 0$ sowie $bc < 0$ für $-\frac{1}{2} < \rho < 0$ gilt.[37] Die Terme $\sum_j l_j''$, $\sum_i l_i''$ sind nach Modellannahme stets negativ, so dass $\operatorname{sgn}(\hat{\eta}'(\nu))$ das gleiche (umgekehrte) Vorzeichen wie bc aufweist, wenn $0 > \sum_j l_j''(\hat{\theta}_2(\nu)) > \sum_i l_i''(\hat{\theta}_1(\nu))$ $(\sum_j l_j''(\hat{\theta}_2(\nu)) < \sum_i l_i''(\hat{\theta}_1(\nu)) < 0)$ gilt.
Interpretiert man den Betrag der zweiten Ableitung als Informationsgehalt[38], den die Items des jeweiligen Zeitpunkts über den betreffenden Parameter besitzen, so lässt sich das Resultat in suggestiver Form wie folgt zusammenfassen:

Satz 3.3.2. *Gegeben sei eine penalisierte Loglikelihood der Form (3.40), deren zweimal stetig differenzierbare Terme negative zweite Ableitungen über den gesamten Parameterbereich aufweisen.*
Ist der lokale Informationsgehalt der Items zum zweiten Zeitpunkt geringer als zum ersten Zeitpunkt, so führt eine Erhöhung des Itemscores zum dritten Zeitpunkt (bei vorausgesetzter Monotonie des Beitrags) unter einer Äquikovarianz-Priori zu einem Anstieg (Rückgang) in der geschätzten Fähigkeitsdifferenz $\hat{\eta} := \hat{\theta}_2 - \hat{\theta}_1$, wenn $\rho > 0$ $(\rho < 0)$ gilt.
Ist der lokale Informationsgehalt zum zweiten Zeitpunkt hingegen höher als zum ersten Zeitpunkt, so führt eine Erhöhung des Itemscores zum dritten Zeitpunkt (bei vorausgesetzter Monotonie des Beitrags) unter einer Äquikovarianz-Priori zu einem Anstieg (Rückgang) in der geschätzten Fähigkeitsdifferenz $\hat{\eta} := \hat{\theta}_2 - \hat{\theta}_1$, wenn $\rho < 0$ $(\rho > 0)$ gilt.

[37] Der Fall $\rho < -\frac{1}{2}$ kann nicht eintreten, da in diesem Fall Σ nicht positiv semidefinit ist.

[38] Diese Interpretation scheint vor dem Hintergrund, dass die Breite eines asymptotischen Konfidenzintervalls ein Vielfaches des Inversen der zweiten Ableitung beträgt, gerechtfertigt.

Mit $c := \frac{1}{2\sigma^2(1-2\rho^2)}$ resultiert schließlich für den Fall eines Moving-Average-Prozesses

$$\frac{\partial^2\gamma}{\partial\theta_1^2} = -2c(1-\rho^2), \quad \frac{\partial^2\gamma}{\partial\theta_2\partial\theta_1} = \frac{\partial^2\gamma}{\partial\theta_3\partial\theta_2} = 2c\rho, \quad \frac{\partial^2\gamma}{\partial\theta_2^2} = -2c, \quad \frac{\partial^2\gamma}{\partial\theta_3\partial\theta_1} = -2c\rho^2,$$

so dass das Vorzeichen der Ableitung von $\hat{\eta}(\nu)$ durch den Ausdruck

$$\left(\sum_j l_j'' + 2c\rho - 2c\right)\left(-2c\rho^2 + 2c\rho\right) - \left(\sum l_i'' + \sum l_j'' - 2c(1-\rho^2) + 4c\rho - 2c\right)2c\rho,$$

der nach kurzer Umformung in den Ausdruck

$$2c\rho\left(-\sum_i l_i'' - \rho\sum_j l_j'' + \frac{1}{\sigma^2}\right) \tag{3.44}$$

übergeht, gegeben ist.

Da $c\rho > 0$ ist, wenn $\rho > 0$ gilt und da ferner die in (3.44) auftretenden zweiten Ableitungen stets negativ ausfallen, folgt, dass $\operatorname{sgn}(\hat{\eta}'(\nu))$ stets positiv ist. Demnach bewirkt für positives ρ eine Erhöhung des Itemscores zum dritten Zeitpunkt einen Anstieg in der geschätzten Fähigkeitsdifferenz.

Für $\rho < 0$ ist die Situation subtiler:

Eine Verringerung in der geschätzten Fähigkeitsdifferenz wird hier nur dann erreicht, wenn der mit $|\rho|$ gewichtete Informationsgehalt zum zweiten Zeitpunkt ($|\sum_j l_j''|$) dem kombinierten Informationsgehalt bestehend aus Prioripräzision ($\frac{1}{\sigma^2}$) und Iteminformation zum ersten Zeitpunkt ($|\sum_i l_i''|$) unterliegt.

Bei einer betragsmäßig nicht zu geringen negativen Korrelation kann demnach der kontraintuitive Fall eintreten, dass eine Erhöhung des Itemscores zum dritten Zeitpunkt mit einem Anstieg in der geschätzten Fähigkeitsdifferenz $\hat{\eta}$ einhergeht - obwohl eine negative Korrelation zwischen Zeitpunkt 3 und Zeitpunkt 2 und eine Nullkorrelation zwischen Zeitpunkt 1 und 3 besteht!

Die vorangegangen Analysen des paradoxen Effekts innerhalb der Klasse longitudinaler IRT-Modelle stellen somit heraus, dass neben der zu erwartenden zentralen Rolle der Korrelation ebenso die zum jeweiligen Zeitpunkt vorliegende Iteminformation das Verhalten des Schätzers mitbestimmt. Letztere ist zwar für faktorenanalytische Modelle identisch mit der Itempräzision (inverse Messfehlervarianz), kann jedoch in anderen IRT-Modellen selber wiederum von der momentanen „Schätzstelle“ $\hat{\theta}$ abhängen. Wie der Fall des Moving-Average-Prozesses abschließend demonstrierte, sollte ferner die Heuristik „Aus positiver (negativer) Korrelation folgt gleichsinniger (gegensinniger) Verlauf“ mit Vorsicht angewandt werden.

4 Möglichkeiten der Vermeidung des paradoxen Effekt

Die bisherigen Betrachtungen des paradoxen Effekts waren darauf bedacht, seine Existenz in einer großen, heterogenen IRT-Modellklasse, die von „üblichen“ Querschnittdesigns über longitudinale Designs bis hin zu „Speed“-Designs reicht, zu etablieren und darüber hinaus die damit verbundenden Konsequenzen darzulegen.
Die in den einzelnen Abschnitten aufgezeigte Omnipräsenz des paradoxen Effekts kann trotz der Heterogenität der diskutierten Modellklassen stets auf eine gemeinsame Ursache zurückgeführt werden - die (lokale) (e)SMDD-Bedingung.
Ein naheliegendes Vorgehen zur Vermeidung des paradoxen Effekts liegt daher in der Formulierung eines der TP_2- Bedingung (bzw. MTP_2) genügenden IRT-Modells. Das Potential hinter dieser Methodik im Hinblick auf die Erlangung fairer Inferenz wird in den folgenden Abschnitten ebenso evaluiert, wie alternative Prozeduren zur Vermeidung des paradoxen Effekts. Die Heterogenität der verschiedenen Vorgehensweisen erlaubt allerdings keine Darstellung in einem gemeinsamen Rahmen, so dass die Diskussion der unterschiedlichen Methoden separat vorgenommen wird.
Nachdem einleitend die asymptotische Fragestellung der Abhängigkeit des paradoxen Effekts von der Testlänge erörtert wird, dienen die anschließenden Kapitel der Besprechung dreier potentieller Methodiken zur Vermeidung des paradoxen Effekts, die sowohl bezüglich ihrer praktischen Umsetzbarkeit wie auch in ihrem Allgemeinheitsgrad stark variieren.
Nach einer kurzen Betrachtung der von Hooker u.a. (2009) skizzierten Methode der Schätzung unter „Paradoxie-vermeidenden“ Restriktionen in Kapitel 4.2, widmet sich Kapitel 4.3 Methoden, die auf einer Umformulierung des IRT-Modells beruhen, um so zu einer fairen Inferenz zu gelangen. Hierunter lässt sich sowohl die bereits erwähnte, auf dem TP_2-Konzept basierende Methode als auch die von Hooker und Finkelman (2010) im Kontext von sogenannten „Item-Bundles“ vorgeschlagene Methode der Marginalisierung des IRT-Modells, welche im Kern auf einer Umformulierung eines MIRT-Modells zu einem eindimensionalen IRT-Modell mit lokal abhängigen Items basiert, subsumieren.
Kapitel 4.4 widmet sich abschließend einer praktisch vielversprechenden, auf Umdefinition des latenten Konstrukts beruhenden Methode, deren Gültigkeit jedoch (bislang) auf den Fall linearer (bzw. affiner) Schätzer beschränkt ist.

4.1 Asymptotische Auflösung der interindividuellen Form des Paradoxons

Gegenstand dieses Abschnitts ist die asymptotische Betrachtung des paradoxen Effekts. Die in Kapitel 2.1 eingeführte Unterscheidung zwischen intra- und interindividueller Interpretation des paradoxen Effekts wird sich hierbei als von zentraler Bedeutung erweisen. Die interindividuelle Interpretation bezieht sich auf den Vergleich zweier Personen, deren korrespondierende Schätzwerte nicht konform mit der von ihrem Antwortmuster zu erwartenden Ordnung ausfallen, während die intraindividuelle Interpretation auf einem fiktiven Szenario - festzumachen anhand der Frage, wie die Schätzung bzw. Klassifikation ausgefallen wäre, wenn die betreffende Person ein bestimmtes Item falsch statt richtig beantwortet hätte - fußt.
Ein Test kann frei von interindividuellen Paradoxien sein (z.B. wenn keine Personenpaare mit geordneten Antwortmustern existieren), jedoch zahlreiche intraindividuelle paradoxe Effekte aufweisen. Das Umgekehrte dagegen ist unmöglich. Daher stellt die Forderung nach der Vermeidung intraindividueller Paradoxien die stärkere Forderung dar als die bloße Vermeidung interindividueller Paradoxien.
Die nachfolgenden Betrachtungen werden für ein prinzipiell beliebiges IRT-Modell - insbesonders fallen alle im Rahmen von Kapitel 3 diskutierten Hauptmodelle hierunter - die asymptotische Irrelevanz der interindividuellen Form des paradoxen Effekts aufzeigen. Der betreffende, im Rahmen binärer IRT-Modelle erbrachte Beweis von Hooker u.a (2009) für die asymptotische Auflösbarkeit des paradoxen Effekts wird dabei wiedergegeben und durch leichte Modifikationen auf ordinale sowie faktorenanalytische Modelle verallgemeinert.
Die Notation $\boldsymbol{u_n^k} \prec \boldsymbol{u_m^k}$ bedeutet, dass die m-te Person auf jedem der ersten k-Items einen mindestens so hohen Score erzielt hat wie die n-te Person und dass zudem mindestens ein Item existiert, bezüglich dessen die Antworten der Personen unterschiedlich ausfielen. Im Kern des folgenden Satzes liegt die Beobachtung, dass für eine feste Anzahl (N) zu diagnostizierender Personen die Wahrscheinlichkeit der Existenz zweier nach obigem Schema geordneter Antwortmuster für steigende Testlänge $k \to \infty$ gegen Null konvergiert.

Satz 4.1.1. *(Theorem 8.1 von Hooker u.a (2009)) Für eine feste Anzahl N zu diagnostizierender Personen sei eine Folge lokal stochastisch unabhängiger, binärer Items $i = 1, 2, \ldots k$ $(k \to \infty)$, die einem gemeinsamen MIRT-Modell mit fester Di-*

mensionalität dim($\boldsymbol{\theta}$) = p entstammen, gegeben.
Wenn Antworten verschiedener Personen unabhängig sind, und wenn die spezifischen Lösungswahrscheinlichkeiten $P_i(\boldsymbol{\theta_m})$ gleichmäßig für alle $i \in \mathbb{N}$ und alle $m \in \{1, 2, \dots N\}$ von Null und Eins entfernt sind, dann konvergiert die Wahrscheinlichkeit, ein Personenpaar (m, n) mit $\boldsymbol{u_n^k} \prec \boldsymbol{u_m^k}$ zu finden, für $k \to \infty$ gegen Null.

Beweis. Bezeichne A_{klj} das Ereignis, dass das Personenpaar (l, j) $(l < j, l, j \in \{1 \dots N\})$ nach k Items geordnete Antwortmuster aufweist und sei A_k das Ereignis, dass im Test der Länge k mindestens ein Personenpaar mit geordneten Antwortmustern existiert. Dann gilt offenbar

$$P(A_k) = P(\cup_{l<j} A_{klj}) \leq \sum_{l<j} P(A_{klj}). \tag{4.45}$$

Wenn die Personen (l, j) bei zwei benachbarten Items gegensätzlich antworten, dann tritt das Ereignis $P(A_{klj})$ nicht auf. Demnach gilt:

$$P(A_{klj}) \leq \prod_{v=1}^{\frac{k}{2}} (1 - P_{2v-1}(\boldsymbol{\theta}_l)(1 - P_{2v-1}(\boldsymbol{\theta}_j))(1 - P_{2v}(\boldsymbol{\theta}_l))P_{2v}(\boldsymbol{\theta}_j)). \tag{4.46}$$

Gemäß Annahme existieren Konstanten $0 < c_u < 1, 0 < c_o < 1$, so dass die Ungleichungen

$$c_u < P_v(\boldsymbol{\theta}_l) < c_o$$

für beliebiges $l \in \{1, \dots N\}$ und beliebiges v gelten.
Nutzt man diese Relationen in (4.46), so folgt:

$$\begin{aligned} P(A_{klj}) &\leq \prod_{v=1}^{\frac{k}{2}} (1 - P_{2v-1}(\boldsymbol{\theta}_l)(1 - P_{2v-1}(\boldsymbol{\theta}_j))(1 - P_{2v}(\boldsymbol{\theta}_l))P_{2v}(\boldsymbol{\theta}_j)) \\ &\leq \prod_{v=1}^{\frac{k}{2}} (1 - c_u^2(1 - c_o^2)) = \prod_{v=1}^{\frac{k}{2}} c, \end{aligned}$$

wobei $c := (1 - c_u^2(1 - c_o^2))$ und $0 < c < 1$ ist.
Einsetzen in (4.45) ergibt für geradzahliges k:

$$P(A_k) \leq \binom{N}{2} c^{\frac{k}{2}}.$$

Da die geometrische Reihe $\sum_v c^v$ für $0 \leq c < 1$ konvergiert, folgt die Konvergenz von $\sum_{v=1}^{\infty} P(A_{2 \cdot v})$.

Im Falle von ungeradzahligem k kann die Ungleichung (4.46) mit allen darauffolgenden Abschätzungen übernommen werden, indem k durch $k-1$ ersetzt wird. Somit ist die Konvergenz der Reihe $\sum_{k=1}^{\infty} P(A_k)$ bewiesen und es folgt $P(A_k) \to 0$ für $k \to \infty$. (Gemäß dem Borel-Cantelli-Lemma (siehe z.B. Taylor, 1997, S. 148) folgt damit auch $P(\cap_{k=1}^{\infty} \cup_{l=k}^{\infty} A_l) = 0$).

□

Man beachte, dass die Annahme einer festen Anzahl N zu diagnostizierender Personen - ebenso wie die Annahme unabhängiger Antwortmuster - ein wesentliches Element in der Herleitung von Satz 4.1.1 darstellte. Wächst die Personenzahl mit der Testlänge, so kann die obige Argumentation - selbst unter Beibehaltung der Existenz der Konstanten c_u und c_o - nicht übernommen werden.
Als zentrale Konsequenz des Satzes folgt die asymptotische „Auflösung“ der interindividuellen Variante des paradoxen Effekts. Diese Auflösung sollte jedoch nicht als Eigenschaft des Schätzers[39] sondern als Implikation der Modellannahmen angesehen werden.
Auch wenn die ursprüngliche Formulierung von Hooker u.a. (2009) auf binäre IRT-Modelle abzielt, kann Satz 4.1.1 leicht auf ordinale bzw. faktorenanalytische Modelle übertragen werden. Für ordinale IRT-Modelle ersetze man lediglich die Kombination richtiger bzw. falscher Antworten in (4.46) durch eine beliebige Kombination geordneter Itemscores, d.h. man ersetze den Term

$$P_{2v-1}(\boldsymbol{\theta}_l)(1 - P_{2v-1}(\boldsymbol{\theta}_j))(1 - P_{2v}(\boldsymbol{\theta}_l))P_{2v}(\boldsymbol{\theta}_j)$$

durch den Term

$$P_{2v-1,i}(\boldsymbol{\theta}_l)P_{2v-1,i+1}(\boldsymbol{\theta}_j)P_{2v,i+1}(\boldsymbol{\theta}_l)P_{2v,i}(\boldsymbol{\theta}_j),$$

wobei $P_{v,i}(\boldsymbol{\theta}_l)$ die Wahrscheinlichkeit bezeichnet, dass Person l den Itemscore i auf Item v erzielt.
Im Falle des faktorenanalytischen Modells bzw. anderer „abweichender“ Modellklassen ersetze man den Term

$$P_{2v-1}(\boldsymbol{\theta}_l)(1 - P_{2v-1}(\boldsymbol{\theta}_j))(1 - P_{2v}(\boldsymbol{\theta}_l))P_{2v}(\boldsymbol{\theta}_j)$$

[39] In der Tat ist in keinem Abschnitt des Beweises ein Schätzer gegenwärtig. Es wird stets mit den Modell inhärenten Wahrscheinlichkeiten für geordnete Antwortmuster argumentiert.

durch den Ausdruck

$$P(u_{l,2v-1} > \tau|\boldsymbol{\theta}_l)P(u_{j,2v-1} \leq \tau|\boldsymbol{\theta}_j)P(u_{l,2v} \leq \tau|\boldsymbol{\theta}_l)P(u_{j,2v} > \tau|\boldsymbol{\theta}_j),$$

worin τ eine reelle Zahl bezeichnet, die konform mit der Annahme der gleichmäßigen Entfernung von 0 bzw. 1 ausfällt.

Insgesamt kann somit zwar für eine sehr allgemeine IRT-Modellklasse die asymptotische Auflösbarkeit des paradoxen Effekts demonstriert werden, jedoch ist diese Auflösbarkeit von relativ geringem theoretischem sowie praktischem Wert. Neben der Tatsache, dass die intraindividuelle Form des Effekts davon unberührt bleibt, ist die asymptotische Argumentation für $k \to \infty$ von zweifelhaftem Wert, da ein und derselbe Test als Folgenglied einer Testsequenz mit festem N und $k \to \infty$, wie auch als Folgenglied einer Testsequenz mit variierendem k *und* variierendem N angesehen werden kann. Asymptotische Eigenschaften beziehen sich auf *Testsequenzen*. Sie sind für die Betrachtung eines einzelnen, konkret vorliegenden Tests irrelevant.

Die vorliegende, asymptotische Betrachtungsweise gibt naturgemäß keine Antworten auf die Frage, wie im Rahmen eines spezifischen Tests mit (bereits) existierenden paradoxen Effekten verfahren werden kann. Die Beantwortung dieser Frage ist Gegenstand der folgenden, von Hooker u.a. (2009) dargelegten Schätzmethode, deren Konsistenzeigenschaften - wie Hooker u.a. (2009, Theorem 8.1) zeigen - unmittelbar aus Satz 4.1.1 folgen.

4.2 Schätzung unter Restriktionen

Im Kern der im Folgenden darzulegenden, von Hooker u.a. (2009) vorgeschlagenen Methodik zur Vermeidung paradoxer Effekt steht die Einführung von Parameterrestriktionen bei der Maximierung der Loglikelihood bzw. der Posterioridichte.

Für jedes Personenpaar (j, l), welches geordnete Antwortmuster - d.h. $\boldsymbol{u}_j \prec \boldsymbol{u}_l$ oder $\boldsymbol{u}_j \succ \boldsymbol{u}_l$ - aufweist, wird die Restriktion $\boldsymbol{\theta}_j \prec \boldsymbol{\theta}_l$ (bzw. $\boldsymbol{\theta}_j \succ \boldsymbol{\theta}_l$) als Nebenbedingung bei der Maximierung eingeführt. Das so entstehende, generalisierte Gleichungssystem[40] kann prinzipiell mit iterativen Methoden gelöst werden. Aufgrund der mit der Anzahl vorhandener Parameterrestriktionen wachsenden Komplexität des Optimierungsproblems, ist jedoch davon auszugehen, dass die Methode höchstens in

[40]Die Übersetzung des Maximierungsproblems in ein (Un)Gleichungssystem kann z.B. mittels des Begriffs des Normalenkegels (*normal cone*) erfolgen. Hierzu sei auf Dontchev und Rockafellar (2009, S.67ff.) verwiesen.

Gegenwart eines geringen Ausmaßes paradoxer Effekte zu adäquaten Ergebnissen führt.

Insbesonders ist die intraindividuelle Variante des paradoxen Effekts mit dieser Methodik nicht vermeidbar, da eine Auflösung der intraindividuellen Variante neben den bereits bestehenden (interindividuell gültigen) Restriktionen auch die Einführung von fiktiven Personen bzw. Antwortmustern erfordern würde, die jedes denkbare Antwortmuster abdecken. Dies stellt selbst für geringe Testlängen ein nicht zu bewältigendes numerisches Problem dar.

Die dargelegte Methodik kann demnach lediglich die interindividuelle Form des Paradoxons in Tests mit einer geringen Anzahl paradoxer Effekte beheben, da lediglich diese Struktur mit einem Optimierungsproblem moderater Größe vereinbar ist. Darüber hinaus impliziert die Anwendung dieser Methodik eine unerwünschte Abhängigkeit zwischen dem Schätzwert der Fähigkeiten einer Person und der Testperformance einer anderen Person. Der Schätzwert einer Person ist somit nicht ausschließlich von dem Antwortmuster der betreffenden Person abhängig, sondern von dem kompletten System an beobachteten Antwortmustern der N verschiedenen Personen $\hat{\boldsymbol{\theta}}_l = f(\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_N)$. Letzteres kann in sich wiederum auf ein neues Paradoxon führen, da die Antwort auf die Frage, ob die Fähigkeit einer Person einen Schwellenwert κ_i überschreitet, somit abhängig von den Antwortmustern anderer Personen ausfällt. Die gleiche Person hätte den Schwellenwert dann z.B. nicht überwunden, wenn eine andere Person ein anderes Antwortmuster gezeigt hätte.

Trotz wünschenswerter statistischer Eigenschaften (siehe Satz 4.1.1) stellt diese Methode folglich keine adäquate Lösung dar. Im nächsten Abschnitt wird daher eine andere Zugangsart zur Auflösung des paradoxen Effekts verfolgt: Anstatt auf der Ebene des Schätzers zu operieren, fokussieren sich die Methodiken des folgenden Abschnitts auf die Rolle der Modellformulierung. Die grundlegenden Schätzprinzipien - d.h. i.d.R unrestringierte Maximierung einer Likelihood bzw. einer Posterioridichte - werden beibehalten und die Vermeidung des paradoxen Effekts erfolgt durch „adäquate“ Spezifikation des IRT-Modells.

4.3 Umformulierung des Item-Response-Modells

Eine erste, bereits in der Einleitung dieses Kapitels erwähnte Methode, die das Konstruktionsprinzip des paradoxen Effekts aus Kapitel 2.3 konterkariert, beruht auf dem SMID-Konzept. Das Gegenstück zur SMDD-Bedingung erlaubt eine Umkehrung des Resultats aus Lemma 2.3.4, d.h. unter der SMID-Bedingung ist die bedingte Maximumsabbildung monoton wachsend. Mit Hilfe des in Satz 2.3.1 dargelegten Beweisprinzips folgt daraus, dass die Schätzer der beiden Dimensionen gleich geordnet sind (in der Terminologie von Satz 2.3.1 ausgedrückt: $y_f \geq y_g$, $x_f \geq x_g$). Man beachte ferner, dass die SMDD-Bedingung aus Lemma 2.3.4 durch die schwächere MDD-Bedingung ersetzt werden kann, wenn zusätzlich angenommen wird, dass die bedingte Maximumsabbildung $\hat{x}(y)$ (lokal) eine echte - d.h. einelementige - Funktion ist.
Analoges gilt natürlich für die MID-Bedingung, so dass folglich die Formulierung eines TP_2-Modells eine potentielle Lösung des Problems darstellt. Wenngleich ein erstes Beispiel hierfür in Satz 3.2.1 im Kontext der bayesianischen Analyse eines zweidimensionalen IRT-Modells festgehalten wurde, muss dennoch konstatiert werden, dass im Regelfall - abgesehen von Fällen einer separierbaren Likelihood bzw. von Fällen, in denen die TP_2-Bedingung durch eine entsprechend „starke“ Prioriverteilung erzwungen wird - die globale TP_2-Bedingung nicht erfüllt wird.
Die folgenden Betrachtungen sollen auf einfache Art nahelegen, warum eine globale TP_2-Bedingung i.d.R inkompatibel mit üblichen IRT-Modellierungsansätzen ist. Im Kern der Betrachtungen steht der in der folgenden Definition formalisierte Gedanke kompensatorischen Grenzwertverhaltens.

Definition 4.3.1. Ein Item i eines zweidimensionalen IRT-Modells besitzt *kompensatorisches Grenzwertverhalten*, wenn eine Antwortkategorie l sowie ein Paar (θ^*, ψ^*) existieren, so dass

$$\lim_{\psi\to\infty} f_{il}(\theta^*, \psi) = \lim_{\theta\to\infty} f_{il}(\theta, \psi^*) = \lim_{\theta\to\infty}\lim_{\psi\to\infty} f_{il}(\theta, \psi) > f_{il}(\theta^*, \psi^*)$$

gilt, wobei $f_{il}(\theta, \psi)$ die (bedingte) Wahrscheinlichkeit von Antwortkategorie l bezüglich Item i kennzeichnet.

Ein einfaches Beispiel kompensatorischen Grenzwertverhaltens ist durch ein Item

eines $M2PL$-Modells mit zugehöriger Item-Response-Funktion

$$f_{i1}(\theta, \psi) = \frac{\exp(a_{i1}\theta + a_{i2}\psi + \beta_i)}{1 + \exp(a_{i1}\theta + a_{i2}\psi + \beta_i)}$$

und positiven Ladungen $a_{i1} > 0, a_{i2} > 0$ gegeben. Ein *beliebiges* Wertepaar (θ, ψ) erfüllt hier die in Definition 4.3.1 dargelegte Bedingung bezüglich (θ^*, ψ^*), da stets

$$1 = \lim_{\psi \to \infty} f_{i1}(\theta, \psi) = \lim_{\theta \to \infty} f_{il}(\theta, \psi) = \lim_{\theta \to \infty} \lim_{\psi \to \infty} f_{il}(\theta, \psi) > f_{il}(\theta, \psi)$$

gilt. Wie leicht ersichtlich, bleibt dieses kompensatorische Grenzwertverhalten bestehen, wenn anstelle der logistischen Verteilungsfunktion eine andere Verteilungsfunktion gewählt wird. Jedes Item i dessen Item-Response-Funktion durch den Ausdruck

$$f_{i1} = F(a_{i1}\theta + a_{i2}\psi)$$

mit einer Verteilungsfunktion F und positiven Ladungen $a_{i1} > 0, a_{i2} > 0$ gegeben ist, weist kompensatorisches Grenzwertverhalten auf. Es sei jeodch bemerkt, dass das in Definition 4.3.1 dargestellte Konzept wesentlich allgemeiner ist, als diese Beispiele vermuten lassen, da es lediglich auf der Existenz *eines* Paares (θ^*, ψ^*) beruht, während in den erörterten Beispielen stets jedes beliebige Paar die Rolle von (θ^*, ψ^*) übernehmen konnte.
Mithilfe des Konzepts des kompensatorischen Grenzwertverhaltens kann nun auf einfache Art demonstriert werden, warum viele IRT-Modelle inkompatibel mit einer globalen TP_2-Bedingung sind.

Lemma 4.3.1. *Weist ein Item i eines (zweidimensionalen) IRT-Modells kompensatorisches Grenzwertverhalten auf, so ist die korrespondierende Kategorie-Response-Funktion f_{il} nicht konform mit einer globalen TP_2-Bedingung.*

Beweis. Im Folgenden bezeichnen l sowie (θ^*, ψ^*) die zur Definition 4.3.1 zugehörige Kategorie bzw. das zur Definition korrespondierende Wertepaar.
Angenommen die Response-Funktion f_{il} ist TP_2. Man wähle monoton wachsende, unbeschränkte Folgen $(\theta_n)_{n\in\mathbb{N}}$, $(\psi_n)_{n\in\mathbb{N}}$. Dann existiert ein Index N, derart, dass für alle $n, m \geq N$ sowohl $\theta_n > \theta^*$ als auch $\psi_m > \psi^*$ gilt. Eine Anwendung der TP_2-Bedingung auf die korrespondierende „Rechtecksfolge“ ergibt

$$f_{il}(\theta^*, \psi^*) f_{il}(\theta_n, \psi_m) \geq f_{il}(\theta^*, \psi_m) f_{il}(\theta_n, \psi^*).$$

Nach Definition 4.3.1 existiert für $m \to \infty$ auf beiden Seiten der Grenzwert und es folgt für alle $n \geq N$ die Beziehung

$$f_{il}(\theta^*, \psi^*) \lim_{\psi \to \infty} f_{il}(\theta_n, \psi) \geq \lim_{\psi \to \infty} f_{il}(\theta^*, \psi) f_{il}(\theta_n, \psi^*).$$

Bildet man hierfür den Grenzwert $n \to \infty$, so folgt

$$f_{il}(\theta^*, \psi^*) \lim_{\theta \to \infty} \lim_{\psi \to \infty} f_{il}(\theta, \psi) \geq \lim_{\psi \to \infty} f_{il}(\theta^*, \psi) \lim_{\theta \to \infty} f_{il}(\theta, \psi^*).$$

Letzteres steht im direkten Widerspruch zu einem kompensatorischen Grenzwertverhalten. □

Die Forderung nach einem globalen TP_2-Modell steht folglich im Widerspruch zu einer grundlegenden kompensatorischen Eigenschaft, die den meisten IRT-Modellen inhärent ist. Mit anderen Worten: Der potentielle Lösungsweg, ein faires TP_2-Modell aufzustellen, stößt, durch die Unmöglichkeit kompensatorisches Testverhalten widerspiegeln zu können, an seine Grenzen.
Die hiervon angedeutete Inkompatibilität der TP_2-Bedingung mit gängigen Modellierungsannahmen lässt sich noch zuspitzen, indem eine Modellklasse der Form

$$l_u(\boldsymbol{\theta}) = \sum_i l_i(\boldsymbol{a}_i^T \boldsymbol{\theta}) \quad (\boldsymbol{a}_i^T \in \mathbb{R}_+^p)$$

betrachtet wird. Eine direkte Berechnung der „gemischten“ Ableitungen (siehe Anhang D) zeigt, dass bei Einführung der Forderung $l_i'' \geq 0$ ein globales MTP_2-Modell resultiert. Ersetzt man demnach die im Rahmen der linear-kompensatorischen Modelle übliche Annahme $l_i'' < 0$ durch ihr „Gegenteil“, so erhält man ein IRT-Modell, welches faire Klassifikationen ermöglicht. Doch welche Konsequenzen ergeben sich aus der postulierten Ungleichung $l_i'' \geq 0$? Zunächst folgt die Konvexität der Funktion l_i. Da die Exponentialfunktion konvex sowie monoton wachsend ist, ergibt sich ferner die Konvexität der Funktion $\exp(l_i)$ (Rockafellar, 1970, Theorem 5.1). Letztere ist in allen Modellen, in denen $\exp(l_i)$ als Wahrscheinlichkeit interpretiert werden kann, zwangsläufig nach oben beschränkt, woraus mittels Korollar 8.6.2 aus Rockafellar (1970) folgt, dass l_i *konstant* ist. Das einzige Modell, welches der Forderung $l_i'' \geq 0$ genügt und zugleich die IRT-Modellen inhärente Beschränktheit der Loglikelihoodbeiträge gewährleistet, ist demnach das Modell, in dem die Antwortwahrscheinlichkeiten nicht von $\boldsymbol{\theta}$ abhängen! Ähnlich wie im Falle kompensatorischen Grenzwertverhaltens resultiert somit ein Verstoß gegen eine basale Eigenschaft eines

IRT-Modells (hier: die triviale Eigenschaft der Beschränktheit), was die Schwierigkeit, ein „vernünftiges“ IRT-Modell mit TP_2-Eigenschaft zu konstruieren, verdeutlicht.

Hervorzuheben ist, dass die bisherigen Lösungsversuche auf der Formulierung eines mehrdimensionalen (TP_2-)IRT-Modells beruhten, in dem „übliche“ Schätzmethoden, d.h. Maximierung einer von *mehreren* unbekannten Parametern abhängigen Funktion, zu einer fairen Einzelfalldiagnose führen. Im Gegensatz hierzu fußt die von Hooker und Finkelman (2010) im Kontext von sogenannten „item bundles“ vorgeschlagene Lösungsmethode auf dem Konzept der Eindimensionalität. Anders als die bisherige, TP_2-basierte Methodik „akzeptiert“ dieser Ansatz eine (lokale) RR_2-Bedingung. Die Vermeidung des paradoxen Effekts wird stattdessen dadurch erzielt, dass das mehrdimensionale (lokale) RR_2-Modell auf ein oder mehrere eindimensionale, lokal abhängige IRT-Modell reduziert wird, in denen die zugehörigen Schätzer jeweils fair ausfallen.

Zur Verdeutlichung dieses Prinzips soll das in Abschnitt 3.1 betrachtete faktorenanalytische Modell herangezogen werden. Wie in Satz 3.1.3 dargelegt, führt die simultane, auf dem gemeinsamen Modell beruhende Schätzung der unbekannten Größen in Abwesenheit einer Einfachstruktur stets zu einem paradoxen Effekt. Grundlage der Schätzung ist hierbei die Verteilung von $\boldsymbol{u}$ unter der Bedingung $\boldsymbol{\theta}$:

$$\boldsymbol{u}|\boldsymbol{\theta} \sim N(A\boldsymbol{\theta}, \Omega).$$

Unter der zusätzlichen (konventionellen) Annahme, dass

$$\boldsymbol{\theta} \sim N(\mathbf{0}, \Sigma)$$

gilt, resultiert die folgende gemeinsame Verteilung für den Vektor $(\boldsymbol{u}, \boldsymbol{\theta})$

$$(\boldsymbol{u}, \boldsymbol{\theta}) \sim N\left(\begin{pmatrix}\mathbf{0}\\ \mathbf{0}\end{pmatrix}, \begin{pmatrix}\Omega + A\Sigma A^T & A\Sigma\\ \Sigma A^T & \Sigma\end{pmatrix}\right)$$

Dieses Modell impliziert p eindimensionale, lokal abhängige Modelle für $\boldsymbol{u}|\theta_j (j = 1, \ldots p)$. Gemäß den speziellen Eigenschaften der multivariaten Normalverteilung (siehe z.B. Mardia u.a., 1979, Kap.3) besitzen die bedingten Verteilungen die Form

$$\boldsymbol{u}|\theta_j \sim N(A\Sigma_{(j)}\sigma_{jj}^{-1}\theta_j, \Sigma^*), \tag{4.47}$$

worin $\Sigma_{(j)}$ die j-te Spalte und σ_{jj} den j-ten Diagonaleintrag der Kovarianzmatrix Σ bezeichnen[41].

[41] Wenngleich die Gestalt der bedingten Kovarianzmatrix Σ^* explizit darstellbar ist, wird darauf verzichtet, da sie für die folgende, den KQ-Schätzer betreffende Argumentation irrelevant ist.

Der KQ-Schätzer für θ_j bezüglich des durch $\boldsymbol{u}|\theta_j$ gegebenen (i.A. lokal abhängigen) IRT-Modells ist bis auf ein positives Vielfaches von der Gestalt

$$\hat{\theta}_j \propto (\Sigma_{(j)}^T A^T A \Sigma_{(j)})^{-1} \Sigma_{(j)}^T A^T \boldsymbol{u}. \tag{4.48}$$

Wenn die j-te Spalte der Kovarianzmatrix Σ ausschließlich aus nichtnegativen Einträgen besteht, dann sind alle Komponenten des Vektors $(\Sigma_{(j)}^T A^T A \Sigma_{(j)})^{-1} \Sigma_{(j)}^T A^T$ aufgrund der Nichtnegativität der Einträge von A nichtnegativ. Folglich besitzt eine Erhöhung eines Itemscores auf einem beliebigen Item stets einen positiven (bzw. nichtnegativen) Effekt. Für die Schätzung der j-ten Dimension entsteht somit kein Paradoxon. Fordert man die Nichtnegativität für jede Spalte von Σ, so wird demnach der paradoxe Effekt global - d.h. für jede Dimension - vermieden.

Das Beispiel hebt die hervorstechende Rolle der Eindimensionalität im Bezug auf die Umgehung des paradoxen Effekts hervor. Anstatt die Schätzung ausgehend von einem mehrdimensionalen, lokal unabhängigen IRT-Modell aufzubauen, erfolgte hier die Formulierung von p lokal abhängigen, eindimensionalen IRT-Modellen, in denen separat jeweils nur ein unbekannter Parameter zu schätzen ist.

Das vorgestellte Beispiel könnte fälschlicherweise den Eindruck erwecken, dass der Übergang zu eindimensionalen Modellen ein generelles Auflösungsprinzip darstellt. Durch ein Gegenbeispiel soll vor einer zu optimistischen Interpretation dieses teilweise wirksamen Auflösungsprinzips gewarnt werden.

Hierzu wird von einem bereits in Kapitel 3.2 umfassend diskutiertem IRT-Modell mit zusätzlicher Speedkomponente ausgegangen. Die Posteriori des Lognormal-Normal-Modells (siehe hierzu auch Abschnitt 3.2) kann unter Verwendung der Abkürzungen $g(\theta) := \prod_i f_{i,u_i}(\theta), g^*(\theta) := g(\theta)\exp(-\frac{1}{2}\sigma^{11}\theta^2)$ wie folgt dargestellt werden:

$$p(\theta, \tau|\boldsymbol{u}, \boldsymbol{t}) \propto g(\theta) \prod_i \exp(-\frac{1}{2}\alpha_i^2(\log(t_i) - \beta_i + \tau)^2) \exp(-\frac{1}{2}q(\theta, \tau))$$

$$p(\theta, \tau|\boldsymbol{u}, \boldsymbol{t}) \propto g^*(\theta) \exp\left(-\frac{1}{2}\left(\sum_i \alpha_i^2(\log(t_i) - \beta_i + \tau)^2 + 2\sigma^{12}\theta\tau + \sigma^{22}\tau^2\right)\right)$$

Der letzte Term ist wiederum proportional zu

$$g^*(\theta) \exp\left(-\frac{1}{2}a\left(\tau + \frac{b(\theta)}{2a}\right)^2\right) \exp\left(+\frac{1}{2}a\left(\frac{b(\theta)}{2a}\right)^2\right)$$

mit $a := \sigma^{22} + \sum_i \alpha_i^2$ und $b(\theta) := 2\sigma^{12}\theta + 2\sum_i \alpha_i^2(\log(t_i) - \beta_i)$.

Demzufolge ist

$$p(\theta|\boldsymbol{u}, \boldsymbol{t}) \propto g^*(\theta) \exp\left(+\frac{1}{2}\left(\frac{b(\theta)}{2a}\right)^2 a\right) \int \exp\left(-\frac{1}{2}a\left(\tau + \frac{b(\theta)}{2a}\right)^2\right) \mathrm{d}\tau$$

Da die Normalisierungskonstanten der Normalverteilungsdichte nicht von ihrem Erwartunsgwert abhängen, gilt folglich:

$$p(\theta|\boldsymbol{u},\boldsymbol{t}) \propto g^*(\theta)\exp\left(+\frac{1}{2}\left(\frac{b(\theta)}{2a}\right)^2 a\right)$$

$$\gamma(\theta|\boldsymbol{u},\boldsymbol{t}) = \sum_i \log(f_{i,u_i}(\theta)) - \frac{1}{2}\sigma^{11}\theta^2 + \frac{b(\theta)^2}{8a}$$

Ableiten nach θ führt für den „marginalisierten“ MAP-Schätzer auf die Schätzgleichung

$$\sum_i \frac{f'_{i,u_i}(\theta)}{f_{i,u_i}(\theta)} - \sigma^{11}\theta + (2a)^{-1}b(\theta)\sigma^{12} = 0. \tag{4.49}$$

Das gleiche Argumentationsprinzip, welches im Kontext der Sensitivität des Modells gegenüber extremen Antwortzeiten etabliert wurde (d.h. Erzielung einer Kontradiktion durch Gegenüberstellung der Unbeschränktheit des $\log(t_i)$-Terms mit der Tatsache, dass eine stetige Funktion in Gegenwart von in einer kompakten Menge liegenden Schätzwerten beschränkt wäre.), führt dann für dieses eindimensionale IRT-Modell zu den gleichen Schlussfolgerungen wie im Falle des mehrdimensionalen IRT-Modells, d.h. in Abhängigkeit des Vorzeichens von σ_{12} kann ein beliebig hoher/geringer Schätzwert erzielt werden. Insbesonders bleibt - wie durch eine Anwendung des Satzes über implizite Funktionen auf (4.49) in Kombination mit entsprechenden „Regularitätsbedingungen“(logkonkave Kategorie-Response-Funktionen + genügend Iteminformation) leicht ersichtlich - die Möglichkeit eines paradoxen Effekts für den Fall negativ korrelierter Dimensionen auch in dieser marginalisierten Form des Modells bestehen.

Die Umformulierung eines IRT-Modells in mehrere eindimensionale, lokal abhängige Modelle garantiert somit nicht zwingend die Vermeidung des paradoxen Effekts. Die so resultierende Klasse eindimensionaler Modelle muss bestimmte, i.A. nichttriviale Bedingungen erfüllen, damit faire Inferenz ermöglicht wird. Eine potentielle Bedingung, die die Auflösung durch das Eindimensionalitätsprinzip erlaubt, wurde von Hooker und Finkelman (2010, Theorem 8.1) dargelegt. Sie ist im nachfolgenden Lemma abschließend kurz paraphrasiert.

Lemma 4.3.2. *Gegeben sei ein Schätzer, der als Lösung eines Gleichungssystem der Form $f(\theta,\boldsymbol{u}) = 0$ darstellbar ist, d.h. die Beziehung $\hat{\theta}(\boldsymbol{u}) \in \{\theta|f(\theta,\boldsymbol{u}) = 0\}$ erfüllt. Wenn f streng monoton fallend in θ für alle Vektoren $\boldsymbol{u}$ ist, und wenn f*

für alle θ die natürliche partielle Ordnung bezüglich $\boldsymbol{u}$ erhält (d.h. aus $\boldsymbol{u}_1 \prec \boldsymbol{u}_2$ folgt $f(\theta, \boldsymbol{u_1}) \leq f(\theta, \boldsymbol{u}_2)$) dann ist die Schätzprozedur $\hat{\theta}$ fair.

Beweis. Siehe Hooker und Finkelman (2010). □

4.4 Neudefinition des interessierenden Konstrukts

Die in diesem Abschnitt darzustellende Methodik unterscheidet sich von den bisher erwähnten Lösungsversuchen auf grundlegende Weise. Ziel der bisherigen Betrachtungen war die paradoxiefreie (und „statistisch sinnvolle“) Schätzung der latenten Fähigkeiten. Diese Schätzungen der verschiedenen Dimensionen können gemäß unterschiedlicher Entscheidungsregeln zur Klassifikation des Individuums beitragen. Eine Möglichkeit besteht hierbei in der Verwendung separater Schwellenwerte $\kappa_i (i = 1, \ldots p)$, deren simultane Überschreitung notwendig für die Erlangung einer positiven Klassifikation ist. Eine alternative Variante zu dieser „multiple hurdle rule“ (Segall, 2010) fußt auf der Gewichtung der Schätzungen $\hat{\theta}_1, \hat{\theta}_2, \ldots \hat{\theta}_p$ und der anschließenden Bildung eines (gewichteten) Schätzers, der die alleinige Grundlage der Klassifikation bildet.
Formaler ausgedrückt: Die Inferenz erfolgt bezüglich einer Linearkombination $\boldsymbol{\lambda}^T \hat{\boldsymbol{\theta}}$ der Dimensionen. Letzteres lässt sich natürlich auch als Neudefinition des Konstrukts ($\theta^* := \boldsymbol{\lambda}^T \boldsymbol{\theta}$) bezüglich dessen die Inferenz stattfindet, interpretieren.
Die im Rahmen des paradoxen Effekts zentrale Frage nach der Möglichkeit fairer Inferenz - in diesem Kontext also die Frage nach der Existenz einer paradoxiefreien Linearkombination - lässt sich für eine bestimmte Schätztypklasse positiv beantworten. Wie nachfolgend demonstriert wird, bietet jede aus der Klasse der linearen Schätzer stammende, und zugleich eine milde Regularitätsbedingung erfüllende Statistik die Möglichkeit, die Inferenz basierend auf einer fairen Linearkombination ihrer Komponenten vornehmen zu können.
Gleichwohl die Hauptanwendung dieses Resultats im faktorenanalytischen Kontext liegen dürfte, da hier die „üblichen“ statistischen Schätzverfahren stets auf lineare Schätzer führen, so sei an dieser Stelle betont, dass die im folgenden Satz dargelegte, faire Linearkombination lediglich auf der Verwendung eines linearen Schätzers basiert. Mit anderen Worten: Existiert in einer prinzipiell beliebigen IRT-Modellklasse ein „statistisch vernünftiger“ *linearer* Schätzer, so kann mittels der Anwendung von

Satz 4.4.1 unter schwachen Regularitätsbedingungen die Möglichkeit einer fairen, auf einem Gewichtungsverfahren basierenden Inferenz eingeräumt werden.
Um die zur weiteren Analyse notwendige formale Notation einzuführen, seien in dem Vektor $\boldsymbol{f(u)} = \sum_i \boldsymbol{e}_i f_i(\boldsymbol{u})$ zusammengeführte, lineare Schätzer $f_i(\boldsymbol{u})$ $(i = 1, \ldots p)$ für die unbekannten Parameter θ_i $(i = 1, \ldots p)$ gegeben. Ferner bezeichne $C := \{\boldsymbol{u} | u_i \geq 0 \; \forall i\}$ den nichtnegativen Orthanten des $\mathbb{R}^k$. Mittels dieser Notation ist die Frage nach der Fairness einer gegebenen Linearkombination $\hat{\psi}(\boldsymbol{u}) := \boldsymbol{\lambda}^T \boldsymbol{f(u)}$ äquivalent zur Frage, ob die durch $\hat{\psi}$ gegebene lineare Abbildung die Beziehung $\hat{\psi}(C) \subset [0, \infty)$ erfüllt - vorausgesetzt die gegebenen Dimensionen stehen in einer positven Beziehung zur Lösung der Items und vorausgesetzt $\boldsymbol{\lambda}$ weist keine negativen Einträge auf, so dass die positive Beziehung auch nach Neudefinition erhalten bleibt.
Der folgende Satz kann als direkter Abkömmling eines klassischen Resultats der konvexen Analysis (Rockafellar, 1970, Theorem 21.1) angesehen werden. Verfolgt man die Ursprünge dieses klassischen Resultats, so stößt man wiederum auf einen der wichtigsten Sätze der modernen Mathematik - den Satz von Hahn-Banach (siehe z.B. Lax, 2002, Kap. 3).

Satz 4.4.1. *Wenn kein Vektor $\boldsymbol{u} \in C$ mit der Eigenschaft $f_i(\boldsymbol{u}) < 0$ für alle i existiert, dann gibt es einen nichtnegativen Vektor $\boldsymbol{\lambda} \neq \boldsymbol{0}$, der die Beziehung*

$$\psi(\boldsymbol{u}) := \boldsymbol{\lambda}^T \boldsymbol{f(u)} \geq 0 \quad \forall \boldsymbol{u} \in C, \tag{4.50}$$

erfüllt. Der Vektor $\boldsymbol{\lambda}$ bildet eine Normale zu einer Hyperebene, die den nichtpositiven Orthanten des $\mathbb{R}^p$ von der konvexen Menge

$$D := \{\boldsymbol{\xi} = (\xi_1, \ldots, \xi_p) \in \mathbb{R}^p \mid \exists \, \boldsymbol{u} \in C, f_i(\boldsymbol{u}) < \xi_i \quad \textit{für } i = 1, \ldots, p\}$$

separiert.

Beweis. Der Satz ist ein Spezialfall von Theorem 21.1 von Rockafellar (1970). □

Man beachte, dass die Ungleichung (4.50) genau den Umstand widerspiegelt, dass der Schätzer für das neu definierte Konstrukt $\psi := \boldsymbol{\lambda}^T \boldsymbol{\theta}$ frei von Paradoxien ist.
Die Menge aller nichtnegativen Vektoren $\boldsymbol{\lambda}$, die der „Fairness-Bedingung“ (4.50) genügen, bildet einen konvexen Kegel (*convex cone*), d.h. sie ist abgeschlossen unter Addition und positiver Skalarmultiplikation. I.A. ist ferner nicht damit zu rechnen,

dass der Vektor $\boldsymbol{\lambda}$ (bis auf Multiplikation) eindeutig ausfällt[42] - wie der folgende Satz aufzeigt:

Satz 4.4.2. *Sei $\psi(\boldsymbol{u}) := \boldsymbol{\lambda}^T \boldsymbol{f}(\boldsymbol{u}) = \sum_i \lambda_i f_i(\boldsymbol{u})$ ($\lambda_i \geq 0 \,\forall\, i$) eine faire Linearkombination der ursprünglichen Schätzkomponenten $(f_i(\boldsymbol{u}))_{i=1,\dots p}$.*
Gilt $\psi(\boldsymbol{u}) \neq 0$ für alle $\boldsymbol{u} \in C \setminus \{\boldsymbol{0}\}$, so existieren p linear unabhängige, nichtnegative Vektoren $(\boldsymbol{\lambda}_i)_{i=1,\dots p}$, deren zugeordnete Schätzer $\psi_i(\boldsymbol{u}) := \boldsymbol{\lambda}_i^T \boldsymbol{f}(\boldsymbol{u})$ die Ungleichung $\psi_i(\boldsymbol{u}) \geq 0$ für alle $\boldsymbol{u} \in C$ erfüllen.
Die Existenz einer „fairen" Linearkombination, die der Bedingung $\boldsymbol{\lambda}^T \boldsymbol{f}(\boldsymbol{u}) > 0$ für alle $\boldsymbol{0} \neq \boldsymbol{u} \in C$ genügt, ist ferner stets gewährleistet, wenn kein Vektor $\boldsymbol{u} \in C \setminus \{\boldsymbol{0}\}$ existiert, der die Ungleichungen $f_i(\boldsymbol{u}) \leq 0$ für alle i erfüllt.

Beweis. Der Beweis ist zweiteilig. Im ersten Teil wird von der Existenz der Linearkombination ausgegangen und der erste Teil des Satzes bewiesen. Im zweiten Teil wird gezeigt, dass unter der Bedingung der Nichtexistenz eines Vektors $\boldsymbol{u} \in C \setminus \{\boldsymbol{0}\}$, der $f_i(\boldsymbol{u}) \leq 0$ für alle $i = 1, \dots p$ erfüllt, die Existenz der Linearkombination gewährleistet ist.
Sei A die darstellende Matrix von $\boldsymbol{f}(\boldsymbol{u})$ bezüglich der Standardbasis des $\mathbb{R}^p$ bzw. des $\mathbb{R}^k$. Nach Voraussetzung gilt $\boldsymbol{\lambda}^T A\boldsymbol{u} > 0$ für alle $\boldsymbol{u} \in C \setminus \{\boldsymbol{0}\}$.
Die Behauptung, es gäbe ein $\delta > 0$, so dass für alle $\boldsymbol{\lambda^*} \in \mathbb{B}_\delta(\boldsymbol{\lambda})$ (ebenfalls) die Ungleichung $\boldsymbol{\lambda^*}^T A\boldsymbol{u} \geq 0$ für alle $\boldsymbol{u} \in C$ gilt, wird im Folgenden durch Widerspruchsbeweis bewiesen.
Angenommen obige Behauptung ist falsch. Zu $\delta_n := \frac{1}{n}$ korrespondiert dann ein $\boldsymbol{\lambda}_n \in \mathbb{B}_{\delta_n}(\boldsymbol{\lambda})$ sowie ein $\boldsymbol{u}_n \in C$ mit $\boldsymbol{\lambda}_n^T A\boldsymbol{u}_n < 0$. Es kann hierbei - da C ein Kegel ist - o.B.d.A $|\boldsymbol{u}_n| = 1$ unterstellt werden, so dass die Folge $(\boldsymbol{u}_n)_{n\in\mathbb{N}}$ beschränkt ist und demnach eine konvergente Teilfolge $(\boldsymbol{u}_{n_v})_{v\in\mathbb{N}}$ besitzt.
Da die Funktion $g(\boldsymbol{x}, \boldsymbol{y}) := \boldsymbol{x}^T A\boldsymbol{y}$ eine *stetige* Bilinearform darstellt, ergibt sich durch Grenzwertübergang:

$$\boldsymbol{\lambda}_{n_v}^T A\boldsymbol{u}_{\boldsymbol{n_v}} < 0 \Rightarrow \boldsymbol{\lambda}^T A\boldsymbol{u}_{\boldsymbol{\infty}} \leq 0.$$

Der Vektor $\boldsymbol{u}_\infty := \lim_{v\to\infty} \boldsymbol{u}_{n_v}$ ist als Grenzwert von Vektoren, die aus der *abgeschlossenen* Menge C stammen, offenbar wiederum ein Element des nichtnegativen

[42] Ein spezifischer Vektor $\boldsymbol{\lambda}$, z.B. ein Vektor, der maximales Gewicht auf die erste Komponente legt, kann durch Übersetzung von (4.50) in ein lineares Programm (siehe z.B. Dantzig, 1962) bestimmt werden.

Orthanten. Wegen der Normgleichung $|\boldsymbol{u}_{n_v}| = 1$ gilt zudem $\boldsymbol{u}_\infty \neq \mathbf{0}$, so dass ein Widerspruch zur Eigenschaft $\psi(\boldsymbol{u}) > 0$ ($\forall \boldsymbol{u} \in C \setminus \{\mathbf{0}\}$) resultiert.

Zur *Existenz* der Linearkombination (zweiter Teil des Beweises):
Man definiere $C_1 := \{\boldsymbol{u} \in C | \sum_i u_i \geq 1\}$. Der „asymptotische Kegel" (*recession cone*) der Menge C_1 ist der nichtnegative Orthant: $0^+(C_1) = C$. Die „Abstiegsrichtungen" (*direction of recession*) von f_i sind diejeniegen Vektoren $\boldsymbol{u} \neq \mathbf{0}$ (modulo positive Skalarmultiplikation), die die Ungleichung $f_i(\boldsymbol{u}) \leq 0$ erfüllen (man wende hierzu beispielsweise Theorem 8.5 aus Rockafellar (1970) an).
Unter der Bedingung, dass es keine gemeinsame Abstiegsrichtung der Funktionen $(f_i)_{i=1,\ldots p}$ gibt, die gleichzeitig im asymptotischen Kegel von C_1 - der wiederum nichts anderes als eine Abstiegsrichtung der Indikatorfunktion (mit Werten $0/\infty$) von C_1 ist - liegt, garantiert Theorem 21.3 von Rockafellar (1970) die Existenz eines nichtnegativen Vektors $\boldsymbol{\lambda}$, der $\boldsymbol{\lambda}^T \boldsymbol{f}(\boldsymbol{u}) \geq \epsilon$ für ein $\epsilon > 0$ und für jeden Vektor $\boldsymbol{u} \in C_1$ erfüllt.
Da jeder Vektor $\boldsymbol{u} \neq \mathbf{0}$ des nichtnegativen Orthanten ein positives Vielfaches eines Vektors aus C_1 ist, folgt die Behauptung aus der Linearität von $\boldsymbol{f}$.

□

Die Existenz eines „vollen Satzes" linear unabhängiger Vektoren besitzt - wie im abschließenden Kapitel (Kapitel 5.2) dieser Arbeit demonstriert werden wird - semantische Konsequenzen, die weit über die rein pragmatische Möglichkeit, faire Inferenz vornehmen zu können, hinausreichen. Da die Erörterung dieser Konsequenzen jedoch aus dem Rahmen dieses Kapitels fällt, sollen an dieser Stelle lediglich einige Bemerkungen bezüglich der Voraussetzung des Satzes 4.4.2 gegeben werden. Letztere fordert, dass $\psi(\boldsymbol{u}) > 0$ für alle Vektoren $\boldsymbol{u} \neq \mathbf{0}$ des nichtnegativen Orthanten gilt und besagt somit, dass der faire (Ausgangs-)Schätzer auf einen Zuwachs des Itemscores eines beliebigen Items reagiert (und nicht konstant bleibt). Unter dieser (schwachen) Voraussetzung garantiert Satz 4.4.2 die Existenz einer aus nichtnegativen Vektoren $\boldsymbol{\lambda}_i$ bestehenden Basis des $\mathbb{R}^p$, deren Elemente zu fairen Linearkombinationen korrespondieren.
Die für die Existenz von $\boldsymbol{\lambda}$ hinreichende (und notwendige) Bedingung aus Satz 4.4.1, kann ferner als milde Regularitätsvoraussetzung angesehen werden, da eine Verletzung dieser Bedingung auf eine pathologische Eigenschaft des Schätzers hindeuten würde. Denn $f_i(\boldsymbol{u}^*) < 0$ (für alle i) an der Stelle $\boldsymbol{u}^* \in C$ impliziert für geeignet

gewählte geordnete Antwortvektoren ein paradoxes Resultat bezüglich *jeder* latenten Dimension. Eine schlechtere Testperformance führt dann zu einem Anstieg *aller* geschätzter Fähigkeiten.
Wie das folgende Lemma demonstriert, kann diese pathologische Bedingung für eine große Klasse linearer Schätzer ausgeschlossen und somit die Möglichkeit fairer Inferenz (via Satz 4.4.1) eingeräumt werden.

Lemma 4.4.1. *Jeder Schätzer der Form* $\boldsymbol{f(u)} := (A^TWA)^{-1}A^TW\boldsymbol{u}$ *erfüllt die Voraussetzung von Satz 4.4.2 (und damit insbesondere auch die Voraussetzung des Satzes 4.4.1), wenn* A *eine nichtnegative (Ladungs-)Matrix vollen Spaltenrangs mit Zeilen* $\boldsymbol{a}_j \neq \mathbf{0}$ *und* W *eine positive Diagonalmatrix bezeichnen.*

Beweis. Angenommen es existiert ein Vektor $\mathbf{0} \neq \boldsymbol{u^*} \in C$, so dass $f_i(\boldsymbol{u^*}) \leq 0$ für alle i ist. Dann folgt, dass das Skalarprodukt von $\boldsymbol{f(u^*)}$ mit jedem nichtnegativen Vektor $\boldsymbol{z}$ nichtpositiv ausfällt. Der Vektor $W\boldsymbol{u^*}$ besitzt aufgrund der Voraussetzung bezüglich W und der Tatsache, dass $\boldsymbol{u^*} \in C$ gilt, nichtnegative Einträge. Ferner gibt es mindestens einen strikt positiven Eintrag. Da kein Spaltenvektor von A^T dem Nullvektor gleicht, und da A aus nichtnegativen Einträgen besteht, weist der nichtnegative Vektor $\boldsymbol{z} := A^TW\boldsymbol{u^*}$ mindestens einen positiven Eintrag auf.
Für dieses $\boldsymbol{z}$ folgt

$$0 \geq \boldsymbol{z^T f(u^*)} = \boldsymbol{z^T}(A^TWA)^{-1}\boldsymbol{z},$$

was im Widerspruch zur positiven Definitheit von A^TWA steht. □

Gleichwohl das Resultat von Lemma 4.4.1 vielversprechend im Hinblick auf das Potential zur Auflösung des paradoxen Effekts wirkt, muss dennoch herausgestellt werden, dass die diesem Abschnitt zugrundeliegende Neudefinition des latenten Konstrukts in vielen Anwendungsfällen - vorwiegend in jenen Fällen, in denen Aussagen über a-priori definierte Dimensionen angestrebt werden - unzufriedenstellend ist. Wenn das Ziel in der Inferenz bezüglich (mehrerer) festgelegter, inhaltlich wohlinterpretierbarer Dimensionen liegt, dann ist eine Umdefinition des Konstrukts im Regelfall undenkbar, da eine sinnvolle, sachlogische Interpretierbarkeit der Linearkombination i.A. nicht gesichert ist.
Abschließend lässt sich somit festhalten, dass trotz einer Reihe möglicher Strategien zur Bewältigung des paradoxen Effekts keine allgemein zufriedenstellende Lösung existiert. Jede der diskutierten Methoden kann in spezifischen Kontexten hilfreich

und unter anderen Umständen wiederum ineffektiv im Hinblick auf die Vermeidung paradoxer Effekte ausfallen:

- Die asymptotische Sichtweise (Kapitel 4.1) tangiert lediglich eine spezielle Unterform des paradoxen Effekts und ist darüber hinaus nicht (unmittelbar) auf Skalen endlicher Länge anwendbar.
- Das Prinzip der Schätzung unter Restriktionen (Kapitel 4.2) scheitert an kombinatorischen bzw. davon implizierten computationellen Hürden.
- Methoden, die auf Umformulierung des IRT-Modells beruhen (Kapitel 4.3), stehen entweder im Widerspruch zu „gängigen“ Annahmen von IRT-Modellen (siehe z.B. Lemma 4.3.1) oder führen nur unter speziellen Umständen zu einer tatsächlich fairen Inferenzmethode (wie das in Kapitel 4.3 dargelegte „Gegenbeispiel“ eines Lognormal-Normal-Modells verdeutlicht)
- Methoden, denen eine Neudefinition des interessierenden Konstrukts zugrundeliegt (Kapitel 4.4), die für sich genommen bereits in vielen Kontexten inakzeptabel sein dürfte, können (bisher) lediglich für die eingeschränkte Klasse linearer (bzw. affiner) Schätzer verwendet werden.

5 Zur Semantik des paradoxen Effekts

Nachdem in unterschiedlichsten Kontexten die Omnipräsenz des paradoxen Effekts demonstriert wurde, stellt sich zwangsläufig die Frage nach seiner immanenten Bedeutung für die psychologische Testtheorie und Individualdiagnostik. Die zwei folgenden Abschnitte widmen sich abschließend der Beantwortung dieser Frage.
In Kapitel 5.1 steht der bereits in einzelnen Abschnitten angedeutete, potentielle Antagonismus zwischen einer fairen Einzelfallbeurteilung und einer zugleich statistisch effizienten Schätzung im Vordergrund, während Kapitel 5.2 den paradoxen Effekt in einen allgemeinen Rahmen, der schließlich in eine generelle Kritik des Latenten-Variablen-Ansatzes mündet, einbettet.
Konträr zu den bisherigen Darstellungen beruhen die meisten Ausführungen dieses Kapitels auf einer *subjektiven* Gesamtbeurteilung des Effekts. Wenngleich der Versuch einer wohlbegründeten Integration der Ergebnisse zum paradoxen Effekt unternommen wird, sei dennoch darauf hingewiesen, dass andere „Lesearten“ des paradoxen Effekts existieren (van der Linden, 2012), die insbesondere nicht die daraus abgeleitete Kritik an (mehrdimensionalen) modellbasierten Schätzern teilen.

5.1 Statistische Effizienz und individuumbasierte Fairness - ein Antagonismus?

Bevor einer potentiellen Kontrastierung von statistischer Effizienz und individuumbasierter Fairness nachgegangen wird, soll an dieser Stelle klar definiert werden, was im Folgenden unter dem Begriff der Fairness zu verstehen ist.
Ein Schätzer $\hat{\boldsymbol{\theta}}(\boldsymbol{u})$ ist fair, wenn er geordneten Antwortmustern stets (im gleichen Sinn) geordnete Schätzungen zuweist, d.h. wenn aus $\boldsymbol{u} \prec \boldsymbol{u}^*$ die Relation $\hat{\boldsymbol{\theta}}(\boldsymbol{u}) \prec \hat{\boldsymbol{\theta}}(\boldsymbol{u}^*)$ folgt. Hierbei wird implizit unterstellt, dass die Erlangung eines „höheren“ Schätzwerts[43] (in irgendeiner Form) mit positiven Konsequenzen für das Individuum einhergeht.
Man beachte, dass dieser Fairnessbegriff maßgeblich anhand der realen *Konsequenzen* für das Individuum festgemacht ist. Es erfolgt an dieser Stelle *ausdrücklich* kein

[43] Analog lässt sich der Begriff der Fairness natürlich auch im Kontext von Hypothesentests bzw. Klassifikationsentscheidungen gebrauchen.

Bezug zur Ebene der latenten Variable, d.h. der naheliegende Gedanke, das Fairnesskriterium - ähnlich wie in der Diskussion des paradoxen Effekts im Rahmen regulär kompensatorischer Modelle aus Kapitel 3.1 - anhand des Vorzeichens der Einträge der Ladungsmatrix zu definieren, wird bewusst verworfen. In Kapitel 5.2 wird gezeigt, dass letzteres Vorgehen - würde man es für die Defintion des Fairnessbegriffs heranziehen - in Konflikt mit dem Rotationsprinzip steht.
Die Verwendung des an den Konsequenzen orientierten Fairnessbegriffs dürfte darüber hinaus wesentlich stärker dem tatsächlichen Fairnessverständnis von Personen entsprechen, da Fairness zumeist mit Bezug auf konkrete Aktionen, die dem Individuum ermöglicht bzw. vorenthalten werden, erlebt wird. Ein Bezug zur latenten Ebene würde in diesem Kontext unnötig abstrakt und nahezu „fehlplatziert“ wirken.[44]
Wie in den einzelnen Abschnitten dargelegt, verstoßen „gängige“ Schätz- und Testverfahren, deren Anwendung in anderen Modellklassen meist zu effizienten sowie inhaltlich „vernünftigen“ Schätzungen führt, gegen dieses Prinzip. Da diese Schätzverfahren i.d.R. asymptotische Optimalitätseigenschaften aufweisen, drängt sich der Eindruck eines Antagonismus zwischen Effizienz und Fairness auf. Eine aufs Äußerste getriebene Überspitzung dieses Gegenspiels liefert das im Anhang C diskutierte Beispiel eines mehrdimensionalen Rasch-Modells, in dem die $H_0 : „\theta_1 \leq \kappa“$ aufgrund einer falschen Antwort auf einem bestimmten Item verworfen werden kann. Bei richtigem Beantworten hingegen wird die H_0 nicht verworfen und dies gilt unabhängig von der Anzahl übriger, korrekt beantworteter Items. Letzteres gipfelt in der Beobachtung, dass die korrespondierende Folge der p-Werte sogar monoton wächst, wenn weitere korrekt beantwortete Aufgaben hinzugefügt werden. Da der entsprechende Hypothesentest nach dem Neyman-Pearson Lemma (Lehmann und Romano, 2005, Kap.3) konstruiert wurde, ist diese (bizarre) Testprozedur statistisch betrachtet (dennoch) optimal. Aufgrund der Tatsache, dass es sich bei der Testprozedur um einen „finite sample“-Test handelt, ergeben sich zusammengenommen mit den in den einzelnen Kapiteln dargelegten Paradoxien, die sich vorwiegend auf Testprozeduren mit *asymptotischen* Optimalitätseigenschaften bezogen, erste Hinweise eines „Trade-Offs“ zwischen Fairness und Effizienz.
Eine weitere aufschlussreiche Möglichkeit sich dem potentiellen Antagonismus zu

[44] Auch wenn das gegebene Fairnesskriterium keine Verbindung zur latenten Ebene hat, so besteht diese natürlich auf indirekter Ebene, da zur Festsetzung der inhaltlichen Bedeutung der latenten Variable (und damit auch ihres korrespondierenden Schätzwerts) die Ladungsmatrix herangezogen wird.

nähern, besteht in einer Umkehr der bisherigen Betrachtungsweise. Anstelle der Frage, ob statistisch optimale Schätz-und Testmethoden elementaren Anforderungen bezüglich der Fairness einer auf ihnen basierenden Diagnostik genügen, tritt die Frage nach den statistischen Eigenschaften fairer Schätzverfahren.
Im Kontext des faktorenanalytischen Modells lässt sich ein Aspekt dieser Fragestellung gut veranschaulichen. Man nehme hierzu an, dass eine Ladungsmatrix A mit ausschließlich positiven Einträgen gegeben sei und dass eine Beschränkung auf lineare Schätzer erfolge. Unter diesen Annahmen lässt sich jeder faire Schätzer in der Form $C\boldsymbol{u}$ mit Einträgen $c_{vl} \geq 0$ darstellen. Die Forderung nach der Unverzerrtheit (Erwartungstreue) des Schätzers führt auf die Relation $CA = I$, wobei I eine $p \times p$ Einheitsmatrix darstellt. Damit die Relation erfüllt ist, muss insbesonders das Skalarprodukt zwischen der ersten Zeile $\boldsymbol{c}_1$ der Matrix C und der zweiten Spalte $\boldsymbol{a}_{(2)}$ der Matrix A verschwinden. Da die erste Zeile von C im Skalarprodukt mit der ersten Spalte von A Eins ergibt, kann die Zeile nicht dem Nullvektor entsprechen. Somit existiert mindestens ein positiver Eintrag, woraus die Unerfüllbarkeit der Forderung $\boldsymbol{c}_1^{\boldsymbol{T}}\boldsymbol{a}_{(2)} = 0$ resultiert. Hieraus folgt, dass in der Klasse der linearen, fairen Schätzer kein unverzerrter Schätzer existiert. Mit anderen Worten: Die Einschränkung auf faire Schätzer hat als Konsequenz, dass ein statistisches Gütekriterium, welches zuvor leicht erfüllbar war[45], nun nicht mehr erreicht werden kann.
In diesem Zusammenhang sei auch bemerkt, dass die von Satz 4.1.1 implizierte Konsistenzeigenschaften nicht zwingend bedeutet, dass der durch Einführung von Schätzrestriktionen (siehe Abschnitt 4.2) gewonnene Schätzer konsistent ist. Die Konsistenz bezieht sich nur auf diejenigen Schätzer, die lediglich dann eine zusätzliche Restriktion einführen, wenn zwei Personen mit geordneten Antwortmustern beobachtet wurden. Für die globale Fairness, d.h. zur Vermeidung der inter- *und* intraindividuellen Variante des paradoxen Effekts, bedarf es jedoch der Einführung von Restriktionen für alle potentiellen Antwortmuster. Über die Konsistenz dieser Schätzklasse gibt das Resultat von Satz 4.1.1 jedoch keine Auskunft.
Diese kurze Rekapitulation einiger Hauptergebnisse zur Kontrastierung zwischen Effizienz und Fairness deutet auf einen antagonistischen Effekt hin. Schätzprozeduren mit statistischen Optimalitätseigenschaften ist der paradoxe Effekt inhärent. Umgekehrt weisen „forciert“ faire Schätzer in vielen Fällen mangelhafte statistische Eigenschaften auf. Im Gegensatz zur ersten Feststellung (effiziente Schätzer sind

[45]Die Gleichung $CA = I$ besitzt ohne die Einschänkung auf nichtnegative Einträge eine Vielzahl an Lösungen, da $k \cdot p - p^2$ Einträge „frei variierbar“ sind.

unfair), die in einem allgemeinen Rahmen gültig ist, muss letztere Feststellung jedoch relativiert werden. Die in Kapitel 4.3 dargelegten, auf Marginalisierung des mehrdimensionalen Modells beruhenden Schätzer können u.U. - wie am Beispiel des faktorenanalytischen Modells erörtert wurde - faire Schätzungen ermöglichen und zugleich über adäquate statistische Eigenschaften verfügen[46]. Ebenso ist die in Kapitel 4.4 diskutierte Schätzung basierend auf dem Konzept einer Linearkombination sowohl fair als auch (bei Verwendung des Gewichteten-Kleinste-Quadrate-Schätzers für $\boldsymbol{f(u)}$) effizient. Beachtet man jedoch, dass die Schätzprozeduren aus Kapitel 4.3 und 4.4 einem gemeinsamen Prinzip folgen, welches auf dem Konzept der Eindimensionalität fußt, so lassen sich die Resultate dieser Arbeit wie folgt im Hinblick auf den potentiellen Antagonismus interpretieren:

- Statistisch effizienten, aus einem mehrdimensionalem IRT-Modell abgeleiteten Schätzern ist der paradoxe Effekt inhärent.
- Faire Schätzer verstoßen in vielen Fällen gegen elementare statistische Gütekriterien.
- Faire Schätzer, die nicht gegen diese elementaren Kriterien verstoßen, basieren (zumeist) indirekt auf dem Konzept der Eindimensionalität.

[46] Die Erwartungstreue kann z.B. unmittelbar anhand von (4.47) und (4.48) nachvollzogen werden.

5.2 Bedeutung für die psychologische Testtheorie

Wenn der paradoxe Effekt ein inhärentes Phänomen mehrdimensionaler IRT-Modelle ist - wofür die Betrachtungen der Kapitel 2.3, 3.1-3.3 starke Indizien liefern - und wenn zudem gemäß den Ergebnissen der Kapitel 4 und 5.1 kein zufriedenstellendes, allgemeingültiges Schätzprinzip zu existieren scheint, das fair ist und zugleich statistischen Qualitätskriterien genügt, dann steht das Gebiet der Individualdiagnostik basierend auf Latenten-Variablen-Modellen vor der Entscheidung, entweder statistisch effiziente Diagnosen zu fällen, die im Kern elementaren Fairnessprinzipien widersprechen, oder faire Diagnosen zu treffen, die fundamentalen Anforderungen an das Ziel einer (quantitativen) Diagnostik - nämlich der Bildung (im statistischen Sinn) akkurater Beurteilungen - zuwiderlaufen.
Ein gängiger Einwand (z.B. van der Linden, 2012, S.27-30) bezüglich dieses Trade-Offs basiert auf der Vorstellung, dass eine präzisere Diagnose gegenüber einer weniger akkuraten Diagnose aufgrund der Tatsache, dass Erstere näher am unbekannten wahren Wert liegt als Letztere, nicht weniger fair seien kann. Implizit wird somit das eingangs definierte Fairnesskriterium verworfen, bzw. für ungültig erklärt. Dieser Einwand ist nach Ansicht des Autors in zweierlei Hinsicht fehlerhaft. Zum Einen beruht er auf einer i.A. unzulässigen Gleichsetzung zwischen dem Konzept der statistischen Effizienz und der Präzision einer Diagnose im Einzelfall. Statistische Begriffe besitzen i.d.R. keine direkte Bedeutung für den Einzelfall, sondern fußen auf dem abstrakten Konzept einer Wahrscheinlichkeit. Als solches sind sie höchstens indirekt - mit Hilfe der Vorstellung einer Folge unabhängiger Wiederholungen des Zufallsvorgangs - interpretierbar. Zum Anderen stellt das zuvor definierte Fairnesskonzept die intuitive und zugleich vernünftige Forderung auf, dass richtiges Beantworten für eine Person niemals negative Konsequenzen haben sollte, während die Gleichsetzung von Fairness mit Akkuratheit die vom Paradoxon gestellte Problematik in tautologischer Manier „wegdefiniert“.
Das Paradoxon stellt folglich (bei Akzeptanz des Fairnesskriteriums) ein genuines Problem der psychologischen Testtheorie dar. Es würde jedoch der Bedeutung des paradoxen Effekts nicht gerecht werden, wenn man ihn „lediglich“ als eine Warnung vor der Verwendung mehrdimensionaler IRT-Modelle zu individualdiagnostischen Zwecken begreift. Der paradoxe Effekt besitzt weitere Implikationen, die sich erst herauskristallisieren, wenn man ihn im Licht bereits bekannter Kritikpunkte des Latenten-Variablen-Paradigmas darstellt. Im Folgenden sollen daher zunächst einige

„klassische“ Kritikansätze des Latenten-Variablen-Paradigmas dargelegt und in Beziehung zum paradoxen Effekt gebracht werden.

Das Rotationsproblem

Als eines der ersten Kernkritikpunkte des Paradigmas lässt sich das sogenannte Rotationsproblem, bzw. allgemeiner die Uneindeutigkeit der Metrik, auf der die latenten Größen gemessen werden, anführen. Die Struktur eines orthogonalen Faktorenmodells der Form

$$\boldsymbol{u} = A\boldsymbol{\theta} + \boldsymbol{\epsilon}, \quad \mathrm{Cov}(\theta_i, \theta_j) = \delta_{ij}$$

bleibt durch Neudefinition des latenten Vektors $\tilde{\boldsymbol{\theta}} := G\boldsymbol{\theta}$

$$\boldsymbol{u} = AG^T\tilde{\boldsymbol{\theta}} + \boldsymbol{\epsilon}, \quad \mathrm{Cov}(\tilde{\theta}_i, \tilde{\theta}_j) = \delta_{ij}$$

unberührt, insofern G orthogonal ist. Verzichtet man auf die Forderung nach orthogonalen Faktoren, so kann ferner jede invertierbare Matrix B die Rolle von G übernehmen. Als unmittelbare Konsequenz ergibt sich, dass die resultierende Ladungsmatrix lediglich bis auf Multiplikation mit einer invertierbaren Matrix identifizierbar ist.
Neben der Tatsache, dass auf diese Art die wahren, zugrundeliegenden Faktoren - insofern diese existieren - quasi nicht „entdeckt“ werden können, birgt diese Uneindeutigkeit des Weiteren das Problem, dass die darauffolgende Skalenkonstruktion elementaren Invarianzprinzipien zuwiderläuft. Geht man beispielsweise zur Formung von Subskalen nach der (durchaus üblichen) Heuristik vor, diejenigen Items in die Skala einfließen zu lassen, deren korrespondierende Ladungen entsprechend hoch ausfallen, resultieren je nach Ausgangsmatrix - A vs. AB (B invertierbar) - verschiedene Skalen. Zwei Items, die ausgehend von der Ladungsmatrix A einer Subskala zugewiesen wurden, können bezüglich der Ausgangsmatrix AB in verschiedene Subskalen selektiert werden. Letzteres verletzt offenbar elementare Invarianzprinzipien[47].
Um das enorme Ausmaß dieser „Willkür“ zu verdeutlichen, sei darauf hingewiesen, dass für jede Auswahl p linear unabhängiger Ladungsvektoren $(\boldsymbol{a}_{i_1}, \boldsymbol{a}_{i_2}, \ldots, \boldsymbol{a}_{i_p})$ von A und jede Vorgabe p linear unabhängiger „Zielvektoren“ $(\tilde{\boldsymbol{a}}_1, \tilde{\boldsymbol{a}}_2, \ldots, \tilde{\boldsymbol{a}}_p)$ stets eine invertierbare Matrix B existiert, die $\tilde{\boldsymbol{a}}_l = B\boldsymbol{a}_{i_l}$ erfüllt. Wählt man als Zielvektoren

[47] Grob gesagt fordert das Invarianzprinzip, dass der Entscheidungsprozess des Forschers bei statistisch ununterscheidbaren Lösungen zu ähnlichen bzw. gleichen Resultaten führen sollte.

die Standardbasis des $\mathbb{R}^p$, so folgt, dass p Items mit potentiell äußerst ähnlichen (aber nicht kollinearen) Ladungsvektoren in p Items überführbar sind, die jeweils ausschließlich eine (eigene) Dimension messen.
Das beschriebene Uneindeutigkeits- bzw. Rotationsproblem besitzt darüber hinaus eine verblüffende Verbindung zum paradoxen Effekt, die an dieser Stelle erörtert werden soll. Man betrachte hierzu einen zweidimensionalen, faktorenanalytischen Test mit korrespondierenden Ladungsvektoren $\boldsymbol{a}_1^T = (1,0), \boldsymbol{a}_2^T = (1,1), \boldsymbol{a}_3^T = (0,1)$. Gemäß Satz 3.1.3 führt eine Anwendung des Kleinste-Quadrate-Schätzers (KQ), der für dieses Zahlenbeispiel - wie sich unmittelbar ausrechnen lässt - die lineare Form $f(\boldsymbol{u}) = cM\boldsymbol{u}$ $(c > 0)$ mit

$$M = \begin{pmatrix} 2 & 1 & -1 \\ -1 & 1 & 2 \end{pmatrix}$$

besitzt, zu Paradoxien. Laut Satz 4.4.1 und Lemma 4.4.1 existiert jedoch mindestens eine „faire“ Linearkombination, d.h. es existieren nichtnegative Gewichtungsvektoren $(\boldsymbol{\lambda}_i)_{i\in I}$, so dass $\boldsymbol{\lambda}_i^T f(\boldsymbol{u}) \geq 0$ für alle $i \in I$ und alle $\boldsymbol{u}$ aus dem nichtnegativen Orthanten gilt.
Für den vorliegenden Fall genügen (beispielsweise) die beiden linear unabhängigen Gewichtungsvektoren $\boldsymbol{\lambda}_1^T := (1,1), \boldsymbol{\lambda}_2^T := (2,1)$ diesen Bedingungen. Definiert man mittels der „Rotationsmatrix“[48] Λ neue Dimensionen

$$\boldsymbol{\psi} := \Lambda\boldsymbol{\theta}, \qquad \Lambda = \begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix},$$

so stellt offenbar $\boldsymbol{\lambda}_i^T f(\boldsymbol{u})$ den KQ-Schätzer für ψ_i dar. Zudem spiegeln beide Komponenten des Schätzers $\hat{\boldsymbol{\psi}} := \Lambda \boldsymbol{f(u)}$ die Ordnung der Responsevektoren wider, d.h. eine bessere Testperformance geht mit höheren (oder gleichen) Schätzwerten auf beiden Komponenten einher. Es scheint somit plausibel, diesen mehrdimensionalen Schätzer als fair zu bezeichnen (und in einer rein manifesten Sichtweise - wie in Abschnitt 5.1 definiert - ist er es auch!). Dieser Plausibilitätsüberlegung steht jedoch folgendes Argument gegenüber: Gemäß Satz 3.1.3 ist der KQ-Schätzer innerhalb der Klasse der regulär kompensatorischen Modelle dann und nur dann fair, wenn die korrespondierende Ladungsmatrix Einfachstruktur aufweist. Die Ausgangsmatrix A lässt sich

[48] Das Wort „Rotation“ wird hier in sehr allgemeiner Form verwendet. Die Matrix Λ bezeichnet keine Rotation im eigentlichen Sinn - ist aber invertierbar und kann somit als Transformationsmatrix fungieren.

jedoch offenbar nicht auf Einfachstruktur bringen. Folglich muss die Annahme, dass es sich um ein regulär kompensatorisches Modell (in der $\boldsymbol{\psi}$-Metrik!) handelt, falsch sein.
In der Tat, wenn B die Ladungsmatrix der Items bezüglich der neu definierten $\boldsymbol{\psi}$-Dimensionen bezeichnet, dann hat B die nicht-kompensatorische Gestalt

$$B = A\Lambda^{-1} = \begin{pmatrix} -1 & 1 \\ 1 & 0 \\ 2 & -1 \end{pmatrix},$$

in der beispielsweise ψ_2 einen negativen Effekt auf die Beantwortung des dritten Items aufweist. Trotz dieses negativen Effekts steigt der Schätzwert für ψ_2 bei Erzielung eines höheren Itemscores auf diesem Item.
Dieser kuriose Effekt, der zugleich den tieferen Grund beinhaltet, warum das in Kapitel 5.1 gegebene Fairnesskriterium nicht mit Bezug auf die latente Ebene (sprich: Ladungsmatrix) definiert wurde, mahnt zur Vorsicht hinsichtlich einer allzu simplifizierenden Interpretation einer Ladungsmatrix. Die Information der Beziehung zwischen den latenten Dimensionen und dem Antwortmuster steckt in der *vollständigen* Ladungsmatrix. Das Verhalten der entsprechenden Schätzwerte ist eine wesentlich kompliziertere Funktion der Itemscores als es „übliche“, heuristische Interpretationen der Ladungsmatrix erahnen lassen. Insbesondere die von Testkonstrukteuren häufig beanspruchte Heuristik, zur Benennung eines Faktors lediglich „hochladende“ Items zu verwenden, und somit die in der Ladungsmatrix vorhandenen Interdependenzen zu ignorieren, kann hierunter subsumiert werden. Ein Blick auf die in der gesamten Matrix vorliegende Information kann jedoch mitunter die nach dieser Heuristik erstellte Benennung stark in Frage stellen, wie im Rahmen der Diskussion des paradoxen Effekts anhand des SPS-J (Kapitel 3.1) bereits dargelegt wurde.
Das soeben beschriebene Phänomen mag zunächst als ein konstruierter Effekt anmuten, dessen Existenz maßgeblich auf den gewählten Itemparameterwerten fußt. In starkem Gegensatz zu dieser Vermutung steht das Resultat aus Satz 4.4.2, welches für eine Teilklasse linearer Schätzer - worunter insbesondere der hier verwendete KQ-Schätzer fällt - die Existenz p linear unabhängiger Gewichtungsvektoren garantiert und somit obige Vorgehensweise stets ermöglicht. Mit anderen Worten: Ausgehend von einer positiven Ladungsmatrix und einem linearen (bzw. affinen) Schätzer, der die Bedingung des Satzes 4.4.2 erfüllt, lassen sich eine Basis bestehend aus nichtne-

gativen Gewichtungsvektoren sowie damit einhergehende, umdefinierte „Konstrukte“ $\psi_1, \ldots \psi_p$ angeben, so dass eine Erhöhung eines Itemscores stets eine Erhöhung (genauer: keine Minderung) der Schätzwerte aller Dimensionen bewirkt. Eine Betrachtung des faktorenanalytischen Modells aus dem Blickwinkel der neu definierten ψ-Dimensionen würde jedoch aufzeigen, dass mindestens ein Item existiert, welches bezüglich mindestens einer ψ-Dimension eine negative Ladung aufweist und somit die Erwartung hervorruft, dass der Schätzwert der zugehörigen ψ-Dimension fallen sollte, wenn der Itemscore des betreffenden Items erhöht wird.

Das Konzept der Eindimensionalität

Wenngleich das Uneindeutigkeitsproblem bislang im Rahmen des faktorenanalytischen Modells behandelt wurde, besitzt es ebenso für IRT-Modelle Relevanz. Dies kann zum Einen aufgrund der Tatsache entnommen werden, dass sich viele IRT-Modelle aus einer kategorisierten, auf Schwellenwerten basierenden Variante des faktorenanalytischen Modell herleiten lassen (Lee, 2007, Kap. 6, 7). Zum Anderen kann obige Argumentation in weiten Teilen auch direkt für ein IRT-Modell, dessen Loglikelihood in der Form $\sum_i l_i(\boldsymbol{a_i^T \theta})$ darstellbar ist, übernommen werden.
Ein potentieller Ausweg aus dieser Problematik bildet - ähnlich wie im Falle des paradoxen Effekts - das Konzept der Eindimensionalität. Ist θ im obigen faktorenanalytischem Modell eindimensional, so beschränkt sich das Uneindeutigkeitsproblem auf eine „schlichte“ Umskalierung $\theta^* = c \cdot \theta$. Letzteres stellt sowohl bei der Interpretation als auch bezüglich des Invarianzprinzips bei Formung der Skala kein nennenswertes Problem dar; muss jedoch für eindimensionale IRT-Modelle, die im Gegensatz zum faktorenanalytischem Modells keine *lineare* Wirkung der latenten Variable postulieren, relativiert werden. In der gängigsten Klasse eindimensionaler, binärer IRT-Modelle - der Klasse der „Monotone-Homogeneity“-Modelle (MHM, Sijtsma und Molenaar, 2002, Kap. 3, 4) - wird die lineare Struktur des faktorenanalytischen Modells durch die Forderung nach *monoton wachsenden* Item-Response-Funktionen ersetzt. Wenn $f_i(\theta)$ die (monoton wachsende) Item-Response-Funktion des i-ten Items kennzeichnet, und wenn ferner die Verteilungsklasse der latenten Variable „hinreichend groß“ ausfällt, dann ist die Metrik der latenten Variable lediglich bis auf streng monoton wachsende Transformationen wohldefiniert, d.h. für beliebiges, streng mo-

noton wachsendes g erlaubt die Wahl von $\theta^* := g(\theta)$ sowie $f_i^*(\theta^*) := f_i(g^{-1}(\theta^*))$ die Konstruktion eines empirisch ununterscheidbaren Modells.

Gleichwohl demnach auch im Kontext eindimensionaler Modelle Identifizierbarkeitsprobleme bestehen, weisen diese dennoch - verglichen mit ihrer Erscheinungsform im Rahmen mehrdimensionaler Modelle - wesentlich mildere Formen[49] auf (in obiger MHM-Klasse bleibt z.B. die Ordnung der Personen bezüglich der Fähigkeit stets bestehen - unabhängig davon, welche streng monotone Transformation gewählt wurde). Da eindimensionale Modelle zudem i.d.R. eine einfache Interpretation (der Wirkung) des latenten Konstrukts ermöglichen, scheint die Konstruktion eindimensionaler Skalen insbesondere im Hinblick auf die Vermeidung paradoxer Klassifikationen ein wünschenswertes Ziel darzustellen. Anhand zahlreicher Quellen, die sich auf diverse Varianten eindimensionaler Modellklassen beziehen, kann jedoch die Erreichbarkeit eines (approximativen) eindimensionalen Modells bezweifelt werden. Für die Klasse der MHM-Modelle zeigt Rosenbaum (1984) z.B. die Eigenschaft der bedingten Assoziiertheit[50] auf, d.h. für (komponentenweise) monoton wachsende Funktionen g_1, g_2, beliebiges h und für eine beliebige Partition des Antwortvektors $\boldsymbol{u} = (\boldsymbol{u_1}, \boldsymbol{u_2})$ gilt $\mathrm{Cov}(g_1(\boldsymbol{u_1}), g_2(\boldsymbol{u_1})|h(\boldsymbol{u_2})) \geq 0$. Das faktorenanalytische, auf Partialkorrelationen basierende Pendant zu dieser Aussage besagt, dass sämtliche Partialkorrelationen nichtnegativ ausfallen und zudem einer restriktiven Tetraeden-Bedingung genügen (Salgueiro u.a., 2008). Beide Aussagen stellen heraus, dass eindimensionale Modelle der „üblichen“ Struktur[51] äußerst restriktive Assoziationsmuster zwischen den Items implizieren, die wiederum ihre Angemessenheit zur Modellierung realer Daten in Frage stellen.

Weitere Resultate, die die Praxisferne eindimensionaler Modelle (indirekt) herausstellen, wurden von Shapiro (1982) sowie von Ten Berge und Socan (2004) erbracht. Während Shapiro aufzeigt, dass die Menge aller Matrizen, die durch Subtraktion einer positiven Diagonalmatrix in eine Matrix mit Rang Eins überführbar sind, eine (Lebesguesche)-Nullmenge bilden und somit die Menge aller Populationskovarianzmatrizen, die der Eindimensionalitätshypothese des faktorenanalytischen Mo-

[49] Schließt man in die Klasse eindimensionaler IRT-Modelle lokal *abhängige*, eindimensionale Modelle mitein, so löst sich diese vermeintliche Abgrenzung jedoch auf, da jedes mehrdimensionale Modell in ein eindimensionales, lokal abhängiges Modell überführt werden kann.

[50] Für eine Definition des Assoziationskonzepts sei auf Esary, Proschan und Walkup (1967) verwiesen. Eine IRT-bezogene Darstellung findet sich ferner bei Holland und Rosenbaum (1986).

[51] Gemeint sind an dieser Stelle eindimensionale Modelle, die lokale Unabhängigkeit sowie einen festgelegten Verlauf der bedingten Antwortwahrscheinlichkeiten postulieren.

dells genügen, „verschwindend klein“ ausfällt, weisen Ten-Berge und Socan auf den Antagonismus zwischen Eindimensionalität und Reliabilität hin, der sich dahingehend manifestiert, dass gängige Maßnahmen zur Reliabilitätserhöhung, wie z.B. Testverlängerung, im Regelfall die Abweichung des Tests von einer Eindimensionalitätsstruktur vergrößern.
Der naheliegende Gedanke, durch Verzicht auf die Spezifikation einer Verlaufsform der Item-Response-Funktion (linear/monoton) zu einer weniger restriktiven Variante eines eindimensionalen IRT-Modells zu gelangen, führt zu einem bemerkenswerten Effekt: Wie Suppes und Zanotti (1981) beweisen, impliziert der Verzicht der Angabe einer spezifischen Verlaufsform für $f_i(\theta)$ - und damit die alleinstehende Annahme eines eindimensionalen, lokal unabhängigen IRT-Modells - ein stets erfüllbares und somit empirisch wertloses Modell. Mit anderen Worten: *Jede* Wahrscheinlichkeitsverteilung für Antwortmuster $\boldsymbol{u}$ ist kompatibel mit einem eindimensionalen, lokal unabhängigen IRT-Modell. Ein Resultat, welches zum Einen die Relevanz von zusätzlichem Vorwissen über die Wirkungsweise der latenten Variable hervorhebt, zum Anderen aber auch auf die (potentielle) Bedeutungslosigkeit des Begriffs der Eindimensionalität verweist.
Es kann somit festgehalten werden, dass die Klasse der IRT-Modelle, die sowohl bezüglich der Thematik paradoxer Klassifikationen als auch bezüglich des Rotationsproblems zu präferieren ist, entweder aufgrund ihrer tautologischen Struktur von keinem empirischen Wert ist, oder, wenn sie von (üblichen) zusätzlichen Modellannahmen begleitet wird, zu restriktiv ausfällt.

Interpretationsspielräume durch mangelnde Identifizierbarkeit

Neben den bisher skizzierten Problemen ein- und mehrdimensionaler IRT-Modelle sieht sich die moderne psychologische Testtheorie jedoch noch einer weiteren, für individualdiagnostische Zwecke fundamentalen Uneindeutigkeitsproblematik, die die Interpretation der latenten Variable tangiert und die im Folgenden o.B.d.A am Beispiel eines eindimensionalen IRT-Modells erläutert werden soll, gegenüber. Im Kern der Thematik steht die Frage nach der inhaltlichen Bedeutung der latenten Variable θ in einem Item-Response-Modell. Die im Bereich der Individualdiagnostik omnipräsente Interpretation der latenten Größe als Fähigkeit, bzw. allgemeiner als

Eigenschaft der zu diagnostizierenden Person und die damit einhergehende Interpretation der Item-Response-Funktion als (bedingte) Wahrscheinlichkeit, das Item zu lösen, scheinen unumstritten. Wie jedoch nach Wissen des Autors erstmals von Holland (1990) dargelegt, stellt dies jedoch nicht die einzig mögliche Interpretation dar. In einer von der üblichen Interpretationsweise (=*stochastic subject view*) streng zu trennenden Sichtweise bezeichnet die latente Variable keine Eigenschaft/Fähigkeit der Person, sondern fungiert vielmehr als „Label“ für eine Schicht, in die die Person implizit durch die Annahmen des IRT-Modells eingeordnet wird. Genauer erörtert beginnt diese Sichtweise mit der zur ursprünglichen Variante diametralen Annahme, dass eine Person durch ein festes, d.h. *nichtzufälliges* Antwortmuster gekennzeichnet ist. Der einzige Zufallsmechanismus entsteht folglich durch die zufällige Auswahl von Personen aus einer Grundgesamtheit (*random sampling view*). Letztere ist gemäß dem postulierten IRT-Modell in Schichten unterteilt, die mit dem Symbol „θ“ identifiziert werden. In Schichten, die durch einen höheren Wert der latenten Größe θ gekennzeichnet sind, findet sich (jedenfalls bei Unterstellung eines MHM-Modells) ein höherer Anteil $f_i(\theta)$ von Personen, die das i-te Item lösen. Ferner besagt die Annahme lokaler Unabhängigkeit, dass bei Kreuzklassifikation der Antwortmuster aller Personen, die derselben Schicht entstammen, ein Unabhängigkeitsmuster resultiert. Diese Interpretationsweise ist semantisch verschieden von der „gängigen“ Sichtweise und dennoch empirisch i.d.R. ununterscheidbar von dieser. Es ist in diesem Kontext nicht verwunderlich, dass diese zweite, empirisch äquivalente Interpretationsweise nahezu in Vergessenheit geraten ist, da sie letztendlich dem Diagnostiker keine nach außen hin leicht zu vertretende Rechtfertigung der Nutzung von IRT-Modellen an die Hand gibt (man kontrastiere hierzu beispielsweise die Interpretation von θ als inhärente Fähigkeit der Person mit der z.T. fragwürdigen Interpretation als „Schichtenlabel“ und stelle sich die Frage, welche individualdiagnostische Relevanz mit der Zuordnung der Person zu einer Schicht, deren „Entstehung“ letztendlich in starker Abhängigkeit zu den postulierten IRT-Modellannahmen steht, verbunden ist.) Da diese zweite Variante einer typisch diagnostischen Interpretationsweise teilweise zuwiderläuft, überrascht es auch wenig, dass der paradoxe Effekt bezüglich dieser Sichtweise sich wesentlich harmloser manifestiert: Durch Umwandeln einer falschen zu einer richtigen Antwort wird die Person „lediglich“ einer anderen Schicht zugeordnet. Wenngleich auch diese Art des paradoxen Effekts je nach Durchführung des diagnostischen Prozesses nicht vollkommen unproblematisch sein muss, so fehlt in

ihr jedoch der direkte Bezug zum Individuum.
Neben dieser die inhaltliche Bedeutung der latenten Variable betreffenden Frage ergeben sich auch auf weiteren Ebenen Identifizierbarkeitsprobleme, die den Forscher implizit dazu zwingen, subjektive Entscheidungen zu treffen, welche nicht in den Daten begründbar, jedoch (trotzdem) von Tragweite, was die Konstruktion sowie Verwendung eines IRT-Modells zu diagnostischen Zwecken tangiert, seien können. Um nur einige Stufen dieses subjektiven Entscheidungsprozess hervorzuheben, sei auf

1) die Wahl der der Auswertung zugrundegelegten IRT-Modellklasse;

2) die Entscheidung, welche latenten Variablen als Primärfaktoren betrachtet werden (und wie viele Faktoren „extrahiert“ werden.);

3) die implizite Entscheidung, die latente Variable nach der „stochastic subject view“ zu interpretieren;

4) den bereits in der Diskussion des Rotationsproblems erwähnten Interpretationsspielraum bei der Benennung der latenten Variablen;

5) die Entscheidung, die latente Größe kausal zu interpretieren, d.h. als verursachend für das gezeigte Antwortmuster anzusehen,

6) die Entscheidung, intraindividuelle Prozesse mit interindividuellen Prozessen „zu vermengen“,

hingewiesen.
Punkt 3) sowie 4) wurden bereits erörtert. Bezüglich 5) ist festzuhalten, dass selbst wenn die latente Variable eines IRT-Modell eine reale Manifestation besitzt[52], diese nicht notwendigerweise in einem kausalen Sinne das Antwortmuster bestimmen muss, da für die Feststellung eines kausalen Zusammenhangs eine (aktive) Manipulation der Ursache (hier: latente Variable) erforderlich wäre.
Punkt 6) verweist auf den Umstand, dass interindividuell gewonnene Daten i.A. keine Schlüsse auf intraindividuelle Prozesse erlauben. Da die der Skalenkonstruktion zugrundeliegende Auswertung auf interindividuellen Daten basiert und da ferner

[52]Ein Punkt, der häufig in der suggestiven Terminologie des Latenten-Variablen-Ansatzes gerne untergeht: Nicht jede latente Variable muss ein „reales Pendant“ besitzen.

keine Person mehrmals unter „identischen“ Umständen (im Sinne einer genuinen Messwiederholung) mit demselben Item konfrontiert wird, kann das im Zuge der Skalenkonstruktion gewonnene Modell gegebenenfalls untauglich für die Beschreibung des für die Individualdiagnostik relevanten, intraindividuellen Prozesses sein.
Es verbleibt somit darzulegen, in welcher Form sich Identifizierbarkeitsprobleme in den unter 1) und 2) aufgeführten Punkten manifestieren.
Bezüglich 1) lässt sich zunächst die Frage aufwerfen, wie diverse Ergebnisse der Skalenkonstruktion (Inferenz der „Fähigkeiten“, Anzahl extrahierter latenter Variablen, Benennung der Variablen, individualdiagnostische Entscheidungen/Klassifikationen) bei Unterstellung einer *anderen* Modellklasse ausgesehen hätten (Frage der Robustheit). Da die konventionell verwendeten Modellklassen (faktorenanalytisches Modell; M2PL-Modell; Partial-Credit-Modell, etc.) - mit Ausnahme des Rasch-Modells - keine gesonderte, inhaltliche oder messtheoretische Motivation besitzen, lässt sich die unter 1) gewählte Modellklasse prinzipiell durch eine andere Modellklasse austauschen - wobei es an dieser Stelle zu einer „Glaubensfrage“ wird, ob die ursprünglichen Inferenzergebnisse unter dieser Modellklasse Bestand hätten.
Auf den ersten Blick mag man gegenüber diesem „Modellsubstitutionsargument“ einwenden, dass statistische Gütekriterien der Modellanpassung dazu dienen können, die unter 1) in Frage kommende Klasse stark einzuschränken. Modellanpassungskriterien operieren jedoch zwangsläufig auf Ebene der *manifesten* Daten. Modellanpassungskriterien können demnach niemals zwei IRT-Modelle trennen, die gleiche (marginale) Antwortmusterverteilungen implizieren - gleichwohl beide Modelle gegebenenfalls zu verschiedenen Deutungen der latenten Variable führen können. Formal gesehen, lässt sich dies wie folgt ausdrücken: Schätzbar - und somit den Modellanpassungstests zugänglich - ist lediglich die gemeinsame Verteilung der beobachtbaren Variablen $P(U_1 = u_1, \dots U_k = u_k)$. Es gibt jedoch beliebig viele Möglichkeiten, einen latenten Vektor $\boldsymbol{\theta}$ mit zugehöriger Verteilungsfunktion G einzuführen (der offensichtlich nicht einmal eine reale Entsprechung besitzen muss), so dass $P(U_1 = u_1, \dots U_k = u_k) = \int P(U_1 = u_1, \dots U_k = u_k | \boldsymbol{\theta}) \mathrm{d}G(\boldsymbol{\theta})$ gilt[53]. Jede Art der Spezifikation „erzählt eine andere Geschichte“, wie die Entstehung der Testdaten zustandegekommen ist. Die Tradition, für $P(U_1 = u_1, \dots U_k = u_k | \boldsymbol{\theta})$ lokale stochastische Unabhängigkeit sowie eine linear-logistische Parametrisierung anzunehmen, gleicht dem Vorgehen einen Punkt auf einer Zahlengerade zu wählen, und diesen

[53]Das hier auftretende Integral ist als Riemann-Stieltjes-Integral zu lesen.

dann für den „Richtigen“ zu erklären.
Ein „Vorgeschmack“, welche realen Differenzen sich aus zwei empirisch ununterscheidbaren Modellen ergeben können, wurde bereits in Kapitel 4.3 gegeben. Hier war es im Rahmen eines mehrdimensionalen faktorenanalytischen Modells, welches paradoxe Klassifikationen induzierte, möglich, ein empirisch ununterscheidbares Modell - siehe (4.47) - anzugeben, das zu fairen Klassifikationen führt. In einem Modell der Form

$$P(U_1 = u_1, \ldots U_k = u_k) = \int P(U_1 = u_1, \ldots U_k = u_k | \, \boldsymbol{\theta}) \mathrm{d}G(\boldsymbol{\theta})$$

wurde $(\theta_2, \ldots, \theta_p)$ herausintegriert, und so ein (empirisch gleichwertiges) Modell der Struktur

$$P(U_1 = u_1, \ldots U_k = u_k) = \int P(U_1 = u_1, \ldots U_k = u_k | \, \theta_1) \mathrm{d}G_1(\theta_1)$$

„geschaffen“. Während der Schätzer für θ_1 im ersten Modell die Erzielung eines höheren Itemscores z.T bestraft, wird im zweiten Modell eine augenscheinlich bessere Performance auch stets belohnt - womit unmittelbar gezeigt wäre, dass empirisch ununterscheidbare Modelle unterschiedliche praktische Implikationen aufweisen können. Letzteres liefert (als Nebenprodukt) zugleich ein Beispiel für die unter Punkt 2) aufgeführte Identifizierbarkeitsproblematik und beschließt somit den Einschub zur Darlegung allgemeiner Kritikansätze des Latenten-Variablen-Paradigmas.

Schlussfolgerungen

Welche unmittelbaren Schlussfolgerungen lassen sich (im Hinblick auf die Beurteilung des paradoxen Effekts) aus diesem Exkurs über generelle Kritikansätze des Latenten-Variablen-Paradigmas gewinnen?
Die vorangegangene Diskussion sollte verdeutlichen, dass der paradoxe Effekt vor dem Hintergrund einer tiefgehenden Unteridentifizierbarkeit von Latenten-Variablen-Modellen zu betrachten ist, die es zudem erschwert, eine kohärente Interpretation des paradoxen Effekts zu geben. Eine schlüssige Interpretation muss neben der Tatsache, dass die scheinbar strikte Trennung zwischen ein- und mehrdimensionalen Modellen bei genauerem Hinsehen nicht aufrechterhalten werden kann, auch das im ersten Abschnitt dieses Kapitels skizzierte Rotationsproblem bei der Verwendung

des Fairnessbegriffs berücksichtigen. Aus Letzterem rechtfertigt sich eine Definition des Fairnessbegriffs, die an den realen Konsequenzen für das Individuum orientiert ist und die jedenfalls in keiner *direkten* Beziehung zur Gestalt der Ladungsmatrix steht. Ein Schätzer ist nach diesem Kriterium fair, wenn er eine (augenscheinlich) bessere Performance belohnt, d.h. wenn die durch eine Erhöhung des Itemscores einhergehende Veränderung im Schätzwert keine negativen Auswirkungen für das Individuum hat. Die Verwendung der Ladungsmatrix zur Definition des Fairnessbegriffs stößt hingegen auf Inkohärenzen: Angenommen die Ladungsmatrix einer Skala, die Mathematikkenntnisse sowie räumliches Vorstellungsvermögen misst, ist positiv. Wird (basierend auf der Nichtnegativität der Ladungen) in diesem Kontext ein Schätzer als fair definiert, wenn er eine bessere Performance auf jedem Item belohnt, so ergibt sich - wie in der Diskussion des Rotationsproblems bereits geschildert wurde - folgendes Problem: Da beide Fähigkeiten positiv zur Beantwortung der Items beitragen, scheint es „legitim“, einen Schätzer für eine gewichtete Summe (positive Gewichte) der beiden Dimensionen als fair zu bezeichnen, wenn er höhere Itemscores in höhere Schätzwerte überführt. Wie das Beispiel des ersten Abschnitts jedoch demonstriert, kann die Ladungsmatrix bezüglich zweier derart gewonnener „Konstrukte“ (jedes Konstrukt ist eine positive Linearkombination der beiden Dimensionen) negative Einträge beinhalten. In dieser Vorgehensweise hängt demnach die Fairness des Schätzers einer Dimension von der Definition der übrigen Dimensionen ab - was offenbar nicht zielführend ist.

Da die Verwendung eindimensionaler Modelle sowohl das geschilderte Rotationsproblem umgeht, als auch in gewisser Form ein „Garant“ für die Vermeidung paradoxer Klassifikationen ist, schien ein näheres Eingehen auf das Konzept der Eindimensionalität angebracht. Hieraus sollte ersichtlich werden, dass Eindimensionalität in ihrer (formalen) Reinform einen empirisch wertlosen Begriff darstellt, der erst durch zusätzliche Modellannahmen (z.B. monoton wachsende Item-Response-Funktionen) an Bedeutung gewinnt. Da zudem unter Verzicht auf lokale stochastische Unabhängigkeit jedes mehrdimensionale Modell in ein eindimensionales Modell überführt werden kann, lässt sich darüber hinaus in der Entscheidung, ob eine Skala als ein- oder mehrdimensional anzusehen ist, eine gewisse Willkür feststellen. Wann immer daher in dieser Arbeit die Verwendung eindimensionaler Modelle als „Ausweg“ nahegelegt wurde, so wurde implizit von einem eindimensionalen, lokal unabhängigen Modell mit Monotonieeigenschaft (d.h. streng monoton verlaufende

Item-Response-Funktionen) ausgegangen. Es gibt jedoch viele eindimensionale, lokal abhängige Modelle, die ebenfalls dem Fairnesskriterium genügen. Da jedes mehrdimensionale Modell in ein eindimensionales, lokal abhängiges Modell überführbar ist, lässt sich somit festhalten, dass die wesentliche Unterscheidung nicht per se die zwischen ein- und mehrdimensionalen Modellen ist, sondern die, ob Inferenz simultan (Maximierung von $P(\boldsymbol{u}|\boldsymbol{\theta})$) bzw. für jede Dimension einzeln (Sukzessive Maximierung von $P(\boldsymbol{u}|\theta_i)$ für $i = 1, \ldots p$) vollzogen wird.

Der gemeinsame Nenner hinter diesen Ausführungen zum paradoxen Effekt ist das fundamentale Identifizierbarkeitsproblem, das dem Latenten-Variablen-Paradigma inhärent ist und welches mitunter den Eindruck des „anything goes“ aufkommen lässt. Es wäre daher aus Sicht des Autors auch abträglich gewesen, diese abschließende Diskussion auf Behebungsmöglichkeiten des paradoxen Effekts auszurichten. Vielmehr sollte der paradoxe Effekt als *ein* „Symptom“ einer tiefergehenden Uneindeutigkeitsproblematik, deren diverse Unterformen (Rotationsproblem; Ambiguität in der Interpretation der latenten Variable; Vagheit in der Definition des Dimensionalitätskonzepts, etc.) zentraler Gegenstand obiger Ausführungen waren, begriffen werden.

A Mathematische Grundlagen

Dieser Anhang rekapituliert mathematische Konzepte, die zur Untersuchung des paradoxen Effekts relevant sind. Er dient dabei zum Einen dem Zweck dem Leser diese Begriffe in geschlossener und übersichtlicher Form darzubieten. Zum Anderen enthält er die notwendigen formalen Aspekte, die erforderlich sind, um gewisse, in den Kapiteln vorgenommene Vereinfachungen - wie z.B. die Existenz der Funktion des bedingten Maximums - durch rigorose Resultate zu ersetzen und weitere Verallgemeinerungen der Resultate zum paradoxen Effekt vornehmen zu können.
Der Konvexitätsabschnitt ist notationell an Rockafellar (1970) angelehnt, während die darauffolgenden Abschnitte auf den Darstellungen in Dontchev und Rockafellar (2009) bzw. Rockafellar und Wets (2009) beruhen. Mit Ausnahme der spezifischen Erörterungen im Rahmen von IRT-Modellen handelt es sich ferner stets um Rekapitulierungen von bereits bekannten Resultaten, deren Beweise den obigen Quellen entnommen werden können.

Konvexe Funktionen

Der Begriff der Konvexität einer Funktion beruht maßgeblich auf dem Begriff einer konvexen Menge.
Eine Menge $C \in \mathbb{R}^n$ heisst konvex, wenn für jedes $x \in C$, jedes $y \in C$ und jedes $0 < \lambda < 1$ auch

$$\lambda x + (1 - \lambda)y \in C$$

gilt.
Anhand dieses Konvexitätsbegriff für Mengen kann die Konvexität einer (potentiell erweitert-reellwertigen) Funktion $f : \mathbb{R}^n \mapsto \overline{\mathbb{R}}$ definiert werden.
Eine erweitert-reellwertige Funktion f heisst konvex, wenn ihr Epigraph $epi f$ konvex ist. Der Epigraph ist dabei durch

$$\text{epi} f = \{(x, \mu) \in R^{n+1} | \, f(x) \leq \mu\}$$

definiert.
Eine erweitert-reellwertige Funktion f heisst konkav, wenn $-f$ konvex ist.
Unter der Konvention $\infty - \infty = \infty$ (Inf-Addition) lässt sich ferner leicht zeigen, dass

eine Funktion genau dann konvex ist, wenn für jede Wahl von $0 < \lambda < 1$, jedes x und jedes y

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) \tag{1.51}$$

gilt. Gilt darüber hinaus für alle Paare $x \neq y$ aus einer konvexen Menge C strikte Ungleichheit in (1.51), so nennt man f strikt konvex in C.

Zwischenbemerkung: Eine reellwertige, auf einer konvexen Menge C definierte Funktion f, die (1.51) auf C erfüllt, kann stets durch die Definition

$$\tilde{f}(x) = \begin{cases} f(x) & x \in C \\ \infty & x \notin C \end{cases}$$

in eine (äquivalente) konvexe Funktion gemäß obiger Definition überführt werden.

Die Relevanz konvexer Funktionen für Modellierungszwecke fußt maßgeblich auf der Tatsache, dass die Menge aller Minima einer konvexen Funktion konvex und im Falle einer strikt konvexen Funktion höchstens einelementig ist. Für strikt konvexe Funktionen impliziert somit die Existenz eines Minimums gleichzeitig dessen Eindeutigkeit. Ferner ist im Falle einer konvexen Funktion die Unterscheidung zwischen lokalem und globalem Minimum hinfällig, da jedes lokale Minimum bereits ein globales Minimum darstellt. Darüber hinaus bieten differenzierbare konvexe Funktionen den Vorteil, den Minimierungsprozess äquivalent als Lösen eines Gleichungssystems durchführen zu können: x ist Stelle eines Minimums von f (f konvex) relativ zu einer offenen, konvexen Teilmenge in der f differenzierbar ist, genau dann, wenn $\nabla f(x) = \mathbf{0}$ ist.
Für die Untersuchung des paradoxen Effekts im Rahmen linear-kompensatorischer Modelle sind ferner die folgenden Ergebnisse von zentraler Bedeutung (siehe Kapitel 2 aus Rockafellar und Wets (2009)):

a) Wenn A eine Lineartransformation $A : \mathbb{R}^m \mapsto \mathbb{R}^n$ darstellt, und wenn f eine konvexe Funktion $f : \mathbb{R}^n \mapsto \overline{\mathbb{R}}$ ist, dann ist die Komposition $g := f \circ A$ konvex.

b) $\sum_{i=1}^{k} f_i$ ist konvex, insofern $f_1, \ldots f_k$ konvexe Funktionen sind, deren Wertebereiche nicht den Wert $-\infty$ umfassen.

c) Ist f eine reellwertige, auf einer offenen konvexen Menge O definierte, zweimal differenzierbare Funktion, so ist f genau dann konvex in O, wenn $\nabla^2 f(x)$ positiv semidefinit für alle x in O ist.

d) Ist f eine reellwertige, auf einer offenen konvexen Menge O definierte, zweimal differenzierbare Funktion, so ist f strikt konvex in O, wenn $\nabla^2 f(x)$ positiv definit für alle x in O ist.

Ein speziell im Rahmen des linear-kompensatorischen Modells (Kapitel 3.1) relevantes Resultat (Gleichung 3.13) betrifft die Funktion

$$l_u(\boldsymbol{\theta}) = \sum_{i=1}^{k+1} l_i(\boldsymbol{a_i^T}\boldsymbol{\theta}).$$

Gemäß a) und b) ist diese Funktion konkav, wenn jedes l_i konkav ist. Es gilt jedoch nur unter Zusatzbedingungen an die Ladungsvektoren, dass die *strikte* Konkavität der l_i-Funktionen gleichzeitig die strikte Konkavität von l_u impliziert:

Lemma A.0.1. *Ist im Rahmen der Modellgleichung (3.13) jedes l_i strikt konkav auf ganz $\mathbb{R}$ und besitzt die Ladungsmatrix A vollen Spaltenrang ($rg(A) = p$), dann ist l_u strikt konkav.*

Beweis. Sei $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$ und $0 < \lambda < 1$. Dann gilt

$$l_u(\lambda\boldsymbol{\theta}_1 + (1-\lambda)\boldsymbol{\theta}_2) = \sum_i l_i(\lambda \boldsymbol{a}_i^T\boldsymbol{\theta}_1 + (1-\lambda)\boldsymbol{a}_i^T\boldsymbol{\theta}_2)$$

Wenn ein Summenindex j existiert, für den $\boldsymbol{a}_j^T\boldsymbol{\theta}_1 \neq \boldsymbol{a}_j^T\boldsymbol{\theta}_2$ ist, dann ergibt die strikte Konkavität von l_j die Behauptung.
Sollte ein derartiger Index nicht existieren, so gilt $\boldsymbol{a}_i^T\boldsymbol{\theta}_1 = \boldsymbol{a}_i^T\boldsymbol{\theta}_2$ für alle i. Demnach ist $\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2$ orthogonal zu jedem Ladungsvektor. Letzteres steht jedoch im Widerspruch zur Annahme $R(A) = p$. □

Die in Kapitel 3.1 omnipräsente Annahme einer Ladungsmatrix vollen Spaltenrangs impliziert somit in Kombination mit $\frac{\partial^2 l_i}{\partial x^2} < 0$ unter Benutzung des Resultats in d) die strikte Konkavität von l_u. Folglich ist der (potentiell nicht existierende) MLE in einem (solchen) linear-kompensatorischen Modell stets eindeutig. Eine analoge Argumentation bezogen auf Komponentenfunktionen der Form $l^*(\theta_1, \ldots \theta_{p-1}) :=$

$l_u(\theta_1, \ldots \theta_{p-1}, \theta_p)$ führt ferner auf die Eindeutigkeit der bedingten Maximumsfunktion bzw. des bedingten ML-Schätzers, dessen Existenz Gegenstand des nächsten Abschnitts ist.

Das Maximum einer Funktion

Während der Begriff der strikten Konkavität primär zur Sicherung der Eindeutigkeit des ML-Schätzers sowie zur Angabe einer vereinfachenden Rechenvorschrift (Lösung eines Gleichungssystems) diente, erfüllt dieser Abschnitt vorrangig den Zweck, Aussagen über die Existenz des Schätzers zu ermöglichen. Des Weiteren bildet er das notwendige Fundament zur Analyse der parametrischen Abhängigkeit des paradoxen Effekts (siehe Kapitel 2.3).
Ein klassischer Satz der Analysis (Rudin, 1976, S.89) besagt, dass eine auf einer kompakten Menge definierte, stetige Funktion stets ihre Extrema annimmt. Dieser Satz ist jedoch für viele Zwecke zu restriktiv. Grundlage eines breiter einsetzbaren Existenzprinzips bildet der Begriff der „Niveaubeschränktheit“ (*level boundedness*), der im Folgenden mit *lb* abgekürzt werden wird.
Ausgehend von einer auf $\mathbb{R}^p$ definierten Funktion[54] $f : \mathbb{R}^p \mapsto \mathbb{R} \cup \{-\infty\}$ sei $\text{lev}_{\geq\alpha} f := \{x |\, f(x) \geq \alpha\}$. Wenn für jedes $\alpha \in \mathbb{R}$ $\text{lev}_{\geq\alpha} f$ beschränkt ist, dann ist f *lb*.
Ist $\text{lev}_{\geq\alpha} f$ abgeschlossen für alle $\alpha \in \mathbb{R}$, dann ist f oberhalbstetig (*upper semicontinuous*) (im Folgenden mit *usc* abgekürzt). Letzteres kann gleichwertig auch mit $f(x) = \limsup_{z \to x} f(z)$ (wobei die konstante Folge $z_n := x$ zulässig ist) ausgedrückt werden.
Basierend auf diesen Begriffen lässt sich das Hauptresultat zur Maximierung wie folgt paraphrasieren (Rockafellar und Wets, 2009, Theorem 1.9):

Jede usc Funktion, die zugleich lb ist, besitzt (mindestens) ein Maximum.

Da im IRT-Kontext viele Modellgleichungen aufgrund der Omnipräsenz der Annahme lokaler stochastischer Unabhängigkeit additiv zusammengesetzt sind, lohnt sich eine kurze Betrachtung des *lb*- bzw. *usc*-Konzepts unter dieser Rechenoperation:

[54] Eine bezüglich einer Teilmenge $C \subset \mathbb{R}^p$ zu maximierende Funktion kann leicht in eine auf ganz $\mathbb{R}^p$ definierte Funktion überführt werden. Man setze lediglich $f = -\infty$ außerhalb von C (siehe hierzu auch das abstrakte Minimierungsprinzip aus Rockafellar und Wets, 2009, S.5f.).

a) Die Summe zweier *usc*-Funktionen ist *usc*.

b) Ist f_1 *lb* und f_2 von oben beschränkt, dann ist $f_1 + f_2$ *lb*.

Zusammen mit diesen Additionseigenschaften liefert das Hauptresultat u.A. die Existenz des MAP-Schätzers in einer Vielzahl von IRT-Modellen. Im Rahmen des in Kapitel 3.2 dargelegten IRT-Modells mit Schnelligkeitskomponente besitzt die zu maximierende Funktion die Gestalt:

$$\gamma^p(\tilde{\boldsymbol{\theta}}) = \sum_i \log(f_{i,u_i}(\theta)) + \sum_i \log(g_i(t_i|\tau)) + \gamma(\theta, \tau).$$

Da die Terme $\log(f_{i,u_i}(\theta))$ und $\log(g_i(t_i|\tau))$ für die Mehrheit der IRT-Modelle stetig und nach oben beschränkt sind, kann mittels *b*) die Existenz eines Maximums angenommen werden, insofern lediglich die log-Priori γ die *lb*-Bedingung erfüllt (und stetig ist). Da Letzteres insbesonders für die Klasse der nichtsingulären Normalverteilungen zutreffend ist (siehe Rockafellar und Wets, 2009, S.89), kann somit auch für die im Kontext der longitudinalen Modellierung (Kapitel 3.3) dargelegten Analysen stets von der Existenz eines Maximums ausgegangen werden.
Ähnliche Resultate gelten für die zur Analyse des paradoxen Effekts zentrale bedingte Maximumsabbildung, die nachfolgend mit $\hat{x}(u)$ (u übernimmt die Rolle des Parameters) bezeichnet wird. Die *usc* bzw. *lb*-Eigenschaft von f überträgt sich unmittelbar auf die Funktion, die entsteht, wenn gewisse Komponenten von f auf feste Werte fixiert werden:

c) Ist $f(x, u)$ ($x \in \mathbb{R}^{p_1}, u \in \mathbb{R}^{p_2}$, $p_1 + p_2 = p$) *usc* (*lb*), dann ist $\tilde{f}_u(x) := f(x, u)$ ebenfalls *usc* (*lb*).

Wenngleich die *lb*-Eigenschaft von f via *c*) somit automatisch auch die Existenz des *bedingten* Maximums sichert, genügt eine modifizierte Form des *lb*-Konzepts, um zu einfachen Stetigkeitsaussagen bezüglich der parametrischen Abhängigkeit der bedingten Maximumsabbildung zu gelangen:
Eine Funktion $f(x, u)$ heisst u-gleichmäßig niveaubeschränkt in x (*uniformly level bounded*), wenn zu jedem $\alpha \in \mathbb{R}$ und jedem $u \in \mathbb{R}^{p_2}$ eine Umgebung V_u von u existiert, so dass $\text{lev}_{\geq\alpha} f \cap (\mathbb{R}^{p_1} \times V_u)$ beschränkt ist. Besitzt eine usc Funktion f diese Eigenschaft und ist $x_k \in \hat{x}(u_k)$ für eine Folge $(u_k)_{k\in\mathbb{N}} \to u$ (mit $\tilde{f}_u(x) \neq -\infty$), so

ist die Folge $(x_k)_{k\in\mathbb{N}}$ beschränkt und jeder ihrer Häufungspunkte liegt in $\hat{x}(u)$ - vorausgesetzt $f(x,u)$ ist stetig an der Stelle u für ein $x \in \hat{x}(u)$ (Rockafellar und Wets, 2009, Theorem 1.17).
Dieses Resultat wurde in Kapitel 2.3 (Lemma 2.3.9) im Zuge der Analyse der parametrischen Abhängigkeit des paradoxen Effekts verwendet. Es nutzt explizit die Darstellung eines Schätzers als Lösung eines Maximierungsproblems. Der Zugang von Hooker u.a. (2009) hingegen betrachtet Schätzer primär als Lösung eines Gleichungssystems. Die Analyse der parametrischen Abhängigkeit in diesem Kontext bedarf der Theorie impliziter Funktionen, die bereits in Kapitel 2.3 zur Erlangung entsprechender Resultate verwendet wurde und im folgenden Abschnitt abschließend knapp skizziert werden soll.

Implizite Funktionen

Ausgehend von einer Funktion[55] $f : \mathbb{R}^p \times \mathbb{R}^n \mapsto \mathbb{R}^p$ $((x,u) \mapsto f(x,u))$ und einem Paar $(\overline{x},\overline{u})$ mit der Eigenschaft $f(\overline{x},\overline{u}) = \mathbf{0}$ steht im Kern der Thematik impliziter Funktionen die Frage nach der lokalen Existenz einer Lösungsfunktion. Spezifischer formuliert lautet die Frage, ob Umgebungen B_1 von $\overline{x}$ und B_2 von $\overline{u}$ existieren, so dass zu jedem $u \in B_2$ genau ein $s(u) \in B_1$ mit $f(s(u),u) = \mathbf{0}$ korrespondiert. Die so definierte Funktion s (falls existent) beschreibt bezogen auf einen lokalen Ausschnitt von f das Verhalten der Lösung x in Abhängigkeit des Parameters u. Sie wird aus diesem Grund auch als *einelementige Lokalisation* der Lösungsabbildung $S(u) := \{x | f(x,u) = \mathbf{0}\}$ um den Punkt $\overline{u}$ für $\overline{x}$ bezeichnet. Ihre Eigenschaften sind offensichtlich für die Analyse der parametrischen Abhängigkeit des paradoxen Effekts von zentralem Interesse. Entsprechende Bedingungen an die Funktion f, die die Stetigkeit bzw. Glattheit der Lokalisation s garantieren, wurden in Kapitel 2.3 explizit in Form von drei bekannten Hauptsätzen wiedergegeben. An dieser Stelle seien lediglich zwei wichtige Hinweise bezüglich der Interpretation einer solchen Lokalisation gegeben:

1. Die Existenz der einelementigen Lokalisation garantiert *nicht*, dass für jeden

[55] Wenngleich die Funktion auf dem kompletten Raum definiert ist, spielen dennoch nur lokale Eigenschaften in der folgenden Analyse eine Rolle, so dass die entsprechenden Resultate unmittelbar für Funktionen, die auf einer offenen Teilmenge definiert sind, ihre Gültigkeit behalten. Der beschrittene Weg wurde lediglich aus Gründen einer vereinfachten Notation gewählt.

Parameterwert $u \in B_2$ genau ein Lösungswert x mit $f(x, u) = \mathbf{0}$ existiert. Die Eindeutigkeit gilt nur bezogen auf eine Nachbarschaft des Punktes $\overline{x}$.

2. Die Existenz einer (stetigen) einelementigen Lokalisation ist hinreichend für die Ausdehnung des paradoxen Effekts - jedoch nicht notwendig. Es kann Situationen geben, in denen keine einelementige Lokalisation existiert, es aber dennoch möglich ist, eine „Lösungsselektion“ (siehe beispielsweise Abschnitt 1.G. aus Dontchev und Rockafellar, 2009) mit den gewünschten Stetigkeitseigenschaften zu bilden. Eine Lösungsselektion ist dabei eine Funktion, die für jedes u (in einer gewissen Umgebung von $\overline{u}$) ein $x \in S(u)$ auswählt. Erfüllt die so entstandene Funktion die Stetigkeitsbedingung, so kann der paradoxe Effekt - bei entsprechender Interpretation dieser Funktion - für Werte u nahe $\overline{u}$ beibehalten werden.

Die in Kapitel 2.3 vorgenommene „Stetigkeitsanalyse“ besagt somit gemäß Punkt 1 *nicht*, dass eine Umgebung von $\boldsymbol{\gamma}_0$ existiert, in der jeder MLE[56] (MAP) in einer ϵ-Umgebung von $\boldsymbol{\theta}_0$ liegt (und somit den paradoxen Effekt teilt), sondern, dass für jedes in einer bestimmten Umgebung liegende $\boldsymbol{\gamma}$ ein zugehöriger MLE (der durch die Auswertung der Lokalisation s an der Stelle $\boldsymbol{\gamma}$ definierte MLE) in der ϵ-Umgebung von $\boldsymbol{\theta}_0$ existiert. Dies schließt jedoch streng genommen nicht aus, dass für derartige $\boldsymbol{\gamma}$ noch weitere MLEs existieren, die außerhalb der korrespondierenden Lokalisation liegen.
Diese (subtile) Unterscheidung kann vermieden werden, wenn es eine $\boldsymbol{\gamma}_0$-Umgebung gibt, in der die Loglikelihood-Funktion strikt konkav in $\boldsymbol{\theta}$ (für jeden Parameterwert $\boldsymbol{\gamma}$ in der $\boldsymbol{\gamma}_0$-Umgebung) ist. In diesem Fall besitzt gemäß den Erörterungen zur Konkavität dieses Anhangs die entsprechende Scoregleichung (die im Likelihood-Kontext die Rolle der oben aufgeführten Funktion f übernimmt) höchstens eine Lösung für jeden Parameterwert.

[56] Man beachte, dass der MLE hier operational als Lösung der Schätzgleichung - und nicht als Maximum einer Funktion - definiert ist.

B Formen der Abhängigkeit zwischen Zufallsvariablen

In diesem Anhang wird die in Kapitel 2.4 erwähnte Beziehung des MDD- bzw. RR_2-Konzepts zu diversen statistischen Assoziationskonzepten mathematisch dargelegt. Gleichwohl für die folgenden Ergebnisse prinzipiell auf Lehmann (1966), Shea (1979) bzw. auf Lehmann und Romano (2005, Kap.3) verwiesen werden könnte, erfolgt hier eine separate Präsentation. Dies geschieht zum Einen, um eine vereinheitlichende Darstellung von Ergebnissen, die - nach Wissen des Autors - in der Literatur „verstreut“ sind, zu geben; zum Anderen, um etwaige (dezente) Modifikationen der Resultate vornehmen zu können, die im Hinblick auf die Diskussion des paradoxen Effekts von gesondertem Interesse sind (z.B. die Darlegung, wann *strikte* Ungleichheit in den entsprechenden Relationen gilt.).

Zunächst (Lemma B.0.2) wird die in Lemma 2.4.2 thematisierte stochastische Ordnung von RR_2-Größen formalisiert.

Lemma B.0.2. *Sei $f_\theta(x)$ eine „reverse rule“-Funktion (RR_2) und sei $\theta' > \theta$. Dann gelten die folgenden Aussagen.*

(a) Wenn $x' \in V_{0+} := \{z | f_{\theta'}(z) - f_\theta(z) \geq 0\}$ und wenn $x < x'$, dann ist $x \in V_{0+}$.

(b) Die Menge V_{0+} liegt links von der Menge $V_- := \{z | f_{\theta'}(z) - f_\theta(z) < 0\}$.

Wenn $f_\theta(x)$ zusätzlich eine durch θ indizierte Familie von Dichten darstellt, dann gilt darüber hinaus:

(c) Wenn $g(x)$ eine reellwertige, monoton wachsende (messbare) Funktion von x bezeichnet, die bezüglich der Familie integrierbar ist, dann gilt $\int g(x)(f_{\theta'}(x) - f_\theta(x))\mathrm{d}\mu \leq 0$ (siehe hierzu auch Lehmann und Romano, 2005, Lemma 3.4.2).

(d) Wenn zusätzlich zu den Bedingungen in c) eine Menge $\tilde{V}_+ \subset V_+ := \{z | f_{\theta'}(z) - f_\theta(z) > 0\}$ positiven Maßes existiert, auf der $g < sup_{V_+} g$ gilt, dann ist die Ungleichung in c) strikt:

$$\int g(x)(f_{\theta'}(x) - f_\theta(x))\mathrm{d}\mu < 0.$$

Beweis. (a) Aufgrund der RR_2-Eigenschaft gilt für $\theta < \theta'$ und $x < x'$

$$f_{\theta'}(x)f_\theta(x') \geq f_\theta(x)f_{\theta'}(x').$$

Letztere Ungleichung kann unter der Bedingung, dass $f_{\theta'}(x') \geq f_\theta(x')$ gilt, nur dann wahr sein, wenn zugleich $f_{\theta'}(x) \geq f_\theta(x)$ ist.

(b) Wenn es ein Paar $(x, \tilde{x})$ mit $x \in V_{0+}$, $\tilde{x} \in V_-$ und der Eigenschaft $\tilde{x} < x$ gäbe, dann würde das Ergebnis in *a*) $\tilde{x} \in V_{0+}$ erzwingen, was im Widerspruch zur Disjunktheit der Mengen V_{0+}, V_- stünde.

(c) Der Beweis folgt der Darstellung in Lehmann und Romano (2005, Lemma 3.4.2). Zunächst gilt

$$\int g(x)(f_{\theta'}(x) - f_\theta(x))\mathrm{d}\mu = \int_{V_+} g(x)(f_{\theta'}(x) - f_\theta(x))\mathrm{d}\mu + \int_{V_-} g(x)(f_{\theta'}(x) - f_\theta(x))\mathrm{d}\mu,$$

wobei $V_+ = \{x | f_{\theta'}(x) - f_\theta(x) > 0\}$ und $V_- = \{x | f_{\theta'}(x) - f_\theta(x) < 0\}$ ist.
Es kann o.B.d.A angenommen werden, dass sowohl $\mu(V_+)$ als auch $\mu(V_-)$ positiv ausfallen. Ist nämlich z.B. $\mu(V_+) = 0$, dann führt $\int (f_{\theta'}(x) - f_\theta(x))\mathrm{d}\mu = 0$ ($f_\theta(x)$ ist für variierendes θ eine Familie von Dichten) auf

$$\int_{V_+} (f_{\theta'}(x) - f_\theta(x))\mathrm{d}\mu + \int_{V_-} (f_{\theta'}(x) - f_\theta(x))\mathrm{d}\mu = \int_{V_-} (f_{\theta'}(x) - f_\theta(x))\mathrm{d}\mu = 0$$

und somit zu $\mu(V_-) = 0$. Da für $\mu(V_+) = \mu(V_-) = 0$ die Behauptung trivialerweise gültig ist, kann dieser Fall ausgeschlossen werden.
Sei also $\mu(V_+) > 0, \mu(V_-) > 0$. Dann sind V_+ und V_- nicht leer und mit den Definitionen $c_a := \sup\{g(x) | x \in V_+\}, c_b := \inf\{g(x) | x \in V_-\}$ gilt

$$-\infty < c_a \leq c_b < \infty,$$

wobei die zweite Ungleichung aus der Monotonie von g und dem Ergebnis aus Teil *b*) des Lemmas folgt.
Folgende Abschätzungen resultieren aus diesen Definitonen:

$$\int g(x)(f_{\theta'}(x) - f_\theta(x))\mathrm{d}\mu \leq c_a \int_{V_+} (f_{\theta'}(x) - f_\theta(x))\mathrm{d}\mu + c_b \int_{V_-} (f_{\theta'}(x) - f_\theta(x))\mathrm{d}\mu$$

Da $f_{\theta'}, f_\theta$ Dichten repräsentieren, gilt

$$\int_{V_+} (f_{\theta'}(x) - f_\theta(x))\mathrm{d}\mu + \int_{V_-} (f_{\theta'}(x) - f_\theta(x))\mathrm{d}\mu = 0.$$

Demnach ist

$$\int g(x)(f_{\theta'}(x) - f_\theta(x))\mathrm{d}\mu \leq (c_a - c_b)\int_{V_+}(f_{\theta'}(x) - f_\theta(x))\mathrm{d}\mu \leq 0.$$

(d) Die Herleitung kann unter Verwendung von

$$\int_{\tilde{V}_+}(c_a - g(x))(f_{\theta'}(x) - f_\theta(x))\mathrm{d}\mu > 0,$$

analog zum Beweis der Behauptung c) erfolgen.

□

Das folgende Lemma fand bereits im Beweis der Existenz des paradoxen Effekts in der Klasse der Typ-2-Schätzer Verwendung. Es wird ferner im Anschluss (s.u.) zur Herleitung der Beziehung des MDD-Konzepts zur negativen Quadrantenabhängigkeit herangezogen.

Lemma B.0.3 (Shea, 1979, S.1125)**.** *f, g seien μ-integrierbare, reellwertige Funktionen, die gegensätzlich monoton verlaufen (d.h. $sgn(f(y) - f(x)) = -sgn(g(y) - g(x))\,\forall\, x, y$). A sei eine meßbare Menge endlichen Maßes. Falls fg integrierbar ist, dann gilt die Ungleichung:*

$$\mu(A)\int_A fg\mathrm{d}\mu \leq \int_A f\mathrm{d}\mu \int_A g\mathrm{d}\mu \tag{2.52}$$

Beweis. Aus der Monotonieeigenschaft folgt, dass $(f(x) - f(y))(g(y) - g(x)) \geq 0$ unabhängig von der Wahl von x und y gilt. Deshalb resultiert:

$$\int_A\int_A (f(x) - f(y))(g(y) - g(x))\mu(\mathrm{d}x)\mu(\mathrm{d}y) \geq 0.$$

Gemäß den Annahmen über f, g und A vereinfacht sich die linke Seite zu

$$-2\mu(A)\int_A fg\mathrm{d}\mu + 2\int_A f\mathrm{d}\mu\int_A g\mathrm{d}\mu,$$

wobei dieser Ausdruck nach kurzer Umformung identisch mit (2.52) ist.
Sei $h(y) := \int_A (f(x) - f(y))(g(y) - g(x))\mu(\mathrm{d}x) \geq 0$. Dann gilt $\int_A h(y)\mu(\mathrm{d}y) = 0$ genau dann, wenn eine (meßbare) Menge $A^* \subset A$ mit $\mu(A^*) = \mu(A)$ existiert, so dass $h(y) = 0$ für alle $y \in A^*$ ist. Ferner gilt $h(y) = 0$ genau dann, wenn eine (meßbare) Menge $A_y \subset A$ existiert, so dass $\mu(A_y) = \mu(A)$ und entweder $f(x) = f(y)$

oder $g(x) = g(y)$ für $x \in A_y$ gilt - wobei die Wahl, welche der beiden Gleichungen gültig ist, von x abhängen kann.
Es gilt $\int_A h(y)\mu(\mathrm{d}y) > 0$ genau dann, wenn eine meßbare Menge $A' \subset A$ positiven Maßes existiert, so dass $h(y) > 0$ für alle $y \in A'$ ist. $h(y) > 0$ ist wiederum äquivalent zur Existenz einer meßbaren Menge A_y mit $\mu(A_y) > 0$ und der Eigenschaft $A_y \subset \{x|f(x) \neq f(y) \wedge g(x) \neq g(y)\}$. □

Mit den bisherigen Resultate ist es nun möglich, die in Korollar 2.4.1 erwähnte Beziehung des RR_2-Konzepts zum Begriff der negativen Quadrantenabhängigkeit zu beweisen. Aus dieser Relation kann dann abschließend (siehe Lemma B.0.4) die Korrelationsbeziehung aus Korollar 2.4.2 gewonnen werden.

Beweisskizze der Implikation: „reverse rule“ ⇒ „negative quadrant dependence“

Beweis. Die nachfolgende Beweisskizze folgt den Ausführungen von Lehmann (1966, Lemma 4).
Sei $f(\theta, \tau)$ eine RR_2-Dichte eines bivariaten Zufallsvektors. Die gemeinsame Verteilungsfunktion kann i.A. wie folgt dargestellt werden:

$$P(\theta \leq y, \tau \leq x) = \int_{-\infty}^{x} P(\theta \leq y|\tau = u)\mathrm{d}F_\tau(u),$$

wobei $P(\theta \leq y|\tau = u)$ gemäß den Ausführungen aus Teil c des Lemmas B.0.2 monoton wachsend in u für jede Wahl von y ist (man wähle für g lediglich die Indikatorfunktion eines entsprechenden Intervalls). Die Indikatorfunktion $I_{\{u \leq x\}}$ ist andererseits monoton fallend in u, so dass das Paar $(I_{\{u \leq x\}}, P(\theta \leq y|\tau = u))$, als Funktion von u betrachtet, gegensätzlich geordnet ist.
Eine Anwendung von Lemma B.0.3 mit $f(u) = P(\theta \leq y|\tau = u), g(u) := I_{\{u \leq x\}}, A := (-\infty, x'](x' > x), \mu := F_\tau$ führt auf den Ausdruck

$$P(\tau \leq x')P(\theta \leq y, \tau \leq x) \leq P(\theta \leq y, \tau \leq x')P(\tau \leq x).$$

Limesbildung beim Übergang $x' \to \infty$ liefert dann das gewünschte Resultat. □

Lemma B.0.4 (Lehmann, 1966, Lemma 2)**.** *Seien $f_1, f_2, f_1 f_2$ reellwertige Funktionen aus $L^1(\mu)$, wobei μ ein Wahrscheinlichkeitsmaß darstelle. Dann gilt:*

$$\int f_1 f_2 \mathrm{d}\mu - \int f_1 \mathrm{d}\mu \int f_2 \mathrm{d}\mu = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (F_{12}(u,v) - F_1(u)F_2(v)) \mathrm{d}u\mathrm{d}v, \qquad (2.53)$$

wobei F_{12} die gemeinsame Verteilungsfunktion von (f_1, f_2) bezeichnet.

Beweis. Seien $(f_1, f_2), (g_1, g_2)$ unabhängig und identisch nach F_{12}-verteilte Zufallsvektoren und sei $\tilde{\mu}$ das Produktmaß $\mu \times \mu$.
Man betrachte die folgende, auf $\mathbb{R}$ definierte, reellwertige Funktion von u:

$$s_1(u) := I_{\{u<f_1\}} - I_{\{u<g_1\}} = \begin{cases} 1 & f_1 > u \geq g_1 \\ -1 & g_1 > u \geq f_1 \\ 0 & \text{sonst} \end{cases}$$

Das Lebesgue-Integral von s ist $f_1 - g_1$.
Das Produkt $(f_1 - g_1)(f_2 - g_2)$, dessen Integral das Doppelte des linken Ausdrucks in (2.53) beträgt, kann demnach wie folgt geschrieben werden:

$$(f_1 - g_1)(f_2 - g_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} s_1(u) s_2(v) \mathrm{d}u\mathrm{d}v,$$

wobei $s_2(v)$ analog zu $s_1(u)$ definiert sei.
Unter der vorläufigen Annahme der Vertauschbarkeit der Integrationsreihenfolge ergibt die Integration dieses Ausdrucks nach $\tilde{\mu}$:

$$\int (f_1 - g_1)(f_2 - g_2) \mathrm{d}\tilde{\mu} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int (I_{\{u<f_1\}} - I_{\{u<g_1\}})(I_{\{v<f_2\}} - I_{\{v<g_2\}}) \mathrm{d}\tilde{\mu}\mathrm{d}u\mathrm{d}v.$$

Oder gleichbedeutend hierzu:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int (I_{\{g_1 \leq u\}} - I_{\{f_1 \leq u\}})(I_{\{g_2 \leq v\}} - I_{\{f_2 \leq v\}}) \mathrm{d}\tilde{\mu}\mathrm{d}u\mathrm{d}v.$$

Das innere Integral gleicht wiederum dem Ausdruck

$$\int I_{\{g_1 \leq u\}} I_{\{g_2 \leq v\}} \mathrm{d}\tilde{\mu} + \int I_{\{f_1 \leq u\}} I_{\{f_2 \leq v\}} \mathrm{d}\tilde{\mu} - \int I_{\{f_1 \leq u\}} I_{\{g_2 \leq v\}} \mathrm{d}\tilde{\mu} - \int I_{\{f_2 \leq v\}} I_{\{g_1 \leq u\}} \mathrm{d}\tilde{\mu}.$$

Da $(f_1, f_2), (g_1, g_2)$ unabhängig und identisch verteilt sind, folgt Gleichung (2.53).
Damit verbleibt lediglich der Beweis der Vertauschbarkeit der Integrationsreihenfolge. Nutzt man für den Ausdruck

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |s_1(u) s_2(v)| \mathrm{d}u\mathrm{d}v = |f_1 - g_1||f_2 - g_2|$$

die Dreiecksungleichung sowie die Integrierbarkeit von f_1, f_2 und $f_1 f_2$ (bzw. der „Kopie“ (g_1, g_2)), so liefert eine Anwendung des Satzes von Fubini das gewünschte Ergebnis. □

Bemerkung: Für die Messbarkeit des Ausdrucks $(I_{\{u<f_1\}} - I_{\{u<g_1\}})(I_{\{v<f_2\}} - I_{\{v<g_2\}})$ bezüglich $\mathrm{d}u \times \mathrm{d}v \times \mathrm{d}\tilde{\mu}$ kann wie folgt argumentiert werden:
Die Indikatorfunktion $I_{\{u<f_1\}}$ ist messbar, wenn die Menge $\{(u, v, x)|u < f_1(x)\}$ als abzählbare Vereinigung messbarer Rechtecke darstellbar ist. Sei r_n eine Aufzählung der rationalen Zahlen und sei $A_n := \{(u, v, x)|u < r_n\} \cap \{(u, v, x)|r_n < f_1(x)\}$. Dann ist A_n als Durchschnitt zweier messbarer Rechtecke wiederum ein messbares Rechteck. Da $\{(u, v, x)|u < f_1(x)\} = \cup_{n=1}^{\infty} A_n$ ist, folgt die Messbarkeit von $\{(u, v, x)|u < f_1(x)\}$.

Alternative Argumentation der Messbarkeit:
Man definiere den stochastischen Prozess $M(u, x) := I_{\{u<f_1(x)\}}$, in dem u als Zeitparameter fungiert.
Da M für jedes u eine messbare Funktion in x darstellt, und da ferner alle Pfade des stochastischen Prozesses rechtsstetig sind, folgt die Produktmessbarkeit (siehe Medvegyev, 2007, S.11)

C Paradoxe Effekte im mehrdimensionalen Rasch-Modell

Während die Diskussion des paradoxen Effekts im Rahmen zweier spezieller Schätztypen (Typ 1: Schätzer als Stelle des Maximums einer Funktion; Typ 2: Schätzer basierend auf Funktionalen der Posteriori) erfolgte, dient dieser Abschnitt einer Erweiterung der Resultate im Hinblick auf Methodiken der klassisch-frequentistischen Schätz-und Testtheorie.

Ziel ist die Demonstration eines paradoxen Effekts beim Testen der Nullhypothese „$\theta_1 \leq \kappa_1$“ im Rahmen der Klasse der mehrdimensionalen Rasch-Modelle (im Folgenden mit „MRM“ abgekürzt). Diese Modellklasse ist von besonderem Interesse für die weitere Untersuchung des paradoxen Effekts, da sie die mehrdimensionale Variante eines probabilistischen Modells darstellt, welches - konträr zu (den meisten) anderen IRT-Modelle - eine fundierte messtheoretische Motivation besitzt (Fischer, 1995) und zudem über eine Anwendung des Neyman-Pearson-Lemmas die Konstruktion eines statistisch optimalen Tests (genauer: „uniformly most powerful“) für prinzipiell beliebigen Stichprobenumfang ermöglicht. Da die meisten ML-Schätzer (bzw. Bayes-Schätzer) lediglich asymptotischen Optimalitätskriterien genügen, stellt die Ausdehnung der Untersuchung des paradoxen Effekts auf die Klasse der mehrdimensionalen Rasch-Modelle folglich einen Zusatzaspekt des Paradoxons heraus, der sich zudem ferner nahtlos in die in Kapitel 5 skizzierte Debatte über den Antagonismus zwischen statistischer Effizienz und individuumbasierter Fairness eingliedern lässt.

Im Folgenden wird anhand eines konkreten Zahlenbeispiels gezeigt werden, dass im Rahmen des MRM paradoxe Effekte beim Hypothesentesten auftreten können. Spezifischer heisst dies, dass falsches Beantworten eines Items zur Ablehnung der H_0: „$\theta_1 \leq \kappa_1$“ führen kann, während korrektes Beantworten die Beibehaltung der H_0 impliziert. Als „Nebenprodukt“ dieser Betrachtung erhält man ferner das verblüffende Resultat, dass selbst beliebig viele richtig beantwortete Items keine positive Klassifikation mehr herbeiführen können. Bevor dies jedoch im Detail demonstriert wird, sollen zunächst einige Vorbemerkungen zur grundlegenden Struktur eines Hypothesentests im Rasch-Modell gegeben werden.

Die Item-Response-Funktionen eines (zur Vereinfachung) zweidimensionalen Rasch-

Modells besitzen die Struktur

$$P(U_i = 1|\boldsymbol{\theta}) = \frac{\exp(a_{i1}\theta_1 + a_{i2}\theta_2 + \beta_i)}{1 + \exp(a_{i1}\theta_1 + a_{i2}\theta_2 + \beta_i)},$$

worin $-\beta_i$ die Schwierigkeit des i-ten Item und a_{ij} die Diskrimination des i-ten Item entlang der j-ten Dimension ($j = 1, 2$) kennzeichnen.
Unter der modellinhärenten Annahme lokaler stochastischer Unabhängigkeit lässt sich somit die Wahrscheinlichkeit eines beobachteten Antwortmusters wie folgt darstellen:

$$\begin{aligned} P(\boldsymbol{u}|\theta_1, \theta_2) &= \frac{\exp\left(\theta_1 \sum_i a_{i1}u_i + \theta_2 \sum_i a_{i2}u_i\right) \exp(\sum_i u_i\beta_i)}{\prod_i \left(1 + \exp(a_{i1}\theta_1 + a_{i2}\theta_2 + \beta_i)\right)} \\ &= \exp\left(\theta_1 \sum_i a_{i1}u_i + \theta_2 \sum_i a_{i2}u_i\right) g(\boldsymbol{u})h(\boldsymbol{\theta}). \end{aligned} \tag{3.54}$$

Es liegt demnach eine zweiparametrige Exponentialfamilie mit suffizienter Statistik $T(\boldsymbol{U}) = (T_1(\boldsymbol{U}), T_2(\boldsymbol{U})) = (\sum_i a_{i1}U_i, \sum_i a_{i2}U_i)$ vor.
Nach „üblicher“ frequentistischer Vorgehensweise werden Störparameter (hier: θ_2) durch Bedingen auf deren suffiziente Statistik eliminiert. Die resultierende bedingte Verteilung ist frei von θ_2 ($M(t_2) := \{\boldsymbol{v}|T_2(\boldsymbol{v}) = t_2\}$):

$$P(\boldsymbol{u}|T_2(\boldsymbol{U}) = t_2, \theta_1) = \frac{g(\boldsymbol{u}) \exp(\theta_1 \sum_i a_{i1}u_i)}{\sum_{\boldsymbol{v} \in M(t_2)} g(\boldsymbol{v}) \exp(\theta_1 \sum a_{i1}v_i)} I_{\{T_2(\boldsymbol{u}) = t_2\}}. \tag{3.55}$$

Die bedingte Verteilung (3.55) bildet wiederum eine Exponentialfamilie. Sie weist darüber hinaus einen monotonen Dichtequotienten in $T_1(\boldsymbol{u})$ auf, so dass das Neyman-Pearson-Lemma anwendbar ist. Der „gleichförmig beste“ (uniformly most powerful) Test (bezüglich der bedingten Verteilung) besitzt daher die Testfunktion (Lehmann und Romano, 2005, Theorem 3.4.1)

$$\phi(\boldsymbol{u}) = \begin{cases} 1 & T_1(\boldsymbol{u}) > C \\ \gamma & T_1(\boldsymbol{u}) = C \\ 0 & T_1(\boldsymbol{u}) < C, \end{cases} \tag{3.56}$$

d.h., wenn der Wert der suffizienten Statistik für θ_1 einen kritischen Wert C überschreitet, dann erfolgt eine Ablehnung der H_0; für Werte $< C$ wird die H_0 beibehalten; und für Werte $T_1 = C$ entscheidet ein Zufallsmechanismus, ob die H_0 verworfen wird. Letzteres geschieht mit Wahrscheinlichkeit γ, wobei γ derart gewählt ist, dass der Test das vorgegebene Signifikanzniveau ausschöpft.

Im Folgenden wird zur Vereinfachung die nichtrandomisierte Form des UMP-Tests betrachtet. Es sei jedoch bemerkt, dass die Resultate qualitativ gesehen nicht von denen des randomisierten Tests abweichen.
Die nichtrandomisierte Form des Tests lehnt genau dann ab, wenn $T_1 > C$ gilt, wobei C derart zu wählen ist, dass das vorgegebene Signifikanzniveau eingehalten wird, d.h.

$$C := \inf\{t \mid P(T_1(\boldsymbol{U}) > t | \theta_1 = \kappa_1, t_2) \leq \alpha\}.$$

Der resultierende Test ist (für jeden beliebigen Stichprobenumfang) die nichtrandomisierte Variante eines gleichförmig besten Tests der H_0 : „$\theta_1 \leq \kappa_1$“ gegen die H_1 : „$\theta_1 > \kappa_1$“ innerhalb des mehrdimensionalen Rasch-Modells.
Das folgende Beispiel wird einen gravierenden paradoxen Effekt dieses „optimalen“ Tests aufzeigen und somit die Diskrepanz zwischen statistischer Angemessenheit und individuumbasierter Fairness noch einmal hervorheben.
Gegeben sei ein mehrdimensionales Rasch-Modells mit den Itemparametern

$$\boldsymbol{a_1^T} = \boldsymbol{a_2^T} = (1,1), \boldsymbol{a_3^T} = (2,1), \boldsymbol{a_4^T} = (0,2)$$

sowie $\beta_i = 0\,(i = 1, \ldots 4.)$. Der Schwellenwert sei ferner $\kappa_1 = 0$. Das beobachtete Antwortmuster der zu diagnostizierenden Person sei ferner $\boldsymbol{u_{obs}^T} = (1,0,1,1)$. Die suffiziente Statistik nimmt demnach die Werte $T(\boldsymbol{u_{obs}}) = (t_1 = 3, t_2 = 4)$ an und die zur Konstruktion des Ablehnbereichs relevante bedingte Verteilung (3.55) nimmt unter $\theta_1 = \kappa_1$ die einfache Form

$$P(\boldsymbol{u}|T_2(\boldsymbol{U}) = t_2, \kappa_1) = \begin{cases} |M(t_2)|^{-1} & \text{falls } T_2(\boldsymbol{u}) = t_2 \\ 0 & \text{sonst} \end{cases}$$

an, so dass die Ermittlung des Ablehnbereichs - bzw. die Berechnung des p-Werts - auf ein Abzählproblem reduziert werden kann.
Folglich gilt es, die Mächtigkeit der Menge $M(t_2)$ sowie die Mächtigkeit der Menge

$$M(t_1) \cap M(t_2) \quad (M(t_1) := \{\boldsymbol{u}|T_1(\boldsymbol{u}) \geq T_1(\boldsymbol{u_{obs}})\})$$

zu ermitteln.
$M(t_2)$ besteht aus den Vektoren

$$\boldsymbol{u_1^T} = (1,1,0,1),\ \boldsymbol{u_2^T} = (1,0,1,1),\ \boldsymbol{u_3^T} = (0,1,1,1),$$

wobei $\boldsymbol{u_1}$ den einzigen Vektor in $M(t_2)$ darstellt, der einen geringeren Wert auf der suffizienten Statistik T_1 aufweist als der beobachtete Wert $t_1 = 3$. Demnach resultiert

ein p-Wert von $\frac{2}{3}$. Hätte die Person stattdessen das letzte Item falsch beantwortet, dann wäre $t_2 = 2$; $M(t_2)$ bestünde aus den Vektoren

$$\boldsymbol{u_1^T} = (0,0,0,1),\ \boldsymbol{u_2^T} = (1,1,0,0),\ \boldsymbol{u_3^T} = (1,0,1,0), \boldsymbol{u_4^T} = (0,1,1,0)$$

und es gäbe zwei Vektoren ($\boldsymbol{u_1}$ und $\boldsymbol{u_2}$), die einen geringeren Wert bezüglich der Teststatistik T_1 aufwiesen. Demnach ergäbe sich ein verringerter p-Wert von 0.5.
Es sei dabei bemerkt, dass die geringe Testlänge in Kombination mit der speziellen Wahl der Itemparameter lediglich einem besseren Verständnis der Berechnungen und des zugrundeliegenden Phänomens dienen. Eine Ausdehnung auf realistische Testlängen und Itemparameter ist möglich. Auf diese Weise wird deutlich, dass das Umwandeln einer richtigen in eine falsche Antwort die Ablehnung der H_0 verursachen kann, obwohl die Dimension θ_1 in keiner direkten Beziehung zum entsprechenden Item steht.
Der beschriebene paradoxe Effekt lässt sich noch akzentuierter darstellen, indem die Testlänge erhöht wird. Man nehme hierfür an, dass die betreffende Person (beliebig viele) weitere Items bearbeitet und diese korrekt löst. Die intuitive Erwartungshaltung legt nahe, dass die korrespondierenden p-Werte gegen Null konvergieren, insofern die hinzugefügten Items positiv auf der ersten Dimension laden. Wie die folgenden Ausführungen verdeutlichen werden, kommt es jedoch zu der drastischen Diskrepanz, dass die Person beliebig viele Items lösen kann, ohne eine bessere Klassifikation zu erhalten; während die Person unter dem „Was wäre wenn“-Szenario (d.h. bei Umwandeln des Responses beim vierten Item) eine positive Klassifikation erreichen könnte.
Um diese Ausführungen zu konkretisieren, seien weitere, von der Person richtig beantwortete Items mit Ladungsvektoren $\boldsymbol{a_i} = \boldsymbol{a_1}$ $(i \geq 5)$ gegeben. Die Gestalt der Mengen[57] $M_i(t_{2i})$ bzw. $M_i(t_{2i}) \cap M_i(t_{1i})$ bei ansteigender Testlänge ist für die Untersuchung der p-Werte von zentraler Bedeutung. Die bereits gekennzeichneten Elemente der Menge $M_4(t_{2,4})$ lassen sich in einer Matrix festhalten:

$$M_4(t_{2,4}) = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{pmatrix}$$

[57]Der Index i wurde eingeführt, um die Abhängigkeit der Mengen von der Testlänge zu reflektieren.

Jede Zeile dieser Matrix taucht - erweitert um eine „1“ - in der Menge $M_5(t_{2,5})$ auf:

$$M_5(t_{2,5}) = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix}$$

Lediglich eine neue Zeile, die zur Falschbeantwortung des neuen Items korrespondiert, unterscheidet die Matrix $M_5(t_{2,5})$ von der vorherigen Matrix $M_4(t_{2,4})$. Diese Prozedur lässt sich beliebig fortsetzen, d.h. die vorherige Matrix ist als Teilmatrix einer um eine Zeile erweiterten Nachfolgermatrix darstellbar. Die Kardinalität der Mengen nimmt demnach konstant um eine Einheit zu. Da innerhalb einer jeden Menge $M_j(t_{2,j})$ lediglich ein Antwortmuster existiert, welches einen geringeren Wert bezüglich der suffizienten Statistik T_1 aufweist als der beobachtete Wert, folgt, dass die korrespondierende p-Wert-Folge $(p_j)_{j\in\mathbb{N}}$ die Gestalt $p_j = \frac{j-2}{j-1}$ $(j \geq 4)$ besitzt und demnach gegen Eins konvergiert.

Im starken Kontrast zu diesem Ergebnis steht die analoge Berechnung für den Fall $\boldsymbol{u_{obs}} = (1, 0, 1, 0)$ und einer (beliebig langen) Sequenz richtig gelöster Items (mit obigen Itemparametern). Die Kardinalität der Menge $M_j(t_{2,j})$ ergibt sich für $j \geq 5$ zu

$$|M_j(t_{2,j})| = \binom{j-1}{j-2} + \binom{j-1}{j-4}. \tag{3.57}$$

Wenn A_j die Items des Tests der Länge j bezeichnet, dann gibt der erste Term in (3.57) die Anzahl der Möglichkeiten an, $j-2$ Items aus der Menge $A_j \backslash \{4\}$ zu wählen. Letzteres entspricht allen Antwortmuster, die zu dem gleichen Wert der Statistik $T_{2,j}$ führen wie der beobachtete Wert $T_{2,j} = j-2$ und zugleich durch falsches Beantworten des vierten Items - das einzige Item mit Diskriminationswert $\neq 1$ bezüglich der zweiten Dimension - gekennzeichnet sind. Der zweite Term zählt die verbliebenen Antwortmöglichkeiten, die ebenfalls auf $T_{2,j} = j-2$ führen, jedoch darüber hinaus durch richtiges Beantworten des vierten Items gekennzeichnet sind. Jedes Antwortmuster in $M_j(t_{2,j})$, in dem das vierte Item richtig beantwortet wird, weist einen geringeren Wert bezüglich $T_{1,j}$ auf als der beobachtete Wert $T_{1,j} = j-1$. Folglich gibt es in $M_j(t_{2,j})$ mindestens $\binom{j-1}{j-4}$ Antwortmuster, die einen geringeren Wert der Teststatistik implizieren. Die Folgenglieder $q_j := 1 - p_j$ (p_j bezeichnet den zur Testlänge j gehörenden p-Wert) erfüllen demnach die Ungleichung

$$\frac{\binom{j-1}{j-4}}{\binom{j-1}{j-2} + \binom{j-1}{j-4}} \leq q_j \leq 1 \quad \forall j \geq 5.$$

Da Zähler und Nenner Polynome in j beschreiben, deren höchste Koeffizienten identisch ausfallen, folgt $q_j \to 1$ bzw. $p_j \to 0$. In diesem fiktiven Szenario ergibt sich folglich ein „natürliches“ Verhalten der p-Werte, d.h. das Lösen weiterer, zur zu testenden Dimension in positiver Relation stehender Items führt letztendlich zur Ablehnung der $\mathrm{H}_0 : „\theta_1 \leq \kappa_1“$ und damit zu einer positiven Klassifikation des Individuums. Konträr zu dieser Beobachtung entsteht durch das Umwandeln *einer* Antwort ein kontraintuitives Verhalten der p-Werte, welches den absurden Effekt impliziert, dass das Individuum durch noch so kompetentes Bearbeiten des Tests keine positive Klassifikation mehr erlangen kann.

D Nachweis der SMDD-Bedingung für „Standard-MIRT-Modelle“

Nachfolgend soll die für das Paradoxon zentrale SMDD-Bedingung in vier gängigen IRT-Modellklassen[58] nachgewiesen werden. Konkret werden

1) M2PL-Modelle (Logistische Linkfunktion)

2) Mehrdimensionale „Normal Ogive“-Modelle (Probitfunktion)

sowie für die Klasse der ordinalen MIRT-Modelle

3) MGRM-Modelle (*multidimensional graded response model*)

4) MGPC-Modelle (*multidimensional generalized partial credit model*)

betrachtet.
Da in allen Fällen die von dem jeweiligen Modell induzierte Loglikelihood die Form[59]

$$l(\boldsymbol{\theta}) = \sum l_i(\boldsymbol{a}_i^T \boldsymbol{\theta}) \tag{4.58}$$

aufweist, erscheinen ein paar generelle Vorbemerkungen zur Struktur (4.58) angebracht.
Bildet man in (4.58) - unter der entsprechenden Differenzierbarkeitsannahme - die gemischten partiellen Ableitungen, so folgt:

$$\frac{\partial^2 l}{\partial \theta_j \partial \theta_l} = \sum_i l_i''(\boldsymbol{a}_i^T \boldsymbol{\theta}) a_{ij} a_{il} \tag{4.59}$$

Unter der Annahme, dass die Diskriminationsvektoren $\boldsymbol{a_i}$ ausschließlich positive Einträge aufweisen, folgt die SMDD-Bedingung für jedes Dimensionenpaar (j, l), insofern $l_i'' \leq 0$ für alle i und $l_{i^*}'' < 0$ für mindestens ein i^* gilt. Letztere Bedingung wird nachfolgend für jedes der aufgeführten Modelle nachgewiesen.

[58] Die Auswahl dieser Modelle ist im Wesentlichen an der Darstellung von Reckase (2009, Kap. 4) orientiert. Dieser Quelle können auch die nachfolgend dargestellten funktionellen Formen entnommen werden.

[59] Zur Vereinfachung der Notation wird auf die $\boldsymbol{u}$-Indizierung, die in den vorangegangenen Kapiteln zur Kennzeichnung der Beziehung der Loglikelihood zum Antwortmuster diente, verzichtet.

1) M2PL-Modell

Mit der Abkürzung $F(x) = \frac{\exp(x)}{1+\exp(x)}$ lässt sich l_i im Falle einer richtigen Antwort ($u_i = 1$) als

$$l_i(x) := \log(F(x))$$

bzw. im Falle einer falschen Antwort als

$$\tilde{l}_i(x) := \log(1 - F(x))$$

darstellen.
Da die logistische Funktion F die Differentialgleichung $F' - F(1 - F) = 0$ erfüllt, folgt:

$$l_i'(x) = 1 - F(x), \quad \tilde{l}_i'(x) = -F(x)$$

Da die Funktionen $1 - F(x)$ bzw. $-F(x)$ streng monoton fallend (und differenzierbar) sind, folgt die Behauptung.

2) Mehrdimensionales „Normal Ogive"-Modell

Mit den Abkürzungen $\phi(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}x^2), \Phi(x) := \int_{-\infty}^{x} \phi(y)\mathrm{d}y$ lässt sich l_i im Falle einer richtigen Antwort ($u_i = 1$) als

$$l_i(x) := \log(\Phi(x))$$

bzw. im Falle einer falschen Antwort als

$$\tilde{l}_i(x) := \log(1 - \Phi(x))$$

darstellen.
Offenbar gilt:

$$\Phi'(x) = \phi(x),\ \Phi''(x) = -x \cdot \phi(x).$$

Es folgt:

$$l_i'' = \frac{-x\Phi(x)\phi(x) - \phi^2(x)}{\Phi^2(x)},\ \tilde{l}_i'' = \frac{x(1 - \Phi(x))\phi(x) - \phi^2(x)}{(1 - \Phi(x))^2}$$

Die Verifikation der SMDD-Eigenschaft mündet demnach für den Fall $u_i = 1$ in den Beweis der Gleichung

$$g(x) := x\Phi(x) + \phi(x) > 0 \quad \forall x.$$

Da $g'(x) = \Phi(x) > 0$ ist, ist g streng monoton wachsend. Demnach folgt die Behauptung, wenn $\lim_{x\to-\infty} g(x) \geq 0$ gezeigt werden kann.
Nach dem Gesetz von l'Hospital (siehe z.B. Theorem 5.13 aus Rudin (1976)) gilt:

$$\lim_{x\to-\infty} x\Phi(x) = \lim_{x\to-\infty} -x^2\phi(x) = \lim_{x\to-\infty} -\frac{x^2}{c\exp(\frac{1}{2}x^2)} = 0.$$

Die Verifikation der SMDD-Eigenschaft für den Fall $u_i = 0$ verläuft analog.

3) MGRM:

In diesem ordinalen MIRT-Modell besitzt der Beitrag eines Items zur Loglikelihood die Form[60]:

$$l_i(x) := \log(\Phi(x+\tau_i) - \Phi(x+\delta_i)) \quad (\tau_i > \delta_i)$$

Offenbar genügt es die Eigenschaft $l_i'' < 0$ für den Fall $\delta_i = 0$ nachzuweisen.
Die erste Ableitung l_i' ist von der Form

$$l_i'(x) = \frac{\phi(x+\tau_i) - \phi(x)}{\Phi(x+\tau_i) - \Phi(x)}$$

Weiteres Differenzieren führt unter Verwendung der Identität $\phi'(x) = -x\phi(x)$ auf:

$$l_i''(x) = \frac{(\Phi(x+\tau_i) - \Phi(x))(x\phi(x) - (x+\tau_i)\phi(x+\tau_i)) - (\phi(x+\tau_i) - \phi(x))^2}{(\Phi(x+\tau_i) - \Phi(x))^2}$$

Wenn somit für alle x die Ungleichung

$$(\Phi(x+\tau_i) - \Phi(x))(x\phi(x) - (x+\tau_i)\phi(x+\tau_i)) - (\phi(x+\tau_i) - \phi(x))^2 < 0 \qquad (4.60)$$

begründet werden kann, dann ist der Beweis vollständig.
Für $x \leq 0 \leq x+\tau_i$ ist die Ungleichung trivialerweise erfüllt. Daher kann im Folgenden angenommen werden, dass $0 < x < x+\tau_i$ oder $x < x+\tau_i < 0$ gilt, so dass das Intervall $[x, x+\tau_i]$ die Null nicht umschließt.
Der erste Faktor der linken Seite von (4.60) kann wie folgt abgeschätzt werden:

$$\begin{aligned}
\Phi(x+\tau_i) - \Phi(x) &= \int_x^{x+\tau_i} \phi(y)\mathrm{d}y = -\int_x^{x+\tau_i} \frac{\phi'(y)}{y}\mathrm{d}y \\
&= -\left[\frac{\phi(y)}{y}\right]_x^{x+\tau_i} - \int_x^{x+\tau_i} \frac{\phi(y)}{y^2}\mathrm{d}y \\
&\leq -\left[\frac{\phi(y)}{y}\right]_x^{x+\tau_i} \qquad (4.61)
\end{aligned}$$

[60]Streng genommen gilt dies nur für die mittleren Antwortkategorien. Der Beitrag der Extremkategorien wurde jedoch indirekt bereits unter 2) behandelt.

Da o.B.d.A der zweite Faktor in (4.60) als nichtnegativ betrachtet werden kann (andernfalls ist die Ungleichung trivialerweise erfüllt), ist die linke Seite in (4.60) stets kleiner oder gleich dem Ausdruck

$$\left(\frac{\phi(x)}{x} - \frac{\phi(x+\tau_i)}{x+\tau_i}\right)(x\phi(x) - (x+\tau_i)\phi(x+\tau_i)) - (\phi(x+\tau_i) - \phi(x))^2.$$

Der Term $\left(\frac{\phi(x)}{x} - \frac{\phi(x+\tau_i)}{x+\tau_i}\right)(x\phi(x) - (x+\tau_i)\phi(x+\tau_i))$ lässt sich ferner als

$$\phi^2(x) + \phi^2(x+\tau_i) - w\left(\frac{x+\tau_i}{x}\right)\phi(x+\tau_i)\phi(x)$$

mit $w(z) := z + \frac{1}{z}$ darstellen.
Unter Beachtung der für $0 < z \neq 1$ gültigen Ungleichung $w(z) > 2$ folgt, dass der Koeffizient des Terms $\phi(x+\tau_i)\phi(x)$ echt kleiner als -2 ausfällt. Aus Letzterem ergibt sich unmittelbar die Behauptung.

4) MGPC:

Für $n \in \{0, 1 \ldots, L\}$ besitzt l_i die Form:

$$l_i(x) = nx - \log\left(\sum_{l=0}^{L} \exp(lx - c_l)\right)$$

Differenzieren führt unmittelbar auf

$$l_i'(x) = n - \frac{\sum_{l=0}^{L} l\exp(lx - c_l)}{\sum_{l=0}^{L} \exp(lx - c_l)}.$$

Die für die Herleitung der SMDD-Eigenschaft relevante Ableitung besitzt demnach die Form:

$$l_i''(x) = -\frac{\left(\sum_{l=0}^{L} \exp(lx - c_l)\right) \cdot \left(\sum_{l=0}^{L} l^2 \exp(lx - c_l)\right) - \left(\sum_{l=0}^{L} l\exp(lx - c_l)\right)^2}{\left(\sum_{l=0}^{L} \exp(lx - c_l)\right)^2}.$$

Es gilt somit die strikte Ungleichung

$$\left(\sum_{l=0}^{L} \exp(lx - c_l)\right) \cdot \left(\sum_{l=0}^{L} l^2 \exp(lx - c_l)\right) > \left(\sum_{l=0}^{L} l\exp(lx - c_l)\right)^2 \quad \forall x \qquad (4.62)$$

zu verifizieren.
Definiert man $d_j := \exp(jx - c_j)$, so lässt sich die linke Seite von (4.62) mit dem Ausdruck

$$\sum_l l^2 d_l^2 + \sum_{i<j} (i^2 + j^2) d_i d_j$$

identifizieren, während die rechte Seite die Gestalt

$$\sum_l l^2 d_l^2 + \sum_{i<j} 2ij d_i d_j$$

aufweist.
Die Differenz ist demnach durch

$$\sum_{i<j} (i^2 + j^2) d_i d_j - \sum_{i<j} 2ij d_i d_j = \sum_{i<j} (i - j)^2 d_i d_j$$

gegeben.
Da stets $d_i > 0$ gilt, folgt die Behauptung.

Bemerkung: Für den Fall binärer Modelle wurde stets unterstellt, dass entsprechende Itemschwierigkeitsparameter auf Null fixiert sind. Dies kann - jedenfalls zur Etablierung der SMDD-Eigenschaft - o.B.d.A angenommen werden, da für eine auf $\mathbb{R}$ definierte Funktion f mit korrespondierender Shiftfunktion $g(x) := f(x+\beta)$ die Relation

$$f'' < 0 \Leftrightarrow g'' < 0$$

gilt.

Literatur

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New York: Wiley.

Andrade, D. F. & Tavares, H. R. (2005). Item response theory for longitudinal data: Population parameter estimation. *Journal of Multivariate Analysis, 95*, 1–22.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis* (2nd ed.). New York: Springer.

Casella, G. & Berger, R. (2002). *Statistical Inference* (2nd ed.). Pacific Grove, CA: Duxbury.

Dantzig, G. B. (1962). *Linear programming and extensions.* Princeton, N.J.: Princeton University Press.

Dini, U. (1877). *Analisi infinitesimale: lezioni dettate nella R.* Università di Pisa: nell'anno accademico 1877-78. Ciro Valenti.

Dontchev, A. L. & Rockafellar, R. T. (2009). *Implicit functions and solution mappings: A view from variational analysis.* Berlin: Springer.

Esary, J. D., Proschan, F. & Walkup, D. W. (1967). Association of random variables, with applications. *Annals of Mathematical Statistics, 38,* 1466-1474.

Finkelman, M., Hooker, G. & Wang, Z. (2010). Prevalence and magnitude of paradoxical results in multidimensional item response theory. *Journal of Educational and Behavioral Statistics, 35,* 744-761.

Fischer, G. H. (1995). Derivations of the Rasch model. In Fischer G. H. & Molenaar, I. W. (Eds.), Rasch models: Foundations, recent developments, and applications, 15-38. New-York: Springer.

Fox, J. P. (2010). *Bayesian item response modeling: theory and applications.* New York: Springer.

Goursat, E. (1903). Sur la théorie des fonctions implicites. *Bull. Soc. Math. France, 31*, 184-192.

Hampel, P. & Petermann, F. (2006). Fragebogen zum Screening psychischer Störungen im Jugendalter (SPS-J). *Zeitschrift für Klinische Psychologie und Psychotherapie, 35,* 204-214.

Harville, D. A. (1997). *Matrix algebra from a statistician's perspective*. New York: Springer.

Holland, P. W. (1990). On the sampling theory foundations of item response theory models. *Psychometrika, 55,* 577-601.

Holland, P. W. & Rosenbaum, P. R. (1986). Conditional association and unidimensionality in monotone latent variable models. *The Annals of Statistics, 14,* 1523-1543.

Hooker, G. (2010). On separable tests, correlated priors, and paradoxical results in multidimensional item response theory. *Psychometrika, 75*, 694-707.

Hooker, G., Finkelman, M. & Schwartzman, A. (2009). Paradoxical results in multidimensional item response theory. *Psychometrika, 74*, 419-442.

Hooker, G. & Finkelman, M. (2010). Paradoxical results and item bundles. *Psychometrika, 75*, 249-271.

Karlin, S. & Rinott. Y. (1980a). Classes of orderings of measures and related correlation inequalities, I: Multivariate totally positive distributions. *Journal of Multivariate Analysis, 10,* 467-498.

Karlin, S. & Rinott, Y. (1980b). Classes of orderings of measures and related correlation inequalities, II: Multivariate reverse rule distributions. *Journal of Multivariate Analysis, 10,* 499-516.

Klein Entink, R. H., Kuhn, J.-T., Hornke, L. F. & Fox, J.-P. (2009). Evaluating cognitive theory: A joint modeling approach using responses and response times. *Psychological Methods*, 14, 54–75.

Lax, P. (2002). *Functional Analysis*. New-York: Wiley.

Lax, P. (2007). *Linear Algebra and its applications* (2nd ed.). New York: Wiley.

Le Cam, L. (1986) *Asymptotic Methods in Statistical Decision Theory*. New York: Springer.

Lee, S.Y. (2007). *Structural equation modeling: A Bayesian approach.* West Sussex, UK: Wiley.

Lehmann, E. L. (1966). Some concepts of dependence. *Annals of Mathematical Statistics, 37,* 1137-1153.

Lehmann, E. L. & Casella, G. (1998). *Theory of point estimation* (2nd ed.). New York: Springer.

Lehmann, E. L. & Romano, J. P. (2005). *Testing statistical hypotheses* (3rd ed.). New York: Springer.

Mardia, K. V., Kent, J. T. & Bibby, J. M. (1979). *Multivariate Analysis.* London: Academic Press.

Marsaglia, G. & Styan, G. P. H. (1974). Equalities and inequalities for ranks of matrices. *Linear and Multilinear Algebra, 2,* 269-292.

Marshall, A. W., Olkin, I. & Arnold, B. C. (2010). *Inequalities: Theory of majorization and its applications* (2nd ed.). New York: Springer.

Medvegyev, P. (2007). *Stochastic integration theory.* Oxford: Oxford University Press.

Ranger, J. M. (2009). *Der Nutzen von Reaktionszeiten bei psychologischen Tests im Rahmen von Item-Response-Modellen.* Dissertation, Universität Giessen.

Reckase, M. D. (2009). *Multidimensional item response theory.* New York: Springer.

Rockafellar, R. T. (1970). *Convex analysis.* Princeton, N.J.: Princeton University Press.

Rockafellar, R. T. & Wets, R. J.-B. (2009). *Variational Analysis* (corrected 3rd printing). Berlin: Springer.

Rosenbaum, P. R. (1984). Testing the conditional independence and monotonicity assumptions of item response theory. *Psychometrika, 49,* 425-435.

Rudin, W. (1976). *Principles of Mathematical Analysis* (3rd ed.). New York: McGraw-Hill.

Rudin, W. (1999). *Reelle und Komplexe Analysis.* München: Oldenbourg Wissenschaftsverlag GmbH.

Salgueiro, M. F., Smith, P.W.F. & McDonald, J.W. (2008). The manifest association structure of the single-factor model: insights from partial correlations. *Psychometrika, 73,* 665-670.

Searle, S. R., Casella, G. & McCulloch C. E. (2006). *Variance Components.* New York: Wiley.

Segall, D. O. (2010). Principles of multidimensional adaptive testing. In W. J. van der Linden & C. E. W. Glas (Eds.), Elements of adaptive testing, 57–76. New York: Springer.

Shapiro, A. (1982). Rank reducibility of a symmetric matrix and sampling theory of minimum trace factor analysis. *Psychometrika, 47,* 187-199.

Shea, G. (1979). Monotone regression and covariance structure. *Annals of Statistics, 7*, 1121-1126.

Sijtsma, K. & Molenaar, I.W. (2002). *Introduction to nonparametric item response theory.* Thousand Oaks, CA: Sage.

Suppes, P. & Zanotti, M. (1981). When are probabilistic explanations possible? *Synthese, 48*, 191-199.

Taylor, J. C. (1997). *An introduction to measure and probability.* New York: Springer.

Ten Berge, J. M. F. & Socan, G. (2004). The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality. *Psychometrika, 69,* 613-625.

Tutz, G. (1991). Sequential models in categorical regression. *Comput. Statist. Data Anal. 11,* 275-295.

van der Linden, W. J. (2007). A hierarchical framework for modeling speed and accuracy on test items. *Psychometrika, 72,* 287-308.

van der Linden, W. J. (2012). On compensation in multidimensional response modeling. *Psychometrika, 77.,* 21-30.

van der Linden, W. J. & Glas, C. A. W. (2010). Statistical tests of conditional independence between responses and response times on test items. *Psychometrika, 75,* 120-139.

van der Linden, W. J. & Guo, F. (2008). Bayesian procedures for identifying aberrant response-time patterns in adaptive testing. *Psychometrika, 73,* 365-384.

Zeger, S. L. & Liang, K. Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics 42*, 121-130.