

# Der Nutzen von Reaktionszeiten bei psychologischen Tests im Rahmen von Item Response Modellen

Inaugural-Dissertation  
zur  
Erlangung des Doktorgrades  
der Philosophie des Fachbereiches Psychologie  
der Justus-Liebig-Universität Gießen

vorgelegt von  
Jochen Matthias Ranger

aus  
Gießen

2009

Dekan: Prof. Joachim Brunstein  
1. Berichtstatter: Prof. Christof Schuster  
2. Berichtstatter: Prof. Heinz Holling

## Danksagung

Mein Dank gilt allen Personen, ohne deren Unterstützung diese Arbeit nicht möglich gewesen wäre. Ich danke vor allem Prof. Schuster für die Betreuung der Arbeit und die zahlreichen Anregungen. Ferner Prof. Holling für die Bereitschaft, die Begutachtung der Arbeit zu übernehmen. Ganz besonders danke ich meiner Ehefrau Leticia, die mir geduldig während der letzten drei Jahre zur Seite stand und meiner Mutter, die mich immer ermunterte.

## Zusammenfassung

Bei psychologischen Tests basiert die Messung von Eigenschaften (Intelligenz, Traits . . .) in der Regel auf einer Auswertung des Antwortmusters einer Person bei standardisierten Testaufgaben. Ignoriert werden die Reaktionszeiten, die eine Person für die Abgabe der Antworten benötigt. In letzter Zeit jedoch verdichten sich Hinweise darauf, dass auch die Reaktionszeiten in Beziehung zur Eigenschaftsausprägung einer Person stehen. In dieser Arbeit wird daher untersucht, wie die Reaktionszeiten neben den Itemantworten einer Person zur Messung der Eigenschaftsausprägung genutzt werden können und welche Vorteile sich durch die Berücksichtigung der Reaktionszeiten ergeben.

Analog zu den IRT-Modellen werden verschiedene Modelle für die Verteilung der Reaktionszeit bei Testitems aufgestellt. Grundgedanke der Modelle ist die Annahme, dass die Verteilung der Reaktionszeit bei einer Testaufgabe von Itemparametern abhängt, jedoch auch von zwei Persönlichkeitseigenschaften, der Personenfähigkeit und einem Residualfaktor. Unter Personenfähigkeit wird dabei ganz allgemein das Vermögen einer Person verstanden, auf ein Item mit einer positiven Antwort zu reagieren. Die genaue Form der Abhängigkeit zwischen der Reaktionszeit und den beiden Persönlichkeitseigenschaften wird in acht unterschiedlichen Modellen spezifiziert. Vier der Modelle gehören zur Klasse der Proportional Hazard Modelle, die restlichen vier Modelle zur Klasse der Accelerated Failure Time Modelle.

Von den vier Proportional Hazard Modellen basieren zwei Modelle auf der Exponentialverteilung, deren Verteilungsparameter über eine lineare oder eine quadratische Funktion auf die beiden Persönlichkeitseigenschaften zurückgeführt wird. Die beiden anderen Proportional Hazard Modelle sind Modelle für dichotomisierte Reaktionszeiten, die ohne eine genaue Verteilungsannahme auskommen. Bei diesen Modellen bestimmt eine lineare oder eine quadratische Funktion der beiden Persönlichkeitseigenschaften eine nicht näher spezifizierte Hazard-Funktion. Alternativ zu den vier Proportional Hazard Modellen werden vier Accelerated Failure Time Modelle vorgeschlagen. Von diesen vier Modellen gehen zwei Modelle von log-normalverteilten Reaktionszeiten aus. Dabei wird der Lokationsparameter der Log-Normalverteilung als lineare oder quadratische Funktion der beiden Persönlichkeitseigenschaften modelliert. Die beiden anderen Accelerated Failure Time Modelle machen keine Aussage über die genaue Form der Reaktionszeitverteilung. Angenommen wird lediglich, dass sich der Erwartungswert der logarithmierten Reaktionszeiten über eine lineare oder eine quadratische Funktion aus den beiden Persönlichkeitseigenschaften ergibt.

Insgesamt werden also in der Arbeit dieselben Reaktionszeitmodelle immer mit einer linearen und einer quadratischen Parametrisierung aufgestellt. Begründet werden kann diese Vorgehensweise damit, dass ein monotoner Einfluss der beiden Persönlichkeitseigenschaften auf die Verteilung der Reaktionszeit nur bei Leistungstests sinnvoll ist. Bei Einstellungs- oder Persönlichkeitstests erscheint in Anbetracht neuerer Befunde eine quadratische Parametrisierung angemessener zu sein.

Bei allen in dieser Arbeit untersuchten Reaktionszeitmodellen wird ein bestimmter Aspekt der Reaktionszeitverteilung auf zwei Persönlichkeitseigenschaften, die Personenfähigkeit und den Residualfaktor, zurückgeführt. Die Vorgehensweise ist dabei analog zu den IRT-Modellen, welche die Auftretenswahrscheinlichkeit einer positiven Itemantwort ebenfalls als Funktion einer Persönlichkeitseigenschaft modellieren. In der aktuellen Arbeit wird angenommen, dass die den Itemantworten zugrunde liegende Persönlichkeitseigenschaft mit der Personenfähigkeit des Reaktionszeitmodells identisch ist. Die Personenfähigkeit ist oft diejenige Größe, die bei diagnostischen Entscheidungen von Interesse ist. In der Arbeit wird daher der Frage nachgegangen, welchen Vorteil der zusätzliche Einbezug der Reaktionszeiten gegenüber einer alleinigen Verwendung der Itemantworten für die Schätzung der Personenfähigkeit einer Person besitzt. Der Schwerpunkt der Betrachtung liegt dabei auf dem Ausmaß, in dem sich die Varianz der Fähigkeitsschätzung durch den Einbezug der Reaktionszeit reduzieren lässt. Um diese Frage beantworten zu können, müssen Annahmen über die gemeinsame Verteilung der Itemantworten und Reaktionszeiten gemacht werden. Für die gemeinsame Verteilung der Itemantworten und Reaktionszeiten einer Person wird angenommen, dass diese Größen bei Bedingung auf die Personenfähigkeit und den Residualfaktor einer Person unabhängig voneinander verteilt sind. Es kann gezeigt werden, dass in diesem Fall die Varianz des Fähigkeitschätzers durch die Berücksichtigung der Reaktionszeiten zwangsläufig reduziert wird und dass die Reduktion bei der Schätzung extremer Fähigkeitsausprägungen groß ausfällt.

Für vier der vorgeschlagenen Modellierungsansätze, für die beiden Proportional Hazard Modelle für dichotomisierte Reaktionszeiten und für die beiden Accelerated Failure Time Modelle für log-normalverteilte Reaktionszeiten, werden Schätzverfahren entwickelt. Diese Verfahren ermöglichen es, simultan die Itemparameter des zweiparametrischen Logit-Modells und des verwendeten Reaktionszeitmodells zu schätzen. Die vorgeschlagenen Schätzverfahren basieren auf dem Marginal-Maximum-Likelihood-Ansatz. In Simulationen wird die Eignung der Schätzansätze bei Stichproben unterschiedlichen Umfangs untersucht. Es zeigt sich, dass für eine präzise Bestimmung der Itemparameter große Stichprobenumfänge notwendig sind. Die Schätzung der restlichen vier Modelle wird nicht separat behandelt. Es kann gezeigt werden, dass diese ebenfalls mit den vorgeschlagenen Verfahren an Beobachtungsdaten angepasst werden können.

Im Gegensatz zu vorliegenden Veröffentlichungen ist das Ziel dieser Arbeit die Entwicklung allgemeingültiger Verfahren, mit denen die Reaktionszeit bei der Testbearbeitung in die Messung der Persönlichkeitseigenschaft einbezogen werden kann. Besonderer Schwerpunkt liegt daher auf der Untersuchung von Reaktionszeitmodellen, die ohne konkreten Verteilungsannahmen auskommen. Hierzu werden als eine Neuerung der Arbeit erstmalig Proportional Hazard Modelle für diskrete Reaktionszeiten vorgeschlagen. Zudem werden unterschiedliche Beziehungen zwischen der zu messenden Persönlichkeitseigenschaft und der Reaktionszeitverteilung untersucht. Die quadratische Parametrisierung der Reaktionszeitmodelle stellt eine zweite Neuerung der Arbeit dar und ist speziell für Persönlichkeitstests gedacht, für die es derzeit nur wenige Reaktionszeitmodelle gibt. Durch die breite Herangehensweise soll es Forschern ermöglicht werden, Daten aus ganz unterschiedlichen Inhaltsbereichen modellieren zu können.

# Inhaltsverzeichnis

<b>1</b>	<b>Einführung</b>	<b>4</b>
1.1	Hinweise auf die Konstruktvalidität der RT . . . . .	4
1.2	Verwendung von RT in der Diagnostik . . . . .	8
1.3	Bewertung der Verwendung der RT . . . . .	11
1.4	Ziele der Arbeit . . . . .	12
<b>I</b>	<b>Modelle für Itemantworten und Reaktionszeiten</b>	<b>14</b>
<b>2</b>	<b>Modellbasiertes Messen</b>	<b>15</b>
<b>3</b>	<b>IRT-Modelle</b>	<b>22</b>
<b>4</b>	<b>RT-Modelle</b>	<b>29</b>
4.1	RT-Modelle auf Basis von Proportional Hazard Modellen . . . . .	33
4.1.1	Proportional Hazard Modelle für exponentialverteilte Reaktionszeiten . . . . .	34
4.1.2	Proportional Hazard Modelle für dichotomisierte Reaktionszeiten . . . . .	44
4.2	RT-Modelle auf Basis von Accelerated Failure Time Modellen . . . . .	54
4.2.1	Accelerated Failure Time Modelle für log-normalverteilte Reaktionszeiten . . . . .	55
4.2.2	Quasi-Likelihood-Ansätze für Accelerated Failure Time Modelle . . . . .	63
4.3	Diskussion . . . . .	68

<b>II</b>	<b>Gemeinsame Modellierung von Antwort und RT</b>	<b>74</b>
<b>5</b>	<b>Schätzung der PH-Modelle</b>	<b>78</b>
5.1	Lineare Parametrisierung . . . . .	78
5.2	Quadratische Parametrisierung . . . . .	89
5.3	Diskussion . . . . .	99
<b>6</b>	<b>Schätzung der AFT-Modelle</b>	<b>103</b>
6.1	Lineare Parametrisierung . . . . .	103
6.2	Quadratische Parametrisierung . . . . .	107
6.3	Diskussion . . . . .	123
<b>7</b>	<b>Beitrag der RT zur Fähigkeitsschätzung</b>	<b>125</b>
7.1	Nutzen der RT für die Schätzung der Fähigkeit . . . . .	126
7.2	Nutzen der RT für die Schätzung der Fähigkeit bei obliquen Faktorstruktur . . . . .	135
7.3	Diskussion . . . . .	146
<b>8</b>	<b>Diskussion der Ergebnisse</b>	<b>150</b>
8.1	Bewertung der Modellannahmen . . . . .	151
8.2	Bewertung des Nutzens für die diagnostische Praxis . . . . .	156
8.3	Ausblick . . . . .	158
<b>III</b>	<b>Anhang</b>	<b>159</b>
<b>A</b>	<b>Momentmethoden-Schätzer der RT-Modelle</b>	<b>160</b>
A.1	Modelle für exponentialverteilte Reaktionszeiten . . . . .	160
A.1.1	Lineare Parametrisierung . . . . .	160
A.1.2	Quadratische Parametrisierung . . . . .	162
A.2	Modelle für log-normalverteilte Reaktionszeiten . . . . .	164
A.2.1	Lineare Parametrisierung . . . . .	164
A.2.2	Quadratische Parametrisierung . . . . .	165

<b>B Informationsfunktionen der RT-Modelle</b>	<b>167</b>
B.1 Modelle für exponentialverteilte Reaktionszeiten . . . . .	167
B.1.1 Lineare Parametrisierung . . . . .	167
B.1.2 Quadratische Parametrisierung . . . . .	168
B.2 Modelle für dichotomisierte Reaktionszeiten . . . . .	169
B.2.1 Lineare Parametrisierung . . . . .	169
B.2.2 Quadratische Parametrisierung . . . . .	169
B.3 Modelle für log-normalverteilte Reaktionszeiten . . . . .	170
B.3.1 Lineare Parametrisierung . . . . .	170
B.3.2 Quadratische Parametrisierung . . . . .	171
<b>C Informationsgewinn durch RT</b>	<b>173</b>
C.1 Fisher-Information bei Kenntnis des Residualfaktors . . . . .	173
C.2 Testinformation bei Korrelation von Personenfähigkeit und Residualfaktor . . . . .	174
<b>Literaturverzeichnis</b>	<b>177</b>



# Kapitel 1

## Einführung

Die Verwendung standardisierter Tests besitzt in der psychologischen Diagnostik eine lange Tradition. Es ist heute weitgehend unumstritten, dass Antworten auf standardisierte Fragen quantitative Aussagen über den Grad individueller Eigenschaftsausprägungen erlauben. Weniger eindeutig fällt die Bewertung des diagnostischen Potentials der Reaktionszeiten (RT) bei psychologischen Tests aus: Einerseits finden sich viel versprechende Hinweise auf die diagnostische Bedeutung der Reaktionszeiten für das Erschließen von Persönlichkeitseigenschaften. Dies veranlasst Forscher immer wieder dazu, den Einbezug der Reaktionszeiten in die Diagnostik vorzuschlagen (Ferrando, 2006; Tyron & Mulloy, 1993; Mayerl, 2003). Andererseits gibt es kaum einen konkreten Versuch, Reaktionszeiten zur Diagnostik zu verwenden. Dies mag vermutlich an dem mit der Aufzeichnung von Reaktionszeiten verbundenen Aufwand liegen. Die Breite der Befunde, die für das diagnostische Potential der Reaktionszeiten bei Testitems sprechen, lassen eine Betrachtung der Reaktionszeiten jedoch durchaus lohnenswert erscheinen.

### 1.1 Hinweise auf die Konstruktvalidität der RT

Der Einsatz von Reaktionszeiten zur Eigenschaftsdiagnostik setzt deren Konstruktvalidität voraus. Tatsächlich finden sich für eine ganze Reihe von Anwendungsbereichen Hinweise darauf, dass die Reaktionszeiten bei der Bearbeitung psychologischer Tests als Indikator für die Ausprägung der zu messenden Eigenschaft dienen können.

#### **Kognitive Fähigkeiten**

Die Nutzbarkeit von Reaktionszeiten zur Messung kognitiver Fähigkeiten wurde schon früh propagiert (Berger, 1982). Einen ersten Fürsprecher findet sich 1908 in Galton, der mentale Fähigkeiten gar mit der Informationsverarbeitungsgeschwindigkeit bei Wahrnehmungsaufgaben gleichsetzte (Jensen, 1982). Diese

strenge Gleichsetzung von Intelligenz und Informationsverarbeitungsgeschwindigkeit wird jedoch von nachfolgenden Forschern aufgegeben. Thorndike, Bregman, Cobb und Woodyard (1927) plädierten beispielsweise für die Messung von Intelligenz auf drei Dimensionen, zu welchen sie neben der Geschwindigkeit auch die Akkuratheit eines Individuums zählten. Eine erste Unterscheidung zwischen Speed- und Power-Tests der geistigen Leistungsfähigkeit geht auf McCall zurück (McFarland, 1928).

Untersuchungen zur Reaktionszeit bei Power-Tests sprechen jedoch gegen eine vollkommene Unabhängigkeit von „Speed“ und „Power“. Mehreren Forschern gelang es, bei Power-Tests Beziehungen zwischen der Lösungsgeschwindigkeit und dem Lösungsvermögen nachzuweisen. Beispielsweise traten in Studien von Rafaeli und Tractinsky (1991), von Lavergne und Vigneau (1997) sowie von MacLennan, Jackson und Bellantino (1988) bei kognitiven Tests Korrelationen zwischen der mittleren Reaktionszeit und der Anzahl der richtig gelösten Aufgaben auf. Ähnliches berichtet Lachman (1993) für Psychologieexamen. Die Ergebnisse einer Untersuchung von Danthiir, Wilhelm und Schacht (2005) legen nahe, dass bei Intelligenztests die Bearbeitungszeit von richtig gelösten Aufgaben mit der individuellen Lösungswahrscheinlichkeit einer Person korreliert ist. Dieser Zusammenhang war bei kristallinen Intelligenzaufgaben besonders akzentuiert. Noch direkter für eine Beziehung zwischen der Fähigkeit einer Person und der Reaktionszeit bei Intelligenztestaufgaben sprechen die Studien von van der Linden (2007) sowie von van den Wollenberg und Creemers (1986). Diese Autoren konnten eine korrelative Beziehung zwischen der Fähigkeit einer Person (gemessen über ein Item-Response-Modell) und einer individuellen Messung der Bearbeitungsgeschwindigkeit belegen. Auch wenn nicht alle Untersuchungen derartige Beziehungen zwischen der Reaktionszeit und der Fähigkeit bei Intelligenztestaufgaben bestätigen – siehe die gegenteiligen Ergebnisse von Bontempo (2000), Neubauer (1990) oder Scrams und Schnipke (1997) – ist die Befundlage Anlass genug, Reaktionszeiten als Indikatoren für die Fähigkeit von Personen zu erwägen.

Allerdings scheint die Beziehung zwischen Reaktionszeit und Fähigkeit komplexer zu sein, als zuerst angenommen: Lohman (1989) vertritt die Position, dass die Itemschwierigkeit die Beziehung zwischen der Reaktionszeit und der Fähigkeit einer Person moderiert: Zwar können fähige Personen leichte Aufgaben schneller lösen als weniger fähige Personen. Bei schweren Aufgaben jedoch benötigten fähige Personen für die Enkodierung mehr Zeit. Diese Vermutung wird durch die Studie von Schaeffer, Reese, Steffen, McKinley und Mills (1993) gestützt. Sollte tatsächlich die Aufgabenschwierigkeit die Beziehung zwischen Reaktionszeit und Fähigkeit moderieren, ließen sich die Reaktionszeiten nur bei leichten Testaufgaben als Indikator für die Fähigkeit einer Person verwenden.

### **Persönlichkeitsmerkmale (Traits)**

In der psychologischen Diagnostik werden Reaktionszeiten bei der Bearbeitung von Persönlichkeitstests relativ selten als Indikator für die Ausprägung von Persönlichkeitsmerkmalen diskutiert. Diese Haltung ist überraschend, scheint doch bei Persönlichkeitstests der für die Beantwortung von Fragen notwendige Zeitbedarf mit der Ausprägung des zu messenden Persönlichkeitsmerkmals

in Beziehung zu stehen. Hierfür jedenfalls finden sich zahlreiche Belege. Zwei Forschungsrichtungen lassen sich ausmachen.

Die erste Forschungsrichtung geht von der Vorstellung aus, dass Personen und Persönlichkeitsitems auf einem Merkmalskontinuum angeordnet sind. Für die Reaktionszeit bei der Beantwortung eines Items ist die Distanz zwischen einer Person und der Lokation des Items bedeutsam: So soll die Reaktionszeit mit größer werdender Distanz zwischen Person und Item abnehmen. Tatsächlich lassen sich Belege dafür finden, dass mit zunehmender Distanz zwischen einer Person und einem Item auf der Merkmalsdimension die Reaktion auf das Item schneller erfolgt: Durchschnittliche Latenzzeiten fallen für extremere Items (Dunn, Lushene & O'Neil, 1972; Hanley, 1962; Kuncel, 1973; Rogers, 1973) und extremere Antwortkategorien (Casey & Tryon, 2001) kürzer aus. Auch Catanzaro (1997) berichtet von einem umgekehrt U-förmigen Zusammenhang zwischen einer Itemantwort und der dazu benötigten Reaktionszeit. Von Akrami, Hedlund und Ekehammar (2007) konnte nach der faktorenanalytischen Auswertung eines Persönlichkeitstests gezeigt werden, dass zwischen dem Faktorwert einer Person auf einer Skala und der mittleren Antwortzeit der Person bei den Items der Skala ein quadratischer Zusammenhang besteht: Personen mit extremen Faktorwerten reagierten schneller auf die Testitems als Personen mit mittlerem Faktorwert. Ähnliche Befunde finden sich auch für Item-Response-Modelle. So ist bei Persönlichkeitsfragebögen die über ein Modell der Item-Response-Theorie (IRT-Modell) bestimmte Distanz zwischen Item und Person mit der Antwortlatenz korreliert (Ferrando, 2006). Erste Versuche existieren, die Reaktionszeit als Funktion des Abstandes von Person und Item zu modellieren (Ferrando & Lorenzo-Sevas, 2007a, 2007b).

Eine andere Forschungsrichtung betont den Einfluss kognitiver Selbstschemata auf die Verarbeitung Schema-relevanter Information. Es konnte gezeigt werden, dass Personen mit ausgeprägten Selbstschemata schneller Schema-konsistenten Adjektiven zustimmen können (Kuiper, 1981). Weitere Studien stützen die Vorstellung einer Schema-gesteuerten Fragebogenbearbeitung. Eine Studie von Fekken und Holden (1992) belegt, dass Personen mit hoher Ausprägung auf einer Merkmalsdimension schneller zu Aussagen der zugehörigen Skala eines Persönlichkeitstests zustimmen als Personen mit geringer Ausprägung. Bei einer Ablehnung kehrt sich dieses Muster jedoch um. Die Autoren erklären dieses Phänomen damit, dass Schema-konsistente Information schneller verarbeitet werden kann als Schema-inkonsistente Information. Von derselben Forschergruppe stammen Veröffentlichungen, die sogar auf die konvergente und externe Validität der Reaktionszeiten hindeuten (Popham & Holden, 1990; Holden, Fekken & Cotton, 1991; Holden & Fekken, 1993). In diesen Studien war die Selbstbeurteilung einer Person auf einer Persönlichkeitsskala in derselben Weise mit der Skalenbearbeitungszeit verknüpft wie die Fremdbeurteilung der Person durch einen nahe stehenden Vertrauten. Eine korrelative Beziehung zwischen Reaktionszeit und Selbstbeurteilung trat jedoch nur auf, wenn Selbsturteil und Reaktionszeit von derselben Skala stammten. Wurde die Selbstbeurteilung auf einer Skala mit der Bearbeitungszeit einer anderen Skala korreliert, trat hingegen kein signifikanter Zusammenhang auf. Siem (1996) gelang es sogar, durch den Einbezug der Reaktionszeiten bei einem Fragebogen die Prognose eines Außenkriteriums zu verbessern. Diese Befunde sprechen dafür, dass Personen mit ausgeprägten Selbstschemata über leichter zugängliche und artikuliertere Vorstellungen dar-

über verfügen, was sie sind.

### **Einstellungen**

Im Bereich der Einstellungsmessung werden Reaktionszeiten als potentielle Indikatoren der Einstellungsausprägung erwogen.<sup>1</sup> Bekannt ist, dass Reaktionszeiten ansteigen, wenn Items dem Standpunkt einer Person sehr nahe kommen und die Person unsicher ist, ob sie die Aussage akzeptieren soll oder nicht (Kuncel, 1973). Zwischen dem Ausmaß der Zustimmung zu einer Aussage und der Reaktionszeit bis zur Abgabe des Urteils tritt eine umgekehrt U-förmige Beziehung auf (Mayerl, 2003). Zudem werden Reaktionszeiten als indikativ für die Einstellungsverfügbarkeit oder Einstellungsstärke angesehen (Bassili, 1996; Fazio, 1995; Mulligan, Grant, Mockabee & Monson, 2003).

### **Entscheidungsaufgaben**

Bei Wahlentscheidungen spiegeln Reaktionszeiten Eigenschaften des kognitiven Entscheidungsprozesses wider. Sind zwei Alternativen ähnlich attraktiv, steigt der mentale Aufwand und der Entscheidungsprozess verlängert sich (Böckenholt, Albert, Aschenbrenner & Schmalhofer, 1991). In Choice-Based-Conjoint-Experimenten lässt sich eine Beziehung zwischen der Reaktionszeit und der Auswahlwahrscheinlichkeit einer Alternative nachweisen (Haaijer, Kamakura & Wedel, 2000). Bei Signal-Detektions-Paradigmen kann gezeigt werden, dass die Reaktionszeit mit wachsender Distanz zwischen Reiz und Entscheidungskriterium abnimmt (Espinoza-Varas & Watson, 1994; Murdock, 1985). Dies deckt sich mit Studien aus der Allgemeinen Psychologie, in denen bei Kategorisierungsaufgaben sinkende Reaktionszeiten mit zunehmender Distanz zwischen Objekt und Kategoriegrenze auftreten – siehe hierzu beispielsweise die Studie von Maddox, Ashby und Gottlob (1998) beziehungsweise Ashby und Maddox (1994). Wird der Zeitbedarf für Urteile über die Ähnlichkeit physikalischer Objekte mit einer multidimensionalen Skalierung ausgewertet, ist die physikalische Ähnlichkeit der Objekte durch die Skalierung der Reaktionszeiten reproduzierbar (Takane & Sergent, 1983). Alle diese Befunde sprechen dafür, dass die für Objektvergleiche notwendige Reaktionszeit von der Distanz der Vergleichsobjekte auf der Vergleichsdimension abhängt. Interessant für die Persönlichkeitsdiagnostik werden die geschilderten Befunde, wenn man die Beantwortung von Fragebögen als Selbstkategorisierung auffasst, die auf einem Vergleich der eigenen Person mit einem Konzept oder einer Itemsaussage beruht. Große Ähnlichkeit zwischen Person und Konzept sollte dann mit einer längeren Reaktionszeit verbunden sein.

---

<sup>1</sup>Implizite Messungen von Einstellungen beruhen oft auf den Latenzzeiten bei Priming- und Assoziationsaufgaben (Fazio & Olson, 2003; Wittenbrink & Schwarz, 2007). Da jedoch diese Verfahren zur interindividuellen Diagnostik nur bedingt verwendbar sind (Cunningham, Preacher & Banaji, 2001), soll die Übersicht auf eine Diskussion der Rolle von Latenzzeiten bei Fragebögen beschränkt bleiben.

## Soziale Erwünschtheit

Die Beziehung zwischen der Reaktionszeit und dem Ausmaß, in dem Itemantworten auf dem Streben nach sozial erwünschten Antworten beruhen, wird seit nahezu 50 Jahren untersucht. Ursache dieses Forschungsinteresses war die Hoffnung, sozial erwünschte Antworten durch eine Aufforderung zu zügiger Fragebogenbearbeitung unterbinden zu können (Oseas, 1966). Tatsächlich sprechen die Befunde eher dafür, dass sozial erwünschte Antworten schneller erfolgen als wahre Antworten. Dies zeigen jedenfalls die Veröffentlichungen von Schneider und Hübner (1980), von Holden und Kroner (1992) sowie von Hsu, Santelli und Hsu (1989). Es existieren allerdings für den Fall offener Fragen auch gegenteilige Befunde (Harrison, Hwalek, Raney & Fritz, 1978; Kraut, 1978). Insgesamt ist die Frage, ob sozial erwünschte Antworten schneller oder langsamer als ehrliche Antworten gegeben werden, noch ungeklärt.

## 1.2 Verwendung von RT in der Diagnostik

Trotz dieser zahlreichen Hinweise auf die diagnostische Verwertbarkeit der Testbearbeitungszeiten spielen die Reaktionszeiten in der Praxis der psychologischen Diagnostik eine untergeordnete Rolle. Allenfalls in der Forschungsliteratur lassen sich einzelne Anregungen und Bemühungen entdecken, wie Reaktionszeiten für die Diagnostik zu Nutze gemacht werden könnten. Dabei sind Anwendungen bei der Diagnose von Einstellungen oder Traits von Anwendungen bei der Diagnose von Fähigkeiten zu unterscheiden.

### 1. Anwendungen bei der Diagnose von Einstellungen und Traits

#### Detektion von Antworttendenzen

Die Detektion sozial erwünschter Antworten ist ein wichtiges Thema in der Persönlichkeitsdiagnostik (Mummendey, 1995). Ein Kriterium zur Detektion sozial erwünschter Antworten könnte die Reaktionszeit bei der Beantwortung der Fragen sein. Tatsächlich lassen sich in einzelnen Studien verschiedene Bearbeitungsmodi („Fake Good“, „Neutral“, „Fake Bad“) anhand der Reaktionszeit unterscheiden (Holden & Kroner, 1992; Hsu et al., 1989). Allerdings ist der genaue Zusammenhang zwischen der Tendenz zu sozial erwünschten Antworten und der Reaktionszeit bei der Beantwortung von Persönlichkeitsitems noch ungeklärt.

#### Detektion von Non-Attitudes

Die Einstellungsstärke bzw. das Vorhandensein dauerhafter Einstellungen wird in der Einstellungsforschung als Moderator der Beziehung von Einstellung und Verhalten erwogen. Als Indikator für die Einstellungsstärke wird in der Einstellungsforschung oft die Reaktionszeit bei der Beurteilung von Aussagen benutzt (Bassili, 1996; Mayerl, 2003; Fazio, 1995). Es kann gezeigt werden, dass die durchschnittliche Reaktionszeit bei der Bearbeitung von Einstellungsfragebögen die Beziehung des Testwertes einer Person zu einem Außenkriterium moderiert und mit der Retestrelabilität assoziiert ist (Bassili, 1995; Mayerl, 2003).

### **Verbesserung der Prognose von Kriterien**

Als Kriterium für die Gültigkeit von Persönlichkeitstests wird häufig die Korrelation des Testwertes einer Person mit einer Fremdbewertung der Testperson durch einen nahe stehenden Vertrauten herangezogen. Diese „Vorhersage“ der Fremdbewertung lässt sich durch Berücksichtigung der für die Beantwortung der Persönlichkeitsitems benötigten Zeit verbessern: In einer Studie von Holden und Fekken (1993) konnte die Reaktionszeit selbst dann noch Varianz der Fremdbewertung aufklären, wenn bereits der Skalenscore als Prädiktor verwendet worden war. Allerdings sind nicht alle Befunde zur inkrementellen Validität der Reaktionszeit so günstig. Zwar gelang es Siem (1996), die Prognose eines Verhaltenskriteriums durch die zusätzliche Verwendung der Reaktionszeit bei der Testbearbeitung zu verbessern, nicht jedoch Stricker und Alderton (1999).

## **2. Anwendungen bei der Diagnose von Fähigkeiten**

### **Detektion von auffälligen Antwortmustern**

Manche Leistungstests besitzen nur geringe Konsequenzen für den Prüfling (diese Tests werden teilweise als „Low-Stakes-Tests“ bezeichnet). Fehlt die Motivation beim Prüfling, werden die Tests schlampig bearbeitet und repräsentieren unter Umständen nicht mehr das wahre Leistungsniveau einer Person. Item- und Personenparameter werden dann teilweise mit großer Verzerrung geschätzt (Schnipke, 1999). Dasselbe Problem ergibt sich, wenn Personen gegen Ende des Tests aus Zeitnot anfangen, die Antworten zu erraten. Diese schnelle Rateantworten lassen sich allerdings anhand verdächtig schneller Reaktionszeiten identifizieren, wie Studien von Wise und DeMars (2006), von van der Linden und van Krimpen-Stoop (2003), von Schnipke (1999) sowie von Kong, Wise und Bholá (2007) demonstrieren. Werden die schnellen Rateantworten anhand auffälliger Reaktionszeiten ausgemacht und ausgesondert, lässt sich die Testvalidität erhöhen und eine Überschätzung der Itemschwierigkeit vermeiden (Wise & DeMars, 2006; Schnipke, 1999; Cohen, Wollack, Bolt & Mroch, 2002; Goegebeur, De Boeck, Wollack & Cohen, 2007). Die Identifikation von schnellen Ratenantworten kann zudem dazu dienen, das Ausmaß der „Speededness“ eines Tests zu messen, also das Ausmaß, in welchem Umfang die Testleistung einer Person durch das Zeitlimit des Tests beeinträchtigt sein könnte (Bontempo, 2000).

### **Die Modellierung des „Speed-Accuracy-Trade-Offs“**

Mit dem Problem des „Speed-Accuracy-Trade-Offs“ wird das Phänomen bezeichnet, dass Individuen bei der Bearbeitung von Tests schnellere Reaktionszeiten auf Kosten der Akkuratheit erreichen können (Lohman, 1989). Die effektive Leistung einer Person (d.h. das aktuell erreichte Akkuratheitsniveau) ist somit vom Geschwindigkeitsniveau bei der Aufgabenbearbeitung und einem möglichen Zeitlimit für die Bearbeitung abhängig – Leistungsfähigkeit und Geschwindigkeit eines Individuums sind folglich in der effektiven Leistung einer Person konfundiert. Von verschiedenen Forschern wurde daher versucht, die effektive Leistung einer Person auf eine

Geschwindigkeitskomponente und eine Leistungsfähigkeitskomponente zurückzuführen (Roskam, 1987; Verhelst, Verstralen & Jansen, 1997; White, 1973) bzw. die Bearbeitungszeit in die Parameterschätzung einzubeziehen (Wang, 2006; van Breukelen, 2005; Jannarone, 1996).

### **Entwurf von Testmodellen für Speed-Tests**

Seit dem Aufkommen der Unterscheidung zwischen Power-Tests und Speed-Tests wurden mehrere Versuche unternommen, den bekannten IRT-Modellen für Power-Tests äquivalente Testmodelle für Speed-Tests zur Seite zu stellen (Rasch, 1960; Maris, 1993; Scheiblechner, 1979; Jansen, 1997). Kernbestandteil dieser Modelle für Speed-Tests ist die Rückführung von Parametern einer Reaktionszeitverteilung auf zwei Komponenten: Eine Komponente für die individuelle Bearbeitungsgeschwindigkeit und eine Komponente für die Aufgabenschwierigkeit.

### **Optimierung der Itemzusammenstellung**

Versuche, die durchschnittliche Bearbeitungszeit von Aufgaben bei der Zusammenstellung von Tests zu berücksichtigen, gehen bis in die 50er Jahre zurück (Ebel, 1953). Einen wesentlichen Impuls erhielt die Umsetzung dieses Gedankens durch die Entwicklung Computer adaptiver Tests. Bei Computer adaptiven Tests werden für Testpersonen individuelle Tests aus Aufgaben mit maximal diagnostischem Potential zusammengestellt. Ausgewählt werden dabei in der Regel nur Aufgaben mit einer für das Individuum angemessenen Schwierigkeit. Wenn zudem prognostizierbar ist, wie lange eine Person für die Bearbeitung einer Aufgabe benötigt, kann diese Information bei der Aufgabenauswahl berücksichtigt werden: Betrachtet werden dann nur alle Aufgabenzusammenstellungen, die ein Individuum in einer vorgegebenen Zeit zu lösen vermag. Ein derartiger Versuch, die prognostizierte Reaktionszeit eines Individuums bei der Testzusammenstellung zu berücksichtigen, wurde von van der Linden, Scrams und Schnipke (1999) unternommen.

### **Setzen von Zeitlimits zur Maximierung der Testinformation**

Bekannt ist, dass die Beschränkung der Testbearbeitungszeit eine Veränderung der Itemparameter nach sich zieht. Bei knapp bemessener Zeit steigt die Itemschwierigkeit an, jedoch verändern sich oft auch die Itemtrennschärfen (Osawa, 2005). Da die Information von Items<sup>2</sup> bei IRT-Modellen von der Trennschärfe und der Itemschwierigkeit abhängt, kann die Iteminformation durch das Setzen eines Zeitlimits beeinflusst werden. In der Tat konnte Osawa (2005) zeigen, dass sich die Information eines Tests durch die Wahl individuell angepasster Bearbeitungslimits erhöhen lässt.

### **Verbesserung der Prognose von Kriterien**

Fähigkeitstests werden in der diagnostischen Praxis eingesetzt, um zukünftige kognitive Leistungen vorherzusagen. Die Vorhersage eines derartigen Außenkriteriums lässt sich verbessern, wenn neben der Akkuratheit der

---

<sup>2</sup>Das Konzept der Iteminformation wird im dritten Kapitel ausführlich behandelt.

Testbearbeitung auch noch ein Maß der Bearbeitungsgeschwindigkeit als Prädiktor berücksichtigt wird (van den Wollenberg & Creemers, 1986). Diese Erkenntnis deckt sich mit äquivalenten Befunden aus dem Bereich der Traitdiagnostik.

### 1.3 Bewertung der Verwendung der RT

Bewertet man die Verwendungspraxis von Reaktionszeiten in der psychologischen Diagnostik, so lässt sich feststellen, dass Reaktionszeiten hauptsächlich in der Leistungsdiagnostik Anwendung finden. Aber auch in der Leistungsdiagnostik beschränkt sich die Nutzung von Reaktionszeiten auf wenige Anwendungsbereiche. Entweder geht es um die Kontrolle von Verzerrungen durch schnelle Rateantworten, die bei mangelnder Testmotivation oder extremen Zeitlimits auftreten können. Oder es ist beabsichtigt, die effektive Fähigkeit, wie sie mit konventionellen IRT-Modellen geschätzt wird, in zwei Komponenten aufzuspalten, einen Geschwindigkeits- und einen Leistungsfaktor. Eine Konfundierung von Leistungsfähigkeit und Bearbeitungsgeschwindigkeit tritt jedoch hauptsächlich bei Tests auf, die unter Zeitdruck durchgeführt werden, und lässt sich durch eine großzügige Bemessung der Testbearbeitungszeit einfach vermeiden. Werden Tests bei motivierten Probanden mit ausreichender Bearbeitungszeit durchgeführt, lassen somit nur wenige der beschriebenen Nutzungsmöglichkeiten von Reaktionszeiten einen Gewinn erwarten. Versuche, Reaktionszeiten auch im Falle optimaler diagnostischer Bedingungen (motivierte Probanden, keine Zeitlimits ...) zur Leistungsdiagnostik zu verwenden, finden sich kaum. Im Bereich der Diagnose von Einstellungen und Traits sind Versuche, die Reaktionszeiten zur Diagnostik zu nutzen, noch seltener zu finden.

Insgesamt steht die eingeschränkte Verwendung von Reaktionszeiten in der Persönlichkeitsdiagnostik im Widerspruch zu der oft geäußerten Hoffnung, dass durch die Berücksichtigung von Reaktionszeiten die Diagnostik allgemein verbessert werden kann. Tyron und Mulloy (1993, S. 231) bemerken:

„Computer measured response time provides information, that is methodologically independent from item content [. . .]. If such information is psychologically relevant, then the reliability and validity of objective personality tests could be improved by using both the scale scores and RT.“

Auch Akrami et al. (2007) sowie Ferrando und Lorenzo-Sevas (2007a) drücken die Hoffnung aus, dass der Einbezug von Reaktionszeiten die Messung von Eigenschaften unterstützen könne. Mayerl (2003) vermutet sogar, dass durch die zusätzliche Berücksichtigung der Reaktionszeit eine Verdopplung der Informationsmenge zu erzielen sei. Trotz dieser positiven Einschätzungen des diagnostischen Potentials der Reaktionszeiten wurde der genaue Beitrag der Reaktionszeiten zur Fähigkeits- oder Traitmessung bisher selten untersucht. Hornke (2000, S. 176) fasst den gegenwärtigen Forschungsstand zusammen:

„At present little is known about the possible diagnostic surplus of item response latencies, testing times in power test environments, and the like.“



Dieser Zustand ist umso überraschender, wenn man bedenkt, dass die psychologische Forschung an der Quantifizierung des Informationsgehalts von Tests durchaus Interesse hat. Dies zeigen Arbeiten zur Informationsstruktur von Tests (Birnbaum, 1968; Samejima, 1973; Penfield, 2005; Verkuilen, 2006; Luo & Andrich, 2005), Veröffentlichungen über die Information unterschiedlicher Antwortformate (Bock, 1972; Dodd & Koch, 1987, 1994; Donoghue, 1994), sowie Untersuchungen über den diagnostischen Nutzen zusätzlicher Aspekte der Antwort, wie zum Beispiel die Unterscheidung falscher Antworten (Thissen, 1976) oder die Erhöhung der Skalenabstufung von Urteilen (Samejima, 1969, S. 42). Der potentielle Informationsbeitrag der Reaktionszeiten wurde jedoch bisher weitgehend ignoriert.

## 1.4 Ziele der Arbeit

Die vorliegende Arbeit dient dem Ziel, den diagnostischen Nutzen von Reaktionszeiten bei der Bearbeitung von Tests zu untersuchen. Thema der Arbeit ist die Frage, wie derartige Reaktionszeiten genutzt werden können, um die Schätzung der mit dem Test gemessenen Persönlichkeitseigenschaft (Fähigkeiten, Traits, Einstellungen . . .) zu verbessern. Konkret geht es dabei um die Frage, in welchem Umfang die Varianz bei der Schätzung der Eigenschaftsausprägung einer Person durch die zusätzliche Berücksichtigung der Reaktionszeit reduziert werden kann. Die Überlegungen sollen dabei nicht an einen bestimmten Inhaltsbereich gebunden sein, auch wenn im Folgenden immer von der Personenfähigkeit einer Person gesprochen wird. Wie aus dieser Übersicht hervorgehen dürfte, scheinen bei ganz unterschiedlichen Inhaltsbereichen diagnostisch nutzbare Beziehungen zwischen der Reaktionszeit bei Tests und der Ausprägung der zu messenden Eigenschaft zu bestehen. Unter dem Begriff der Personenfähigkeit wird daher nicht die kognitive Leistungsfähigkeit einer Person verstanden, sondern die „Fähigkeit“, auf ein Testitem mit einer positiven Antwort zu reagieren. Dies kann bei Leistungstests die kognitive Fähigkeit sein, eine Aufgabe richtig zu lösen, bei Einstellungstests die Neigung einer Person, einer Aussage zuzustimmen und bei Persönlichkeitstests die Tendenz, eine vorgegebene Beschreibung der eigenen Person zu bejahen.

Die Arbeit gliedert sich in zwei Teile. Der erste Teil der Arbeit umfasst die Kapitel zwei bis vier, in denen die Grundlagen für die weiteren Überlegungen gelegt werden. Im zweiten Kapitel wird das Prinzip des modellbasierten Messens vorgestellt. Grundlage des modellbasierten Messens ist ein Modell, das beschreibt, wie beobachtbares Verhalten von einer zugrunde liegenden Persönlichkeitseigenschaft abhängt. Dieses Modell kann dann genutzt werden, um vom beobachtbaren Verhalten auf die Ausprägung der Persönlichkeitseigenschaft zurückzuschließen. Beispiele für modellbasiertes Messen sind die IRT-Modelle, mit denen die Auftretenswahrscheinlichkeit von Itemantworten auf die Fähigkeit einer Person zurückgeführt wird. Im dritten Kapitel wird ein populärer Vertreter der IRT-Modelle, das zweiparametrische Logit-Modell, vorgestellt. Dort wird auch gezeigt, wie mit diesem Modell die Fähigkeitsausprägung einer Person bestimmt werden kann. Modelle zur Beschreibung der Reaktionszeitverteilung bei Testitems finden sich im vierten Kapitel. Grundlage der Modelle ist die Rückführung der Reaktionszeitverteilung auf die Personenfähigkeit eines Individuums und

eine zweite Eigenschaft, den so genannten Residualfaktor einer Person. Diese Reaktionszeitmodelle können ebenfalls dazu genutzt werden, die Fähigkeitsausprägung einer Person zu bestimmen. Das vierte Kapitel schließt den ersten Teil der Arbeit ab.

Der zweite Teil der Arbeit umfasst die Kapiteln fünf bis acht. Im fünften und sechsten Kapitel werden Modelle für die gemeinsame Verteilung von Itemantworten und Reaktionszeiten bei Tests entwickelt. Dabei wird auf die bereits vorgestellten IRT- und Reaktionszeitmodelle zurückgegriffen. Insgesamt werden vier Modelle näher betrachtet. Für jedes der vorgeschlagenen Modelle werden Methoden beschrieben, wie diese an reale Daten angepasst werden können. Diese Methoden basieren auf der Marginal-Maximum-Likelihood-Schätzung. Die Eignung der Methoden wird in Simulationsstudien erprobt. Im siebten Kapitel erfolgt dann eine Auseinandersetzung mit der eigentlichen Fragestellung der Arbeit, der Frage nach dem Nutzen der Reaktionszeiten für die Diagnostik der Fähigkeit einer Person. Untersucht wird dabei schwerpunktmäßig, in welchem Umfang die Berücksichtigung der Reaktionszeiten die Präzision bei der Fähigkeitsschätzung verbessern können. Es wird sich dabei zeigen, dass die Reaktionszeit selbst bei relativ schwachen Annahmen zur Reduktion der Varianz des Fähigkeitsschätzers genutzt werden kann. Eine Bewertung der Ergebnisse des zweiten Teils der Arbeit schließt sich im achten Kapitel an.

## Teil I

# Modelle für Itemantworten und Reaktionszeiten

## Kapitel 2

# Modellbasiertes Messen in der Psychologie

Die Messung von Persönlichkeitseigenschaften ist ein wichtiger Bestandteil der psychologischen Diagnostik. Trotz der Notwendigkeit zu bedeutsamen Vergleichen von Personen sind der Nachweis und die Begründung von Skaleneigenschaften bei psychologischen Messungen oft schwierig. Tatsächlich gründeten sich die ersten Messungen von Leistung und Einstellung auf den Glauben an ihre Angemessenheit (Molenaar, 1995, S. 4). Eine wesentliche Verbesserung dieser unbefriedigenden Situation wurde durch die Einführung des modellbasierten Messens erreicht (Embretson & Reise, 2000, S. 13). Grundidee des modellbasierten Messens ist die Aufstellung eines Modells, das beobachtbares Verhalten auf eine dem manifesten Verhalten zugrunde liegende Persönlichkeitseigenschaft zurückführt. Dieses Modell besteht aus einer quantitativen Beschreibung der Beziehung zwischen einer Verhaltensbeobachtung  $x$  und der Ausprägung der Persönlichkeitseigenschaft  $\theta$  bei einer Person (Cliff & Keats, 2003, S. 16). Bei der Verhaltensbeobachtung handelt es sich in der Regel um die Antwort auf eine standardisierte Frage eines psychologischen Tests, die bei den IRT-Modellen mit der Fähigkeit einer Person in Beziehung gesetzt wird. Es kann aber zum Beispiel auch die Reaktionszeit einer Person modelliert werden. In der Regel besteht das Messmodell aus einer stochastischen Beziehung der Form

$$x = \varphi(\theta) + e, \tag{2.1}$$

wobei  $x$  die manifeste Beobachtung ist,  $\theta$  die Ausprägung der latenten Persönlichkeitseigenschaft einer Person und  $e$  ein zufälliger Effekt, der nicht mit der Personenvariablen  $\theta$  assoziiert ist (Lord & Novick, 1968, S. 24) und einen Erwartungswert von Null besitzt. Die Funktion  $\varphi(\theta)$  beschreibt dann die Abhängigkeit des manifesten Verhaltens von der latenten Persönlichkeitseigenschaft. Je nach Detailliertheitsgrad des Modells wird die genaue Gestalt der Regressionsfunktion  $\varphi(\theta)$  und die Verteilungsform des zufälligen Residualterms  $e$  offen gelassen oder genau spezifiziert. Legt man die Verteilung des zufälligen Residualterms fest, lässt sich die Verteilung der manifesten Beobachtung  $x$  angeben.

### Eigenschaftsschätzung beim modellbasierten Messen

Testmodelle führen Aspekte des beobachtbaren Verhaltens einer Person auf eine hypothetische, latente Persönlichkeitseigenschaft zurück. Die Abhängigkeit des Verhaltens von der latenten Persönlichkeitseigenschaft wird dabei mit einer quantitativen Beziehung beschrieben. Diese Gesetzmäßigkeit kann ausgenutzt werden, um mit Hilfe der Beobachtungen einer Person die Ausprägung der Persönlichkeitseigenschaft zu schätzen. Eine Methode der Schätzung ist die Methode der Maximum-Likelihood-Schätzung (Baker & Seock-Ho, 2004, Kap. 3), bei der einer Testperson die angesichts ihrer gezeigten Verhaltensdaten „plausibelste“ Eigenschaftsausprägung zugeschrieben wird.

Wird eine Person unter  $G$  verschiedenen Bedingungen wiederholt getestet, fällt der Beobachtungsvektor  $\mathbf{x} = [x_1, \dots, x_G]$  an. Da die latente Eigenschaft  $\theta$  als alleinige systematische Ursache des manifesten Verhaltens betrachtet wird, gelten die verschiedenen Verhaltensbeobachtungen  $x_1$  bis  $x_G$  bei Bedingung auf die Persönlichkeitseigenschaft  $\theta$  als voneinander unabhängig. Dies kann mit dem „Common Cause Criterion“ begründet werden, welches besagt, dass bei Berücksichtigung der gemeinsamen Ursachen der Zusammenhang zweier Phänomene vollständig erklärt wird (Suppes & Zanotti, 1981). Aufgrund der bedingten Unabhängigkeit der Verhaltensweisen einer Person kann die Auftretenswahrscheinlichkeit eines Verhaltensmusters  $\mathbf{x} = [x_1, \dots, x_G]$  beschrieben werden als:

$$f(\mathbf{x}; \theta) = \prod_{g=1}^G f(x_g; \theta). \quad (2.2)$$

Die Funktion (2.2) wird auch als Likelihood-Funktion von  $\theta$  bezeichnet, da die Funktion  $f(\mathbf{x}; \theta)$  bei gegebenem Verhaltensmuster  $\mathbf{x}$  für jede Eigenschaftsausprägung  $\theta$  angibt, wie plausibel das Verhaltensmuster bei der Eigenschaftsausprägung auftreten würde. Auf dieser Likelihood-Funktion basiert das Prinzip der Maximum-Likelihood-Schätzung der Persönlichkeitseigenschaft  $\theta$ .

Bei der Maximum-Likelihood-Schätzung wird als Schätzer der Eigenschaftsausprägung diejenige Ausprägung  $\hat{\theta}$  gewählt, welche für ein gegebenes Verhaltensmuster die Likelihood-Funktion maximiert:

$$\hat{\theta} = \arg \max_{\theta} \left[ \prod_{g=1}^G f(x_g; \theta) \right]. \quad (2.3)$$

Bei der Maximum-Likelihood-Schätzung besitzt die Eigenschaft einer Person somit den Status eines fixen Verteilungsparameters. Es ist bekannt, dass die Maximum-Likelihood-Schätzung bei wachsender Anzahl von Beobachtungen zu einer konsistenten Schätzung führt, wenn bestimmte Regularitätsannahmen erfüllt sind (Lehmann, 1983). Da die Regularitätsannahmen für alle in der Arbeit behandelten Modelle erfüllt sind, soll auf eine umfassende Darstellung verzichtet werden. Für Maximum-Likelihood-Schätzer kann die asymptotische Varianz berechnet werden. Diese Größe ist von besonderer Bedeutung, da sie Auskunft über die Präzision gibt, mit welcher die Eigenschaftsausprägung einer Person gemessen werden kann.

### Die Präzision der Eigenschaftsschätzung im Rahmen der ML-Theorie

Bei psychologischen Messungen entspricht die geschätzte Eigenschaftsausprägung  $\hat{\theta}$  in der Regel nicht der wahren Eigenschaftsausprägung  $\theta$  einer Person, so dass jede Messung mit einem Standardfehler behaftet ist. Die fehlerhafte Messung ist mit einem Verlust verbunden. Der Verlust, der sich bei einem bestimmten Messfehler ergibt, wird durch die so genannte Verlustfunktion definiert, welche jedem Messfehler einen bestimmten Verlust zuordnet. Oft wird in der statistischen Schätztheorie eine quadratische Verlustfunktion betrachtet (Casella & Berger, 2002, S. 349). Der erwartete quadratische Verlust, der so genannte „Mean Squared Error“, bei der Messung eines bestimmten Eigenschaftsniveaus  $\theta$  hängt von dem Bias und der Varianz des Schätzers  $\hat{\theta}$  ab:

$$\begin{aligned} \text{MSE} &= \text{Var} + \text{Bias}^2 \\ E[(\hat{\theta} - \theta)^2] &= \sigma^2 [\hat{\theta}] + [E(\hat{\theta}) - \theta]^2. \end{aligned} \quad (2.4)$$

Die asymptotische Verteilung eines Maximum-Likelihood-Schätzers folgt einer Normalverteilung mit dem Erwartungswert  $E(\hat{\theta}) = \theta$ . Maximum-Likelihood-Schätzer sind also asymptotisch unbiased (Casella & Berger, 2002, S. 472). Bei großen Stichproben ist daher die Varianz des Eigenschaftsschätzers  $\hat{\theta}$  von entscheidender Bedeutung. Die Varianz eines Maximum-Likelihood-Schätzers kann für die asymptotische Verteilung des Schätzers bestimmt werden. Sie hängt von der so genannten Fisher-Information ab. Die Fisher-Information ist eine Größe aus der mathematischen Statistik, die für einen Parameter einer Verteilungsdichte berechnet werden kann (Casella & Berger, 2002). Die Fisher-Information lässt sich über die so genannte Score-Funktion der Zufallsvariablen definieren. Die Score-Funktion einer Zufallsvariablen  $x$ , deren Verteilung durch die parametrisierte Wahrscheinlichkeitsdichte  $f(x; \theta)$  beschrieben wird, lautet allgemein (Casella & Berger, 2002):

$$S(\theta, x) = \frac{\partial}{\partial \theta} \log f(x; \theta). \quad (2.5)$$

Die Varianz der Score-Funktion wird als Fisher-Information der Zufallsvariablen über den Parameter  $\theta$  bezeichnet. Da der Erwartungswert der Score-Funktion Null ergibt, gilt für die Varianz der Score-Funktion bzw. die Fisher-Information:

$$I(\theta) = \sigma^2(S(\theta, x)) = E \left[ \left( \frac{\partial}{\partial \theta} \log f(x; \theta) \right)^2 \right]. \quad (2.6)$$

Alternativ lässt sich die Fisher-Information bei Gültigkeit bestimmter Regularitätsbedingungen berechnen als:

$$I(\theta) = -E \left[ \frac{\partial^2}{\partial \theta^2} \log f(x; \theta) \right]. \quad (2.7)$$

Das Maß von Fisher ist bei unabhängigen Zufallsvariablen additiv. Sind  $G$  Beobachtungen  $x_1, x_2, \dots, x_G$  unabhängig verteilt, ergibt sich die Gesamtinformation der Beobachtungen über  $\theta$  als Summe der Fisher-Informationen der einzelnen Beobachtungen (Sorensen & Gianola, 2002, S. 129):

$$I(\theta) = \sum_{g=1}^G I_g(\theta) \quad (2.8)$$

$$\text{mit } I_g(\theta) = E \left[ \left( \frac{\partial}{\partial \theta} \log f(x_g; \theta) \right)^2 \right].$$

Die Bedeutung der Fisher-Information resultiert unter anderem daraus, dass mit Hilfe der Fisher-Information die asymptotische Varianz des Maximum-Likelihood-Schätzers  $\hat{\theta}$  berechnet werden kann. Dies leitet sich aus der Cramer-Rao-Ungleichung ab. Es lässt sich zeigen, dass für einen konsistenten und asymptotisch normalverteilten Schätzer  $\hat{\theta}$  des Parameters  $\theta$  die Fisher-Information bei Gültigkeit bestimmter Regularitätsbedingungen in Bezug zur Untergrenze der Varianz des Schätzers steht (Casella & Berger, 2002):

$$\sigma^2(\hat{\theta}) \geq \frac{1}{I(\theta)}. \quad (2.9)$$

Die Untergrenze wird von Maximum-Likelihood-Schätzern mit zunehmender Zahl von Beobachtungen erreicht, Maximum-Likelihood-Schätzer sind asymptotisch effizient (Casella & Berger, 2002, S. 472). Verwendet man in Ungleichung (2.9) die Fisher-Information, die sich bei dem Messmodell (2.2) ergibt, kann die Beziehung ausgenutzt werden, um die Varianz des Eigenschaftsschätzers bei einem Test zu bestimmen. Auch wenn bei Tests endlicher Länge die asymptotische Varianz durch den Maximum-Likelihood-Schätzer nicht zwangsläufig erreicht wird und bei kurzen Tests die tatsächliche Varianz von der über Beziehung (2.9) berechneten, asymptotischen Varianz sogar erheblich abweichen kann, wird in den weiteren Kapiteln dennoch die Varianz des Schätzers mit der inversen Fisher-Information gleichgesetzt. Dies kann damit begründet werden, dass bereits bei Tests mittlerer Länge in der Regel der Unterschied zwischen der tatsächlichen Varianz und der reziproken Fisher-Information gering ausfällt.

Die Gesetzesmäßigkeit (2.9) über die asymptotische Varianz des Maximum-Likelihood-Schätzers kann auf den Fall mehrerer zu schätzender Parameter erweitert werden. Dadurch lässt sich die Beziehung auf ein zweidimensionales Messmodell anwenden. Bei zweidimensionalen Messmodellen wird die Verteilung einer beobachtbaren Verhaltensweise  $x$  mit zwei Persönlichkeitseigenschaften  $\theta$  und  $\omega$  und der zweiparametrischen Verteilung  $f(x; \theta, \omega)$  erklärt. Auch in diesem Fall hängt die asymptotische Varianz der Maximum-Likelihood-Schätzer von  $\theta$  und  $\omega$  von der Fisher-Information ab. Die Fisher-Information lässt sich im zweiparametrischen Fall ebenfalls über die Score-Funktion definieren.

Für einen Parametervektor ist die Score-Funktion ein Vektor. Bei einer zweiparametrischen Verteilungsdichte  $f(x; \theta, \omega)$  ist die Score-Funktion beispielsweise:

$$\mathbf{S}(\theta, \omega, x) = [S_1(\theta, \omega, x), S_2(\theta, \omega, x)]'. \quad (2.10)$$

Die Elemente des Vektors sind dabei die partiellen Ableitungen der logarithmierten Likelihoodfunktion nach den beiden Persönlichkeitseigenschaften  $\theta$  und  $\omega$ . In diesem Fall ist die Varianz-Kovarianz-Matrix der Elemente der Score-Funktion die Informationsmatrix  $\mathbf{I}(\theta, \omega)$ . Da der Erwartungswert der Score-Funktion ein Nullvektor ist, lässt sich die Varianz-Kovarianz-Matrix der Score-Funktion bzw. die Informationsmatrix schreiben als:

$$\begin{aligned} \mathbf{I}(\theta, \omega) &= E[\mathbf{S}(\theta, \omega, x)\mathbf{S}(\theta, \omega, x)'] \\ &= E \begin{bmatrix} S_1(\theta, \omega, x)^2 & S_1(\theta, \omega, x)S_2(\theta, \omega, x) \\ S_1(\theta, \omega, x)S_2(\theta, \omega, x) & S_2(\theta, \omega, x)^2 \end{bmatrix}. \end{aligned} \quad (2.11)$$

Alternativ lässt sich die Informationsmatrix über den Erwartungswert der Hesse-Matrix der Log-Likelihood-Funktion bei Ableitung nach den Personenparametern bestimmen. Es gilt:

$$\mathbf{I}(\theta, \omega) = -E \begin{bmatrix} \frac{\partial}{\partial \theta^2} \log f(x; \theta, \omega) & \frac{\partial}{\partial \theta \partial \omega} \log f(x; \theta, \omega) \\ \frac{\partial}{\partial \theta \partial \omega} \log f(x; \theta, \omega) & \frac{\partial}{\partial \omega^2} \log f(x; \theta, \omega) \end{bmatrix}. \quad (2.12)$$

Bei unabhängigen Beobachtungen können die Fisher-Informationen mehrerer Zufallsvariablen  $x_1, x_2, \dots, x_G$  zur Fisher-Information des Tests addiert werden:

$$\mathbf{I}(\theta, \omega) = \sum_{g=1}^G \mathbf{I}_g(\theta, \omega) \quad (2.13)$$

$$\text{mit } \mathbf{I}_g(\theta, \omega) = E[\mathbf{S}(\theta, \omega, x_g)\mathbf{S}(\theta, \omega, x_g)'].$$

Die Inverse der Testinformation (2.13) stellt eine Untergrenze für die Varianz-Kovarianz-Matrix von (asymptotisch) erwartungstreuen Schätzern der Parameter  $\theta$  und  $\omega$  dar, die von Maximum-Likelihood-Schätzern mit steigender Beobachtungszahl erreicht wird. Bei großer Anzahl von Beobachtungen kann daher die Inverse der Informationsmatrix mit der Kovarianzmatrix der Maximum-Likelihood-Schätzer praktisch gleichgesetzt werden. Die Beziehung von Elementen der Informationsmatrix zur asymptotischen Varianz der Parameterschätzer ist jedoch komplexer als im einparametrischen Fall. Wird die Informationsmatrix (2.13) symbolisch geschrieben als

$$\mathbf{I}(\theta, \omega) = \begin{bmatrix} I_{11} & I_{12} \\ I_{12} & I_{22} \end{bmatrix}, \quad (2.14)$$

so lautet die Inverse der Informationsmatrix



$$\mathbf{I}^{-1}(\theta, \omega) = \begin{bmatrix} \frac{I_{22}}{\Delta} & -\frac{I_{12}}{\Delta} \\ -\frac{I_{12}}{\Delta} & \frac{I_{11}}{\Delta} \end{bmatrix}, \quad (2.15)$$

wobei  $\Delta = I_{11}I_{22} - I_{12}^2$  die Determinante der Informationsmatrix (2.14) bezeichnet. Es kann gezeigt werden (Sorensen & Gianola, 2002, S. 183), dass im zweiparametrischen Fall für die asymptotische Varianz des Maximum-Likelihood-Schätzers  $\hat{\theta}$  folgende Beziehung gilt:

$$\sigma^2(\hat{\theta}) = \frac{I_{22}}{\Delta} = \frac{1}{I_{11} - I_{12}^2/I_{22}} \geq I_{11}^{-1} = \frac{1}{I(\theta)}. \quad (2.16)$$

In Gleichung (2.16) entspricht  $1/I(\theta)$  der Varianz des Schätzers von  $\theta$ , die man nach Gleichung (2.9) erhielte, wenn die Ausprägung des Parameters  $\omega$  bekannt wäre und nur der Parameter  $\theta$  geschätzt werden müsste. Aus der Ungleichung (2.16) geht hervor, dass die Varianz des Maximum-Likelihood-Schätzers von  $\theta$  größer ist, wenn zusätzlich der Parameter  $\omega$  geschätzt werden muss – die Erhöhung der Dimensionalität führt also zu einer Verringerung der Präzision der Schätzung. Das Element  $I_{11}$  der Testinformationsmatrix (2.14) besitzt in der Testtheorie eine hervorgehobene Bedeutung, auch wenn es im zweiparametrischen Falle nicht mehr direkt in Bezug zur asymptotischen Varianz des Schätzers  $\hat{\theta}$  steht. Gleichung (2.16) kann entnommen werden, dass ein hoher Wert von  $I_{11}$  eine notwendige Bedingung für eine geringe Varianz des Schätzers  $\hat{\theta}$  ist. Aus diesem Grund kann eine grobe Abschätzung der Varianz von  $\hat{\theta}$  bereits alleine anhand dieses Elementes vorgenommen werden. Element  $I_{11}$  der Fisher-Information des Tests addiert sich aus den Informationsbeiträgen der einzelnen Zufallsvariablen (siehe 2.13). Dies bedeutet, dass die Berücksichtigung einer Beobachtung nur dann die Varianz des Eigenschaftsschätzers reduzieren kann, wenn die Beobachtung mit einem hohen Wert zu  $I_{11}$  beiträgt. Diese Eigenschaft kann zur Itemselektion ausgenutzt werden.

### Zusammenfassung

Modellbasiertes Messen basiert auf der Annahme eines mathematischen Modells, das Aspekte des beobachtbaren Verhaltens auf eine latente Persönlichkeitseigenschaft zurückführt. Wird durch das Modell die genaue Verteilung des beobachtbaren Verhaltens beschrieben, kann die Ausprägung der Persönlichkeitseigenschaft anhand des beobachteten Verhaltens über die Maximum-Likelihood-Methode geschätzt werden. Die Varianz des Schätzers lässt sich dann über die Fisher-Information berechnen. Die Bestimmung der Fisher-Information ist somit ein wichtiger Zwischenschritt bei der Berechnung der (asymptotischen) Varianz von Schätzern. Da die Fisher-Information in inverser Beziehung zur Varianz steht, können manche Fragestellungen zur Varianz von Schätzern bereits mit Hilfe der Fisher-Information beantwortet werden. Dies mag ein Grund dafür sein, dass im Bereich der psychologischen Diagnostik die Fisher-Information eine hervorgehobene Bedeutung besitzt, die weit über den Status eines Zwischenergebnisses hinausgeht. Aus diesem Grund werden in der aktuellen Arbeit neben der Varianz von Schätzern immer auch die einzelnen Beiträge der einer

Schätzung zugrunde liegenden Beobachtungen zur Fisher-Information des Tests betrachtet.

Möchte man die Reaktionszeit zur Schätzung der Personenfähigkeit eines Individuums verwenden, sind nach der Vorgehensweise des modellbasierten Messens Reaktionszeitmodelle zu formulieren, welche die Reaktionszeit bei der Bearbeitung von Testaufgaben auf die unbekannte Ausprägung der Personenfähigkeit des Individuums zurückführen. Solche Reaktionszeitmodelle sollen im Folgenden eingeführt werden. Über die Fisher-Information der Reaktionszeitmodelle kann dann bestimmt werden, wie genau die Personenfähigkeit mit Hilfe der Modelle geschätzt werden kann. Bevor jedoch Reaktionszeitmodelle beschrieben werden, wird zuerst ein bekanntes Messmodell für binäre Itemantworten, das so genannte zweiparametrische Logit-Modell, dargestellt.

## Kapitel 3

# IRT-Modelle und ihre Fisher-Information

Die Item-Response-Theorie beschäftigt sich mit Messmodellen für die Itemantworten bei standardisierten Tests. Während die ersten IRT-Modelle für ein binäres Antwortformat aufgestellt wurden (Rasch, 1960; Birnbaum, 1968), liegt mittlerweile eine Vielzahl von Modellen für mehrkategoriale oder ordinale Antworten vor. Für eine Übersicht siehe zum Beispiel Fischer und Molenaar (1995) oder van der Linden und Hambleton (1997). Eines dieser Modelle, das zweiparametrische Logit-Modell, soll im folgenden Abschnitt näher vorgestellt werden.

### Das zweiparametrische Logit-Modell

Kernbestandteil von IRT-Modellen ist die Beschreibung einer quantitativen Beziehung zwischen der Wahrscheinlichkeit einer Itemantwort, einer Persönlichkeitseigenschaft und mehreren Itemeigenschaften, welche in der Item-Response-Funktion oder auch Itemcharakteristik ausgedrückt wird (Molenaar, 1995). Das zweiparametrische Logit-Modell modelliert die Wahrscheinlichkeit einer positiven Antwort bei Item  $g$  als Funktion der Aufgabenschwierigkeit  $b_g$ , einer aufgabenspezifischen Trennschärfe  $a_g$  und der Personenfähigkeit  $\theta$  mit Hilfe der logistischen Verteilung (Birnbaum, 1968). Die Itemcharakteristik des zweiparametrischen Logit-Modells lautet:

$$P_g(\theta) = \frac{\exp [a_g(\theta - b_g)]}{1 + \exp [a_g(\theta - b_g)]}. \quad (3.1)$$

Abbildung 3.1 visualisiert den Einfluss der Personenfähigkeit  $\theta$  auf die Wahrscheinlichkeit einer positiven Antwort bei Aufgabe  $g$  und verdeutlicht die Bedeutung unterschiedlicher Trennschärfekoeffizienten  $a_g$ . Die Itemschwierigkeit  $b_g$  bestimmt die Lage der Kurve, so dass auf eine Demonstration des Effektes unterschiedlicher Itemschwierigkeiten verzichtet werden kann.

Um die Wahrscheinlichkeit eines Antwortmusters einer Person bei mehreren Aufgaben zu berechnen, wird in der Regel angenommen, dass bei Bedingung auf

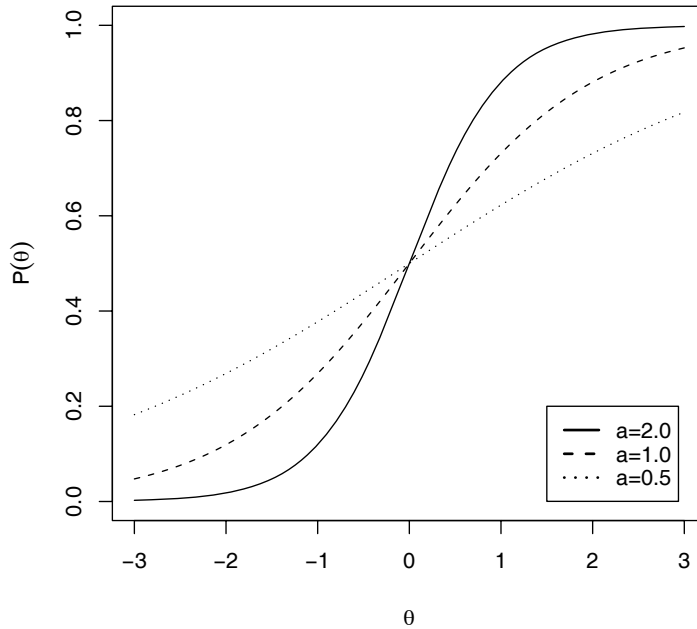


Abbildung 3.1: Itemcharakteristik des zweiparametrischen Logit-Modells mit Itemschwierigkeit  $b_g = 0$  und verschiedenen Trennschärfekoeffizienten  $a_g$  (siehe Gleichung 3.1)

die Personenfähigkeit  $\theta$  die Itemantworten verschiedener Aufgaben voneinander unabhängig sind (lokale stochastische Unabhängigkeit). Die gemeinsame Auftretenswahrscheinlichkeit der verschiedenen Itemantworten einer Person ergibt sich dann einfach als Produkt der Einzelwahrscheinlichkeiten.

Das zweiparameterige Logit-Modell wurde ursprünglich für Leistungstests entwickelt. Der Anwendungsbereich ist jedoch nicht alleine auf den Leistungsbereich (Intelligenztests, Aufnahmeprüfungen ...) beschränkt, sondern umfasst auch den Bereich der Einstellungs- und Traitmessung. Modelliert wird in diesem Fall die Wahrscheinlichkeit einer Zustimmung zu einer Aussage. Aus diesem Grund sollte unter dem Begriff der Personenfähigkeit nicht die kognitive Fähigkeit einer Person verstanden werden, sondern die Fähigkeit, auf ein Item mit einer positiven Antwort zu reagieren.

### Die Schätzung der Personenfähigkeit

Über die Itemcharakteristik wird die Auftretenswahrscheinlichkeit von Itemantworten als Funktion der Personenfähigkeit  $\theta$  und weiteren Itemparametern angegeben. Diese Beziehung kann ausgenutzt werden, um anhand eines bei einer Testperson beobachteten Antwortmusters die Ausprägung der Personenfähig-

keit zu schätzen. Eine Methode der Schätzung ist die Maximum-Likelihood-Schätzung, deren allgemeines Prinzip bereits im zweiten Kapitel dargestellt wurde. Nach dem zweiparametrischen Logit-Modell lautet die Auftretenswahrscheinlichkeit eines Antwortmusters  $\mathbf{x} = [x_1, \dots, x_G]$  bei den  $G$  Items eines Tests unter der Annahme der bedingten Unabhängigkeit

$$P(\theta) = \prod_{g=1}^G P_g(\theta)^{x_g} (1 - P_g(\theta))^{1-x_g}, \quad (3.2)$$

wobei  $P_g(\theta)$  die in Gleichung (3.1) beschriebene Itemcharakteristik des  $g$ -ten Items ist und  $x_g = 1$  eine positive Antwort und  $x_g = 0$  eine negative Antwort kodiert. Setzt man die Itemparameter als bekannt voraus, hängt die Auftretenswahrscheinlichkeit eines Antwortmusters nur von der Fähigkeit einer Person ab. Als Schätzer der Personenfähigkeit wird diejenige Ausprägung  $\hat{\theta}$  gewählt, welche für ein gegebenes Antwortmuster die Likelihood-Funktion maximiert:

$$\hat{\theta} = \arg \max_{\theta} \left[ \prod_{g=1}^G P_g(\theta)^{x_g} (1 - P_g(\theta))^{1-x_g} \right]. \quad (3.3)$$

Die Vorgehensweise bei der Bestimmung des Maximums ist in Baker und Seock-Ho (2004) beschrieben. Bei Anwendung der im zweiten Kapitel dargestellten Beziehungen kann die Varianz des Schätzers berechnet werden. Diese Größe ist von besonderer Bedeutung, da sie Auskunft über den erwarteten (quadrierten) Fehler gibt, der mit der Fähigkeitsschätzung verbunden ist (siehe Gleichung 2.4). Die Bestimmung der Varianz des Fähigkeitsschätzers erfordert die Berechnung der Fisher-Information des Tests. Diese ergibt sich für einen Test aus  $G$  Items aufgrund der Unabhängigkeit der Itemantworten einer Person nach Gleichung (2.8) als:

$$I(\theta) = \sum_{g=1}^G I_g(\theta). \quad (3.4)$$

Der Beitrag der  $g$ -ten Aufgabe zur Testinformation, die so genannte Iteminformation<sup>1</sup>  $I_g(\theta)$ , lässt sich bei binärem Antwortformat nach Baker und Seock-Ho (2004) berechnen als:

$$I_g(\theta) = \frac{\left[ \frac{\partial}{\partial \theta} P_g(\theta) \right]^2}{P_g(\theta)(1 - P_g(\theta))}. \quad (3.5)$$

Die genaue Form der Iteminformation hängt von der angenommenen Itemcharakteristik  $P_g(\theta)$  ab. Beim zweiparametrischen Logit-Modell (siehe Gleichung 3.1) lautet die Iteminformation:

---

<sup>1</sup>Die Fisher-Information einer Itemantwort wird in der Testtheorie auch als Iteminformation bezeichnet.

$$I_g(\theta) = a_g^2 \frac{\exp[a_g(\theta - b_g)]}{[1 + \exp[a_g(\theta - b_g)]]^2} = a_g^2 P_g(\theta)(1 - P_g(\theta)). \quad (3.6)$$

Die Fisher-Information eines Items besitzt in der Testtheorie eine hervorgehobene Bedeutung und wird dort unter anderem zur Itemselektion verwendet. Der Verlauf der Iteminformation über den Bereich der Personenfähigkeit ist in Abbildung 3.2 dargestellt. Die Iteminformationsfunktion ist unimodal und besitzt ein Maximum von  $(1/4)a_g^2$  im Punkt  $b_g$ . Die Schwierigkeit  $b_g$  bestimmt lediglich die Lage der Kurve, während die Trennschärfe  $a_g$  die Steilheit und den Funktionswert im Maximum der Kurve festlegt. Mit zunehmender Distanz zur Itemschwierigkeit  $b_g$  fällt die Iteminformation schnell ab.

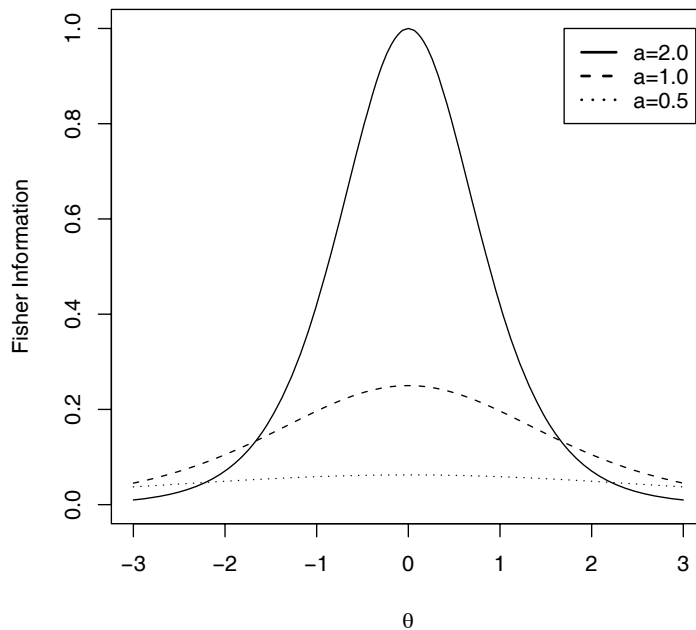


Abbildung 3.2: Iteminformationsfunktion des zweiparametrischen Logit-Modells für Items mit Itemschwierigkeit  $b_g = 0$  und verschiedenen Trennschärfekoeffizienten  $a_g$  (siehe Gleichung 3.6)

Innerhalb der Testtheorie wird der Fisher-Information eines Items oder der Fisher-Information des ganzen Tests großes Interesse geschenkt – siehe zum Beispiel die Artikel von Verkuilen (2006), Luo und Andrich (2005) oder Bickel, Buyske, Chang und Ying (2001), die ausschließlich diesen Größen gewidmet sind. Das große Interesse mag daran liegen, dass die Iteminformation eine Größe ist, die zur Itemselektion eingesetzt werden kann. Geht es jedoch um die Abschätzung des Standardfehlers der Fähigkeitsschätzung, ist die für einen Maximum-Likelihood-Schätzer relevante Größe die reziproke Fisher-Information des Tests:

$$I^{-1}(\theta) = \frac{1}{\sum_{g=1}^G I_g(\theta)}. \quad (3.7)$$

Der Beitrag eines Items zur Reduktion der Varianz des Fähigkeitsschätzers steht daher in umgekehrt monotoner Beziehung zu dessen Fischer-Information. Je höher die Iteminformation eines Items ist, desto stärker nimmt die Varianz des Fähigkeitsschätzers  $\hat{\theta}$  ab, wenn das Item in den Test aufgenommen wird. In Abbildung 3.2 ist zu erkennen, dass die Iteminformation von der wahren Fähigkeitsausprägung abhängt und mit wachsender oder sinkender Fähigkeitsausprägung abfällt. Aus diesem Verhalten der Iteminformation kann geschlossen werden, dass beim zweiparametrischen Logit-Modell die Varianzreduktion am größten ausfällt, wenn Itemschwierigkeit  $b_g$  und Fähigkeitsausprägung  $\theta$  übereinstimmen. Mit zunehmendem Abstand der Fähigkeitsausprägung  $\theta$  zur Itemschwierigkeit  $b_g$  leistet ein Item nur einen geringen Beitrag zur Reduktion der Varianz des Fähigkeitsschätzers. Items besitzen also nur in einem begrenzten Bereich um ihre Itemschwierigkeit diagnostischen Nutzen.

### Randbemerkungen

Die Iteminformation gilt als wichtige Größe in der Testtheorie. Die Bedeutung leitet sich daraus ab, dass die Informationskurven der Items eines Tests bestimmen, mit welcher Präzision bestimmte Ausprägungen der Personenfähigkeit gemessen werden können. Betrachtet man den Verlauf der Iteminformation des zweiparametrischen Logit-Modells, so stellt man fest, dass die Iteminformation mit steigender Distanz zur Itemlokation  $b_g$  abfällt, so dass Items nicht in allen Bereichen genau so gut messen.

Dieser nicht optimale Verlauf der Iteminformation ist dem binären Format der Itemantworten geschuldet. Dies zeigen folgende Überlegungen: Das zweiparametrische Logit-Modell lässt sich derart interpretieren, dass eine Aufgabe dann positiv beantwortet wird, wenn eine unbeobachtbare, kontinuierlich verteilte Zufallsvariable  $x^*$  eine Schwelle überschreitet, siehe Takane und de Leeuw (1987) für eine analoge Begründung des zweiparametrischen Probit-Modells. Durch Beobachtung der dichotomen Itemantwort steht somit lediglich fest, ob die Ausprägung der latenten Variablen über oder unter dem Schwellenwert liegt, ohne dass die genaue Ausprägung bekannt ist. Das zweiparametrische Logit-Modell ergibt sich, wenn man als latente Variable  $x^*$  eine logistisch verteilte Zufallsvariable verwendet, deren Erwartungswert auf die Personenfähigkeit  $\theta$  zurückgeführt werden kann. Angenommen wird, dass die latente Variable  $x^*$  einer Person den personenspezifischen Erwartungswert  $\mu(x^*|\theta; a^*, b) = a^*(\theta - b)$  und die Varianz  $\pi^2\sigma^2/3$  besitzt. Bei den Parametern  $a^*$  und  $b$  handelt es sich um Itemparameter. Die Verteilungsdichte der latenten Zufallsvariablen einer Person bei einem bestimmten Item lautet dann:

$$f(x^*|\theta; a^*, b, \sigma) = \frac{1}{\sigma} \frac{\exp\left[\frac{x^* - a^*(\theta - b)}{\sigma}\right]}{\left[1 + \exp\left[\frac{x^* - a^*(\theta - b)}{\sigma}\right]\right]^2}. \quad (3.8)$$

Bei jeder Konfrontation von einer Person mit einer Aufgabe nimmt die latente Zufallsvariable  $x^*$  eine bestimmte Realisation an. Sobald die Ausprägung von  $x^*$  die Schwelle Null übersteigt, beantwortet eine Person die Aufgabe positiv. Die Wahrscheinlichkeit einer positiven Antwort berechnet sich für eine Person mit einem bestimmten Fähigkeitsniveau folglich als:

$$\begin{aligned} P(\theta) &= \int_0^\infty f(x^*|\theta; a^*, b, \sigma) dx^* = \int_{-\frac{a^*(\theta-b)}{\sigma}}^\infty \frac{\exp[u]}{[1 + \exp[u]]^2} du \\ &= \left[ \frac{\exp[u]}{1 + \exp[u]} \right]_{-\frac{a^*(\theta-b)}{\sigma}}^\infty = \frac{\exp\left[\frac{a^*}{\sigma}(\theta - b)\right]}{1 + \exp\left[\frac{a^*}{\sigma}(\theta - b)\right]}. \end{aligned} \quad (3.9)$$

Nach Definition von  $a$  als  $a^*/\sigma$  resultiert die bekannte Item-Charakteristik (3.1) des zweiparametrischen Logit-Modells:

$$P(\theta) = \frac{\exp[a(\theta - b)]}{1 + \exp[a(\theta - b)]}. \quad (3.10)$$

Aussagen über die Iteminformation beziehen sich immer auf die Itemcharakteristik (3.10), also auf die Iteminformation, welche die dichotome Itemantwort über den Personenparameter  $\theta$  besitzt. Es stellt sich die Frage, welche Information über  $\theta$  in der un beobachteten, kontinuierlichen Variablen  $x^*$  steckt. Dies ist natürlich nur eine theoretische Frage, da die genaue Ausprägung der Zufallsvariablen  $x^*$  nicht beobachtbar ist. Die Fisher-Information der Zufallsvariablen  $x^*$  ist nach (2.7) definiert als:

$$I(\theta) = -E \left[ \frac{\partial^2}{\partial \theta^2} \log f(x^*|\theta; a^*, b, \sigma) \right]. \quad (3.11)$$

Nach Einsetzen der Wahrscheinlichkeit (3.9) ergeben weitere Umformungen die Iteminformationsfunktion:

$$\begin{aligned} I(\theta) &= 2 \frac{a^{*2}}{\sigma^2} \int_{-\infty}^\infty \frac{1}{\sigma} \frac{\left[ \exp\left[\frac{x^* - a^*(\theta-b)}{\sigma}\right] \right]^2}{\left[ 1 + \exp\left[\frac{x^* - a^*(\theta-b)}{\sigma}\right] \right]^4} dx^* \\ &= 2 \frac{a^{*2}}{\sigma^2} \int_0^\infty \frac{u}{[1 + u]^4} du = \frac{1}{3} \frac{a^{*2}}{\sigma^2} = \frac{1}{3} a^2. \end{aligned} \quad (3.12)$$

Anhand von Gleichung (3.12) kann man sehen, dass die Iteminformation höher ausfällt, wenn die kontinuierliche, nicht-dichotomisierte Variable beobachtet werden könnte. Die Iteminformation von Item  $g$  ist bei Kenntnis der latenten Variablen konstant  $1/3a_g^2$ . Gegenüber der Iteminformation des zweiparametrischen Logit-Modells bei dichotomen Antworten stellt dies ein Informationsgewinn dar: Das Maximum der Informationsfunktion des zweiparametrischen Logit-Modells tritt an der Stelle  $\theta = b_g$  auf und beträgt lediglich den Wert  $1/4a_g^2$ .



Feststellen lässt sich somit Folgendes: Optimalerweise würde nicht eine dichotomisierte Variable beobachtet, sondern eine Variable, die „vollständige“ Information über den Abstand der Person zum Item besäße. Eine derartige Variable könnte sich in der Reaktionszeit bei der Bearbeitung der Items finden. Um aber die Reaktionszeiten einer Person zur Diagnostik nutzen zu können, müssen zuerst Testmodelle für die Reaktionszeiten entwickelt werden. Dies ist Thema des nächsten Kapitels.

## Kapitel 4

# RT-Modelle und ihre Fisher-Information

Item-Response-Modelle geben an, wie die Wahrscheinlichkeit einer positiven Antwort von der Fähigkeit einer Person abhängt. Auch wenn seit den Anfängen der Item-Response-Theorie eine Vielzahl unterschiedlicher IRT-Modelle vorgeschlagen wurde, beschränkt sich doch die Anzahl der gängigen IRT-Modelle auf eine Handvoll.<sup>1</sup> Den IRT-Modellen analoge Versuche, die Reaktionszeiten für die Beantwortung von Testaufgaben auf Eigenschaften der Testperson und des Items zurückzuführen, wurden bereits in den 50er Jahren begonnen (Furieux, 1952). Mittlerweile liegen einige Arbeiten mit sehr unterschiedlichen Vorstellungen über das angemessene Reaktionszeitmodell vor. Die Unterschiede beginnen bereits bei der zu wählenden Reaktionszeitverteilung: Als mögliche Verteilung wurden die Exponentialverteilung (Scheiblechner, 1979), die Weibull-Verteilung (Tatsuoka & Tatsuoka, 1978; Roskam, 1997; Wang, 2006), die Gamma-Verteilung (Jansen, 1997; Maris, 1993; Verhelst et al., 1997; Rasch, 1960), die Log-Normalverteilung (van Breukelen, 2005; Schnipke & Scrams, 1997; van der Linden, 2006; Thissen, 1984; Ferrando & Lorenzo-Sevas, 2007a), die Inverse Gauß-Verteilung (Pashley, 1987) oder die Poisson-Erlang-Verteilung (Pieters & van der Ven, 1982) vorgeschlagen. Auch bezüglich der Modellierung der Parameter der Reaktionszeitverteilung durch Personen- und Itemeigenschaften findet sich nur wenig Übereinstimmung: Zwar gehen viele Modelle von abnehmenden Reaktionszeiten mit der Dominanz einer Person über ein Item aus, was zumindest für Intelligenztests eine sehr plausible Annahme ist, jedoch finden sich auch Ausnahmen (Wang, 2006; Ferrando & Lorenzo-Sevas, 2007a, 2007b). Die Heterogenität der Ansätze dürfte insgesamt der Unterschiedlichkeit der Anwendungsbereiche der Modelle geschuldet sein: Die Anwendungsbereiche der einzelnen Modelle erstrecken sich von der Modellierung der Reaktionszeiten basaler Aufgaben (wie zum Beispiel mentaler Rotationsaufgaben) über die Modellierung der Reaktionszeiten bei Konzentrationstests bis hin zur Modellierung der Reaktionszeiten beim Lösen komplexer Testaufgaben. Vermutlich kann kein allgemeingültiges

---

<sup>1</sup>Zumindest wenn man die in den Standardlehrbüchern behandelten IRT-Modelle zu den gängigen IRT-Modellen rechnet.

Standardmodell gefunden werden, welches diesen breiten Inhaltsbereich abzudecken vermag.

Auch sind die Modelle, die für die Reaktionszeit zu finden sind, losgelöst von Annahmen über kognitive Prozesse: Präzise Theorien über den Lösungsprozess, aus denen sich die gemeinsame Verteilung von Reaktionszeit und Itemantwort ableiten ließe, sind nicht existent. Hier unterscheidet sich die Vorgehensweise in der Testtheorie von der Allgemeinen Psychologie, die sich unter anderem mit präzisen Modellen für basale Wahrnehmungsaufgaben beschäftigt, siehe zum Beispiel Luce (1991). Insgesamt ist der Forschungsstand bezüglich der Modellierung von Reaktionszeiten bei Tests wenig zufrieden stellend.

Im Folgenden sollen eigene Testmodelle aufgestellt werden, die zwar teilweise auf publizierten Modellen aufbauen, doch größeren Allgemeinheitsgrad besitzen. Die vorgestellten Reaktionszeitmodelle stehen in der Tradition der Item-Response-Theorie der psychologischen Diagnostik, die nicht auf expliziten Annahmen über die den Antworten zugrunde liegenden Prozessen basiert (Eysenck, 1967). Ansatzpunkte der Modelle sind folgende Überlegungen und Bestrebungen:

### **Rückführung der RT auf Personen- und Itemparameter**

Die Reaktionszeitverteilung wird als Funktion zweier Persönlichkeitseigenschaften und dreier Itemeigenschaften modelliert. Angenommen wird, dass die Reaktionszeitverteilung bei einem Item von der Personenfähigkeit  $\theta$  einer Person abhängt. Diese Annahme wird dadurch gestützt, dass zahlreiche empirische Studien für eine Beziehung zwischen den Reaktionszeiten und der zu messenden Eigenschaft bei Tests sprechen – siehe hierzu die Übersicht im ersten Kapitel (S. 4ff). Zusätzlich wird ein zweiter personenspezifischer Einflussfaktor  $\omega$  angenommen, der als Residualfaktor unabhängig von der Personenfähigkeit einen Einfluss auf die Reaktionszeitverteilung ausübt. Der Residualfaktor fasst alle über die Personenfähigkeit hinausgehenden, systematischen Einflüsse der Person auf die Reaktionszeitverteilung zusammen. Die Einführung eines zweiten Faktors scheint deshalb notwendig, da empirische Belege dafür sprechen, dass Reaktionszeiten neben einem inhaltspezifischen Faktor noch auf einem weiteren Faktor laden. Es ist beispielsweise bekannt, dass auch Persönlichkeitseigenschaften wie Extraversion (Danthiir et al., 2005), Lesefähigkeit (Jansen, 1997), Impulsivität (Vigil & Codorniu, 2004) oder die Motivation einer Person (Callegaro, Yang, Bholá & Dillman, 2005) die Reaktionszeit bei Testitems beeinflussen. Die Gesamtheit dieser Einflüsse wird daher in dem skalenspezifischen Faktor  $\omega$  zusammengefasst. Die Annahme eines zusätzlichen systematischen Personeneinflusses begründet letztendlich auch die in der Reaktionszeitforschung gängige Praxis der doppelten Standardisierung – siehe zum Beispiel Holden et al. (1991) – bei der „inhaltsfremde“ systematische Persönlichkeitseinflüsse aus den Reaktionszeiten entfernt werden. Auch wenn beide Persönlichkeitseigenschaften die Reaktionszeitverteilung gleichberechtigt beeinflussen, gilt das Hauptaugenmerk der Betrachtung der Personenfähigkeit. Dies liegt daran, dass mit einem Test oft nur die Personenfähigkeit diagnostiziert werden soll.

Stabile Itemunterschiede in der Bearbeitungszeit, die beispielsweise auf der Itemlänge oder der Frageformulierung beruhen mögen – für eine Übersicht über

derartige Einflussquellen siehe z.B. Ferrando und Demestre (2007) – sollen in dem Itemparameter  $k$  (Konstante) widerspiegelt werden. Zu beachten ist allerdings, dass die Konstante  $k$  zwar die Regulation des allgemeinen Reaktionszeitniveaus bei Items übernimmt, jedoch zur mittleren Reaktionszeit nur einen indirekten Bezug besitzt und bei den verschiedenen Modellen unterschiedliches bedeutet. Aus diesem Grund kann die Konstante sowohl positive als auch negative Werte annehmen. Neben dem Itemparameter  $k$  wird pro Item für jede der beiden Persönlichkeitseigenschaften  $\theta$  und  $\omega$  ein Einflussgewicht angenommen, welches die Stärke des Einflusses der Persönlichkeitseigenschaften auf die Reaktionszeitverteilung bestimmt. Dieses Einflussgewicht wird bei der Personenfähigkeit als  $l$  und bei dem Residualfaktor als  $c$  bezeichnet. Da je nach vorgeschlagenem Reaktionszeitmodell unterschiedliche Aspekte der Reaktionszeitverteilung modelliert werden, nehmen die Einflussgewichte bei einigen Reaktionszeitmodellen positive, bei anderen Reaktionszeitmodellen negative Werte an. Bei einem Teil der Modelle wird zudem ein weiterer Itemparameter, die Itemlokation  $d$ , eingeführt. Die Funktion der Itemlokation  $d$  ist die Anordnung eines Items auf der Fähigkeitsdimension. Die genaue Bedeutung von  $d$  wird im nächsten Abschnitt verdeutlicht. Eine Übersicht über die verwendeten Bezeichnungen findet sich in Tabelle 4.1.

Tabelle 4.1: Übersicht über die Parameter der RT-Modelle

PARAMETER	SYMBOL	BEDEUTUNG	WERTEBEREICH
Person	$\theta$	Personenfähigkeit	$(-\infty; \infty)$
	$\omega$	Residualfaktor	$(-\infty; \infty)$
Item	$k$	Konstante	$(-\infty; \infty)$
	$l$	Einflussgewicht von $\theta$	$(-\infty; \infty)$
	$c$	Einflussgewicht von $\omega$	$(-\infty; \infty)$
	$d$	Itemlokation	$(-\infty; \infty)$

### Funktionale Verknüpfung von Personen- und Itemparametern

Zwei funktionale Verknüpfungen der Item- und Personenparameter zu einem Prädiktor  $\eta$  werden untersucht. Zuerst eine lineare Verknüpfung der Form  $\eta = k + l\theta + c\omega$ . Diese Verknüpfung ist analog zur Modellierung bei den IRT-Modellen zu sehen. Ihr liegt die Idee zugrunde, dass die Verteilung der Reaktionszeit vom Ausmaß beeinflusst wird, in welchem eine Person eine Aufgabe zu „dominieren“ vermag. Bei diesen Modellen bedeuten zunehmende Ausprägungen der Persönlichkeitseigenschaften abnehmende Reaktionszeiten, sofern entsprechende Vorzeichen für die Itemparameter gewählt werden. In ihrer Struktur entspricht die lineare Parametrisierung einem linearen zweidimensionalen Faktormodell. Der Anwendungsbereich dieser Modellierung ist vermutlich hauptsächlich in der Leistungsdiagnostik zu sehen. Viele der bisher publizierten RT-Modelle gehen von einer derartigen Beziehung aus, siehe zum Beispiel Furneaux (1952), Thissen (1984) oder Scrams und Schnipke (1997). Alternativ zur linearen Verknüpfung wird eine quadratische Verknüpfung der Form  $\eta = k + l(\theta - d)^2 + c\omega$  untersucht. Bei der Interpretation der quadratischen Parametrisierung ist es hilfreich, den

Personenparameter  $\theta$  als Lokation einer Person und den Itemparameter  $d$  als Lokation eines Items auf der Fähigkeitsdimension zu betrachten. Die quadratische Verknüpfung bedeutet dann, dass die Reaktionszeit von der Distanz zwischen Item und Person auf der Fähigkeitsdimension beeinflusst wird. Diese Art der Beziehung zielt auf die Modellierung des „Inverted-U“-Effektes ab. Mit dieser Bezeichnung ist der Befund gemeint, dass bei einer Ambivalenz von Aussagen (geringe Distanz zwischen Item und Person) die Reaktionszeiten groß ausfallen, jedoch mit zunehmender Distanz von Item und Person abnehmen. Dieses Abhängigkeitsmuster schlägt sich in einem umgekehrt quadratischen Zusammenhang zwischen mittlerer Reaktionszeit und Abstand von Item und Person nieder. Erste Hinweise auf eine derartige Unfolding-Beziehung zwischen Eigenschaftsausprägung und Latenzzeit liegen vor (Akrami et al., 2007; Ferrando & Lorenzo-Sevas, 2007a). Anwendungsmöglichkeiten dieser Modellierung dürften hauptsächlich bei Tests für Einstellungen und Persönlichkeitsmerkmale zu finden sein. Die Entwicklung quadratischer „Unfolding-Modelle“ für Reaktionszeiten ist eine der Neuerung dieser Arbeit.

### Art des RT-Modells

Die funktionalen Verknüpfungen von Personen- und Itemparametern zu einem Prädiktor werden verwendet, um bestimmte Parameter oder Eigenschaften von Reaktionszeitverteilungen zu modellieren. Welche Eigenschaft jedoch modelliert wird, hängt von dem gewählten Reaktionszeitmodell ab. Insgesamt werden vier Verteilungsmodelle vorgeschlagen, bei denen jeweils beide Parametrisierungen (linear/quadratisch) Anwendung finden. Betrachtet man die bisherige Literatur zur Modellierung von Reaktionszeiten bei Tests, stellt man fest, dass kaum Bezüge zur Analyse von „Survival Times“ hergestellt werden. Alle publizierten Modelle beruhen auf der Annahme einer bestimmten Verteilungsform der Reaktionszeiten und stellen oft so genannte „Accelerated Failure Time“ Modelle (AFT-Modelle) dar. Ein Versuch, das populäre „Proportional Hazard“ Modell (PH-Modelle) von Cox (1972) aus der Überlebenszeitanalyse auf den Bereich der Testbearbeitungszeit zu übertragen, wurde hingegen noch nicht unternommen. Diese Lücke soll in der aktuellen Arbeit geschlossen werden. Zuerst werden zwei einfache Proportional Hazard Modelle vorgestellt, die von exponentialverteilten Reaktionszeiten ausgehen. Anschließend wird gezeigt, wie Proportional Hazard Modelle ohne die Annahme einer bestimmten Reaktionszeitverteilung aufgestellt werden können, wenn statt der kontinuierlichen Reaktionszeiten dichotomisierte Zeiten betrachtet werden. Dies stellt eine weitere Neuerung der Arbeit dar. Eine zweite populäre Modellklasse zur Analyse von Überlebenszeiten sind die Accelerated Failure Time Modelle. Aus diesem Grund werden zuerst zwei Accelerated Failure Time Modelle vorgeschlagen, die auf log-normalverteilten Reaktionszeiten basieren. Wie auch bei den Proportional Hazard Modellen lassen sich Accelerated Failure Time Modelle ohne Annahme einer bestimmten Reaktionszeitverteilung formulieren. Dies wird anhand zweier weiterer Modelle gezeigt. Dies ist eine weitere Neuerung der Arbeit. Durch die Entwicklung verteilungsfreier Reaktionszeitmodelle können unterschiedliche Vorschläge zur Reaktionszeitverteilung bei Testitems integriert werden.

Durch die Kombination der vier Verteilungsmodelle (Exponentialverteilung, dichotomisierte Zeiten, Log-Normalverteilung, semiparametrischer AFT-Ansatz)

mit den beiden Parametrisierungen (linear, quadratisch) ergeben sich insgesamt acht Reaktionszeitmodelle. Eine Übersicht findet sich in Tabelle 4.2. Die Darstellung der verschiedenen Reaktionszeitmodelle ist in zwei Blöcke unterteilt. Der erste Block (Kapitel 4.1) enthält die Proportional Hazard Modelle. Im zweiten Block (Kapitel 4.2) sind die Accelerated Failure Time Modelle zusammengefasst. Bis auf wenige Ausnahmen sind Accelerated Failure Time Modelle und Proportional Hazard Modelle sich einander ausschließende Modellierungsstrategien (Hougaard, 1999).

Tabelle 4.2: Übersicht über die untersuchten Reaktionszeitmodelle

KLASSE	VERTEILUNG	PARAMETRISIERUNG	KAPITEL
PH Modell	Exponential	Linear	4.1.1
		Quadratisch	4.1.1
	Dichotomisierte RT	Linear	4.1.2
		Quadratisch	4.1.2
AFT Modell	Log-Normal	Linear	4.2.1
		Quadratisch	4.2.1
	Semiparametrisch	Linear	4.2.2
		Quadratisch	4.2.2

## 4.1 RT-Modelle auf Basis von Proportional Hazard Modellen

Proportional Hazard Modelle nehmen eine dominante Position in der Analyse von „Failure Times“ ein (Klein & Moeschberger, 1997) und sind das am häufigsten verwendete Modell zur Analyse von Überlebenszeiten in der medizinischen Forschung (Bradburn, Clark, Love & Altman, 2003). Die Proportional Hazard Modelle charakterisiert, dass primär die Hazard-Rate als Funktion von Prädiktoren modelliert wird. Ist  $T$  die Zeit bis Eintritt eines Ereignisses, so ist die Hazard-Rate definiert als (Klein & Moeschberger, 1997, S. 27):

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P[t \leq T < t + \Delta t | T \geq t]}{\Delta t}. \quad (4.1)$$

Die Hazard-Rate wird teilweise in Begriffen umschrieben wie die „aktuelle Wahrscheinlichkeit“ eines Individuums, das Ereignis im nächsten Moment zu erleben – eine derartige Interpretation gibt zum Beispiel Bradburn et al. (2003). Obwohl die Hazard-Rate keine Dichtefunktion darstellt, definiert sie eindeutig eine Überlebenszeitverteilung (Marubini & Valsecchi, 1995, S. 145). Bei Kenntnis der Hazard-Rate kann die Wahrscheinlichkeit  $S(t)$  eines Individuums, den Zeitpunkt  $t$  zu überleben, bestimmt werden. Umformungen ergeben die bekannte Beziehung (Klein & Moeschberger, 1997):

$$S(t) = \exp \left[ - \int_0^t h(x) dx \right]. \quad (4.2)$$

Mit Proportional Hazard Modellen wird die Hazard-Rate einer Person als multiplikative Verknüpfung einer Grundrate  $h_B(t)$  und einer Funktion von Prädiktoren  $f(\mathbf{x})$  modelliert. Da Hazard-Raten positiv sein müssen, wird oft folgendes Modell angenommen, das in seiner Formulierung auf Cox (1972) bzw. Cox (1975) zurückgeht:

$$h_{\mathbf{x}}(t) = h_B(t) \exp[\mathbf{x}'\boldsymbol{\beta}]. \quad (4.3)$$

In Gleichung (4.3) beinhaltet der Vektor  $\mathbf{x}' = [x_1, \dots, x_p]$  die Ausprägungen einer Person auf den Prädiktoren und der Vektor  $\boldsymbol{\beta}' = [\beta_1, \dots, \beta_p]$  die Regressionsgewichte. Die Bezeichnung der Modelle als Proportional Hazard Modelle leitet sich davon ab, dass sich die Hazard-Raten zweier Personen proportional zueinander verhalten. Das Verhältnis der Hazard-Raten zweier Personen mit Prädiktorausprägungen  $\mathbf{x}_1$  und  $\mathbf{x}_2$  ist:

$$\frac{h_{\mathbf{x}_1}(t)}{h_{\mathbf{x}_2}(t)} = \exp[(\mathbf{x}_1 - \mathbf{x}_2)'\boldsymbol{\beta}]. \quad (4.4)$$

Die Quantitäten  $\exp(\beta_i)$  werden als „Hazard Ratios“ bezeichnet und sind ein Kennwert für die Interpretation der Effekte der Prädiktoren. Im einfachsten Fall ist  $x$  eine dichotome Indikatorvariable, z.B. das Geschlecht einer Person. In diesem Fall ist  $\exp(\beta_i)$  gleich dem Verhältnis der Hazard-Raten von Männern und Frauen. Alternativ kann der Effekt der Prädiktoren auf die Überlebensfunktion  $S(t)$  verdeutlicht werden. Ist  $S_B(t)$  die zur Grundhazard-Rate  $h_B(t)$  gehörende Überlebensfunktion (siehe 4.2), verändert sich die Überlebensfunktion bei Personen mit der Prädiktorausprägung  $\mathbf{x}$  zu:

$$S_{\mathbf{x}}(t) = [S_B(t)]^{\exp[\mathbf{x}'\boldsymbol{\beta}]}. \quad (4.5)$$

Die Proportional Hazard Modelle lassen sich danach unterscheiden, ob eine bestimmte Form für die Grundhazard-Rate  $h_B(t)$  und damit gleichzeitig eine bestimmte Überlebenszeitverteilung  $S_B(t)$  angenommen wird (parametrische Proportional Hazard Modelle) oder ob die genaue Form der Grundhazard-Rate  $h_B(t)$  unspezifiziert bleibt (semiparametrische Proportional Hazard Modelle). Auf beide Fälle soll im Folgenden eingegangen werden, zuerst auf ein parametrisches Proportional Hazard Modell für exponentialverteilte Reaktionszeiten, dann auf ein semiparametrisches Proportional Hazard Modell für dichotomisierte Reaktionszeiten.

#### 4.1.1 Proportional Hazard Modelle für exponentialverteilte Reaktionszeiten

Die Exponentialverteilung ist eine einparametrische Verteilung mit Dichte (Klein & Moeschberger, 1997):

$$f(t; \lambda) = \lambda \exp[-\lambda t] \quad \text{mit } \lambda > 0. \quad (4.6)$$

Die Hazard-Rate bei exponentialverteilten Reaktionszeiten ist konstant und besitzt den Wert  $h(t) = \lambda$ , weshalb die Verteilung oft als „gedächtnislos“ bezeichnet wird und deswegen für „ermüdungsfreie“ Prozesse geeignet ist. Trotz dieser Eigenschaften ist die Exponentialverteilung eine beliebte Verteilung für die Modellierung von Überlebenszeiten (N. Johnson, Kotz & Balakrishnan, 1994, Kap. 19). Im Kontext der Testtheorie wurde die Exponentialverteilung erstmalig von Rasch (1960) vorgeschlagen als Verteilung des Zeitbedarfs für das Lesen eines Wortes. Eine weitere Verwendung findet sich bei Scheiblechner (1979). Die Exponentialverteilung wird durch den Parameter  $\lambda$  vollständig festgelegt. Sollen Reaktionszeiten bei Testaufgaben modelliert werden, ist es nahe liegend, diesen Parameter über eine Funktion auf Item- und Persönlichkeitseigenschaften zurückzuführen. Für die Abhängigkeit des Verteilungsparameters  $\lambda$  von den Persönlichkeitseigenschaften  $\theta$  und  $\omega$  können allerdings unterschiedliche Formen angenommen werden. Zwei davon sollen näher betrachtet werden.

### Lineare Parametrisierung

Bei der linearen Parametrisierung wird der Logarithmus des Parameters  $\lambda$  auf eine lineare Funktion der Persönlichkeitseigenschaften  $\theta$  und  $\omega$  zurückgeführt. Die Reaktionszeit einer Person mit Personenfähigkeit  $\theta$  und Residualfaktor  $\omega$  ist dann bei Aufgabe  $g$  mit Konstante  $k_g$  und Einflussgewichten  $c_g$  und  $l_g$  exponentialverteilt mit folgendem Verteilungsparameter  $\lambda_g$ :

$$\log(\lambda_g) = k_g + l_g\theta + c_g\omega. \quad (4.7)$$

Wie Gleichung (4.7) unschwer zu entnehmen ist, handelt es sich bei dem Modell um ein parametrisches Proportional Hazard Modell. Umformungen des Modells ergeben die Gleichung  $\lambda_g = \exp[k_g + l_g\theta + c_g\omega] = \exp[k_g] \exp[l_g\theta + c_g\omega]$ , was analog zu der Grundgleichung der Proportional Hazard Modelle (4.3) ist. Die Größe  $\exp[k_g]$  repräsentiert folglich die (in diesem Fall zeitinvariante) Grundrate  $h_B(t)$ ,  $\exp[c_g]$  und  $\exp[l_g]$  stellen die „Hazard Ratios“ dar.

Das Modell (4.7) lässt sich auch in Begriffen des generalisierten linearen Modells interpretieren, wobei der Einfluss der Persönlichkeitseigenschaften  $\theta$  und  $\omega$  auf den Erwartungswert der Reaktionszeit deutlich wird. Der Erwartungswert  $E(t_g|\theta, \omega)$  der Reaktionszeit einer Person bei Aufgabe  $g$  steht zu den Persönlichkeitseigenschaften  $\omega$  und  $\theta$  in der Beziehung:

$$E(t_g|\theta, \omega) = 1/\lambda_g(\theta, \omega) = \exp[-k_g - l_g\theta - c_g\omega]. \quad (4.8)$$

Der Erwartungswert  $E(t_g|\theta, \omega)$  kann über die Link-Funktion  $\log [1/E(t_g|\theta, \omega)]$  auf den linearen Prädiktor  $\log [1/E(t_g|\theta, \omega)] = k_g + l_g\theta + c_g\omega$  zurückgeführt werden. Die Annahme einer derartigen Link-Funktion findet sich häufig bei Modellen für Überlebenszeiten, als Beispiel sei hierfür auf McCullagh und Nelder (1989) oder Kalbfleisch und Prentice (1980, Kap. 2) verwiesen. In Abbildung 4.1 ist die Beziehung (4.8) zwischen dem Erwartungswert der Reaktionszeit und den Persönlichkeitseigenschaften  $\theta$  und  $\omega$  für verschiedene Einflussgewichte  $l_g$  und  $c_g$  und die Konstante  $k_g = -2$  verdeutlicht.



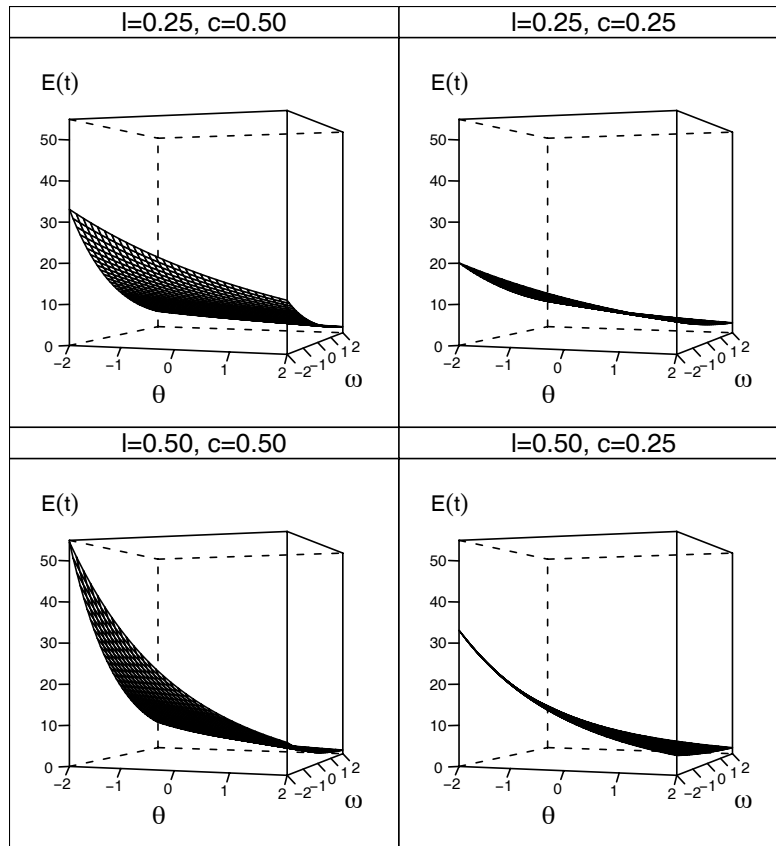


Abbildung 4.1: Exponentialverteilte RT: Mittlere Reaktionszeit in Abhängigkeit von den Persönlichkeitseigenschaften  $\theta$  und  $\omega$  bei linearer Parametrisierung mit Parameter  $k_g = -2$  für verschiedene Einflussgewichte  $l_g$  und  $c_g$  (siehe Gleichung 4.8)

Aus Gleichung (4.8) ist ersichtlich, dass  $\exp[-k]$  der Erwartungswert einer Person mit Personenfähigkeiten  $\theta = 0$  und Residualfaktor  $\omega = 0$  ist und somit das Grundniveau der Reaktionszeiten reguliert: Geht man davon aus, dass die Persönlichkeitseigenschaften  $\theta$  und  $\omega$  in einer Population von potentiellen Testteilnehmern bivariat standardnormalverteilt sind, bestimmt der Parameter  $k_g$  also die mittlere Reaktionszeit einer Durchschnittsperson. Wie Abbildung 4.1 zeigt, fällt bei positivem Einflussgewicht  $l_g$  die mittlere Reaktionszeit mit zunehmender Personenfähigkeit  $\theta$  ab. Die Stärke des Abfalls hängt von der Höhe des Einflussgewichtes  $l_g$  ab, was die Bezeichnung „Einflussgewicht“ rechtfertigt. Diese Art der Beziehung ist bei Leistungstests plausibel, wenn fähige Personen Aufgaben schneller lösen können. Auch der Residualfaktor  $\omega$  steht in monotoner Beziehung zur mittleren Reaktionszeit. Im Kontext von Leistungstests könnte der Residualfaktor beispielsweise der Extraversionwert einer Person sein, welcher mit der Reaktionszeit bei Tests zu korrelieren scheint (Danthiir et al., 2005).

Eine weitere Eigenschaft des Modells fällt auf, wenn man Abbildung 4.1 näher betrachtet. Mit Zunahme der Personenfähigkeit und des Residualfaktors konvergiert die mittlere Reaktionszeit gegen Null. Dieser Aspekt des Modells ist nicht plausibel, da zumindest für das Lesen der Aufgabe und die motorische Antwortreaktion eine bestimmte minimale Zeit benötigt wird. Ein derartiges Verhalten der mittleren Reaktionszeit ist analog zum realitätsfremden Verhalten der Lösungswahrscheinlichkeit beim zweiparametrischen Logit-Modell zu sehen (siehe Abbildung 3.1): Dort konvergiert die Lösungswahrscheinlichkeit mit abnehmender Fähigkeit gegen Null. Oft kann die korrekte Lösung jedoch zumindest erraten werden, so dass die minimale Lösungswahrscheinlichkeit die Ratewahrscheinlichkeit ist. Ein Ausweg besteht darin, eine Untergrenze für die Reaktionszeit einzuführen. Jedoch erschwert die Einführung eines derartigen „Shift-Parameters“ die Parameterschätzung erheblich – siehe hierzu die Schilderung der Problematik durch Dolan, van der Maas und Molenaar (2002). Es ist vermutlich dieser Grund, weshalb keines der zu Beginn des vierten Kapitels zitierten Reaktionszeitmodelle eine Reaktionszeituntergrenze enthält. Ebenso unplausibel wie das Fehlen einer Reaktionszeituntergrenze ist das unbegrenzte Anwachsen der mittleren Reaktionszeit mit abnehmender Personenfähigkeit und abnehmendem Residualfaktor, auch wenn dieser Aspekt in der Praxis weniger problematisch sein dürfte. Die Einführung einer Obergrenze für die Reaktionszeit ist jedoch aus denselben Gründen mit Schwierigkeiten verbunden wie die Modellierung von Reaktionszeituntergrenzen. Einzuwenden ist allerdings, dass wohl niemand von einem Modell eine exakte Übereinstimmung mit allen Aspekten der Realität fordern würde, sondern lediglich eine für die beabsichtigte Anwendung angemessene Approximation der Realität.

Ein eindimensionales Modell für exponentialverteilte Reaktionszeiten bei Testaufgaben wurde bereits von Scheiblechner (1979) vorgeschlagen, jedoch mit einer anderen Modellierung des Parameters  $\lambda$ . Scheiblechner führten Gedanken über das Konzept der spezifischen Objektivität zur additiven Zusammensetzung des Parameters der Exponentialverteilung:  $\lambda_g = \theta + d_g$ . Somit ist das Modell von Scheiblechner ebenfalls ein Proportional Hazard Modell, jedoch keines mit einer in der Überlebenszeitanalyse populären Parametrisierung. Zudem ist bei dem Modell die Positivität der Hazard-Rate nicht garantiert.

Wie sich einfach zeigen lässt (siehe hierzu die Ausführungen in Anhang A.1.1) ist das vorgeschlagene Reaktionszeitmodell (4.7) identifiziert – die Itemparameter lassen sich über die Momentmethode bestimmen, wenn man davon ausgeht, dass die untersuchten Personen aus einer Population potentieller Testteilnehmer stammen, in welcher die Persönlichkeitseigenschaften  $\theta$  und  $\omega$  normalverteilt sind. Somit steht zur Modellschätzung eine konsistente Methode zur Verfügung (Greene, 2003; Casella & Berger, 2002). Die Itemparameter lassen sich bis auf die bei latenten Variablenmodellen mögliche Rotation der latenten Dimensionen eindeutig bestimmen. Nimmt man ein einfaktorielles Modell (ohne den Residualfaktor  $\omega$ ) an, können die Itemparameter bereits für ein einziges Item geschätzt werden. Bei zweifaktoriellen Modellen sind jedoch mindestens drei Items notwendig.

### Schätzung der Personenparameter

Kennt man die Itemparameter der Items eines Tests, kann anhand der Reaktionszeiten einer Person die Ausprägung der Personenfähigkeit  $\theta$  und des Residualfaktors  $\omega$  geschätzt werden. Dies ist mit der im zweiten Kapitel vorgestellten Maximum-Likelihood-Methode möglich. Nimmt man wie bei den IRT-Modellen an, dass die Reaktionszeiten  $\mathbf{t} = [t_1, \dots, t_G]$  für die Beantwortung von  $G$  Testitems bei Bedingung auf die Persönlichkeitseigenschaften  $\theta$  und  $\omega$  einer Person voneinander unabhängig sind, ergibt die Likelihood-Funktion für die Ausprägung der unbekanntenen Personenparameter

$$L(\theta, \omega) = \prod_{g=1}^G f(t_g; \lambda_g(\omega, \theta)), \quad (4.9)$$

wobei  $f(t_g; \lambda_g(\theta, \omega))$  die Verteilungsdichte (4.6) der Exponentialverteilung mit Parameter  $\lambda_g(\theta, \omega) = \exp[k_g + l_g\theta + c_g\omega]$  ist. Die Maximum-Likelihood-Schätzung basiert darauf, diejenigen Parameterausprägungen  $\hat{\theta}$  und  $\hat{\omega}$  zu bestimmen, welche die Likelihood-Funktion für das beobachtete Zeitmuster  $\mathbf{t}$  einer Person maximieren. Die Frage, mit welcher Präzision die Persönlichkeitseigenschaften bestimmt werden können, führt zur Fisher-Information – siehe hierzu die Erläuterungen im zweiten Kapitel zur Rolle dieser Größe bei der Berechnung der asymptotischen Varianz von Maximum-Likelihood-Schätzern.

Der Beitrag einer Reaktionszeitbeobachtung zur Fisher-Information des Tests über die Persönlichkeitseigenschaften  $\theta$  und  $\omega$  entspricht der Varianz der Score-Funktion der Beobachtung (siehe 2.11). Diese Größe wird oft als Fisher-Information eines Items oder auch nur als Item-Information bezeichnet. Verwendet man zur Bildung der Score-Funktion die Verteilungsdichte der Exponentialverteilung (4.6) mit linearer Parametrisierung (4.7), ergibt sich für ein Item folgende Fisher-Informationsmatrix:

$$\mathbf{I}_g(\theta, \omega) = \begin{bmatrix} l_g^2 & l_g c_g \\ l_g c_g & c_g^2 \end{bmatrix}. \quad (4.10)$$

In Anhang B.1.1 sind die einzelnen Rechenschritte dargestellt, die zu Beziehung (4.10) führen. Abbildung 4.2 zeigt den Verlauf der Fisher-Information einer Reaktionszeit über die Personenfähigkeit  $\theta$ , d.h. den Verlauf des Elementes  $I_{g[11]}(\theta, \omega) = l_g^2$  der Informationsmatrix  $\mathbf{I}_g(\theta, \omega)$ , für Aufgaben mit verschiedenen Einflussgewichten  $l_g$ . Wie man einfach anhand von Gleichung (4.10) erkennen kann, ist die Information bezüglich  $\theta$  über den Bereich der Persönlichkeitseigenschaften konstant.

Das Element  $I_{g[11]}(\theta, \omega)$  der Informationsmatrix eines Items bestimmt maßgeblich, ob die Berücksichtigung einer Reaktionszeit zur Schätzung der Personenfähigkeit beiträgt. Anhand der Gleichungen (2.13) und (2.16) kann man erkennen, dass nur bei hoher Ausprägung des Elementes  $I_{g[11]}(\theta, \omega)$  mit einer starken Reduktion der Varianz des Fähigkeitsschätzers zu rechnen ist. Nach Beziehung (4.10) ist das Element  $I_{g[11]}(\theta, \omega)$  der Fisher-Information eines Items für verschiedene Ausprägungen der Persönlichkeitseigenschaften  $\theta$  und  $\omega$  konstant.

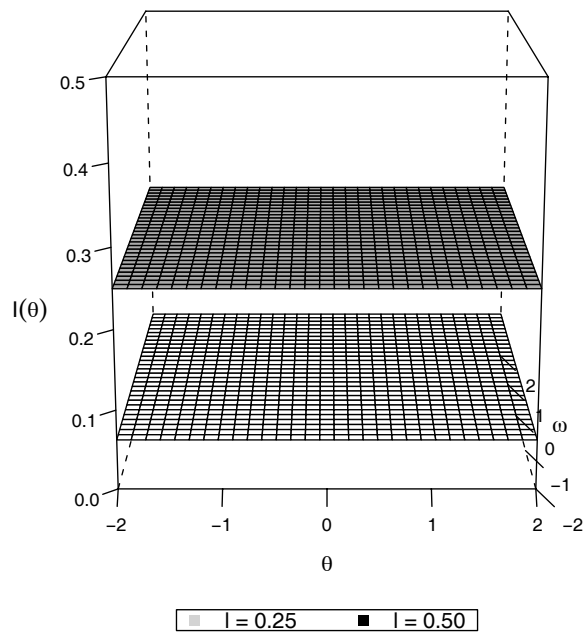


Abbildung 4.2: Fisher-Information über  $\theta$  bei exponentialverteilter Reaktionszeit und linearer Parametrisierung für verschiedene Einflussgewichte  $l_g$  (Element  $I_{g[11]}(\theta, \omega)$  der Informationsmatrix 4.10)

Diese Eigenschaft ist vorteilhaft: Soll die Personenfähigkeit  $\theta$  geschätzt werden, ist der Beitrag eines Items zum Element  $I_{11}(\theta, \omega)$  der Fisher-Information des Tests unabhängig von der Ausprägung der zu diagnostizierenden Personenfähigkeit  $\theta$ . Somit besitzen exponentialverteilte Reaktionszeiten mit linearer Parametrisierung für alle Fähigkeitsbereiche dasselbe diagnostische Potential. Da auch die übrigen Elemente der Informationsmatrix  $\mathbf{I}_g(\theta, \omega)$  konstant sind, fällt die aus der Berücksichtigung der Reaktionszeit eines Items resultierende Varianzreduktion des Fähigkeitsschätzers für alle Fähigkeitsausprägungen  $\theta$  identisch aus. Dies ist beispielsweise beim zweiparametrischen Logit-Modell (siehe Abbildung 3.2) nicht der Fall. Um bei diesem Modell einen für eine Person optimalen Test zu erhalten, müssen die Schwierigkeiten der vorgelegten Aufgaben an die Fähigkeit einer Person angepasst werden, siehe hierzu die Ausführungen von Bickel et al. (2001). Die tatsächliche Varianzreduktion, die sich durch Berücksichtigung eines zusätzlichen Items ergibt, hängt jedoch nicht alleine von dem betrachteten Element  $I_{g[11]}(\theta, \omega)$  der Informationsmatrix  $\mathbf{I}_g(\theta, \omega)$  ab, sondern auch von den Nebendiagonalelementen der Informationsmatrix – siehe hierzu Beziehung (2.16). Je größer die Elemente in der Nebendiagonalen sind, also je höher die Reaktionszeiten auf dem Residualfaktor laden, desto geringer fällt die Varianzreduktion aus. Eine detailliertere Betrachtung der resultierenden Varianzreduktion für einen Test aus mehreren Items sei jedoch auf das siebte Kapitel

verschoben.

### Quadratische Parametrisierung

Wie in Abbildung 4.1 zu sehen war, fällt bei der linearen Parametrisierung die mittlere Reaktionszeit mit steigenden Ausprägungen der Persönlichkeitseigenschaften  $\theta$  und  $\omega$  ab. Diese monotone Beziehung zwischen der mittleren Reaktionszeit und den individuellen Ausprägungen der Persönlichkeitseigenschaften ist nicht für alle Inhaltsbereiche plausibel. Im Bereich der Diagnostik von Persönlichkeitsmerkmalen wurde zwischen der Ausprägung des zu diagnostizierenden Traits und der mittleren Reaktionszeit beim Bewerten von Fragebogenitems ein quadratischer Zusammenhang nachgewiesen: In einer Untersuchung von Akrami et al. (2007) konnten beispielsweise Personen mit extremer Traitausprägung schneller auf die Fragen des NEO-PI antworten als Personen mit mittlerer Ausprägung des Traits. Auch die im ersten Kapitel beschriebenen Befunde sprechen dafür, dass die Reaktion auf eine Aussage desto schneller erfolgt, je eindeutiger ein Individuum die Aussage bewerten kann. Geht man von der Vorstellung aus, dass sowohl Items als auch Personen auf dem Merkmalskontinuum lokalisiert sind, bedeutet dies, dass Reaktionen desto weniger Zeit benötigen, je größer die Distanz zwischen einer Person und einem Item ist. Eine derartige Unfolding-Beziehung zwischen Item und Person kann über eine quadratische Funktion<sup>2</sup> hergestellt werden. Geht man von exponentialverteilten Reaktionszeiten aus, bietet sich folgende Modellierung des Verteilungsparameters  $\lambda_g$  an:

$$\log(\lambda_g) = k_g + l_g(\theta - d_g)^2 + c_g\omega. \quad (4.11)$$

In Gleichung (4.11) steht die Personenfähigkeit  $\theta$  in Unfolding-Beziehung zum Parameter  $\lambda_g$ : Mit wachsender Distanz zwischen einer Person und einem Item auf der Fähigkeitsskala wächst der Parameter  $\lambda_g$  an. Der Residualfaktor jedoch geht weiterhin linear in Gleichung (4.11) ein und steht somit in monotoner Beziehung zum Parameter  $\lambda_g$ . Welche psychologische Rechtfertigung besitzt das vorgeschlagene Modell? Die Parametrisierung in Modell (4.11) kann damit begründet werden, dass die Zustimmung oder Ablehnung von Aussagen in Persönlichkeits- oder Einstellungstests desto schneller möglich ist, je weiter Person und Item auf der Merkmalsdimension entfernt sind (siehe hierzu die Befunde in Kapitel 1.1). Wachsende Distanz zwischen Person und Item vergrößert die Komponente  $(\theta - d_g)^2$  und führt bei positivem Einflussgewicht  $l_g$  über die Beeinflussung der Hazard-Rate zu kürzeren Reaktionszeiten. Diese Abhängigkeit der Reaktionszeiten von  $\theta$  entspricht folglich der Beziehung der Lösungswahrscheinlichkeit zu  $\theta$  bei den Idealpunktmodellen der Item-Response-Theorie, siehe zum Beispiel Verkuilen (2006) oder Luo (1998). Anzunehmen ist jedoch, dass weitere systematische Einflussfaktoren die Reaktionszeiten einer Person bestimmen.

---

<sup>2</sup>Im IRT-Bereich wird die „Unfolding“-Beziehung zwischen Personen und Items teilweise auch mit Cosinus-Hyperbolicus-Funktionen modelliert (Andrich & Luo, 1993). Eine derartige Modellierung wurde ebenfalls für die Reaktionszeiten untersucht. Da jedoch bei den später dargestellten Accelerated Failure Time Modellen eine quadratische Funktion angenommen wird, soll auch an dieser Stelle der quadratischen Funktion der Vortritt überlassen werden. Die Unterschiede zwischen den beiden Modellen sind jedoch insgesamt gering: Beide Modelle dürften sich empirisch nur schwer voneinander unterscheiden lassen.

Bekannt ist Beispielsweise, dass Einflüsse wie Lesefähigkeit oder Neurotizismus die Reaktionszeit bei Testitems beeinflussen. Diese weiteren Einflüsse werden im zweiten Faktor, dem Residualfaktor  $\omega$ , zusammengefasst. Da jedoch Neurotizismus und Lesefähigkeit in monotoner Beziehung zur Reaktionszeit stehen, erscheint es den empirischen Befunden angemessener, für den Residualfaktor ebenfalls eine monotone Beziehung zur Reaktionszeit anzunehmen. Dies führt dazu, den Residualfaktor in Gleichung (4.11) weiterhin als linearen Term aufzunehmen.

Auch das Modell mit quadratischer Parametrisierung ist ein Proportional Hazard Modell, wie nach einigen Umformungen leicht ersichtlich wird. Schreibt man  $\lambda_g = \exp [k_g + l_g(\theta - d_g)^2 + c_g\omega]$  als  $\lambda_g = \exp [k_g] \exp [l_g(\theta - d_g)^2 + c_g\omega]$ , kann man durch einen Vergleich mit Gleichung (4.3) erkennen, dass  $\exp[k_g]$  wiederum die Grundhazard-Rate  $h_B(t)$  widerspiegelt, die nun jedoch einer Person mit Personenfähigkeit  $\theta = d_g$  und Residualfaktor  $\omega = 0$  zu eigen ist. Mit wachsendem Abstand zwischen Person und Item auf dem Fähigkeitskontinuum steigt bei positivem Einflussgewicht  $l_g$  die Hazard-Rate an. Ebenso wächst die Hazard-Rate mit steigender Ausprägung des Residualfaktors  $\omega$ .

Die mittlere Reaktionszeit bei Aufgabe  $g$  ergibt sich für Personen mit Personenfähigkeit  $\theta$  und Residualfaktor  $\omega$  als:

$$E(t_g|\theta, \omega) = 1/\lambda_g(\theta, \omega) = \exp [-k_g - l_g(\theta - d_g)^2 - c_g\omega]. \quad (4.12)$$

Der Effekt verschiedener Ausprägungen der Personenfähigkeit  $\theta$  und des Residualfaktors  $\omega$  auf die mittlere Reaktionszeit ist in Abbildung 4.3 für verschiedene positive Einflussgewichte  $l_g$  und  $c_g$  dargestellt. Die Komponente  $\exp[-k_g]$  bestimmt die mittlere Reaktionszeit, die an der Stelle  $\theta = d_g$  für eine Person mit Residualfaktor  $\omega = 0$  auftritt. Die Konstante  $k_g$  reguliert somit wiederum das Grundniveau der Reaktionszeiten bei einem Item. Mit zunehmendem Abstand zwischen einem Item mit Itemlokation  $d_g$  und einer Person mit Personenfähigkeit  $\theta$  fällt die mittlere Reaktionszeit ab (wenn die Ausprägung des Residualfaktors  $\omega$  konstant gehalten wird) – mittlere Reaktionszeit und Personenfähigkeit stehen somit in einer umgekehrt U-förmigen Beziehung zueinander. Der Residualfaktor beeinflusst weiterhin die mittlere Reaktionszeit monoton. Auch bei diesem Modell konvergiert die mittlere Reaktionszeit gegen Null, nun jedoch mit zunehmender Distanz einer Person zur Itemlokation auf der Fähigkeitsdimension bzw. bei abnehmendem Residualfaktor. Dieses Verhalten des Modells deckt sich vermutlich nicht mit der Realität – siehe hierzu die Anmerkungen zur linearen Parametrisierung – erleichtert jedoch die Schätzung des Modells erheblich.

Das Modell ist identifiziert. Die Parameter lassen sich (bis auf die bei latenten Variablenmodellen mögliche Rotation der latenten Dimensionen) eindeutig über die Momentmethode bestimmen, wenn man gewillt ist, für die Verteilung der Persönlichkeitseigenschaften  $\theta$  und  $\omega$  über eine Population von potentiellen Testteilnehmern eine Normalverteilung anzunehmen. Genauere Details zur Schätzung finden sich in Anhang A.1.2.

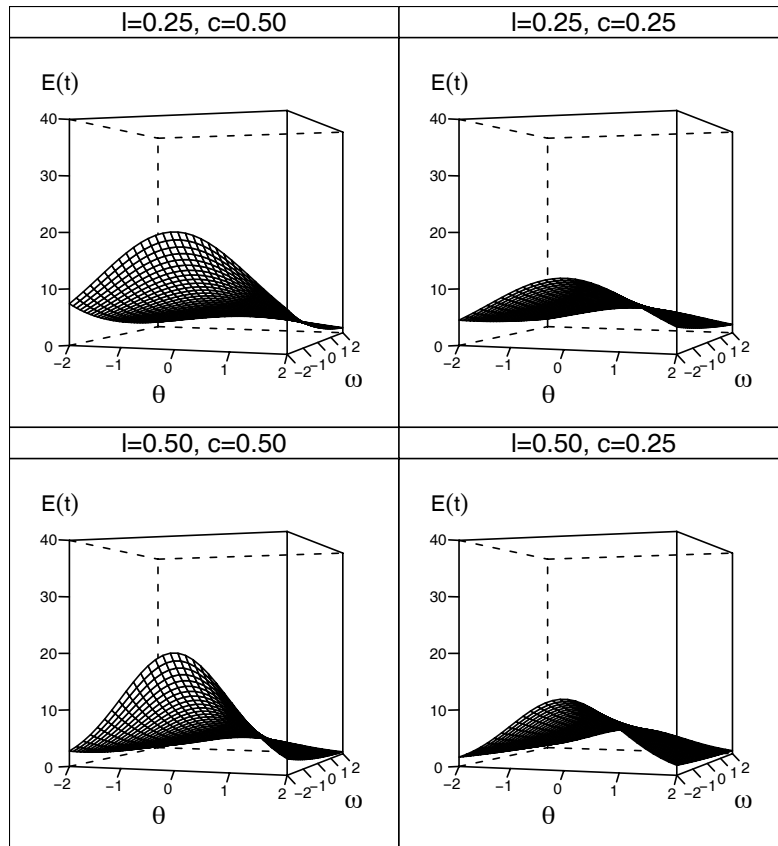


Abbildung 4.3: Exponentialverteilte RT: Mittlere Reaktionszeit in Abhängigkeit von den Persönlichkeitseigenschaften  $\theta$  und  $\omega$  bei quadratischer Parametrisierung mit Parametern  $k_g = -2$  und  $d_g = 0$  für verschiedene Einflussgewichte  $l_g$  und  $c_g$  (siehe Gleichung 4.12)

### Schätzung der Personenparameter

Wurden die Itemparameter eines Tests in einer Vorstudie bestimmt, kann bei Beobachtung der Reaktionszeiten einer Person die Ausprägung der Personenfähigkeit und des Residualfaktors mit der Maximum-Likelihood-Methode geschätzt werden. Bei Beobachtung von  $G$  unabhängigen Reaktionszeiten  $\mathbf{t} = [t_1, \dots, t_G]$  ist die Likelihood-Funktion für die Personenparameter

$$L(\theta, \omega) = \prod_{g=1}^G f(t_g; \lambda_g(\theta, \omega)), \quad (4.13)$$

wobei  $f(t_g; \lambda_g(\theta, \omega))$  die Verteilungsdichte (4.6) der Exponentialverteilung mit Parameter  $\lambda_g(\theta, \omega) = \exp[k_g + l_g(\theta - d_g)^2 + c_g\omega]$  ist. Diese Funktion ist bei

gegebenem Reaktionszeitmuster  $\mathbf{t}$  für die Persönlichkeitseigenschaften zu maximieren. Die so gefundenen Werte  $\hat{\theta}$  und  $\hat{\omega}$  sind dann die Maximum-Likelihood-Schätzer der Eigenschaften einer Person.

Um die Varianz der Personenparameterschätzer bestimmen zu können, muss wiederum die Matrix der Fisher-Information des Tests berechnet werden. Der Beitrag eines Items zu dieser Matrix ist bei der quadratischen Parametrisierung (4.11) die Iteminformationsmatrix:

$$\mathbf{I}_g(\theta, \omega) = \begin{bmatrix} 4l_g^2(\theta - d_g)^2 & 2l_g c_g(\theta - d_g) \\ 2l_g c_g(\theta - d_g) & c_g^2 \end{bmatrix}. \quad (4.14)$$

Genauere Details zur Bestimmung dieser Matrix finden sich in Anhang B.1.2. Der Verlauf der Fisher-Information eines Items über die Personenfähigkeit  $\theta$ , das Element  $I_{g[11]}(\theta, \omega)$  der Informationsmatrix (4.14), ist in Abbildung 4.4 abgetragen. Die Itemlokation  $d_g$  bestimmt nur die Lokation der Funktion, so dass auf eine Demonstration des Effektes unterschiedlicher Ausprägungen von  $d_g$  verzichtet werden kann.

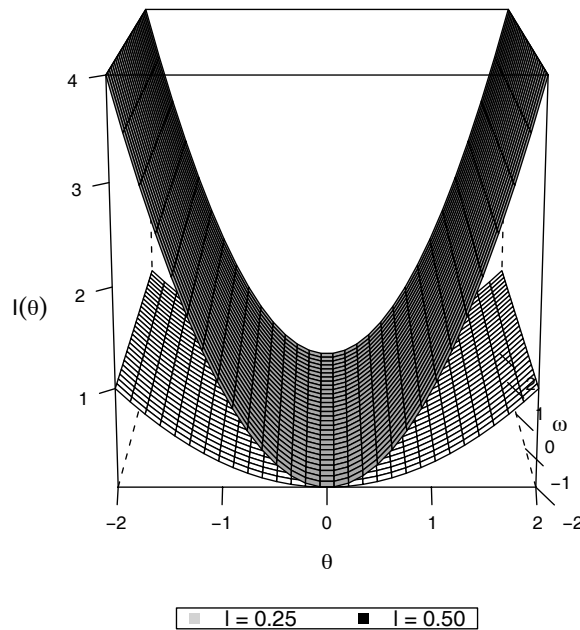


Abbildung 4.4: Fisher-Information über  $\theta$  bei exponentialverteilter Reaktionszeit und quadratischer Parametrisierung mit Parameter  $d_g = 0$  für verschiedene Einflussgewichte  $l_g$  (Element  $I_{g[11]}(\theta, \omega)$  der Informationsmatrix 4.14)

Die Fisher-Information eines Items über  $\theta$  besitzt ihr Minimum in  $d_g$  und nimmt dort den Wert Null an. Diese Eigenschaft teilt das Modell mit vielen Idealpunktmodellen im IRT-Bereich, die ebenfalls an der Stelle der Itemlokation



eine Iteminformation von Null besitzen (Verkuilen, 2006). Mit zunehmendem Abstand zwischen einer Person mit Fähigkeit  $\theta$  und einem Item mit Itemlokation  $d_g$  steigt die Information unbegrenzt an. Dieses Verhalten kann erklärt werden, wenn man sich bestimmte Eigenschaften des Modells vor Augen hält. Kernannahme des Modells ist die quadratische Beziehung (4.11) zwischen der Personenfähigkeit  $\theta$  und dem Verteilungsparameter  $\lambda$  der Exponentialverteilung. Aufgrund dieser Beziehung führen kleine Veränderungen der Personenfähigkeit  $\theta$  zu einer großer Veränderung von  $\lambda(\theta, \omega)$ , wenn die Distanz zwischen einer Person und einem Item auf dem Fähigkeitskontinuum groß ist. Starke Veränderungen des Parameters  $\lambda(\theta, \omega)$  jedoch führen zu starken Veränderungen der Log-Likelihood-Funktion – im Bereich des wahren Parameters  $\theta$  ist die Log-Likelihood-Funktion somit stark „gekrümmt“. Die Fisher-Information kann wiederum als Maß der Krümmung der Log-Likelihood-Funktion interpretiert werden (Sorensen & Gianola, 2002, S. 132). Mit steigender Krümmung nimmt folglich auch die Iteminformation zu. Über die Dimension des Residualfaktors jedoch ist die Iteminformation konstant.

Notwendige Bedingung für eine geringe Varianz des Fähigkeitsschätzers  $\hat{\theta}$  ist die Forderung, dass die einzelnen Items eines Tests hohe Iteminformationen über  $\theta$ , also große Elemente  $I_{g[11]}(\theta, \omega)$  besitzen. Dies geht aus Beziehung (2.16) hervor. Aus dieser Beziehung kann geschlossen werden, dass bei einem Test die Varianz des Fähigkeitsschätzers groß ausfällt, wenn eine Person auf der Fähigkeitsdimension in der Nähe der Itemlokationen angesiedelt ist. In diesen Bereichen fällt das Element  $I_{g[11]}(\theta, \omega)$  der Informationsmatrix extrem gering aus, wie Abbildung 4.4 zeigt. Hingegen sollten sich im Vergleich zu den Itemlokationen extreme Fähigkeitsniveaus präzise diagnostizieren lassen. Allerdings wird die Varianz der Fähigkeitsschätzung durch die positiven Nebendiagonalelementen in der Item-Informationsmatrix (4.14) erhöht: Je höher die Einflussgewichte des Residualfaktors  $c_g$  sind, desto geringer ist die Präzision der Fähigkeitsschätzung.

### 4.1.2 Proportional Hazard Modelle für dichotomisierte Reaktionszeiten

Die in Kapitel 4.1.1 beschriebenen Modelle basieren auf exponentialverteilten Reaktionszeiten, deren Hazard-Rate zeitinvariant ist. Diese Festlegung auf eine konkrete Reaktionszeitverteilung schränkt die Anwendbarkeit der Modelle erheblich ein. Ein Vorteil von Proportional Hazard Modellen ist die Möglichkeit, diese Modelle auch ohne Annahme einer bestimmten Verteilungsform bzw. Grundhazard-Rate anwenden zu können. Die Annahme einer bestimmten Verteilungsform lässt sich nämlich umgehen, wenn man dichotomisierte oder kategorisierte Reaktionszeiten betrachtet.<sup>3</sup> Diese Idee geht auf McCullagh (1980) zurück und wird im Folgenden dargestellt.

Anstelle des genauen Eintrittszeitpunktes einer Antwort kann lediglich festgehalten werden, ob die Antwort vor oder nach einem bestimmten, vorgegebenen

<sup>3</sup>Alternativ existieren weitere Ansätze, Annahmen über die genaue Form der Hazard-Rate zu vermeiden. Zu diesen zählen die abschnittsweise Approximation der Hazard-Funktion mit einer konstanten Hazard-Rate oder die Approximation über ein Polynom höherer Ordnung (Skrondal & Rabe-Hesketh, 2004, S. 39ff) sowie die Erweiterung des Partial Likelihood Ansatzes auf latente Variablenmodelle (McGilchrist, 1993).

Zeitpunkt  $t_c$  erfolgt. Die ursprünglich kontinuierliche Reaktionszeit  $t$  wird dabei durch „dichotomisierte“ Reaktionszeiten  $y$  ersetzt. Jede Person wird genau einer von zwei Klassen zugeordnet:  $y = 1$  wenn  $t < t_c$  versus  $y = 0$  wenn  $t > t_c$ . Die Art der Antwort, ob richtig oder falsch bzw. Zustimmung oder Ablehnung, spielt dabei keine Rolle. Nimmt man für die Verteilung der Reaktionszeit ein Proportional Hazard Modell an, lässt sich die Wahrscheinlichkeit, nach dem Zeitpunkt  $t_c$  zu antworten, einfach herleiten. Ausgangspunkt ist dabei die Annahme (4.3), dass eine nicht näher spezifizierte Grundhazard-Rate  $h_B(t)$  durch mehrere Prädiktoren  $\mathbf{x}$  beeinflusst wird gemäß der Beziehung:

$$h_{\mathbf{x}}(t) = h_B(t) \exp(\mathbf{x}'\boldsymbol{\beta}). \quad (4.15)$$

Die Parameter  $\boldsymbol{\beta}$  stellen die Regressionsgewichte der einzelnen Prädiktoren dar. Die „Überlebensfunktion“  $S_{\mathbf{x}}(t)$ , welche die Wahrscheinlichkeit einer Reaktion nach Zeitpunkt  $t$  für eine Person mit Prädiktorausprägung  $\mathbf{x}$  angibt, erhält man bei Ausnutzung der bekannten Beziehung der Überlebensfunktion zur Hazard-Rate (siehe 4.2) als:

$$S_{\mathbf{x}}(t) = \exp \left[ - \int_0^t h_{\mathbf{x}}(z) dz \right] = \exp \left[ - [\exp(\mathbf{x}'\boldsymbol{\beta})] \int_0^t h_B(z) dz \right]. \quad (4.16)$$

Gleichung (4.16) gibt an, welche Wahrscheinlichkeit besteht, nach einem bestimmten Zeitpunkt  $t$  zu antworten. Bei Dichotomisierung der Reaktionszeit an Zeitpunkt  $t_c$  berechnet sich die Wahrscheinlichkeit, nach dem Zeitpunkt  $t_c$  zu reagieren, somit als:

$$\begin{aligned} S_{\mathbf{x}}(t_c) &= \exp \left[ - \left[ \exp(\mathbf{x}'\boldsymbol{\beta}) \int_0^{t_c} h_B(z) dz \right] \right] \\ &= \exp \left[ - \left[ \exp \left( \log \left( \int_0^{t_c} h_B(z) dz \right) + \mathbf{x}'\boldsymbol{\beta} \right) \right] \right] \\ &= \exp \left[ - \left[ \exp \left( \beta_0(t_c) + \mathbf{x}'\boldsymbol{\beta} \right) \right] \right], \end{aligned} \quad (4.17)$$

wobei die Konstante  $\beta_0(t_c) = \log \left( \int_0^{t_c} h_B(z) dz \right)$  ist. Wie Gleichung (4.17) zu entnehmen ist, wird die genaue Form der Grundhazard-Rate in der Konstanten  $\beta_0(t_c)$  absorbiert. Diese Konstante lässt sich schätzen, so dass auf eine detaillierte Annahme über die Form der Grundhazard-Rate verzichtet werden kann. Eine Anwendung des Modells auf die Modellierung von Reaktionszeiten bei Tests ist nahe liegend.

### Lineare Parametrisierung

Die in den vorhergehenden Abschnitten beschriebenen Modelle für exponentialverteilte Reaktionszeiten lassen sich auch als Proportional Hazard Modelle für

dichotomisierte Reaktionszeiten darstellen. Die Modelle für exponentialverteilte Reaktionszeiten können somit auf den Fall dichotomisierter Reaktionszeiten übertragen werden. Jedoch sind die Proportional Hazard Modelle für dichotomisierte Reaktionszeiten allgemeiner, da bei diesen Modellen keine bestimmte Form der Grundhazard-Rate  $h_B(t)$  angenommen wird. Zunächst soll der Fall der linearen Parametrisierung näher beschrieben werden. Geht man von einer Dominanzbeziehung zwischen Item und Person aus, ergibt sich als Proportional Hazard Modell bei Anwendung von (4.15) die Beziehung

$$h_{\theta,\omega}(t) = h_{B_g}(t) \exp[l_g\theta + c_g\omega], \quad (4.18)$$

wobei die Parameter  $l_g$  und  $c_g$  die Einflussgewichte der beiden Persönlichkeitseigenschaften bei Aufgabe  $g$  darstellen. Bei der Funktion  $h_{B_g}(t)$  handelt es sich um die Grundhazard-Rate bei Aufgabe  $g$ , die sich von den Grundhazard-Raten der übrigen Items unterscheiden kann. Die Wahrscheinlichkeit  $P_g(\theta, \omega)$ , eine Aufgabe  $g$  innerhalb eines bestimmten Zeitlimits  $t_c$  zu beantworten, lässt sich nach (4.17) berechnen als

$$\begin{aligned} P_g(\theta, \omega) &= 1 - \exp \left[ - \exp \left[ \log \left( \int_0^{t_c} h_{B_g}(z) dz \right) + l_g\theta + c_g\omega \right] \right] \\ &= 1 - \exp \left[ - \exp \left[ k_g + l_g\theta + c_g\omega \right] \right], \end{aligned} \quad (4.19)$$

wobei  $k_g = \log \left( \int_0^{t_c} h_{B_g}(z) dz \right)$  ist. Im Vergleich zum vorherigen Abschnitt 4.1.1 verändert der Parameter  $k_g$  seine Bedeutung. Die resultierende Konstante  $k_g$  hängt einerseits von der Grundhazard-Rate  $h_{B_g}(t)$  eines Items ab, andererseits von dem Zeitlimit  $t_c$  und lässt sich somit streng genommen nicht mehr als „reine“ Itemeigenschaft interpretieren. Interessanterweise setzt die Verwendung von dichotomisierten Zeiten nicht voraus, dass alle Items dieselbe Hazard-Rate besitzen. Dadurch ist die Modellierung sehr flexibel, da für die einzelnen Items unterschiedliche Reaktionszeitverteilungen möglich sind.

Zur Demonstration der Eigenschaften von Modell (4.19) soll für den Fall exponentialverteilter Reaktionszeiten die Abhängigkeit der Wahrscheinlichkeit, innerhalb eines Intervalls  $[0, t_c]$  zu reagieren, von den Persönlichkeitseigenschaften  $\theta$  und  $\omega$  demonstriert werden. Im ersten Abschnitt des Kapitels wurden exponentialverteilte Reaktionszeiten mit der zeitinvarianten Grundhazard-Rate  $\exp[-2]$  betrachtet – siehe hierzu Abbildung (4.1) und das zugehörige Beispiel. Bei Annahme dieser zeitinvarianten Grundhazard-Rate ergibt sich somit für eine Person mit Personenfähigkeit  $\theta$  und Residualfaktor  $\omega$  die Hazard-Rate  $h_{\theta,\omega}(t) = \exp[-2] \exp[l_g\theta + c_g\omega]$ . In diesem Fall kann nach Gleichung (4.19) die Wahrscheinlichkeit einer Antwort vor dem Zeitlimit  $t_c$  berechnet werden als

$$P_g(\theta, \omega) = 1 - \exp \left[ - \exp \left[ \log(t_c) - 2 + l_g\theta + c_g\omega \right] \right] = 1 - \exp \left[ - \exp \left[ k_g + l_g\theta + c_g\omega \right] \right], \quad (4.20)$$

wobei  $k_g = \log(\int_0^{t_c} \exp[-2z] dz) = \log(t_c) - 2$  ist. Die Abhängigkeit der Wahrscheinlichkeit  $P_g(\theta, \omega)$  von den Persönlichkeitseigenschaften  $\theta$  und  $\omega$  ist in Abbildung 4.5 für das Zeitlimit  $t_c = 10$  und die resultierende Konstante  $k_g = \log(10) - 2 = 0.3$  sowie für verschiedene Einflussgewichte  $l_g$  und  $c_g$  visualisiert.

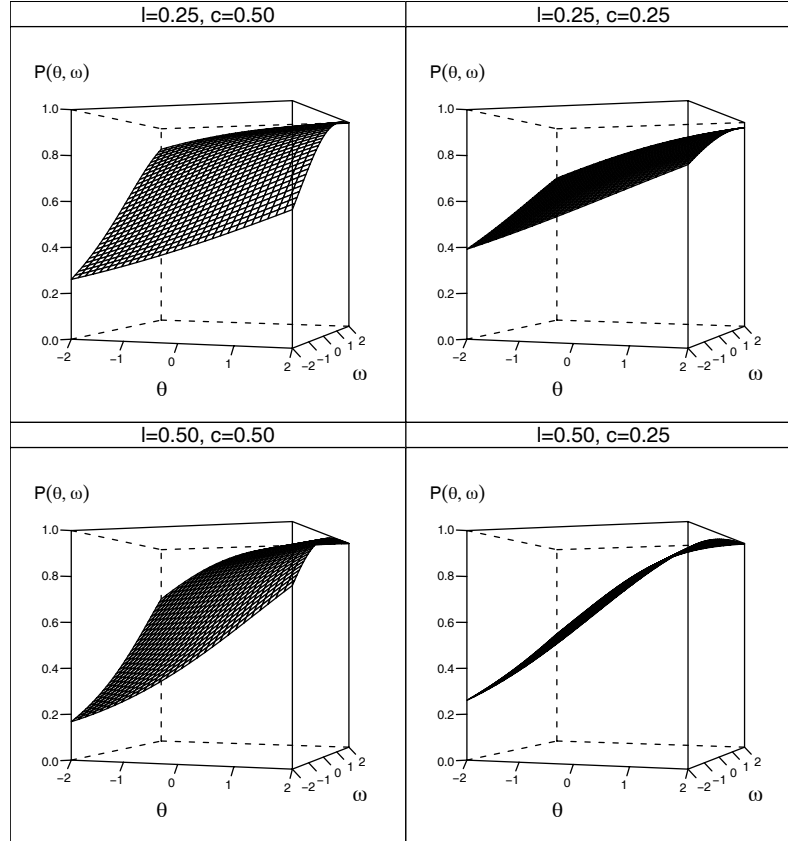


Abbildung 4.5: Dichotomisierte RT: Wahrscheinlichkeit einer Antwort vor  $t_c = 10$  in Abhängigkeit von den Persönlichkeitseigenschaften  $\theta$  und  $\omega$  bei linearer Parametrisierung mit Parameter  $k_g = 0.3$  für verschiedene Einflussgewichte  $c_g$  und  $l_g$  (siehe Gleichung 4.20)

Die Persönlichkeitseigenschaften  $\theta$  und  $\omega$  beeinflussen die Bearbeitungsgeschwindigkeit einer Person monoton: Größere Ausprägungen erhöhen die Wahrscheinlichkeit, vor Zeitlimit  $t_c$  zu reagieren. Das Einflussgewicht  $l_g$  bestimmt dabei den Einfluss der Personenfähigkeit  $\theta$  auf die Wahrscheinlichkeit einer Antwort vor  $t_c$  – je größer dieser Parameter ausfällt, desto stärker verändert sich die Antwortwahrscheinlichkeit mit der Personenfähigkeit. Das Einflussgewicht  $c_g$  regelt in derselben Weise die Enge der Beziehung zwischen dem Residualfaktor  $\omega$  und der Wahrscheinlichkeit einer Antwort vor  $t_c$ . Die Bezeichnung von  $l_g$  und  $c_g$  als Einflussgewicht ist somit gerechtfertigt. Der Parameter  $k_g$  bestimmt die Antwortwahrscheinlichkeit einer Person mit Eigenschaftsausprägung  $\omega = 0$  und

$\theta = 0$ . Sind die Persönlichkeitseigenschaften bivariat standardnormalverteilt, bestimmt  $k_g$  somit die Wahrscheinlichkeit einer durchschnittlichen Person. Somit reguliert auch bei dichotomisierten Reaktionszeiten die Konstante  $k_g$  das allgemeine Reaktionszeitniveau bei einer Aufgabe, diesmal jedoch in Form der Wahrscheinlichkeit einer Antwort vor einem bestimmten Zeitpunkt. Das Modell ist identifiziert. Die Itemparameter können über den EM-Ansatz geschätzt werden, wenn man eine bestimmte Verteilung der Persönlichkeitseigenschaften  $\theta$  und  $\omega$  über die Population der potentiellen Testteilnehmer annimmt. Mehr zur Schätzung des Modells findet sich im fünften Kapitel.

### Schätzung der Personenparameter

Sind die Itemparameter bekannt, können bei Beobachtung der dichotomisierten Reaktionszeiten einer Person deren Eigenschaftsausprägungen geschätzt werden. Ein mögliches Schätzverfahren hierfür ist wiederum die Maximum-Likelihood-Schätzung. Wurde für eine Person bei  $G$  Items das Reaktionszeitmuster  $\mathbf{y} = [y_1, \dots, y_G]$  beobachtet, wobei  $y_g$  den Umstand kodiert, ob eine Reaktion vor dem Zeitlimit  $t_c$  erfolgte oder nicht, ist die Likelihood bestimmter Personenparameterausprägungen  $\theta$  und  $\omega$  definiert als:

$$L(\theta, \omega) = \prod_{g=1}^G P_g(\theta, \omega)^{y_g} (1 - P_g(\theta, \omega))^{1-y_g}. \quad (4.21)$$

In Gleichung (4.21) ist für  $P_g(\theta, \omega)$  die entsprechende Beziehung (4.19) zu verwenden. Bei der Maximum-Likelihood-Schätzung werden einer Person diejenigen Parameterausprägungen zugeordnet, welche die Likelihood-Funktion (4.21) maximieren. Um die Varianz der Personenparameterschätzung zu approximieren, muss die Fisher-Information des Tests betrachtet werden. Der Beitrag eines Items zur Fisher-Information des Tests wird durch dessen Iteminformationsmatrix wiedergegeben. Die Bestimmung der Iteminformationsmatrix von Modell (4.19) ist in Anhang B.2.1 beschrieben. Als Ergebnis resultiert die Iteminformationsmatrix:

$$\mathbf{I}_g(\theta, \omega) = \frac{[\exp[k_g + l_g\theta + c_g\omega]]^2}{\exp[\exp[k_g + l_g\theta + c_g\omega]] - 1} \begin{bmatrix} l_g^2 & l_g c_g \\ l_g c_g & c_g^2 \end{bmatrix}. \quad (4.22)$$

Von besonderer Relevanz für die Bestimmung der Varianz des Fähigkeitschätzers  $\hat{\theta}$  ist das Element  $I_{g[11]}(\theta, \omega)$  der Informationsmatrix (4.22). Der Verlauf dieses Elementes über verschiedene Ausprägungen der Personenfähigkeit  $\theta$  und des Residualfaktors  $\omega$  findet sich für unterschiedliche Einflussgewichte  $l_g$  in Abbildung 4.6. Als Ausprägung des Parameters  $k_g$  wurde der bereits in Beispiel (4.20) resultierende Parameterwert  $k_g = -2 + \log(10) = 0.30$  gewählt. Das Einflussgewicht des Residualfaktors ist  $c_g = 0.25$ .

Wie aus Abbildung 4.6 ersichtlich ist und aufgrund der Bildung von Grenzwerten einfach belegt werden kann, ist die Informationsfunktion  $I_{g[11]}(\theta, \omega)$

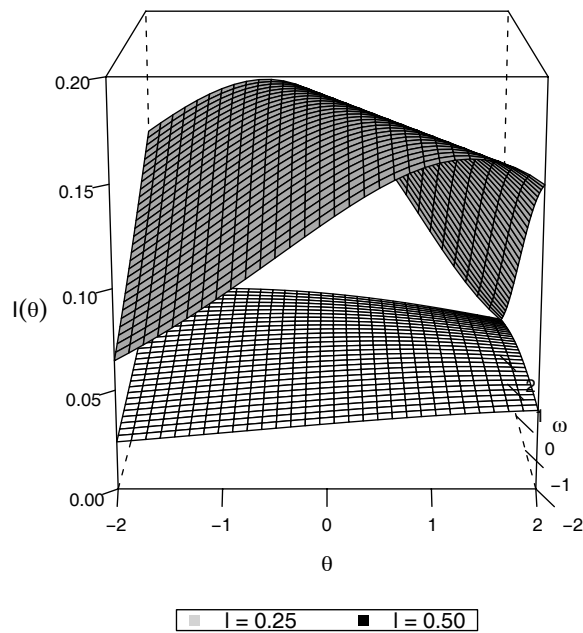


Abbildung 4.6: Fisher-Information über  $\theta$  bei dichotomisierten Reaktionszeiten und linearer Parametrisierung mit Parametern  $k_g = 0.3$  und  $c_g = 0.25$  für verschiedene Einflussgewichte  $l_g$  (Element  $I_{g[11]}(\theta, \omega)$  der Informationsmatrix 4.22)

beschränkt. Mit unbegrenzt zunehmender und unbegrenzt abnehmender Personenfähigkeit  $\theta$  strebt die Informationsfunktion gegen Null bei Konstanthaltung des Residualfaktors  $\omega$ . Die Maxima der Funktion liegen auf der Geraden<sup>4</sup>  $\log(1.5936) = k_g + l_g\theta + c_g\omega$ . Der Bereich maximaler Information hängt somit nicht allein von der Grundhazard-Rate eines Items ab, sondern kann auch über die Wahl des Zeitlimits  $t_c$ , an welchem dichotomisiert wird, beeinflusst werden. Diese Eigenschaft ist günstig, da sie bei freier Wählbarkeit der Zeitlimits  $t_c$  ein beliebig Verschieben des Maximums erlaubt. Dies entspräche bei IRT-Modellen dem Fall, dass die Itemschwierigkeit einer Aufgabe beliebig verändert werden könnte.

In Abbildung (4.6) ist die Iteminformation nach Dichotomisierung von ursprünglich exponentialverteilten Reaktionszeiten visualisiert. Vergleicht man die Informationsfunktion der dichotomisierten Reaktionszeiten über  $\theta$  (siehe Abbildung 4.6) mit der Informationsfunktion der ursprünglich exponentialverteilten Reaktionszeiten über  $\theta$  (siehe Abbildung 4.2), stellt man einen Informationsverlust fest, der auf der Dichotomisierung beruht. Dieser Verlust ist jedoch nicht

<sup>4</sup>Die Extremwerte der Funktion lassen sich nicht vollständig explizit bestimmen. Im letzten Berechnungsschritt muss das Gleichungssystem numerisch gelöst werden, wobei sich die Konstante  $\log(1.5936)$  ergibt.

überraschend und kann mit allgemeinen Überlegungen von Orchard und Woodbury (1972) bzw. mit einer Anwendung dieser Überlegungen auf dichotomisierte Beobachtungen von Schader und Schmid (1988) erklärt werden. Im siebten Kapitel soll auf das Phänomen „fehlender Information“ näher eingegangen werden.

In der Testtheorie ist die Iteminformation eine eigenständige Größe, die zur Itemselektion eingesetzt wird. Möchte man jedoch die Varianz des Fähigkeitschätzers bestimmen, so ist die Inverse der Testinformationsmatrix zu betrachten. Hohe Iteminformationen über  $\theta$ , also große Elemente  $I_{g[11]}(\theta, \omega)$  der einzelnen Items sind eine notwendige Bedingung für eine geringe Varianz des Schätzers  $\hat{\theta}$ . Dies legt Beziehung (2.16) nahe. Aus Abbildung 4.6 kann daher geschlossen werden, dass die Reaktionszeit eines Items nur für einen begrenzten Bereich von Eigenschaftsausprägungen diagnostisches Potential besitzt, nämlich nur in der Umgebung der Geraden  $\log(1.5936) = k_g + c_g\omega + l_g\theta$ . Allerdings sind die Nebendiagonalelemente der Informationsmatrix nicht Null, sondern besitzen positive Gewichte, wodurch der Beitrag von Items zur Varianzreduktion reduziert wird. Mehr hierzu jedoch im siebten Kapitel, wenn die Präzision der Fähigkeitsschätzung verschiedener Tests betrachtet wird.

### Quadratische Parametrisierung

Bei linearer Parametrisierung steigt die Wahrscheinlichkeit einer Antwort vor einem bestimmten Zeitlimit  $t_c$  mit der Personenfähigkeit  $\theta$  an. Diese Beziehung ist sicherlich bei Leistungstests plausibel. Denkbar ist jedoch auch eine Unfolding-Beziehung, die im Einstellungs- und Traitbereich angemessener sein mag (siehe hierzu die Übersicht im ersten Kapitel). Es ist daher nahe liegend, auch die quadratische Modellierung aus Kapitel 4.1.1 auf das Proportional Hazard Modell für dichotomisierte Reaktionszeiten zu übertragen. Legt man das Proportional Hazard Modell mit quadratischer Beziehung zugrunde, lässt sich die Hazard-Rate einer Person mit Personenfähigkeit  $\theta$  und Residualfaktor  $\omega$  bei Item  $g$  mit Itemlokation  $d_g$  und Einflussgewichten  $l_g$  und  $c_g$  in Begriffen einer itemspezifischen Grundhazard-Rate  $h_{B_g}(t)$  angeben als:

$$h_{\theta,\omega}(t) = h_{B_g}(t) \exp[l_g(\theta - d_g)^2 + c_g\omega]. \quad (4.23)$$

Die Grundhazard-Rate  $h_{B_g}(t)$  entspricht der Hazard-Rate einer Person mit Personenfähigkeit  $\theta = d_g$  und Residualfaktor  $\omega = 0$ . Mit zunehmendem Abstand von Person und Item auf dem Fähigkeitskontinuum wächst die Hazard-Rate  $h_{\theta,\omega}(t)$  an. Die Hazard-Rate  $h_{\theta,\omega}(t)$  ist zudem als Funktion von  $\theta$  symmetrisch zum Symmetriezentrum  $d_g$ .

Bei Dichotomisierung der Reaktionszeit gilt für die Wahrscheinlichkeit, auf Item  $g$  vor dem vorgegebenen Zeitlimit  $t_c$  zu reagieren, die Beziehung

$$\begin{aligned} P_g(\theta, \omega) &= 1 - \exp \left[ - \exp \left[ l_g(\theta - d_g)^2 + c_g\omega \right] \int_0^{t_c} h_{B_g}(z) dz \right] \\ &= 1 - \exp \left[ - \exp \left[ k_g + l_g(\theta - d_g)^2 + c_g\omega \right] \right], \end{aligned} \quad (4.24)$$

wobei die Konstante  $k_g$  für  $\log \left[ \int_0^{t_c} h_{B_g}(z) dz \right]$  steht. Der Verlauf der Antwortwahrscheinlichkeit über das Kontinuum der Personenfähigkeit und des Residualfaktors wird in Abbildung 4.7 für verschiedene Einflussgewichte  $l_g$  und  $c_g$  visualisiert. Gewählt wurden als weitere Parameter die Werte  $d_g = 0$  und  $k_g = 0.3$ . Eine Parameterausprägung von  $k_g = 0.3$  ergibt sich beispielsweise bei exponentialverteilten Reaktionszeiten mit Hazard-Rate  $h_{B_g}(t) = \lambda_g = \exp[-2]$  und einer Dichotomisierung an  $t_c = 10$ . Deutlich tritt eine U-förmige Beziehung zwischen der Antwortwahrscheinlichkeit und der Distanz von Person und Item auf der Fähigkeitsdimension separiert sind, desto größer ist die Wahrscheinlichkeit, dass diese Person vor dem Zeitpunkt  $t_c$  antwortet.

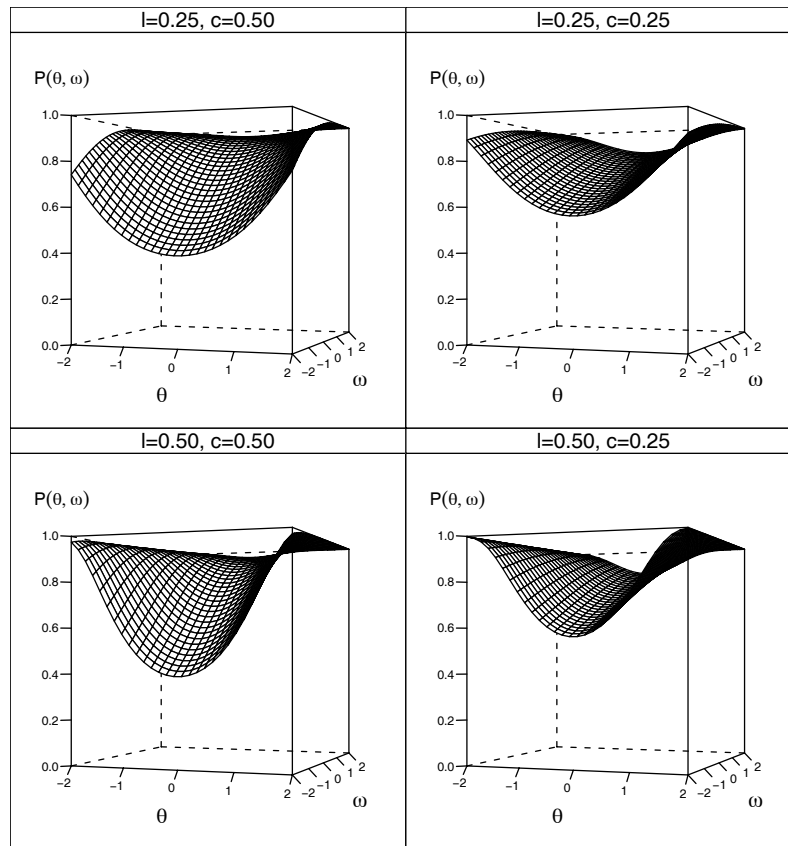


Abbildung 4.7: Dichotomisierte RT: Wahrscheinlichkeit einer Antwort vor  $t_c = 10$  in Abhängigkeit von den Persönlichkeitseigenschaften  $\theta$  und  $\omega$  bei quadratischer Parametrisierung mit Parametern  $k_g = 0.3$  und  $d_g = 0.0$  für verschiedene Einflussgewichte  $c_g$  und  $l_g$  (siehe Gleichung 4.24)

Das Modell ist identifiziert. Wenn man annimmt, dass die beobachtete Personenstichprobe aus einer Population stammt, in welcher die Verteilung der Persönlichkeitseigenschaften eine bestimmte Verteilungsform besitzt, können die



Itemparameter über den EM-Ansatz bestimmt werden. Details zur Schätzung finden sich im fünften Kapitel.

### Schätzung der Personenparameter

Wiederum kann bei bekannten Itemparametern anhand der dichotomisierten Reaktionszeiten einer Person deren Eigenschaftsausprägung bestimmt werden. Ein mögliches Schätzverfahren ist der Maximum-Likelihood-Ansatz. Wurde für eine Person bei  $G$  Items das Reaktionszeitmuster  $\mathbf{y} = [y_1, \dots, y_G]$  beobachtet, wobei  $y_g$  kodiert, ob eine Reaktion bei Item  $g$  vor Zeitlimit  $t_c$  erfolgte oder nicht, ist die Likelihood bestimmter Personenparameterausprägungen  $\theta$  und  $\omega$  definiert als:

$$L(\theta, \omega) = \prod_{i=g}^G P_g(\theta, \omega)^{y_g} (1 - P_g(\theta, \omega))^{1-y_g}. \quad (4.25)$$

Für  $P_g(\theta, \omega)$  ist dabei die Antwortwahrscheinlichkeit (4.24) zu verwenden. Bei der Maximum-Likelihood-Schätzung werden einer Person diejenigen Parameterausprägungen zugeschrieben, welche die Gleichung (4.25) maximieren. Die Abschätzung der Varianz des Maximum-Likelihood-Schätzers erfordert eine Berechnung der Fisher-Informationsmatrix des Tests. Jedes Item trägt zur Testinformationsmatrix mit seiner Iteminformationsmatrix bei. Diese Matrix besitzt bei Modell (4.24) die Gestalt:

$$\mathbf{I}_g(\theta, \omega) = \frac{[\exp[k_g + l_g(\theta - d_g)^2 + c_g\omega]]^2}{\exp[k_g + l_g(\theta - d_g)^2 + c_g\omega] - 1} \begin{bmatrix} 4l_g^2(\theta - d_g)^2 & 2l_g c_g(\theta - d_g) \\ 2l_g c_g(\theta - d_g) & c_g^2 \end{bmatrix}. \quad (4.26)$$

Details zur Berechnung finden sich in Anhang B.2.2. Der Verlauf der Iteminformationsfunktion über die Personenfähigkeit  $\theta$ , das Element  $I_{g[11]}(\theta, \omega)$  der Iteminformationsmatrix (4.26), ist in Abbildung 4.8 visualisiert.

Anhand Abbildung 4.8 lässt sich nur schwer erkennen, daß die Funktion  $I_{g[11]}(\theta, \omega)$  bimodal ist. Die Funktion besitzt ein Minimum von Null an der Stelle  $\theta = d_g$ . Diese Eigenschaft ist häufig bei Unfolding-Modellen zu finden. Verkuilen (2006) konnte sogar zeigen, dass dies für alle binären Unfolding-Modelle mit  $P(\theta = 0) < 1$  gilt. Mit unbegrenzt wachsender oder fallender Personenfähigkeit nähert sich die Funktion  $I_{g[11]}(\theta, \omega)$  bei konstantem Residualfaktor  $\omega$  ebenfalls dem Wert Null. Auch bei zunehmendem und abnehmendem Residualfaktor konvergiert  $I_{g[11]}(\theta, \omega)$  gegen Null, wenn man die Personenfähigkeit konstant hält. Dieser Aspekt ist zwar dem Schaubild nicht unmittelbar zu entnehmen, ergibt sich aber aufgrund von Grenzwertbildungen.

Die Fisher-Information des Tests bildet die Basis für die Berechnung der Varianz des Fähigkeitsschätzers  $\hat{\theta}$  einer Person. Das betrachtete Element  $I_{g[11]}(\theta, \omega)$  der Iteminformationsmatrix steht in Beziehung zur Varianz der Schätzung. Wie aus der generellen Beziehung (2.16) abgeleitet werden kann, können nur Items

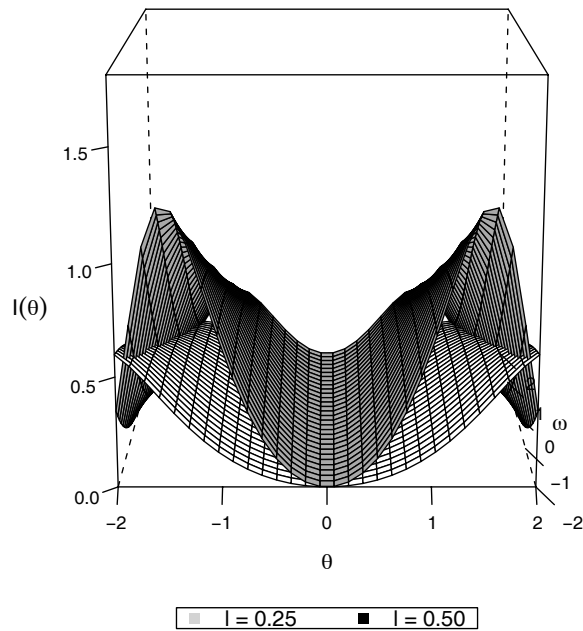


Abbildung 4.8: Fisher-Information über  $\theta$  bei dichotomisierten Reaktionszeiten und quadratischer Parametrisierung mit Parametern  $k_g = 0.30$ ,  $d_g = 0.00$ ,  $c_g = 0.25$  für verschiedene Einflussgewichte  $l_g$  (Element  $I_{g[11]}(\theta, \omega)$  der Informationsmatrix 4.26)

mit hohem Element  $I_{g[11]}(\theta, \omega)$  zu einer starken Reduktion der Varianz beitragen. Diese Überlegung führt zum Schluss, dass Items nicht zur Schätzung von Fähigkeitsausprägungen im Bereich ihrer Itemlokation  $d_g$  geeignet sind. Erst mit einem bestimmten Abstand zur Itemlokation steigt das diagnostische Potential eines Items an, um jedoch für extreme Fähigkeitsausprägungen wieder abzufallen. Der Bereich, in dem das Modell die Fähigkeitsschätzung verbessern kann, ist also im Vergleich zur linearen Parametrisierung wesentlich enger. Zudem wird wiederum die Varianz der Fähigkeitsschätzung dadurch vergrößert, dass zusätzlich der Residualfaktor geschätzt werden muss und die Nebendiagonalelemente der Iteminformationsmatrix nicht Null sind. Für eine detailliertere Betrachtung der Varianz des Fähigkeitsschätzers bei einem Test aus mehreren Items sei auf das siebte Kapitel verwiesen, wo das genaue Ausmaß der Varianzreduktion durch Berücksichtigung der Reaktionszeiten untersucht wird.

## 4.2 RT-Modelle auf Basis von Accelerated Failure Time Modellen

Accelerated Failure Time Modelle bilden die zweite große Klasse der Modelle für Überlebenszeiten. Grundannahme dieser Modelle ist eine log-lineare Beziehung zwischen der Überlebenszeit  $t$  und den Prädiktoren  $\mathbf{x}$  in der Form (Marubini & Valsecchi, 1995)

$$\log(t) = \mathbf{x}'\boldsymbol{\beta} + e, \quad (4.27)$$

wobei  $\boldsymbol{\beta}$  ein Vektor von Regressionsparametern und  $e$  eine nach einer bestimmten Form verteilte Zufallsvariable ist. Im Gegensatz zu generalisierten linearen Modellen wird bei Accelerated Failure Time Modellen also keine Transformation des Erwartungswertes der Reaktionszeiten modelliert, sondern der Erwartungswert der transformierten Reaktionszeiten.

Einfache Umformungen ergeben für die Überlebensfunktion  $S_{\mathbf{x}_1}(t)$  einer Person mit Prädiktorausprägung  $\mathbf{x} = \mathbf{x}_1$  im Vergleich zur Überlebensfunktion  $S_{\mathbf{0}}(t)$  einer Person mit Prädiktorausprägung  $\mathbf{x} = \mathbf{0}$  die Beziehung:

$$S_{\mathbf{x}_1}(t) = S_{\mathbf{0}}(\exp[-\mathbf{x}'_1\boldsymbol{\beta}] t). \quad (4.28)$$

Eine Person mit Prädiktorausprägung  $\mathbf{x} = \mathbf{x}_1$  besitzt also zum Zeitpunkt  $t$  dieselbe Überlebenswahrscheinlichkeit wie eine Person mit Prädiktorausprägung  $\mathbf{x} = \mathbf{0}$  zum Zeitpunkt  $\exp[-\mathbf{x}'_1\boldsymbol{\beta}] t$ . Die Funktion der Prädiktoren besteht somit in einer Streckung bzw. Stauchung der Zeitachse – eine Eigenschaft, von der sich der Name „Accelerated Failure Time Modell“ ableitet. Die Größe  $\exp(\beta_i)$  wird teilweise als Beschleunigungsfaktor bezeichnet. Es kann gezeigt werden (Klein & Moeschberger, 1997), dass bei Accelerated Failure Time Modellen für die Hazard-Rate folgende Beziehung gilt:

$$h_{\mathbf{x}}(t) = h_{\mathbf{0}}(\exp[-\mathbf{x}'\boldsymbol{\beta}]t) \exp[-\mathbf{x}'\boldsymbol{\beta}]. \quad (4.29)$$

Dabei steht  $h_{\mathbf{0}}(t)$  für die Hazard-Rate einer Person mit Prädiktorausprägung  $\mathbf{x} = \mathbf{0}$ . Folgt die Reaktionszeitverteilung einer Weibull-Verteilung, entspricht das Accelerated Failure Time Modell einem Proportional Hazard Modell, jedoch mit einer anderen Parametrisierung: Negative Regressionskoeffizienten im Proportional Hazard Modell entsprechen positiven Regressionskoeffizienten im Accelerated Failure Time Modell (Hougaard, 1999). Bei allen anderen Verteilungsformen schließen sich Proportional Hazard Modelle und Accelerated Failure Time Modelle aus.<sup>5</sup>

---

<sup>5</sup>Allerdings kann gezeigt werden, dass sowohl Accelerated Failure Time Modelle als auch Proportional Hazard Modelle zur Klasse der generellen Transformationsmodelle gehören (Doksum, 1987; Doksum & Gasko, 1990). Bei kleinen Regressionseffekten  $\beta$  können zudem Accelerated Failure Time Modelle durch Proportional Hazard Modelle mit zeitabhängigen Prädiktoren approximiert werden (Therneau & Grambsch, 2000, S. 149)

Aus der Annahme einer bestimmten Verteilungsform für die Residualkomponente  $e$  des log-linearen Modells in (4.27) ergibt sich eine bestimmte Verteilungsform der Reaktionszeiten. Normalverteilte Residuen führen zu log-normalverteilten Reaktionszeiten, während beispielsweise Extreme-Value-Type-I verteilte Residuen Weibull-verteilte Reaktionszeiten implizieren (Klein & Moeschberger, 1997, S. 46). Aufgrund ihrer großen Popularität sollen im Folgenden Modelle für log-normalverteilte Reaktionszeiten besprochen werden. Anschließend wird die Frage untersucht, wie sich Verteilungsannahmen vermeiden lassen.

#### 4.2.1 Accelerated Failure Time Modelle für log-normalverteilte Reaktionszeiten

Die zweiparametrische Log-Normalverteilung besitzt die Verteilungsdichte (N. Johnson et al., 1994):

$$f(t; \eta, \sigma) = \frac{1}{\sqrt{2\pi}\sigma t} \exp \left[ -\frac{1}{2} \frac{[\log(t) - \eta]^2}{\sigma^2} \right] \quad \text{mit } t > 0. \quad (4.30)$$

Log-normalverteilte Daten lassen sich bei einer Vielzahl von Anwendungsbereichen finden, siehe zum Beispiel N. Johnson et al. (1994), Limpert, Stahel und Abbt (2001) oder Crow und Shimizu (1988). Auch sind unterschiedliche Informationsverarbeitungsprozesse mit der Log-Normalverteilung vereinbar (Ulrich & Miller, 1993; van Breukelen, 1995). Für spezielle Anwendungen im Bereich der Testtheorie sei auf van Breukelen (2005), van der Linden (2006), Schnipke und Scrams (1997), Thissen (1984) oder Ferrando und Lorenzo-Sevas (2007a) verwiesen. Den Anwendungen in der Testtheorie ist gemein, dass sie die Parameter der Log-Normalverteilung auf Personen- und Itemparameter zurückführen. Kennzeichnend für die Ansätze ist, dass sowohl bei Thissen (1984) als auch bei van der Linden (2006) sowie bei Ferrando und Lorenzo-Sevas (2007a) der Parameter  $\sigma$  entweder als konstant über alle Items gilt oder als itemspezifisch, niemals aber über Personen variiert. Der Parameter  $\eta$  wird dagegen als Funktion von personen- und itemspezifischen Parametern modelliert. Diese Tradition, den Parameter  $\sigma$  als itemspezifisch anzunehmen und nicht auf Personeneigenschaften zurückzuführen, soll bei den folgenden Vorschlägen beibehalten werden.

#### Lineare Parametrisierung

Ausgangspunkt der weiteren Überlegungen ist die Annahme, dass die Verteilung der Reaktionszeit bei Aufgabe  $g$  einer Log-Normalverteilung folgt, deren Lokationsparameter  $\eta_g$  von der Personenfähigkeit  $\theta$  und dem Residualfaktor  $\omega$  einer Person abhängt:

$$f(t_g|\theta, \omega) = \frac{1}{\sqrt{2\pi}\sigma_g t_g} \exp \left[ -\frac{1}{2} \frac{[\log(t_g) - \eta_g(\theta, \omega)]^2}{\sigma_g^2} \right] \quad \text{mit } t_g > 0. \quad (4.31)$$

Der Skalenparameter  $\sigma_g$  wird in Gleichung (4.31) als Itemparameter aufgefasst, während der Lokationsparameter  $\eta_g(\theta, \omega)$  auf die Personenfähigkeit  $\theta$ , den Residualfaktor  $\omega$  und eine Itemkomponente, die Konstante  $k_g$ , zurückgeführt werden soll.<sup>6</sup> Angenommen wird eine lineare Beziehung zwischen dem Lokationsparameter  $\eta_g(\theta, \omega)$  und den Persönlichkeitseigenschaften  $\theta$  und  $\omega$ , so dass der bekannte lineare Prädiktor resultiert, diesmal jedoch mit einer anderen Link-Funktion:

$$\eta_g(\theta, \omega) = k_g + l_g\theta + c_g\omega. \quad (4.32)$$

Der Parameter  $k_g$  ist die Konstante einer Aufgabe, die Parameter  $l_g$  und  $c_g$  sind die Einflussgewichte der Persönlichkeitseigenschaften. Bei Modell (4.32) handelt es sich, wie bei einem Vergleich mit Gleichung (4.27) zu erkennen ist, um ein Accelerated Failure Time Modell mit zufälligen Personeneffekten (Skrondal & Rabe-Hesketh, 2004). Der Einfluss der Persönlichkeitseigenschaften auf die Reaktionszeitverteilung wird deutlich, wenn man die Veränderung des Erwartungswertes der Reaktionszeiten mit variierender Personenfähigkeit  $\theta$  und variierendem Residualfaktor  $\omega$  betrachtet. Für log-normalverteilte Zufallsvariablen hängt der Erwartungswert von den Parametern  $\eta_g$  und  $\sigma_g$  ab. Es gilt:

$$E(t_g|\theta, \omega) = \exp \left[ \eta_g(\theta, \omega) + \frac{\sigma_g^2}{2} \right] = \exp \left[ k_g + l_g\theta + c_g\omega + \frac{\sigma_g^2}{2} \right]. \quad (4.33)$$

Bei Accelerated Failure Time Modellen wird der Erwartungswert der transformierten Reaktionszeiten modelliert. Dies unterscheidet den Ansatz von generalisierten linearen Modellen, bei denen eine Transformation des Erwartungswertes auf einen linearen Prädiktor zurückgeführt wird. Der Erwartungswert der Reaktionszeiten kann daher bei Accelerated Failure Time Modellen nicht als alleinige Funktion des linearen Prädiktors  $\eta_g(\theta, \omega)$  geschrieben werden. Abbildung 4.9 gibt die Abhängigkeit der mittleren Reaktionszeit von den Persönlichkeitseigenschaften  $\theta$  und  $\omega$  für verschiedene Einflussgewichte  $l_g$  und  $c_g$  wieder. Gewählt wurden für die übrigen Parameter die Werte  $\sigma_g = 1$  und  $k_g = 2$ . Die Funktionen des Parameters  $k_g$  sowie der Parameter  $l_g$  und  $c_g$  sind offensichtlich. Der Parameter  $k_g$  bestimmt (neben  $\sigma_g^2$ ) die mittlere Reaktionszeit einer Person mit Eigenschaftsausprägungen  $\theta = 0$  und  $\omega = 0$ . Die Einflussgewichte  $l_g$  und  $c_g$  regulieren, wie stark die mittlere Reaktionszeit von den Ausprägungen der Persönlichkeitseigenschaften abhängt. Die Persönlichkeitseigenschaften repräsentieren jedoch im Gegensatz zu den Proportional Hazard Modellen nicht mehr positive Einflüsse auf die Bearbeitungsgeschwindigkeit, sondern negative Einflussfaktoren. Diese Interpretation ergibt sich daraus, dass die mittlere Reaktionszeit mit steigenden Eigenschaftsausprägungen anwächst. Diese Eigenschaft ist allerdings

---

<sup>6</sup>Streng genommen sind die Parameter  $\eta$  und  $\sigma^2$  bei der Log-Normalverteilung keine Lokations- und Skalenparameter. Dass diese Parameter dennoch als Lokations- und Skalenparameter bezeichnet werden, liegt daran, dass oft die logarithmierten Reaktionszeiten betrachtet werden, die normalverteilt sind mit Erwartungswert  $\mu = \eta$  und Varianz  $\sigma^2$ . Bei der Normalverteilung aber erfüllen die Parameter  $\mu$  und  $\sigma^2$  die Funktion von Lokations- und Skalenparametern.

nicht überraschend, wenn man die verschiedenartige Parametrisierung bei Proportional Hazard Modellen und Accelerated Failure Time Modellen bedenkt, siehe hierzu Hougaard (1999). Wie bereits erwähnt, entspricht ein negatives Einflussgewicht beim Proportional Hazard Modell einem positiven Einflussgewicht beim Accelerated Failure Time Modell. Auch bei dem aktuellen Reaktionszeitmodell konvergiert die mittlere Reaktionszeit gegen Null, diesmal allerdings mit abnehmender Personenfähigkeit und konstant gehaltenem Residualfaktor bzw. mit abnehmendem Residualfaktor und konstant gehaltener Personenfähigkeit. In diesem Aspekt vermag das Modell die Realität sicherlich nicht adäquat zu beschreiben, jedoch ist dabei zu bedenken, dass vermutlich alle Modelle nur unvollkommene Approximationen der Realität darstellen. Eine Einführung von Reaktionszeituntergrenzen ist zwar möglich und führt zur dreiparametrischen Log-Normalverteilung, erschwert die Modellschätzung aber erheblich.

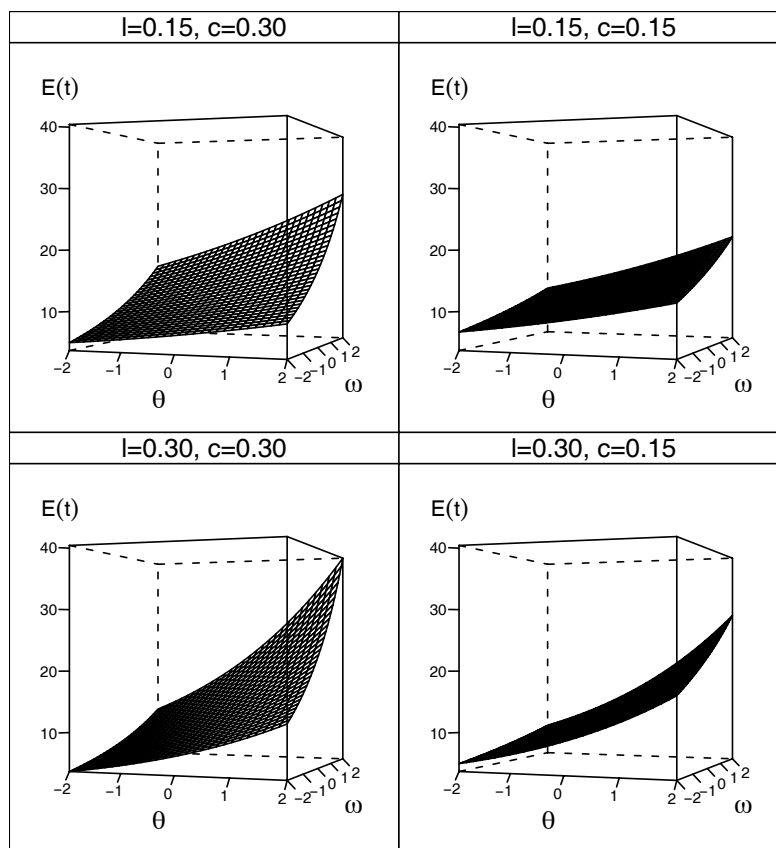


Abbildung 4.9: Log-normalverteilte RT: Mittlere Reaktionszeit in Abhängigkeit von den Persönlichkeitseigenschaften  $\theta$  und  $\omega$  bei linearer Parametrisierung mit Parametern  $k_g = 2$  und  $\sigma_g^2 = 1$  für verschiedene Einflussgewichte  $l_g$  und  $c_g$  (siehe Gleichung 4.33)

Die Schätzung der Itemparameter von Modell (4.32) ist durch die Anwen-

dung von Standardverfahren möglich, wenn man bestimmte Eigenschaften von log-normalverteilten Zufallsvariablen ausnutzt. Eine log-normalverteilte Zufallsvariable besitzt die Eigenschaft, dass sie durch Logarithmierung in eine normalverteilte Zufallsvariable transformiert wird mit Erwartungswert  $\eta_g(\theta, \omega)$  und Varianz  $\sigma_g^2$ . Nimmt man an, dass in einer Population möglicher Testteilnehmer die Personenfähigkeit  $\theta$  und der Residualfaktor  $\omega$  bivariat standardnormalverteilt sind mit Korrelation  $\rho = 0$ , resultiert als Marginalverteilung der logarithmierten Reaktionszeit von Item  $g$  nach Ausintegration der Persönlichkeitseigenschaften die Beziehung:

$$\log(t_g) \sim N(k_g, \sigma_g^2 + l_g^2 + c_g^2). \quad (4.34)$$

Für eine genaue Begründung dieser Beziehung sei auf Anhang A.2.1 verwiesen. Es folgt, dass einfache Maximum-Likelihood-Schätzer bzw. Momentmethoden-Schätzer für die Konstante  $k_g$  existieren. Die Einflussgewichte  $l_g$  und  $c_g$  der verschiedenen Aufgaben werden über die Kovarianzen der logarithmierten Reaktionszeiten bei verschiedenen Items festgelegt, welche sich nach der Beziehung  $\sigma(\log(t_g), \log(t_j)) = l_g l_j + c_g c_j$  berechnen lassen. Wie unschwer zu erkennen ist, entspricht das Modell in seiner Kovarianzstruktur einem orthogonalen faktorenanalytischen Modell – für eine genaue Darstellung der Fundamentalgleichung der Faktorenanalyse siehe zum Beispiel R. Johnson und Wichern (2002, S. 479). Da sich die Kovarianzmatrix der logarithmierten Reaktionszeiten konsistent schätzen lässt, ergeben sich bei einer Faktorisierung der Kovarianzmatrix konsistente Schätzer für die Einflussgewichte. Die genaue Schätzung der Modellparameter ist an dieser Stelle bewusst kurz dargestellt und wird im sechsten Kapitel erneut aufgegriffen.

### Schätzung der Personenparameter

Wenn die Itemparameter bekannt sind, lassen sich die Persönlichkeitseigenschaften einer Person anhand des Reaktionszeitmusters der Person schätzen. Ein Verfahren zur Bestimmung der Persönlichkeitseigenschaften ist die Maximum-Likelihood-Schätzung. Werden für eine Person die Reaktionszeiten  $\mathbf{t} = [t_1, \dots, t_G]$  bei  $G$  Testaufgaben beobachtet, lautet die Likelihood-Funktion für die Personenparameter  $\theta$  und  $\omega$  bei Annahme der bedingten Unabhängigkeit

$$L(\theta, \omega) = \prod_{g=1}^G f(t_g; \eta_g(\theta, \omega), \sigma_g), \quad (4.35)$$

wobei  $f(t_g; \eta_g(\theta, \omega), \sigma_g)$  die Verteilungsdichte der Log-Normalverteilung (4.30) mit Lokationsparameter  $\eta_g(\theta, \omega) = k_g + l_g \theta + c_g \omega$  und Skalenparameter  $\sigma_g$  ist. Als Schätzer werden die Ausprägungen  $\hat{\theta}$  und  $\hat{\omega}$  der Persönlichkeitseigenschaften bestimmt, welche die Likelihood-Funktion maximieren. Die Genauigkeit der Fähigkeitsschätzung hängt von der Varianz des Fähigkeitsschätzers ab. Diese Varianz kann abgeschätzt werden, wenn die Fisher-Information des Tests bestimmt wird. Der Beitrag eines einzelnen Items zur Fisher-Information des Tests besitzt bei Modell (4.32) folgende Iteminformationsmatrix (eine genaue Herleitung findet sich in Anhang B.3.1):

$$\mathbf{I}_g(\theta, \omega) = \frac{1}{\sigma_g^2} \begin{bmatrix} l_g^2 & l_g c_g \\ l_g c_g & c_g^2 \end{bmatrix}. \quad (4.36)$$

Der Verlauf des Elementes  $I_{g[11]}(\theta, \omega)$  der Iteminformationsmatrix (4.36) über die Fähigkeits- und Residualfaktordimension ist in Abbildung 4.10 für  $\sigma_g^2 = 1$  und verschiedene Ausprägungen des Einflussgewichts  $l_g$  illustriert. Wie im Falle exponentialverteilter Reaktionszeiten ist die Funktion  $I_{g[11]}(\theta, \omega)$  konstant für unterschiedliche Werte von  $\theta$  und  $\omega$ . Die Höhe von  $I_{g[11]}(\theta, \omega)$  steigt mit zunehmendem Einflussgewicht  $l_g$  und fällt mit wachsender Varianz  $\sigma_g^2$ . Mit zunehmender „Kommunalität“ und sinkender Spezifität (um die Terminologie der Faktorenanalyse zu verwenden) wächst somit auch die Iteminformation über  $\theta$ .

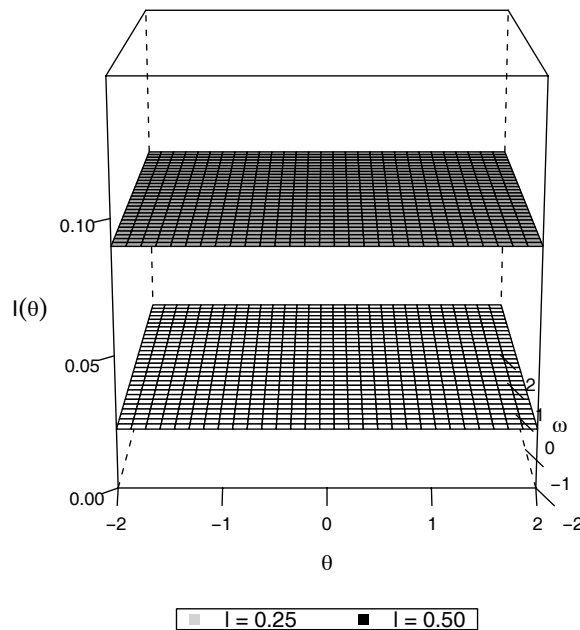


Abbildung 4.10: Fisher-Information über  $\theta$  bei log-normalverteilten Reaktionszeiten und linearer Parametrisierung mit Parameter  $\sigma_g^2 = 1$  für verschiedene Einflussgewichte  $l_g$  (Element  $I_{g[11]}(\theta, \omega)$  der Informationsmatrix 4.36)

Bewertet man die Bedeutung von Abbildung 4.10 für die Varianz des Fähigkeitsschätzers  $\hat{\theta}$ , kommt man zum Schluss, dass der Informationsverlauf günstig ist. Aufgrund des konstanten Verlaufes trägt ein Item bei allen Ausprägungen von  $\theta$  und  $\omega$  zur Reduktion der Varianz der Schätzer bei – im Gegensatz zum zweiparametrischen Logit-Modell (Abbildung 3.2) gibt es somit keinen Wertebereich der Personenfähigkeit, in denen ein Testitem keine diagnostische Bedeutung besitzt. Dadurch kann die Reaktionszeit die Itemantwort bei der Schätzung der Personenfähigkeit optimal ergänzen – eine Eigenschaft, auf die



im siebten Kapitel näher eingegangen wird. Der genaue Beitrag eines Items zur Varianzreduktion hängt jedoch noch von den Nebendiagonalelementen der Iteminformationsmatrix ab, hohe Nebendiagonalelemente verringern den Nutzen der Reaktionszeit.

### Quadratische Parametrisierung

Auch bei log-normalverteilten Reaktionszeiten kann für die mittlere Reaktionszeit eine Unfolding-Beziehung zur Personenfähigkeit angenommen werden, wie sie im Bereich der Einstellungsmessung plausibel ist und wofür sich erste Hinweise finden (Akrami et al., 2007; Ferrando & Lorenzo-Sevas, 2007a, 2007b). Eine derartige Beziehung kann über eine quadratische Funktion modelliert werden. Dabei wird der Lokationsparameter  $\eta_g$  der Log-Normalverteilung (4.30) bei Aufgabe  $g$  auf die Personenfähigkeit  $\theta$  und den Residualfaktor  $\omega$  einer Person zurückgeführt über die Funktion:

$$\eta_g(\theta, \omega) = k_g + l_g(\theta - d_g)^2 + c_g\omega. \quad (4.37)$$

Bei dem Parameter  $k_g$  handelt es sich um die Konstante einer Aufgabe,  $d_g$  ist die Itemlokation und  $l_g$  und  $c_g$  sind die Einflussgewichte der Persönlichkeitseigenschaften. Die Parametrisierung von Modell (4.37) kann mit denselben Argumenten gerechtfertigt werden wie die quadratischen Proportional Hazard Modelle. Bei log-normalverteilten Reaktionszeiten ist der Erwartungswert der Reaktionszeit  $t_g$  eine Funktion der Parameter  $\eta_g$  und  $\sigma_g$ . Es gilt:

$$E(t_g|\theta, \omega) = \exp \left[ \eta_g(\theta, \omega) + \frac{\sigma_g^2}{2} \right] = \exp \left[ k_g + l_g(\theta - d_g)^2 + c_g\omega + \frac{\sigma_g^2}{2} \right]. \quad (4.38)$$

Der Einfluss der Persönlichkeitseigenschaften auf die mittlere Reaktionszeit ist in Abbildung 4.11 wiedergegeben. Gewählt wurden für diese Abbildung die Itemparameter  $k_g = 2$  und  $\sigma_g^2 = 1$ . Möchte man eine umgekehrt U-förmige Beziehung zwischen der mittleren Reaktionszeit und der Personenfähigkeit  $\theta$ , so ist für das Einflussgewicht  $l_g$  ein negativer Wert zu wählen.

Die Komponente  $k_g$  reguliert bei negativem Einflussgewicht  $l_g$  den maximalen Erwartungswert der Reaktionszeit einer Person mit Personenfähigkeit  $\theta = d_g$  und Residualfaktor  $\omega = 0$ . Damit bestimmt der Parameter das Grundniveau der Reaktionszeiten, das neben der Itemschwierigkeit von Itemspezifika wie der Itemlänge abhängen dürfte – für eine Übersicht über mögliche Einflussfaktoren auf das Grundniveau der Reaktionszeiten siehe beispielsweise Dunn et al. (1972) oder Ferrando und Demestre (2007). Die Itemlokation  $d_g$  bestimmt als Lokationsparameter die Lage der Funktion, das Einflussgewicht  $c_g$  die Stärke des Abfalls der Reaktionszeiten bei sinkendem Residualfaktor. Auch bei diesem Modell konvergiert die mittlere Reaktionszeit bei größer werdendem Abstand von Person und Item auf der Fähigkeitsdimension gegen Null (wenn die Ausprägung des Residualfaktors konstant gehalten wird). Denselben Grenzwert besitzt

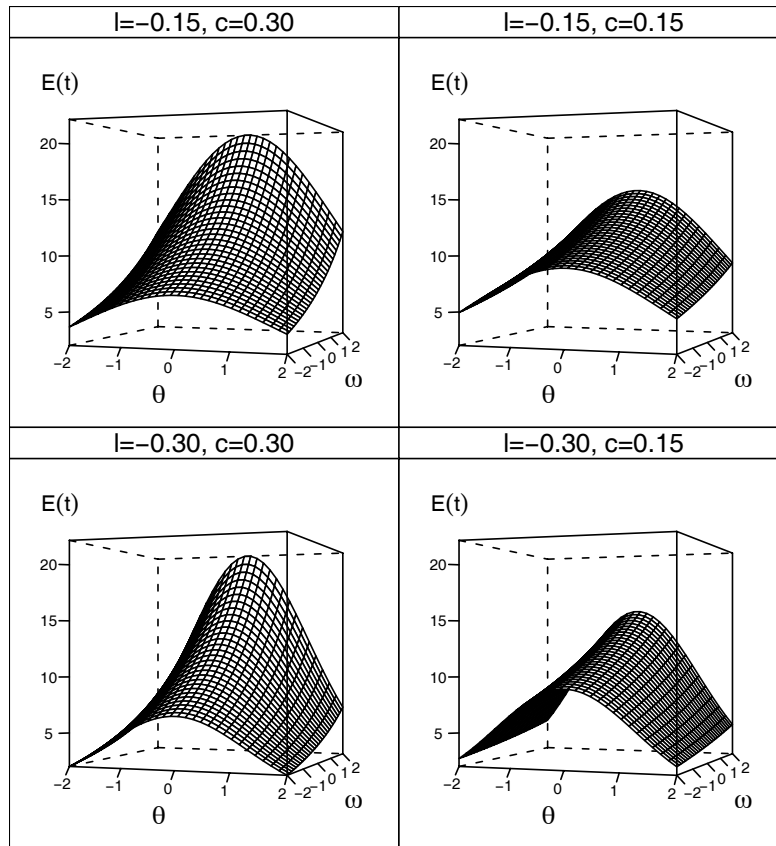


Abbildung 4.11: Log-normalverteilte RT: Mittlere Reaktionszeit in Abhängigkeit von den Persönlichkeitseigenschaften  $\theta$  und  $\omega$  bei quadratischer Parametrisierung mit Parametern  $k_g = 2$  und  $\sigma_g^2 = 1$  für verschiedene Einflussgewichte  $l_g$  und  $c_g$  (siehe Gleichung 4.38)

die mittlere Reaktionszeit bei abnehmendem Residualfaktor  $\omega$ . Diese unrealistische Eigenschaft des Modells kann kaum vermieden werden, wenn nicht die Modellschätzung durch die Einführung von Reaktionszeituntergrenzen erheblich erschwert werden soll.

Wie sich zeigen lässt, ist das Modell identifiziert. Als einfache Methode der Parameterschätzung bietet sich eine Schätzung nach der Momentmethode an. Für eine genaue Darstellung der Parameterschätzung sei auf Anhang A.2.2 verwiesen. Da log-normalverteilte Reaktionszeiten nach Logarithmierung normalverteilt sind mit Erwartungswert  $E(\log(t_g)) = \eta_g$ , kann Modell (4.37) auch als nonlineares faktorenanalytisches Modell für die logarithmierten Reaktionszeiten interpretiert werden. Nonlineare faktorenanalytische Modelle wurden von M. Davidson (1977) oder von Maraun und Rossi (2001) behandelt.

### Schätzung der Personenparameter

Bei Beobachtung der Reaktionszeiten einer Person können diese Beobachtungen genutzt werden, um die Ausprägungen der Persönlichkeitseigenschaften zu bestimmen. Wenn die Itemparameter des RT-Modells bekannt sind, kann hierzu der Maximum-Likelihood-Ansatz genutzt werden. Ausgangspunkt der Schätzung ist die Likelihood-Funktion, die bei gegebenem Reaktionszeitmuster  $\mathbf{t} = [t_1, \dots, t_G]$  jeder Ausprägung von  $\theta$  und  $\omega$  eine Plausibilität zuordnet. Die Likelihood-Funktion des Modells lautet bei bedingter Unabhängigkeit der Reaktionszeiten

$$L(\theta, \omega) = \prod_{g=1}^G f(t_g; \eta_g(\theta, \omega), \sigma_g), \quad (4.39)$$

wobei für  $f(t_g; \eta_g(\theta, \omega), \sigma_g)$  die Verteilungsdichte der Log-Normalverteilung (4.30) mit Lokationsparameter  $\eta(\theta, \omega) = k_g + l_g(\theta - d_g)^2 + c_g$  und Skalenparameter  $\sigma_g$  eingesetzt werden muss. Gewählt werden die Parameterausprägungen  $\hat{\theta}$  und  $\hat{\omega}$ , welche die Likelihood-Funktion für ein gegebenes Reaktionszeitmuster  $\mathbf{t}$  maximieren. Die Genauigkeit der Schätzung hängt dabei von der Fisher-Information der Reaktionszeiten ab. Bei quadratischer Parametrisierung berechnet sich der Beitrag eines Items zur Testinformation, die Iteminformationsmatrix, als:

$$\mathbf{I}_g(\theta, \omega) = \frac{1}{\sigma_g^2} \begin{bmatrix} 4l_g^2(\theta - d_g)^2 & 2l_g c_g(\theta - d_g) \\ 2l_g c_g(\theta - d_g) & c_g^2 \end{bmatrix}. \quad (4.40)$$

Genauere Details zur Herleitung von Beziehung (4.40) finden sich in Anhang B.3.2. Abbildung 4.12 gibt den Verlauf des Elementes  $I_{g[11]}(\theta, \omega)$  der Iteminformationsmatrix (4.40) für die Itemparameter  $d_g = 0$ ,  $\sigma_g^2 = 1$  und verschiedene Einflussgewichte  $l_g$  wieder. Die Itemlokation  $d_g$  bestimmt alleine die Lage des Minimums von  $I_{g[11]}(\theta, \omega)$ , so dass auf eine Demonstration des Effektes unterschiedlicher Werte verzichtet werden kann.

Das Element  $I_{g[11]}(\theta, \omega)$  der Iteminformationsmatrix nimmt mit wachsender Varianz  $\sigma_g^2$  ab. Mit zunehmendem Abstand von Person und Item auf dem Fähigkeitskontinuum steigt  $I_{g[11]}(\theta, \omega)$  an. Für Fähigkeiten im Bereich der Itemlokation nimmt  $I_{g[11]}(\theta, \omega)$  jedoch unabhängig von der Ausprägung des Residualfaktors nur geringe Werte an. Bewertet man dieses Verhalten in Hinblick auf die mögliche Varianzreduktion der Fähigkeitsschätzung durch Einbezug der Reaktionszeit, bedeutet dies, dass für Fähigkeiten im Bereich von  $d_g$  kaum Verbesserungen bei der Schätzung zu erzielen sind. Allerdings nehmen die Funktionswerte von  $I_{g[11]}(\theta, \omega)$  mit extremer werdender Fähigkeiten schnell zu. Auch wenn die tatsächliche Varianzreduktion noch von den Nebendiagonalelementen der Informationsmatrix  $\mathbf{I}_g(\theta, \omega)$  abhängt, deutet dies doch darauf hin, dass besonders extreme Fähigkeitsausprägungen durch die Verwendung von Reaktionszeiten präzise geschätzt werden können. Dadurch unterscheidet sich das Modell erheblich vom zweiparametrischen Logit-Modell, bei welchem sich im Vergleich zur Itemlokation extreme Fähigkeitsausprägungen nicht präzise bestimmen lassen.

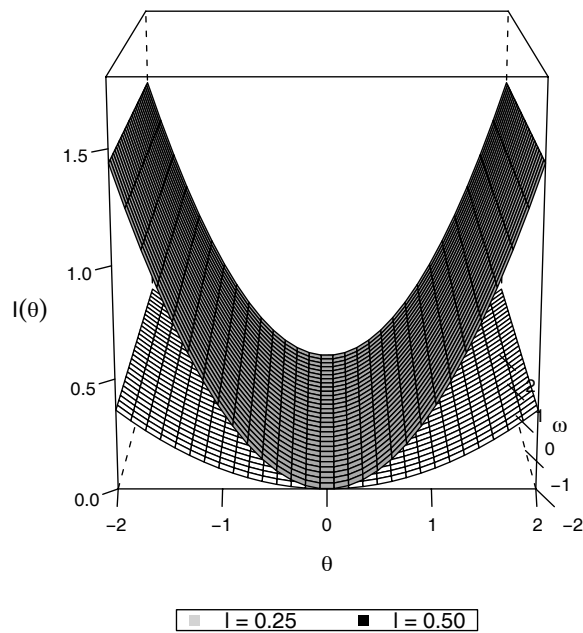


Abbildung 4.12: Fisher-Information über  $\theta$  bei log-normalverteilten Reaktionszeiten und quadratischer Parametrisierung mit Parametern  $d_g = 0$  und  $\sigma_g^2 = 1$  für verschiedene Einflussgewichte  $l_g$  (Element  $I_{g[11]}(\theta, \omega)$  der Informationsmatrix 4.40)

#### 4.2.2 Quasi-Likelihood-Ansätze für Accelerated Failure Time Modelle

Bei Accelerated Failure Time Modellen werden in der Regel Annahmen über die Verteilung der Reaktionszeit gemacht. Diese Festlegung auf eine bestimmte Verteilungsform ist jedoch eine Einschränkung, die in der Praxis gerne vermieden wird. Aus diesem Grund wurden Versuche unternommen, Schätzmethoden für verteilungsfreie Accelerated Failure Time Modelle zu entwickeln. Erste Vorschläge gehen in die 90er Jahre zurück und mittlerweile gibt es eine Vielzahl unterschiedlicher Schätzmethoden – für eine Übersicht siehe Jin (2005). Die vorgeschlagenen Schätzmethoden sind jedoch mit erheblichem Aufwand und mathematischen Schwierigkeiten verbunden, so dass ein Einzug dieser Verfahren in den Standardmethodenkanon bislang ausgeblieben ist. Zudem ist eine Berücksichtigung von Zufallseffekten problematisch. Die Schwierigkeit beim Aufstellen von verteilungsfreien Accelerated Failure Time Modellen besteht jedoch vorwiegend in der Behandlung von zensierten Zeiten. Zensierte Zeiten treten bei Personen auf, bei denen nicht der genaue Eintrittszeitpunkt eines Ereignisses bekannt ist, sondern nur feststeht, dass das Ereignis nach einem bestimmten Zeitpunkt eingetreten ist. Bei der Modellierung von Reaktionszeiten bei Tests

spielen zensierte Zeiten nur eine untergeordnete Rolle. In der Regel werden alle Items von den Versuchspersonen beantwortet. Nur bei so genannten Speed-Tests ist dies nicht der Fall. Sind keine zensierten Reaktionszeiten vorhanden, wird der Quasi-Likelihood-Ansatz auf die Modellierung der Reaktionszeiten bei Tests anwendbar.

Der Ansatz der Quasi-Likelihood-Schätzung geht auf Wedderburn (1974) zurück und dient dem Schätzen von Regressionsmodellen. Notwendig sind zwei Annahmen: Erstens die Annahme, dass sich der bedingte Erwartungswert einer Zufallsvariablen  $y$  auf eine bis auf unbekannte Parameter  $\beta$  bekannte Funktion mehrerer Prädiktoren  $\mathbf{x}$  zurückführen lässt:

$$E(y|\mathbf{x}) = f(\mathbf{x}, \beta). \quad (4.41)$$

Die Funktion  $f(\mathbf{x}, \beta)$  hängt von den unbekanntem Parametern  $\beta$  ab, die geschätzt werden sollen. Die zweite Annahme ist das Postulat, dass die Varianz von  $y$  eine bekannte Funktion des Erwartungswertes  $V(E(y|\mathbf{x}))$  ist. Als „Quasi-Log-Likelihood-Funktion“ der Beobachtung  $y$  definiert Wedderburn (1974):<sup>7</sup>

$$QL(\beta) = \int^{E(y|\mathbf{x})} \frac{y - m}{V(m)} dm. \quad (4.42)$$

Das Rational dieser Definition ist der Umstand, dass die Ableitung der Quasi-Log-Likelihood-Funktion nach dem Parameter  $\beta_i$ , die so genannte Quasi-Score-Funktion, äquivalente Eigenschaften besitzt wie die Score-Funktion bei der Maximum-Likelihood-Schätzung. Die Quasi-Score-Funktion erhält man durch Ableitung der Quasi-Log-Likelihood-Funktion nach dem Parameter  $\beta_i$  als:

$$QS(\beta_i, y) = \frac{\partial}{\partial \beta_i} QL(\beta) = \frac{y - E(y|\mathbf{x})}{V(E(y|\mathbf{x}))} \frac{\partial}{\partial \beta_i} E(y|\mathbf{x}). \quad (4.43)$$

Hierbei lässt sich  $E(y|\mathbf{x})$  auf  $\mathbf{x}$  über die Beziehung  $E(y|\mathbf{x}) = f(\mathbf{x}, \beta)$  zurückführen. Enthält die Regressionsfunktion mehrere unbekannte Parameter  $\beta' = [\beta_1, \beta_2, \dots]$ , ist die Score-Funktion ein Vektor, dessen einzelne Elemente aus den partiellen Ableitungen (4.43) bestehen. Es kann gezeigt werden (Wedderburn, 1974), dass die Elemente (4.43) der Quasi-Score-Funktion Eigenschaften besitzen, die auch die Score-Funktion bei der Maximum-Likelihood-Schätzung aufweist (siehe 2.7). Diese Eigenschaften sind:

$$E[QS(\beta_i, y)] = 0 \quad (4.44)$$

$$\sigma^2[QS(\beta_i, y)] = -E \left[ \frac{\partial^2}{\partial \beta_i^2} QL(\beta) \right] = \frac{1}{V(E(y|\mathbf{x}))} \left[ \frac{\partial}{\partial \beta_i} E(y|\mathbf{x}) \right]^2 \quad (4.45)$$

Die Ableitungen werden dabei immer an der Stelle des wahren Parameters  $\beta_i$  berechnet. Bei unabhängig verteilten Beobachtungen  $\mathbf{y} = [y_1, \dots, y_G]$  werden

<sup>7</sup>Wedderburn (1974) verzichtet in seinem Artikel bei der Definition der Quasi-Likelihood-Funktion auf eine untere Grenze des Integrals. Diese Darstellung wird für die Arbeit übernommen.

die Score-Funktionen der  $G$  Beobachtungen wie bei der Maximum-Likelihood-Schätzung addiert. Der Schätzwert des Quasi-Maximum-Likelihood-Schätzers  $\hat{\beta}_i$  ist der Wert, für welchen die Summe der Score-Funktionen der einzelnen Beobachtungen Null ergibt. Die Varianz des Schätzers  $\hat{\beta}_i$  kann über die „Quasi-Information“ berechnet werden. Als „Quasi-Information“ einer Zufallsvariablen  $y_g$  wird die Varianz-Kovarianz-Matrix der Elemente der Score-Funktion bezeichnet. Nach Beziehung (4.45) gilt für die Quasi-Information der Beobachtung  $y_g$ :

$$\mathbf{QI}_g(\beta) = -E \left[ \frac{\partial^2}{\partial \beta \partial \beta'} QL(\beta, y_g) \right]. \quad (4.46)$$

Bei mehreren Prädiktoren handelt es sich bei der Quasi-Information einer Beobachtung also um eine Matrix, deren Elemente aus dem Erwartungswert der Hesse-Matrix der Quasi-Likelihood-Funktion bestehen. Die Quasi-Information von  $G$  unabhängigen Zufallsvariablen  $\mathbf{y} = [y_1, \dots, y_G]$  ergibt sich durch Addition der Quasi-Informationen der einzelnen Beobachtungen:

$$\mathbf{QI}(\beta) = \sum_{g=1}^G \mathbf{QI}_g(\beta) \quad (4.47)$$

$$\text{mit } \mathbf{QI}_g(\beta) = -E \left[ \frac{\partial^2}{\partial \beta \partial \beta'} \mathbf{QL}(\beta, y_g) \right].$$

Die Begründung des Konzeptes der Quasi-Information beruht darauf, dass die Varianz-Kovarianz-Matrix der Quasi-Maximum-Likelihood-Schätzer der Regressionsgewichte  $\hat{\beta}$  approximativ  $\mathbf{QI}^{-1}(\beta)$  ist, siehe hierzu Wedderburn (1974). Es kann gezeigt werden, dass die Quasi-Information geringer ausfällt als die Fisher-Information bei Bekanntheit der exakten Verteilung der Zufallsvariablen  $y_g$ . Lediglich im Falle von Verteilungen aus der Exponentialverteilungsfamilie stimmt die Quasi-Likelihood-Methode mit der Maximum-Likelihood-Methode überein (Wedderburn, 1974). Da Accelerated Failure Time Modelle Regressionsmodelle sind, kann der Ansatz der Quasi-Maximum-Likelihood-Schätzung einfach auf diese Modelle übertragen werden. Dies soll anhand der beiden Accelerated Failure Time Modelle für log-normalverteilte Reaktionszeiten, dem Modell mit linearer Parametrisierung und dem Modell mit quadratischer Parametrisierung, demonstriert werden. Die Rolle der unbekannt Parameter  $\beta$  übernehmen dabei, wenn die Itemparameter der Aufgaben bekannt sind, die unbekannt Eigenschaften  $\theta$  und  $\omega$  einer Person. Allerdings handelt es sich bei dem Quasi-Maximum-Likelihood-Ansatz um einen allgemeineren Ansatz, der nicht an die Gültigkeit einer bestimmten Verteilungsform gebunden ist.

### Lineare Parametrisierung

Der Quasi-Maximum-Likelihood-Ansatz spielt bei der Analyse von Überlebenszeiten keine große Rolle. In Studien zu Überlebenszeiten sind die Beobachtungen teilweise zensiert, was besondere verteilungsfreie Schätzverfahren erfordert – für

eine Übersicht siehe zum Beispiel Jin (2005). Geht man jedoch von unzensurierten Reaktionszeiten aus, lässt sich der Ansatz der Quasi-Maximum-Likelihood-Schätzung auf die Accelerated Failure Time Modelle anwenden. Kernbestandteil des Accelerated Failure Time Modells (4.27) mit linearer Parametrisierung (4.32) war die Annahme, dass die logarithmierte Reaktionszeit bei Aufgabe  $g$  der Beziehung

$$\log(t_g) = k_g + l_g\theta + c_g\omega + e_g \quad (4.48)$$

folgt, wobei  $e_g$  eine zufällig verteilte Residualkomponente ist. Fordert man für die Residualkomponente  $e_g$  eine bestimmte Verteilung und sind die Itemparameter bekannt, kann die Schätzung der Personenparameter  $\theta$  und  $\omega$  über einen herkömmlichen Maximum-Likelihood-Ansatz erfolgen. Dies wurde in Kapitel 4.2.1 für log-normalverteilte Reaktionszeiten gezeigt. Bei einer Quasi-Maximum-Likelihood-Schätzung bleibt die genaue Verteilungsform von  $e_g$  hingegen unspezifiziert. Bedingung ist lediglich, dass der Erwartungswert der Residualkomponente  $e_g$  Null ist, was durch eine geeignete Konstante  $k_g$  jederzeit erreicht werden kann. Nimmt man ferner an, dass bei den einzelnen Aufgaben die Varianz  $\sigma^2$  der Residualkomponente  $e_g$  für alle Personen und Aufgaben identisch<sup>8</sup> ausfällt und endlich ist, ergibt sich als Quasi-Score-Funktion (siehe 4.43) einer Aufgabe  $g$  der Vektor:

$$\mathbf{QS}(\theta, \omega, \log(t_g)) = \frac{\log(t_g) - (k_g + l_g\theta + c_g\omega)}{\sigma^2} \begin{bmatrix} l_g \\ c_g \end{bmatrix}. \quad (4.49)$$

Die unbekannt Parameter der Quasi-Score-Funktion sind die Persönlichkeitseigenschaften  $\theta$  und  $\omega$  – die Itemparameter  $k_g$ ,  $l_g$ ,  $c_g$  und  $\sigma^2$  werden als bekannt vorausgesetzt. Begründet werden kann dies damit, dass sich die Itemparameter mit hinreichender Genauigkeit in Vorstudien schätzen lassen. Eine Schätzung der Itemparameter ist bei dem Reaktionszeitmodell (4.48) über die Methoden der Momente möglich. Tatsächlich können die Itemparameter mit demselben Ansatz geschätzt werden, wie die Itemparameter des Modells für log-normalverteilte Reaktionszeiten, wenn man bestimmte Annahmen über die Verteilung der Persönlichkeitseigenschaften macht. Annahmen über die Verteilung der Residualkomponenten der einzelnen Items sind hierfür nicht notwendig. Näheres zur Schätzung findet sich in Anhang A.2.1.

Bei Bedingung auf die Persönlichkeitseigenschaften  $\theta$  und  $\omega$  werden die Reaktionszeiten einer Person bei den einzelnen Aufgaben als voneinander unabhängig angenommen. Gleichung (4.49) erfüllt somit alle Bedingungen, die bei einer Anwendung des Quasi-Likelihood-Ansatzes von Wedderburn (1974) gegeben sein müssen. Tatsächlich handelt es sich sogar um einen einfachen Spezialfall des Quasi-Likelihood-Ansatzes, da die Varianz einer Beobachtung  $V(E(y|\mathbf{x})) = \sigma^2$  nicht von dem bedingten Erwartungswert der Beobachtung abhängt.

<sup>8</sup>Die Annahme identischer Varianzen von  $e_g$  bei unterschiedlichen Aufgaben stellt keine Einschränkung dar, sondern ist durch eine Lineartransformation der logarithmierten Reaktionszeiten jederzeit erreichbar. Besitzen die Residualkomponenten  $e_g$  bei verschiedenen Items unterschiedliche Varianzen  $\sigma_g^2$ , können die logarithmierten Reaktionszeiten bei Kenntnis von  $\sigma_g^2$  durch die Standardabweichung  $\sigma_g$  dividiert werden.

### Schätzung der Personenparameter

Einen Quasi-Maximum-Likelihood-Schätzer für die Persönlichkeitseigenschaften  $\theta$  und  $\omega$  einer Testperson erhält man bei bekannten Itemparametern dadurch, dass die Ausprägungen der Eigenschaften gewählt werden, für welche die Summe der Quasi-Score-Funktionen der einzelnen Aufgaben Null ergibt. Technisch gesehen handelt es sich bei diesem Schätzwert um einen Least-Squares-Schätzer (Wedderburn, 1974).

Der Informationsbeitrag einer Aufgabe zur Testinformation wurde bei der Quasi-Likelihood-Schätzung über Gleichung (4.46) definiert. Bei linearer Parametrisierung des Erwartungswertes der Reaktionszeit ergibt sich für die Quasi-Information eines Items die Informationsmatrix:

$$\mathbf{QI}_g(\theta, \omega) = \frac{1}{\sigma^2} \begin{bmatrix} l_g^2 & l_g c_g \\ l_g c_g & c_g^2 \end{bmatrix}. \quad (4.50)$$

Dies entspricht genau der Fisher-Information bei log-normalverteilten Reaktionszeiten (siehe Gleichung 4.36), so dass auf eine graphische Abbildung der Quasi-Informationsfunktion verzichtet werden kann. Die Übereinstimmung von Fisher-Information und Quasi-Information ist keine Überraschung, da bei normalverteilten Residualkomponenten der Maximum-Likelihood-Ansatz mit dem Quasi-Likelihood-Ansatz identisch ist. Bei anderen Verteilungen der Residualkomponente  $e_g$  hingegen kann der Informationsbeitrag der Quasi-Information  $\mathbf{QI}_g(\theta, \omega)$  geringer ausfallen als die Fisher-Information  $\mathbf{I}_g(\theta, \omega)$ . Bezüglich der Quasi-Information eines Items kann nur wiederholt werden, was bereits über die Fisher-Information des Modells für log-normalverteilte Reaktionszeiten gesagt wurde. Aufgrund der Konstanz der Quasi-Information für unterschiedliche Ausprägungen der Persönlichkeitseigenschaften fällt die Varianzreduktion bei Berücksichtigung der Reaktionszeit für alle Fähigkeitsausprägungen identisch aus. Im Gegensatz zum zweiparametrischen Logit-Modell lassen sich extreme Eigenschaftsausprägungen mit derselben Präzision diagnostizieren wie weniger extreme.

### Quadratische Parametrisierung

Der Quasi-Likelihood-Ansatz lässt sich auch auf Modell (4.37) mit quadratischer Beziehung zwischen dem Erwartungswert der logarithmierten Reaktionszeit und der Distanz von Person zu Item auf der Fähigkeitsdimension übertragen. Die Bedingungen des Quasi-Likelihood-Ansatzes sind in diesem Fall natürlich genauso erfüllt, wie bei dem Modell (4.32) mit linearer Parametrisierung. Die Quasi-Score-Funktion von Aufgabe  $g$  lautet bei dem Modell:

$$\mathbf{QS}(\theta, \omega, \log(t_g)) = \frac{\log(t_g) - (k_g + l_g(\theta - d_g)^2 + c_g\omega)}{\sigma^2} \begin{bmatrix} 2l_g(\theta - d_g) \\ c_g \end{bmatrix}. \quad (4.51)$$

Bezüglich der Parameter gilt, was bereits im Falle der linearen Parametrisierung gesagt wurde: Die Itemparameter gelten als bekannt, lediglich die Ausprägung der Persönlichkeitseigenschaften  $\theta$  und  $\omega$  einer Person sind zu schätzen.



Dass die Itemparameter als bekannt vorausgesetzt werden, kann damit gerechtfertigt werden, dass diese konsistent geschätzt werden können. Mehr hierzu im sechsten Kapitel.

### Schätzung der Personenparameter

Die Präzision, mit der anhand der Reaktionszeiten einer Person die Ausprägungen der Eigenschaften  $\theta$  und  $\omega$  bestimmt werden können, hängt von der Quasi-Information eines Items ab. Als Quasi-Informationsfunktion eines Items ergibt sich bei Anwendung von Formel (4.46) die Informationsmatrix:

$$\mathbf{QI}_g(\theta, \omega) = \frac{1}{\sigma^2} \begin{bmatrix} 4l_g^2(\theta - d_g)^2 & 2l_g c_g(\theta - d_g) \\ 2l_g c_g(\theta - d_g) & c_g^2 \end{bmatrix}. \quad (4.52)$$

Wiederum findet sich eine Äquivalenz zur Fisher-Information bei log-normalverteilten Variablen (4.40), so dass auf eine Abbildung des Informationsverlaufs verzichtet werden kann. Die Überlegungen, die zum diagnostischen Potential der Reaktionszeiten bei Gültigkeit des log-normalverteilten Modells mit quadratischer Parametrisierung angestellt wurden, lassen sich somit ohne Einschränkung auf die Quasi-Likelihood-Schätzung übertragen. Im Gegensatz zu zweiparametrischen Logit-Modellen können extreme Fähigkeitsniveaus präzise geschätzt werden.

## 4.3 Diskussion

Die Ausführungen lassen sich hinsichtlich zweier Aspekte diskutieren: (1) Der Art der aufgestellten Modelle und (2) der Bedeutung der Modelle für die Schätzung der Persönlichkeitseigenschaften.

### Die Modellierung von Reaktionszeiten

Die Verteilung der Reaktionszeit bei der Bearbeitung von Testaufgaben kann sicherlich nicht mit einem einzigen Standardansatz modelliert werden. Unterschiedliche Anforderungen von Aufgaben lassen verschiedene Verteilungsformen erwarten. Es ist daher nicht überraschend, wenn im Verlauf der Jahre die unterschiedlichsten Verteilungsformen für die Reaktionszeiten bei Tests vorgeschlagen wurden. Zwei dieser Vorschläge, die Annahme log-normalverteilter und exponentialverteilter Reaktionszeiten wurden aufgegriffen. Bei beiden Modellen handelt es sich um Accelerated Failure Time Modelle, das Modell für exponentialverteilte Reaktionszeiten ist gleichzeitig ein Proportional Hazard Modell.<sup>9</sup>

Der Zwang, sich auf eine bestimmte Verteilungsform festlegen zu müssen, schränkt die Generalität der Ansätze erheblich ein und setzt zudem eine genaue Kenntnis des Gegenstandsbereiches voraus. Ein wichtiges Anliegen bei

---

<sup>9</sup>Der Proportional Hazard Ansatz ist mit dem Accelerated Failure Time Ansatz bis auf den Fall Weibull-verteilter Reaktionszeiten unvereinbar. Bei der Exponentialverteilung handelt es sich jedoch um einen Spezialfall der Weibull-Verteilung, weshalb das Modell gleichzeitig beiden Klassen angehört.

der Modellierung der Reaktionszeiten war es deshalb, auch verteilungsfreie Methoden vorzustellen. Weniger starke Annahmen machen Proportional Hazard Modelle für dichotomisierte Reaktionszeiten, die nur eine multiplikative Beziehung zwischen einer Funktion von Prädiktoren und einer Grundhazard-Rate voraussetzen. Ein populäres Proportional Hazard Modell für die Analyse von Überlebenszeiten ist das Modell von Cox (1972). Werden die Reaktionszeiten dichotomisiert, lässt sich dieses Modell auch ohne genaue Verteilungsannahmen schätzen. Allerdings wird die größere Generalität des Proportional Hazard Modells für dichotomisierte Zeiten mit einem Preis erkauft: Parameterschätzer sind bei Proportional Hazard Modellen für dichotomisierte Zeiten weniger präzise als bei einer Maximum-Likelihood-Schätzung mit genauen Verteilungsannahmen. Dies sieht man, wenn man die Iteminformationsfunktionen der Modelle für exponentialverteilte Reaktionszeiten mit den Iteminformationsfunktionen der äquivalenten Modelle für dichotomisierte Reaktionszeiten vergleicht. Auch wird die Modellevaluation erschwert, wenn man auf genauen Annahmen über die Verteilungsform der Reaktionszeiten verzichtet (Bradburn et al., 2003). Vergleichbare semiparametrische Ansätze für Accelerated Failure Time Modelle wurden zwar in den letzten Jahren vorgeschlagen – für eine Übersicht siehe Jin (2005) – jedoch lassen sich diese Modelle nur schwer schätzen und erlauben im Moment noch keinen Einbezug von Random-Personen-Effekten. Wenn jedoch keine zensierten Reaktionszeiten auftreten, ist der wesentlich einfachere Quasi-Maximum-Likelihood-Ansatz heranziehbar. Auch bei der Quasi-Maximum-Likelihood-Schätzung kann die Verteilungsform der Reaktionszeiten unbestimmt bleiben. Wie jedoch schon bei den Proportional Hazard Modellen für dichotomisierte Reaktionszeiten kann es auch in diesem Fall bei bestimmten Reaktionszeitverteilungen zu einem Informationsverlust kommen.

Als erstes Ergebnis dieser Arbeit kann somit festgehalten werden, dass für eine Nutzung der Reaktionszeit keine detaillierten Annahmen über die Reaktionszeitverteilung notwendig sind. Wenn man bereit ist, einen Informationsverlust hinzunehmen, benötigt man zur Schätzung der Fähigkeit einer Person nur zwei Voraussetzungen: (1) Die Gültigkeit der Regressionsbeziehung, sei es zwischen Prädiktoren und der Hazard-Rate oder zwischen Prädiktoren und dem Erwartungswert der logarithmierten Reaktionszeiten und (2) die Unabhängigkeit der Reaktionszeiten einer Person bei Bedingung auf die Persönlichkeitseigenschaften  $\theta$  und  $\omega$ .

Was in den vorgestellten Ansätzen modelliert wurde, war die Verteilung der Reaktionszeit einer Person, also die Zeit bis zur Beantwortung einer Frage. Dabei wurde die Itemantwort ignoriert – es wurde nicht zwischen den Reaktionszeiten für eine richtige Aufgabenlösung und den Reaktionszeiten für eine falsche Aufgabenlösung unterschieden. Das liegt zum einen daran, dass eine Unterscheidung zwischen richtigen und falschen Antworten nur bei Leistungstests möglich ist, die Modelle jedoch auch für Tests im Eigenschafts- oder Einstellungsbereich anwendbar sein sollen. Zum anderen kann die Gleichbehandlung richtiger und falscher Antworten damit begründet werden, dass zumindest bei einigen Leistungstests keine Verteilungsunterschiede zwischen Reaktionszeiten bei richtigen und falschen Antworten zu finden waren (Swygert, 2003). Teilweise wird jedoch die Auffassung vertreten, dass richtige und falsche Antworten auf unterschiedlichen Prozessen beruhen (Lavergne & Vigneau, 1997) und nur richtige Antworten das Resultat adäquater Problemlöseprozesse sind. Sollte dies tatsäch-

lich der Fall sein, stehen unter Umständen nur die Reaktionszeiten bei richtigen Itemantworten in Beziehung zur Personenfähigkeit eines Individuums. Prinzipiell können mit den vorgestellten Ansätzen auch nur die Reaktionszeiten für richtige Itemantworten modelliert werden. Eine Modellierung nur der Reaktionszeiten bei richtigen Itemantworten lässt sich dadurch erreichen, dass die Reaktionszeiten für falsche Itemantworten als zensierte Reaktionszeiten aufgefasst werden. Beantwortet eine Person eine Aufgabe zu einem bestimmten Zeitpunkt  $t$  falsch, lässt sich dies so deuten, dass eine hypothetische, unbeobachtbare richtige Antwort zumindest nicht vor dem Zeitpunkt  $t$  möglich war. Mit Ausnahme der Quasi-Likelihood-Modelle können bei allen vorgestellten Modellen zensierte Reaktionszeiten berücksichtigt werden, was allerdings immer restriktive Annahmen über den Zensurprozess voraussetzt. Die Frage, ob alle Reaktionszeiten modelliert werden sollen, oder nur die Reaktionszeiten bei richtigen Itemantworten, ist jedoch eine empirische Frage, die nicht generell beantwortet werden kann.

Alle Reaktionszeitmodelle dieser Arbeit sowie alle in der Fachliteratur vorgeschlagenen Reaktionszeitmodelle für Fragebogenitems sind losgelöst von Vorstellungen über die intrapsychische Bearbeitung der Aufgaben durch die Individuen. Bisher wurde kein detailliertes Modell des Lösungsprozesses aufgestellt, mit welchem die angenommene Reaktionszeitverteilung begründet werden könnte. Darin unterscheiden sich die Ansätze wesentlich von den Modellierungsversuchen in der Allgemeinen Psychologie. Dort ist zwar die Frage nach der gemeinsamen Verteilung der Art einer Reaktion auf einfache Reize und des dafür benötigten Zeitbedarfs ein ebenfalls oft untersuchtes Thema, siehe zum Beispiel die Zusammenstellung von Luce (1991). Die Herangehensweise und der Anspruch in der Allgemeinen Psychologie ist jedoch anders als die Vorgehensweise in der psychologischen Diagnostik: Anhand von sehr detaillierten Annahmen über den kognitiven Wahrnehmungsprozess werden Verteilungen abgeleitet, deren Gültigkeit als Bestätigung des postulierten Modells aufgefasst wird. Die Diagnostik von Fähigkeiten steht dabei im Hintergrund. Eine Übertragung dieser detaillierten Modelle für basale Wahrnehmungsaufgaben auf den Bereich der Fähigkeitsdiagnostik ist nur schwer möglich – Modelle für basale Wahrnehmungs- und Reaktionsaufgaben gelten kaum für einen anderen Inhaltsbereich wie zum Beispiel das Lösen komplexer Intelligenztestaufgaben. Zudem sind die notwendigen Voraussetzungen für die Parameterschätzung bei den vorgeschlagenen allgemeinpsychologischen Modellen (mehrere hundert Beobachtungen pro Person) bei diagnostischen Anwendungen nicht gegeben.

### **Bedeutung der Modellierung für die Fähigkeitsschätzung**

Ausgangspunkt der dargestellten Modelle war die Annahme, dass die Reaktionszeit einer Person von den Eigenschaften einer Aufgabe abhängen, aber auch von den individuellen Persönlichkeitseigenschaften  $\theta$  und  $\omega$ . Werden die Aufgabenparameter bei den einzelnen Modellen in Vorstudien bestimmt, lassen sich anhand des Reaktionszeitmusters einer Person ihre Ausprägungen auf den Persönlichkeitseigenschaften schätzen. Wie bei den IRT-Modellen steht dann die Fisher-Information in Beziehung zur asymptotischen Varianz der Schätzer  $\hat{\theta}$  und  $\hat{\omega}$ . Hohe Werte der Fisher-Information über die einzelnen Parameter, also hohe Diagonalelemente in der Informationsmatrix, sind bei mehrdimensionalen

Modellen eine notwendige Voraussetzung für eine Schätzung mit geringer Varianz. Da für den diagnostischen Prozess in der Regel nur die Personenfähigkeit  $\theta$  von Interesse ist, soll die Diskussion auf die Bedeutung der Reaktionszeiten für die Schätzung der Personenfähigkeit eingeschränkt werden.

Bei den beschriebenen Reaktionszeitmodellen wurde in der Regel ein Verteilungsparameter als Funktion der Persönlichkeitseigenschaften  $\omega$  und  $\theta$  modelliert. In diesem Fall hängt die Information der Reaktionszeiten über die Ausprägung der Persönlichkeitseigenschaften von zwei Elementen ab: Von der Information einer Verteilung über den entsprechenden Verteilungsparameter und von der Art der Rückführung des Verteilungsparameters auf die Eigenschaften  $\theta$  und  $\omega$ . Untersucht wurden zwei Verteilungsformen, bei denen jeweils ein Verteilungsparameter modelliert wurde (der Lokationsparameter  $\eta$  bei log-normalverteilten Reaktionszeiten bzw. die Intensität  $\lambda$  bei den exponentialverteilten Reaktionszeiten). Als mögliche Regressionsfunktionen wurden dabei eine lineare und eine quadratische Funktion diskutiert. Es zeigte sich, dass die genaue Art der Reaktionszeitverteilung für die Form der Informationsfunktion bedeutungslos ist.

Bei linearer Parametrisierung fällt die Iteminformationsmatrix sowohl für das Exponentialverteilungsmodell wie auch für das Log-Normalverteilungsmodell konstant aus. Das Vorliegen einer konstanten Iteminformationsfunktion ist sicherlich ein günstiger Fall, da folglich keine maßgeschneiderten Tests erforderlich sind, sondern für alle Ausprägungen der Personenfähigkeit mit einem Test dieselbe Messpräzision erzielt werden kann. Konstante Iteminformationsfunktionen sind bei IRT-Modellen nur bei extrem konstruierten Itemcharakteristiken, siehe zum Beispiel Bickel et al. (2001), oder bei komplexen Parametertransformationen, siehe zum Beispiel Samejima (1983), zu finden.

Bei quadratischer Parametrisierung resultieren für die Personenfähigkeit Iteminformationsfunktionen, wie sie auch bei Unfolding-Modellen im IRT-Bereich auftreten. Die Iteminformation ist minimal, wenn die Ausprägung der Personenfähigkeit mit der Itemlokation auf der Fähigkeitsdimension zusammenfällt. Allerdings wächst bei quadratischer Parametrisierung die Information mit zunehmender Distanz zwischen Item und Person. Dieser Aspekt unterscheidet diese Modelle erheblich von dem im dritten Kapitel dargestellten, zweiparametrischen Logit-Modell, bei welchem extreme Merkmalsausprägungen mit geringer Iteminformation einhergingen, siehe hierzu Abbildung 3.2. Dies deutet darauf hin, dass die Reaktionszeitmodelle besonders für die Diagnostik extremer Fähigkeitsniveaus geeignet sind.

Auch Proportional Hazard Modelle für dichotomisierte Reaktionszeiten können zur Schätzung der Eigenschaften  $\theta$  und  $\omega$  genutzt werden. Die Iteminformationsfunktionen dieser Modelle weisen jedoch ungünstigere Verläufe auf. Sowohl bei der linearen Parametrisierung als auch bei der quadratischen Parametrisierung sind die Informationsfunktionen der Personenfähigkeit beschränkt und konvergieren bei konstantem Niveau des Residualfaktors mit zunehmender und abnehmender Personenfähigkeit gegen Null. Dies deutet darauf hin, dass es Fähigkeitsbereiche gibt, die alleine anhand der Reaktionszeiten nur mit großer Varianz diagnostiziert werden können. Vergleicht man die Iteminformation des semiparametrischen Proportional Hazard Modells für dichotomisierte Reaktionszeiten mit der Iteminformation des parametrischen Modells für exponentialverteilte Reaktionszeiten, stellt man für die semiparametrische Modellie-

rung gegenüber der parametrischen Modellierung einen Informationsverlust fest, der auf der Dichotomisierung beruht. Die weniger starken Annahmen der verteilungsfreien Modelle werden folglich mit geringerer Iteminformation bezahlt. Tatsächlich lässt sich dieses Phänomen mit dem Missing Information Prinzip von Orchard und Woodbury (1972) erklären. Allerdings kann der Informationsverlust reduziert werden, wenn die Reaktionszeiten nicht dichotomisiert, sondern in mehrere Kategorien klassifiziert werden. Das Proportional Hazard Modell für dichotomisierte Reaktionszeiten lässt sich nämlich einfach auf den Fall mehrerer Kategorien erweitern. Durch die feinere Auflösung der Reaktionszeiten kann der Informationsverlust reduziert werden. Die Proportional Hazard Modelle für dichotomisierte Reaktionszeiten teilen zwar mit den IRT-Modellen den ungünstigen Verlauf der Iteminformationsfunktion, besitzen aber eine vorteilhafte Eigenschaft: Die Lage des Maximums der Informationsfunktion über die Personenfähigkeit hängt vom Parameter  $k_j$  ab, welcher wiederum durch die Wahl des Dichotomisierungszeitpunktes  $t_c$  bestimmt wird (siehe 4.17). Durch die Wahl des Zeitpunktes  $t_c$  kann das Informationsmaximum nach Belieben verschoben werden. Dies lässt sich für die Entwicklung adaptiver Tests ausnützen. Ein weiterer Vorteil der Verwendung dichotomisierter Reaktionszeiten ist die größere Robustheit (Heathcote, Brown & Mewhort, 2002).

Bei Accelerated Failure Time Modellen wird in der Regel für die Residualkomponente  $e$  bei einem Item eine bestimmte Verteilungsform angenommen. Zur Schätzung der Eigenschaften einer Person ist eine konkrete Verteilungsannahme jedoch nicht erforderlich, unter milden Bedingungen kann die Ausprägung der Eigenschaften einer Person mit Hilfe des Quasi-Maximum-Likelihood-Ansatzes geschätzt werden. Auch innerhalb des Quasi-Maximum-Likelihood-Ansatzes ist ein Informationsmaß definiert, das in Beziehung zur Varianz des Schätzers steht, die Quasi-Information. Im Falle log-normalverteilter Reaktionszeiten entspricht dieses Maß der Fisher-Information. Dies muss jedoch bei anderen Verteilungsformen nicht zwangsläufig der Fall sein: Wedderburn (1974) konnte zeigen, dass die Quasi-Information niemals größer als die Fisher-Information wird, so dass es auch bei dieser Modellierung zu einem Informationsverlust kommt, wenn die genaue Verteilungsform der Reaktionszeiten nicht bekannt ist. Im Vergleich zum „verteilungsfreien“ Ansatz der Proportional Hazard Modelle für dichotomisierte Reaktionszeiten weist die Quasi-Information einen günstigeren Verlauf über den Wertebereich der Personenfähigkeit auf. Für extreme Ausprägungen nimmt die Quasi-Information nicht ab. Die Schätzung der Persönlichkeitseigenschaften profitiert erheblich von dem günstigeren Verlauf der Quasi-Information. Bei linearer Parametrisierung können über den Quasi-Maximum-Likelihood-Ansatz alle Fähigkeitsniveaus mit derselben Präzision geschätzt werden. Bei quadratischer Parametrisierung können zwar Fähigkeiten im Bereich der Itemlokation nur mit geringer Präzision geschätzt werden, die Präzision der Messung steigt aber für extreme Ausprägungen an.

Bisher lag der Schwerpunkt der Diskussion auf den Informationsfunktionen der einzelnen Items. Dies begründet sich damit, dass die Berechnung der asymptotischen Varianz des Eigenschaftsschätzers auf der Fisher-Information der Items basiert. Manche Aussagen über die Varianz des Fähigkeitsschätzers lassen sich daher bereits mit dem Verlauf der Iteminformationsfunktionen belegen. Eine notwendige Voraussetzung für eine geringe Varianz des Fähigkeitsschätzers sind hohe Funktionswerte der Iteminformationen über die Personenfähigkeit,

also hohe Diagonalelemente  $I_{g[11]}(\theta, \omega)$  der Informationsmatrizen der einzelnen Items. Aus diesem Grund kann bereits alleine anhand der Verläufe der Iteminformationsfunktionen abgeschätzt werden, in welchen Bereichen der Fähigkeit Reaktionszeiten eine präzise Messung erlauben könnten und in welchen Bereichen Reaktionszeiten kein diagnostisches Potential besitzen. Vergleicht man diese Bereiche mit den Bereichen des zweiparametrischen Logit-Modells, stellt man fest, dass sie oft komplementär sind. Reaktionszeiten besitzen für diejenigen Fähigkeitsausprägungen diagnostisches Potential, welche mit den Itemantworten nur sehr unpräzise geschätzt werden können. Diese Aussage gilt in besonderer Weise für die Modelle für kontinuierliche Reaktionszeiten. Aufgrund dieser Eigenschaft ist zu vermuten, dass besonders die Schätzung extremer Fähigkeiten davon profitiert, wenn neben den Itemantworten die Reaktionszeiten bei der Schätzung berücksichtigt werden. Dies soll in einem späteren Kapitel noch genauer untersucht werden. Zuvor jedoch muss noch die Frage nach der gemeinsamen Modellierung von Itemantwort und Reaktionszeit geklärt werden. Diesem Thema ist das nächste Kapitel gewidmet.

## Teil II

# Gemeinsame Modellierung von Itemantworten und Reaktionszeiten

Bei der Durchführung von psychologischen Tests fallen nicht nur Itemantworten an. Durch die Verfügbarkeit von computergestützten Tests ist es mittlerweile auch möglich, die Bearbeitungszeiten einer Person auf Itemebene präzise zu erfassen. Es ist daher nicht verwunderlich, dass die Reaktionszeiten bei Tests in den Fokus psychologischer Forschung gerückt sind. In den vorherigen Kapiteln wurden die Verteilungen der Itemantworten und der Reaktionszeiten bei den Items eines psychologischen Tests isoliert voneinander behandelt. Die Verteilung der Itemantworten  $\mathbf{x} = [x_1, \dots, x_G]$  einer Person bei einem Test aus  $G$  Items hing dabei von einer Persönlichkeitseigenschaft ab, der so genannten Personenfähigkeit  $\theta$ . Die Verteilung der Reaktionszeiten  $\mathbf{t} = [t_1, \dots, t_G]$  bei den Testaufgaben wurde auf zwei Persönlichkeitseigenschaften zurückgeführt, auf die Personenfähigkeit  $\theta$  und den Residualfaktor  $\omega$ . Ein zweifaktorielles Modell erschien deshalb notwendig, da sich Hinweise darauf finden lassen, dass Reaktionszeiten neben einem inhaltspezifischen Faktor noch auf einem weiteren Faktor laden. Kein Wort fiel jedoch bisher zur gemeinsamen Verteilung der Reaktionszeiten und der Itemantworten. Dies ist Gegenstand dieses Teils der Arbeit.

Psychologische Tests sind in der Regel für eine Population potentieller Testteilnehmer entwickelt, beispielsweise für Schulkinder einer bestimmten Altersgruppe etc. Innerhalb dieser Referenzpopulation besitzen die Persönlichkeitseigenschaften  $\theta$  und  $\omega$  eine gemeinsame Verteilung  $f(\theta, \omega)$ . Bei Anwendung des Tests auf eine bestimmte Person wird somit eine „Zufallsstichprobe“ aus der Referenzpopulation gezogen: Die genauen Ausprägungen der Persönlichkeitseigenschaften  $\theta$  und  $\omega$  bei einer Person sind dabei Realisierungen von Zufallsvariablen. Die realisierten Ausprägungen dieser Eigenschaften bestimmen die Verteilung der Itemantworten und Reaktionszeiten bei dem Test. Allgemein gehorcht die Verteilung der Reaktionszeiten und Itemantworten einer Person mit Personenfähigkeit  $\theta$  und Residualfaktor  $\omega$  einer bedingten Verteilung  $f(\mathbf{t}, \mathbf{x}|\theta, \omega)$ . Die gemeinsame Verteilung aller Größen in der Population lässt sich bei zufälliger Auswahl einer Testperson folglich mit der Verteilung  $f(\mathbf{t}, \mathbf{x}|\theta, \omega)f(\theta, \omega)$  beschreiben. Um an die Modelle der früheren Kapitel anknüpfen zu können, sind weitere Annahmen notwendig.

Das „Common Cause Criterion“ besagt, dass bei Berücksichtigung der gemeinsamen Ursachen der Zusammenhang zweier Phänomene vollständig erklärt wird (Suppes & Zanotti, 1981). Aus diesem Grund werden beispielsweise bei den IRT-Modellen die Verteilungen der einzelnen Itemantworten bei Bedingung auf die Personenfähigkeit  $\theta$  als unabhängig angenommen. Überträgt man diesen Gedanken auf die gemeinsame Modellierung der Verteilung von Reaktionszeiten und Itemantworten, folgt daraus, dass bei Bedingung auf die Personenfähigkeit  $\theta$  und den Residualfaktor  $\omega$  die Reaktionszeiten und Itemantworten aller Items voneinander unabhängig sind. Diese Annahme wird bei der Modellierung der gemeinsamen Verteilung von Reaktionszeiten und Itemantworten häufig getroffen – siehe beispielsweise van der Linden (2007), Thissen (1984) oder Ferrando und Lorenzo-Sevas (2007a). Geht man von der bedingten Unabhängigkeit von Reaktionszeiten und Itemantworten aus, gilt für die gemeinsame Verteilung von Reaktionszeiten, Itemantworten, Personenfähigkeit und Residualfaktor einer Person bei zufälliger Auswahl der Person aus einer Population von potentiellen Testpersonen die Beziehung:



$$\begin{aligned}
f(\mathbf{t}, \mathbf{x}, \theta, \omega) &= f(\mathbf{t}|\theta, \omega)f(\mathbf{x}|\theta, \omega)f(\theta, \omega) \\
&= \left[ \prod_{g=1}^G f(t_g|\theta, \omega) \right] \left[ \prod_{g=1}^G f(x_g|\theta, \omega) \right] f(\theta, \omega). \tag{II.2}
\end{aligned}$$

Nimmt man ferner an, dass die Verteilung der Itemantworten nur von der Personenfähigkeit  $\theta$  abhängt, vereinfacht sich diese Beziehung zu:

$$f(\mathbf{t}, \mathbf{x}, \theta, \omega) = \left[ \prod_{g=1}^G f(t_g|\theta, \omega) \right] \left[ \prod_{g=1}^G f(x_g|\theta) \right] f(\theta, \omega). \tag{II.3}$$

Für die Verteilungen  $f(t_g|\theta, \omega)$  und  $f(x_g|\theta)$  kann ein geeignet erscheinendes Modell aus dem dritten und vierten Kapitel eingesetzt werden. Gleichung (II.3) beschreibt dann ein Messmodell, das eine Schätzung der Personenfähigkeit sowohl anhand der Itemantworten als auch anhand der Reaktionszeiten erlaubt. Allerdings hängt die genaue Verteilung  $f(\mathbf{t}, \mathbf{x}, \theta, \omega)$  von den Itemparametern der IRT- und Reaktionszeitmodelle ab, die unbekannt sind und vor einer Anwendung des Messmodells erst bestimmt werden müssen. Die Schätzung der Itemparameter ist für das zweiparametrische Logit-Modell in vielen Software-Paketen umgesetzt. Wie die Itemparameter der Reaktionszeitmodelle bestimmt werden können, wurde bereits im vierten Kapitel beziehungsweise in Anhang A gezeigt. Nicht behandelt wurde jedoch, wie die Itemparameter von IRT- und Reaktionszeitmodell gemeinsam geschätzt werden können. Dies soll in den folgenden Kapiteln demonstriert werden.

Allerdings werden nicht für alle vorgeschlagenen Reaktionszeitmodelle eigenständige Schätzverfahren entwickelt, sondern nur für vier: Für die beiden Modelle mit log-normalverteilten Reaktionszeiten sowie für die beiden Proportional Hazard Modelle für dichotomisierte Reaktionszeiten. Die Schätzung der Modelle für exponentialverteilte Reaktionszeiten wird nicht betrachtet, da sich diese Modelle auch mit Hilfe der Schätzverfahren der Proportional Hazard Modelle für dichotomisierte Reaktionszeiten an die Daten anpassen lassen. Auch auf die beiden verteilungsfreien Accelerated Failure Time Modelle wird nicht gesondert eingegangen. Es wird an geeigneter Stelle gezeigt, dass sich deren Itemparameter über die für die Modelle mit log-normalverteilten Reaktionszeiten vorgeschlagenen Momentmethoden-Schätzer bestimmen lassen.

Alle vorgeschlagenen Schätzverfahren basieren auf der Marginal-Maximum-Likelihood-Methode. Teilweise wird zusätzlich ein Momentmethoden-Schätzer beschrieben, der verwendet werden kann, um Startwerte für die Marginal-Maximum-Likelihood-Methode zu bestimmen. Der Marginal-Maximum-Likelihood-Ansatz kann inzwischen als Standardverfahren in der probabilistischen Testtheorie bezeichnet werden. Für eine Umsetzung der Schätzung wird auf den EM-Algorithmus von Dempster, Laird und Rubin (1977) zurückgegriffen. Genaue Details zur Umsetzung dieses generellen Ansatzes finden sich in den nächsten Kapiteln. Im sechsten Kapitel wird die Schätzung der Proportional Hazard Modelle für dichotomisierte Reaktionszeiten detailliert dargestellt. Im siebten

Kapitel folgt eine genaue Beschreibung der Parameterschätzung bei den Accelerated Failure Time Modellen für log-normalverteilte Reaktionszeiten. Die Eignung der vorgeschlagenen Schätzverfahren wird in Simulationen untersucht, in denen die Präzision der Parameterschätzung für verschiedene Stichprobenumfänge beleuchtet wird. Es wird sich zeigen, dass für eine genaue Bestimmung der Itemparameter große Stichprobenumfänge notwendig sind. Dies ist allerdings nicht überraschend, da bereits die Schätzung der IRT-Modelle nach großen Stichprobenumfängen verlangt.

## Kapitel 5

# Itemparameterschätzung bei Proportional Hazard Modellen

Während im ersten Abschnitt der Arbeit die Verteilung von Itemantworten und Reaktionszeiten bei psychologischen Tests isoliert voneinander betrachtet wurden, ist das Thema des zweiten Abschnittes die gemeinsame Modellierung dieser beiden Fähigkeitsmanifestationen. Geht man davon aus, dass bei Bedingung auf die Eigenschaften einer Person die Reaktionszeiten und Itemantworten bei den  $G$  Items eines Tests voneinander unabhängig sind, kann die gemeinsame Verteilung von Reaktionszeiten, Itemantworten, Personenfähigkeit und Residualfaktorwert einer Person geschrieben werden als:

$$f(\mathbf{t}, \mathbf{x}, \theta, \omega) = \left[ \prod_{g=1}^G f(t_g | \theta, \omega) \right] \left[ \prod_{g=1}^G f(x_g | \theta) \right] f(\theta, \omega). \quad (5.1)$$

Eine konkrete Verteilung ergibt sich durch Festlegung des IRT- und RT-Modells. In diesem Kapitel werden in Gleichung (5.1) die Proportional Hazard Modelle für dichotomisierte Reaktionszeiten mit dem im dritten Kapitel beschriebenen, zweiparametrischen Logit-Modell kombiniert. Der Vorteil des Proportional Hazard Modells für dichotomisierte Reaktionszeiten ist der große Allgemeingrad, da keine bestimmte Form für die Reaktionszeitverteilung angenommen werden muss. Somit kann der Ansatz auch auf die im vierten Kapitel beschriebenen Modelle für exponentialverteilte Reaktionszeiten angewandt werden.

### 5.1 Lineare Parametrisierung

Proportional Hazard Modelle nehmen in der Analyse von Überlebenszeiten eine hervorgehobene Position ein. Grundgedanke der Proportional Hazard Modelle

ist die Idee, dass eine Grundhazard-Rate durch eine Funktion von Prädiktoren über eine multiplikative Beziehung verändert wird. Eine Anwendung dieses Gedankens auf die Modellierung von Reaktionszeiten bei Tests findet sich in Modell (4.18). Laut Modell (4.18) gehorcht die Hazard-Rate einer Person mit Personenfähigkeit  $\theta$  und Residualfaktor  $\omega$  der Beziehung:

$$h_{\theta,\omega}(t) = h_{B_g}(t) \exp [l_g \theta + c_g \omega]. \quad (5.2)$$

Dabei ist  $l_g$  das Einflussgewicht der Personenfähigkeit und  $c_g$  das Einflussgewicht des Residualfaktors  $\omega$  auf die Hazard-Rate einer Person. Bei der Hazard-Rate handelt es sich um die „Wahrscheinlichkeit“, dass eine Person im nächsten Moment reagiert, wenn sie bis zu einem bestimmten Zeitpunkt noch nicht reagiert hat. Sowohl eine hohe Ausprägung der Personenfähigkeit  $\theta$  als auch eine hohe Ausprägung des Residualfaktors  $\omega$  führen bei positiven Einflussgewichten  $l_g$  und  $c_g$  zu schnellen Reaktionen – zur weiteren Interpretation des Modells sei auf die Ausführungen im vierten Kapitel verwiesen.

Bei Proportional Hazard Modellen für dichotomisierte Reaktionszeiten werden die Reaktionszeiten in zwei Klassen eingeteilt. Die ursprünglich kontinuierliche Zeitmessung  $t_g$  wird dabei in eine binäre Zufallsvariable  $y_g$  transformiert. Kriterium der Dichotomisierung ist der Umstand, ob die Reaktionszeit  $t_g$  eine Schwelle  $t_c$  unterschreitet ( $y_g = 1$ ) oder überschreitet ( $y_g = 0$ ). Wird die Reaktionszeit dichotomisiert, lässt sich nach den Überlegungen in Abschnitt 4.1.2 die Wahrscheinlichkeit, auf ein Item  $g$  vor dem Zeitlimit  $t_c$  zu reagieren, angeben als

$$\begin{aligned} P_{y_g}(\theta, \omega; c_g, l_g, k_g) &= 1 - \exp \left[ - \int_0^{t_c} h_{B_g}(z) \exp [l_g \theta + c_g \omega] dz \right] \\ &= 1 - \exp \left[ - \exp [k_g + l_g \theta + c_g \omega] \right], \end{aligned} \quad (5.3)$$

wobei der Parameter  $k_g = \log \left( \int_0^{t_c} h_{B_g}(z) dz \right)$  die genaue Form der Grundhazard-Rate absorbiert.<sup>1</sup> Die Wahrscheinlichkeit einer positiven Itemantwort soll einem zweiparametrischen Logit-Modell folgen, siehe Gleichung (3.1). Aus Gründen einer einfacheren Schätzbarkeit soll jedoch eine etwas andere Form der Parametrisierung gewählt werden:

$$P_{x_g}(\theta; a_g, b_g) = \frac{\exp [a_g \theta + b_g]}{1 + \exp [a_g \theta + b_g]}. \quad (5.4)$$

Die Itemparameter der neuen Parametrisierung können einfach in die Itemparameter der ursprünglichen Parametrisierung (3.1) umgerechnet werden. Die gemeinsame Wahrscheinlichkeit, bei einer Person mit Eigenschaftsausprägungen  $\theta$  und  $\omega$  eine bestimmte Itemantwort vor oder nach Zeitlimit  $t_c$  zu beobachten,

<sup>1</sup>Da im aktuellen Kapitel die Itemparameter von Interesse sind, weicht die Darstellung der Antwort- und Lösungswahrscheinlichkeit von der bisherigen Darstellung leicht ab: In Gleichung (5.3) und (5.4) wird im Gegensatz zu den früheren Kapiteln durch die explizite Nennung der Itemparameter die Abhängigkeit von den Itemparametern besonders hervorgehoben.

berechnet sich bei bedingter Unabhängigkeit von Reaktionszeit und Itemantwort für Item  $g$  als:

$$P(x_g, y_g | \theta, \omega; a_g, b_g, c_g, l_g, k_g) = [P_{x_g}(\theta; a_g, b_g)]^{x_g} [1 - P_{x_g}(\theta; a_g, b_g)]^{(1-x_g)} \\ [P_{y_g}(\theta, \omega; c_g, l_g, k_g)]^{y_g} [1 - P_{y_g}(\theta, \omega; c_g, l_g, k_g)]^{(1-y_g)}. \quad (5.5)$$

In Gleichung (5.5) kodiert die Zufallsvariable  $x_g$  die Art der Itemantwort, die Zufallsvariable  $y_g$  den Umstand, ob auf ein Item vor oder nach dem Zeitlimit  $t_c$  reagiert wurde. Bei Beobachtung von  $G$  Aufgaben ergibt sich aufgrund der bedingten Unabhängigkeit der Reaktionszeiten und der Itemantworten für die Wahrscheinlichkeit, bei einer Person mit Eigenschaftsausprägungen  $\theta$  und  $\omega$  ein komplettes Antwortmusters  $\mathbf{x} = [x_1, \dots, x_G]$  und  $\mathbf{y} = [y_1, \dots, y_G]$  zu beobachten, die Beziehung

$$P(\mathbf{x}, \mathbf{y} | \theta, \omega; \mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{l}, \mathbf{k}) = \prod_{g=1}^G P(x_g, y_g | \theta, \omega; a_g, b_g, c_g, l_g, k_g), \quad (5.6)$$

wobei  $\mathbf{a} = [a_1, \dots, a_G]$ ,  $\mathbf{b} = [b_1, \dots, b_G]$ ,  $\mathbf{c} = [c_1, \dots, c_G]$ ,  $\mathbf{l} = [l_1, \dots, l_G]$  und  $\mathbf{k} = [k_1, \dots, k_G]$  Vektoren sind, welche die Itemparameter der  $G$  Aufgaben enthalten.

### Schätzung der Itemparameter

Zieht man eine Stichprobe von  $N$  Testpersonen und beobachtet deren Itemantworten und dichotomisierte Reaktionszeiten, ergibt sich als Log-Likelihood-Funktion:

$$\text{LL}(\boldsymbol{\theta}, \boldsymbol{\omega}, \mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{l}, \mathbf{k}) = \log \prod_{i=1}^N P(\mathbf{x}_i, \mathbf{y}_i | \theta_i, \omega_i; \mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{l}, \mathbf{k}). \quad (5.7)$$

Die Vektoren  $\boldsymbol{\theta} = [\theta_1, \dots, \theta_N]$  und  $\boldsymbol{\omega} = [\omega_1, \dots, \omega_N]$  enthalten die unbekanntenen Persönlichkeitseigenschaften der  $N$  Testpersonen. Die Parameterschätzung wird dadurch erschwert, dass es sich bei den Personenparametern  $\theta_i$  und  $\omega_i$  um so genannte „Incidental Parameter“ handelt, deren Anzahl mit dem Stichprobenumfang anwächst, siehe hierzu die klassische Referenz von Neyman und Scott (1948). Dadurch ist eine konsistente Schätzung der Itemparameter bei fixem Testumfang und steigender Versuchspersonenanzahl nicht garantiert. Eine Lösung des Problems stellt der Marginal-Maximum-Likelihood-Ansatz dar (Kiefer & Wolfowitz, 1956). Für einen Überblick über verschiedene alternative Schätzansätze siehe Rabe-Hesketh, Skrondal und Pickles (2001) oder Tuerlinckx, Rijmen, Verbeke und De Boeck (2006).

Ausgangspunkt des Marginal-Maximum-Likelihood-Ansatzes ist die Annahme einer bestimmten Verteilung der „Incidental Parameter“. Beispielsweise kann davon ausgegangen werden, dass innerhalb einer Population von potentiellen

Testteilnehmern die Personenfähigkeit  $\theta$  und der Residualfaktor  $\omega$  bivariat normalverteilt sind mit Varianz-Kovarianz-Matrix  $\Sigma$  und Mittelwertsvektor  $\mu$ . Da die Skala der latenten Personenparameter unbestimmt ist, wird in der Regel eine bivariate Standardnormalverteilung angenommen. Der Mittelwertsvektor ist dann  $\mu = [0, 0]$ , die Varianz-Kovarianz-Matrix  $\Sigma$  ist eine Identitätsmatrix. Die marginale Wahrscheinlichkeit, ein bestimmtes Antwortmuster  $[\mathbf{x}, \mathbf{y}]$  zu beobachten, erhält man nach Ausintegration der Personenparameter  $\theta$  und  $\omega$  aus Gleichung (5.6):

$$P(\mathbf{x}, \mathbf{y}; \mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{l}, \mathbf{k}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{g=1}^G P(x_g, y_g | \theta, \omega; a_g, b_g, c_g, l_g, k_g) \frac{1}{(2\pi)\sqrt{\det(\Sigma)}} \exp\left[-\frac{1}{2}[\theta, \omega]\Sigma^{-1}[\theta, \omega]'\right] d\theta d\omega. \quad (5.8)$$

Es ist diese marginale Wahrscheinlichkeit eines Antwortmusters, die beim Marginal-Maximum-Likelihood-Ansatz verwendet wird. Die marginale Log-Likelihood-Funktion der Itemparameter von  $G$  Items lautet bei Beobachtung der Antwortmuster von  $N$  Personen:

$$\text{MLL}(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{l}, \mathbf{k}) = \log \prod_{i=1}^N P(\mathbf{x}_i, \mathbf{y}_i; \mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{l}, \mathbf{k}). \quad (5.9)$$

Diese Funktion ist bezüglich der Itemparameter  $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{l}$  und  $\mathbf{k}$  zu maximieren. Es ergeben sich jedoch bei der Marginal-Maximum-Likelihood-Schätzung zwei Komplikationen: Erstens ist die marginale Wahrscheinlichkeit (5.8) nicht in geschlossener Form angebar. Und zweitens sind nach Ausintegration der Personenvariablen  $\theta$  und  $\omega$  die Reaktionszeiten und Itemantworten nicht mehr voneinander unabhängig, so dass die marginale Log-Likelihood-Funktion (5.9) gleichzeitig für alle Itemparameter maximiert werden muss. Eine Lösung des ersten Problems besteht in der numerischen Approximation des nicht direkt berechenbaren Integrals. Dadurch kann die marginale Log-Likelihood-Funktion näherungsweise bestimmt werden. Die Maximierung der approximierten marginalen Log-Likelihood-Funktion erfordert numerische Suchverfahren. Jedoch steigt die Dimensionalität des Suchraumes mit der Itemanzahl schnell an, so dass eine Maximierung sehr zeitintensiv wird. Als numerisches Verfahren zur Maximierung von (5.9) wird daher der EM-Algorithmus verwendet.

Die Anwendung des EM-Algorithmus zur Schätzung von IRT-Modellen geht auf Bock und Aitkin (1981) zurück. Der EM-Algorithmus ist ein iteratives Verfahren, das aus einer Abfolge einzelner Maximierungsschritte besteht (McLachlan & Krishnan, 1997). Bei jedem Maximierungsschritt werden ausgehend von den vorläufigen Itemparametern  $\mathbf{a}_0, \mathbf{b}_0, \mathbf{c}_0, \mathbf{l}_0$  und  $\mathbf{k}_0$  neue Itemparameter  $\mathbf{a}_1, \mathbf{b}_1, \mathbf{c}_1, \mathbf{l}_1$  und  $\mathbf{k}_1$  bestimmt, bei denen der Wert der marginalen Log-Likelihood-Funktion gegenüber dem zu den vorläufigen Itemparametern gehörenden Funktionswert ansteigt. Diese neuen Itemparameter dienen dann als Eingangsgrößen eines weiteren Maximierungsschrittes. So wächst die marginale Log-Likelihood-Funktion mit jedem Maximierungsschritt an. Die Abfolge von Maximierungsschritten wird

abgebrochen, wenn die bei einem Maximierungsschritt erzielte Zunahme der marginalen Log-Likelihood-Funktion ein vorgegebenes Kriterium unterschreitet. Es kann gezeigt werden, dass bei Gültigkeit bestimmter Regularitätsannahmen die Sequenz der Funktionswerte der marginalen Likelihood-Funktion zumindest gegen einen stationären Punkt konvergiert (McLachlan & Krishnan, 1997).

Jeder einzelne Maximierungsschritt des EM-Algorithmus kann wiederum in zwei Subschritte aufgeteilt werden, in eine Phase der Erwartungswertbildung (den so genannten E-Schritt) und in eine Phase der Maximierung (den so genannten M-Schritt). Während des E-Schrittes wird, ausgehend von vorläufigen Werten für die gesuchten Itemparameter, der auf die Beobachtungen bedingte Erwartungswert der Log-Likelihood-Funktion (5.7) bestimmt. Dieser wird im M-Schritt bezüglich der gesuchten Itemparameter maximiert. Auf die genauen Details der Schätzung wird in den nächsten beiden Abschnitten eingegangen. Diese Abschnitte richten sich vor allem an Leser, die an der technischen Umsetzung der Parameterschätzung interessiert sind und können beim ersten Lesen der Arbeit übersprungen werden.

### E-Schritt

Eine Schwierigkeit der Itemparameterschätzung von IRT-Modellen liegt in dem Umstand, dass die Likelihood-Funktion bei Beobachtung der Itemantworten und Reaktionszeiten einer Stichprobe von Personen nicht alleine von den Itemparametern abhängt, sondern auch von den Personenparametern der Testpersonen, die unbekannt sind. Während des E-Schrittes wird daher der auf die Beobachtungen bedingte Erwartungswert der Log-Likelihood-Funktion bestimmt. Hierzu wird zuerst einmal die Log-Likelihood-Funktion betrachtet, die bei Kenntnis der Personenparameter zur Schätzung der Itemparameter verwendet würde. Bei einer Person mit Personenfähigkeit  $\theta$  und Residualfaktor  $\omega$  tritt nach Gleichung (5.6) ein bestimmtes Antwortmuster  $\mathbf{x} = [x_1, \dots, x_G]$  und  $\mathbf{y} = [y_1, \dots, y_G]$  mit der Wahrscheinlichkeit

$$\begin{aligned}
P(\mathbf{x}, \mathbf{y} | \theta, \omega; \mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{l}, \mathbf{k}) &= \prod_{g=1}^G P(x_g, y_g | \theta, \omega; a_g, b_g, c_g, l_g, k_g) \\
&= \prod_{g=1}^G \left[ P_{y_g}(\theta, \omega; c_g, l_g, k_g) \right]^{y_g} \left[ 1 - P_{y_g}(\theta, \omega; c_g, l_g, k_g) \right]^{(1-y_g)} \\
&\quad \left[ P_{x_g}(\theta; a_g, b_g) \right]^{x_g} \left[ 1 - P_{x_g}(\theta; a_g, b_g) \right]^{(1-x_g)} \\
&= \prod_{g=1}^G \left[ 1 - \exp \left[ - \exp[k_g + l_g \theta + c_g \omega] \right] \right]^{y_g} \left[ \exp \left[ - \exp[k_g + l_g \theta + c_g \omega] \right] \right]^{(1-y_g)} \\
&\quad \left[ \frac{\exp[a_g \theta + b_g]}{1 + \exp[a_g \theta + b_g]} \right]^{x_g} \left[ 1 - \frac{\exp[a_g \theta + b_g]}{1 + \exp[a_g \theta + b_g]} \right]^{(1-x_g)}
\end{aligned} \tag{5.10}$$

auf. In dieser Gleichung kodiert  $x_g$  die Itemantwort und  $y_g$  den Umstand, ob

eine Reaktion vor oder nach einem Zeitlimit  $t_c$  erfolgte. Wären die Ausprägungen der Personenparameter  $\theta$  und  $\omega$  der einzelnen Versuchspersonen bekannt, ergäbe sich die „vollständige“ Log-Likelihood-Funktion<sup>2</sup> des Modells (5.5) bei Beobachtung der Itemantworten und Reaktionszeiten von  $N$  Personen mit den bekannten Personenparametern  $\boldsymbol{\theta} = [\theta_1, \dots, \theta_n]$  und  $\boldsymbol{\omega} = [\omega_1, \dots, \omega_n]$ , als:

$$\text{LL}(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{l}, \mathbf{k}) = \log \prod_{i=1}^N P(\mathbf{x}_i, \mathbf{y}_i | \theta_i, \omega_i; \mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{l}, \mathbf{k}). \quad (5.11)$$

Da jedoch die Ausprägungen der Persönlichkeitseigenschaften  $\boldsymbol{\theta}$  und  $\boldsymbol{\omega}$  der Personen nicht bekannt sind, kann diese Funktion nicht zur Schätzung der Itemparameter verwendet werden. Tatsächlich handelt es sich bei den Personenparametern  $\boldsymbol{\theta}$  und  $\boldsymbol{\omega}$  einer Person um Realisierungen von Zufallsvariablen. Versuchspersonen werden zufällig aus einer Population potentieller Testpersonen ausgewählt. Innerhalb dieser Population sind die Persönlichkeitseigenschaften nach einer bestimmten Verteilung  $f(\theta, \omega)$  verteilt. In der Regel wird angenommen, dass es sich bei der Verteilung  $f(\theta, \omega)$  um eine bivariate Standardnormalverteilung mit Korrelationskoeffizienten  $\rho = 0$  handelt. Bei Beobachtung des Antwortmusters  $\mathbf{x} = [x_1, \dots, x_G]$  und  $\mathbf{y} = [y_1, \dots, y_G]$  kann die bedingte Verteilung der Persönlichkeitseigenschaften, die Verteilung  $f(\theta, \omega | \mathbf{x}, \mathbf{y}; \mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{l}, \mathbf{k})$ , angegeben werden. Sie lautet:

$$f(\theta, \omega | \mathbf{x}, \mathbf{y}; \mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{l}, \mathbf{k}) = \frac{P(\mathbf{x}, \mathbf{y} | \theta, \omega; \mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{l}, \mathbf{k}) f(\theta, \omega)}{P(\mathbf{x}, \mathbf{y}; \mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{l}, \mathbf{k})}. \quad (5.12)$$

Der E-Schritt des EM-Algorithmus besteht darin, den auf die beobachteten Antwortmuster bedingten Erwartungswert der „vollständigen“ Log-Likelihood-Funktion zu bestimmen (McLachlan & Krishnan, 1997, S. 22). Dies setzt jedoch eigentlich voraus, dass die Itemparameter bekannt sind. Gelöst wird dieses Dilemma dadurch, dass von vorläufigen Itemparametern  $\mathbf{a}_0, \mathbf{b}_0, \mathbf{c}_0, \mathbf{l}_0, \mathbf{k}_0$  ausgegangen wird. Bei Annahme dieser Itemparameter ergibt sich als bedingter Erwartungswert der Log-Likelihood-Funktion die Beziehung:

$$\text{ELL}(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{l}, \mathbf{k}) = \sum_{i=1}^N \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \log \left[ P(\mathbf{x}_i, \mathbf{y}_i | \theta, \omega; \mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{l}, \mathbf{k}) \right] \frac{P(\mathbf{x}_i, \mathbf{y}_i | \theta, \omega; \mathbf{a}_0, \mathbf{b}_0, \mathbf{c}_0, \mathbf{l}_0, \mathbf{k}_0) f(\theta, \omega)}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(\mathbf{x}_i, \mathbf{y}_i | \theta, \omega; \mathbf{a}_0, \mathbf{b}_0, \mathbf{c}_0, \mathbf{l}_0, \mathbf{k}_0) f(\theta, \omega) d\theta d\omega} d\theta d\omega. \quad (5.13)$$

Das Ergebnis des Integrals ist jedoch nicht in geschlossener Form angebar. Zur Berechnung derartiger Integralen findet sich allerdings in der Literatur eine Vielzahl von Approximationsverfahren – siehe zum Beispiel Bos (2002), Clarkson und Zhan (2002), Evans und Swartz (1995), Gonzalez, Tuerlinckx, De Boeck und Cools (2006), Schilling und Bock (2005) oder die Übersicht in

<sup>2</sup>Streng genommen ist Funktion (5.11) nicht die vollständige Log-Likelihood-Funktion, da die Verteilung der Persönlichkeitseigenschaften  $f(\theta, \omega)$  fehlt. Diese spielt für die Itemparameterschätzung jedoch keine Rolle, so dass sie ignoriert werden kann.



Skrondal und Rabe-Hesketh (2004). Darunter fallen Verfahren der Monte-Carlo-Integration, der Quasi-Monte-Carlo-Integration, der LaPlace-Approximation sowie die in den meisten Softwarepaketen umgesetzten Quadraturverfahren. Ein Quadraturverfahren wird auch in der aktuellen Arbeit zur Approximation der erwarteten Log-Likelihood-Funktion eingesetzt. Die Klasse der Quadraturverfahren umfasst jedoch wiederum eine Vielzahl verschiedenartiger Ansätze. Bei multidimensionaler Integration mit der Normalverteilungsdichte als Gewichtsfunktion dominieren jedoch in der probabilistischen Testtheorie kartesische Quadraturregeln nach der Gauss-Hermite-Quadratur. Zwar ist eine kartesische Quadraturregel, wie sie beispielsweise von Bock, Gibbons und Muraki (1988) für die EM-Schätzung von IRT-Modellen vorgeschlagen wurde, teilweise einer sphärischen Quadraturregel unterlegen (Rabe-Hesketh, Skrondal & Pickles, 2005; Stroud, 1971), jedoch zahlt sich dieser Vorteil hauptsächlich bei einer Integration über viele Dimensionen aus. Adaptive Quadraturregeln, die in neueren Softwarepaketen umgesetzt sind – siehe zum Beispiel Rabe-Hesketh et al. (2005), Schilling und Bock (2005) oder Lessafre und Spiessens (2001) – dürften hingegen erst bei großem Itemumfang erforderlich sein (Rabe-Hesketh et al., 2005).

Bei Verwendung einer kartesischen Quadraturregel mit den Quadraturpunkten der Gauß-Hermite-Quadratur ergibt sich das multiple Integral (5.13) nach Stroud (1971) als:

$$\text{ELL}(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{l}, \mathbf{k}) = \sum_{i=1}^N \sum_{q_1=1}^{Q_1} \sum_{q_2=1}^{Q_2} \log \left[ P(\mathbf{x}_i, \mathbf{y}_i | \theta_{q_1}, \omega_{q_2}; \mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{l}, \mathbf{k}) \right] \quad (5.14)$$

$$\frac{P(\mathbf{x}_i, \mathbf{y}_i | \theta_{q_1}, \omega_{q_2}; \mathbf{a}_0, \mathbf{b}_0, \mathbf{c}_0, \mathbf{l}_0, \mathbf{k}_0)}{\sum_{q_3=1}^{Q_3} \sum_{q_4=1}^{Q_4} P(\mathbf{x}_i, \mathbf{y}_i | \theta_{q_3}, \omega_{q_4}; \mathbf{a}_0, \mathbf{b}_0, \mathbf{c}_0, \mathbf{l}_0, \mathbf{k}_0) w_{q_3} w_{q_4}} w_{q_1} w_{q_2}.$$

Die Größen  $\theta_{q_1}$ ,  $\omega_{q_2}$ ,  $\theta_{q_3}$  und  $\omega_{q_4}$  sind dabei die Quadraturpunkte,  $w_{q_1}$ ,  $w_{q_2}$ ,  $w_{q_3}$  und  $w_{q_4}$  die zu den Quadraturpunkten gehörenden Gewichte. Die Werte hierfür können Standardlehrbüchern zur numerischen Mathematik entnommen werden, siehe zum Beispiel Davis und Rabinowitz (1975). Gleichung (5.14) kann folgendermaßen interpretiert werden: Pro Antwortmuster  $[\mathbf{x}_i, \mathbf{y}_i]$  wird angenommen, dass dieses bei Personen mit den diskreten Ausprägungen  $\theta_{q_1}$  und  $\omega_{q_2}$  der Persönlichkeitseigenschaften in einer "relativen Häufigkeit" von

$$w_{\mathbf{x}_i, \mathbf{y}_i, \theta_{q_1}, \omega_{q_2}} = \frac{P(\mathbf{x}_i, \mathbf{y}_i | \theta_{q_1}, \omega_{q_2}, \mathbf{a}_0, \mathbf{b}_0, \mathbf{c}_0, \mathbf{l}_0, \mathbf{k}_0)}{\sum_{q_3}^{Q_3} \sum_{q_4}^{Q_4} P(\mathbf{x}, \mathbf{y} | \theta_{q_3}, \omega_{q_4}, \mathbf{a}_0, \mathbf{b}_0, \mathbf{c}_0, \mathbf{l}_0, \mathbf{k}_0) w_{q_3} w_{q_4}} w_{q_1} w_{q_2} \quad (5.15)$$

auftrat. Eine derartige Interpretation gibt beispielsweise Baker und Seock-Ho (2004) oder auch Fahrmeir und Tutz (1994, S. 244). Der Rechenaufwand reduziert sich etwas, wenn anstelle der Antwortmuster der einzelnen Personen die verschiedenen Antwortmuster betrachtet werden. Summiert wird dann in Gleichung (5.14) nicht über die  $n$  Personen, sondern über die unterschiedlichen Antwortmuster, die allerdings noch mit ihrer Auftrittshäufigkeit gewichtet werden müssen.

Die Verwendung der Gauss-Hermite-Quadratur ist nicht ganz unproblematisch und wurde beispielsweise von McCulloch und Searle (2001) eher kritisch bewertet. Zwar ist mit steigender Anzahl der Quadraturpunkte die Zielfunktion mit immer größerer Genauigkeit approximierbar (Bock & Aitkin, 1981), so dass mit steigender Anzahl von Quadraturpunkten die Itemparameter bei zunehmendem Stichprobenumfang konsistent geschätzt werden können (Fahrmeir & Tutz, 1994, S. 242), doch lassen sich (brauchbare) Obergrenzen für den Approximationsfehler nicht einfach berechnen. Aus Untersuchungen zur Schätzung von IRT-Modellen ist bekannt, dass bei langen Tests die Genauigkeit erst ab einer größeren Anzahl von Quadraturpunkten akzeptabel ist (Lessafre & Spiesens, 2001), bei kurzen Tests jedoch bereits drei Quadraturpunkte ausreichen (Schilling & Bock, 2005). Empfehlungen für die notwendige Anzahl von Quadraturpunkten bei einer zweidimensionalen Integration im Kontext generalisierter linearer gemischter Modelle reichen daher von 5 bis 20 Quadraturpunkten pro Dimension (Tuerlinckx et al., 2006). Nach Versuchen mit Datensätzen verschiedenen Umfangs erschien für das aktuelle Modell eine Approximation mit  $7 \times 7$  Quadraturpunkten ausreichend – diese Anzahl dürfte einen Kompromiss zwischen einer angemessenen Genauigkeit bei gleichzeitig noch vertretbarem Rechenaufwand darstellen. Für praktische Anwendungen mit längeren Tests sollte jedoch die Anzahl der Quadraturpunkte erhöht werden.

### M-Schritt

Die erwartete Log-Likelihood-Funktion (5.14) ist eine Funktion der unbekanntem Itemparameter. Bei der Maximierung der erwarteten Log-Likelihood-Funktion kann man davon ausgehen, dass ein bestimmtes Antwortmuster  $[\mathbf{x}_i, \mathbf{y}_i]$  bei Personen mit den Personenparametern  $\theta_{q_1}$  und  $\omega_{q_2}$  mit einer relativer Häufigkeit von  $w_{\mathbf{x}_i, \mathbf{y}_i, \theta_{q_1}, \omega_{q_2}}$  auftritt. Die Maximierung ist also identisch zur Maximierung bei bekannten Personenparametern, wenn zusätzlich die Gewichtung mit  $w_{\mathbf{x}_i, \mathbf{y}_i, \theta_{q_1}, \omega_{q_2}}$  berücksichtigt wird. Gesucht wird die Lösung für:

$$\begin{aligned} \arg \max_{\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{l}, \mathbf{k}} \left[ \text{ELL}(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{l}, \mathbf{k}) \right] = \\ \arg \max_{\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{l}, \mathbf{k}} \left[ \sum_{i=1}^N \sum_{q_1=1}^{Q_1} \sum_{q_2=1}^{Q_2} \log \left[ P(\mathbf{x}_i, \mathbf{y}_i | \theta_{q_1}, \omega_{q_2}; \mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{l}, \mathbf{k}) \right] w_{\mathbf{x}_i, \mathbf{y}_i, \theta_{q_1}, \omega_{q_2}} \right]. \end{aligned} \quad (5.16)$$

Gegenüber der marginalen Log-Likelihood-Funktion (5.9) besitzt die erwartete Log-Likelihood-Funktion (5.14) jedoch eine sehr vorteilhafte Eigenschaft: Wie aus Gleichung (5.16) ersichtlich ist, kann die Funktion separat für die Itemparameter jedes Items maximiert werden, was die Maximierung erheblich erleichtert. Dies ist der entscheidende Vorteil des EM-Algorithmus gegenüber einer direkten Maximierung der marginalen Likelihood-Funktion. Trotz dieser Vereinfachung ist das Maximum von  $\text{ELL}(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{l}, \mathbf{k})$  nicht explizit bestimmbar.

Notwendige Bedingung für ein Maximum ist, dass der Gradientenvektor der erwarteten Log-Likelihood-Funktion an der Stelle des potentiellen Maximums

eine Nullstelle besitzt. Der Gradientenvektor  $\mathbf{G}_y$  für die Parameter des Reaktionszeitenmodells bei Aufgabe  $g$  lautet:

$$\begin{aligned}
\frac{\partial}{\partial c_g} \text{ELL}(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{l}, \mathbf{k}) &= \sum_{i=1}^N \sum_{q_1=1}^{Q_1} \sum_{q_2=1}^{Q_2} \left[ \frac{y_{ig} - P_{y_g}(\theta_{q_1}, \omega_{q_2}; c_g, l_g, k_g)}{P_{y_g}(\theta_{q_1}, \omega_{q_2}; c_g, l_g, k_g)} \right] \\
&\quad \times \exp[k_g + l_g \theta_{q_1} + c_g \omega_{q_2}] \omega_{q_2} w_{\mathbf{x}_i, \mathbf{y}_i, \theta_{q_1}, \omega_{q_2}} \\
\frac{\partial}{\partial l_g} \text{ELL}(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{l}, \mathbf{k}) &= \sum_{i=1}^N \sum_{q_1=1}^{Q_1} \sum_{q_2=1}^{Q_2} \left[ \frac{y_{ig} - P_{y_g}(\theta_{q_1}, \omega_{q_2}; c_g, l_g, k_g)}{P_{y_g}(\theta_{q_1}, \omega_{q_2}; c_g, l_g, k_g)} \right] \\
&\quad \times \exp[k_g + l_g \theta_{q_1} + c_g \omega_{q_2}] \theta_{q_1} w_{\mathbf{x}_i, \mathbf{y}_i, \theta_{q_1}, \omega_{q_2}} \\
\frac{\partial}{\partial k_g} \text{ELL}(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{l}, \mathbf{k}) &= \sum_{i=1}^N \sum_{q_1=1}^{Q_1} \sum_{q_2=1}^{Q_2} \left[ \frac{y_{ig} - P_{y_g}(\theta_{q_1}, \omega_{q_2}; c_g, l_g, k_g)}{P_{y_g}(\theta_{q_1}, \omega_{q_2}; c_g, l_g, k_g)} \right] \\
&\quad \times \exp[k_g + l_g \theta_{q_1} + c_g \omega_{q_2}] 1 w_{\mathbf{x}_i, \mathbf{y}_i, \theta_{q_1}, \omega_{q_2}}.
\end{aligned} \tag{5.17}$$

Analog ergibt sich der Gradientenvektor  $\mathbf{G}_x$  für die Itemparameter des zweiparametrischen Logit-Modells als:

$$\begin{aligned}
\frac{\partial}{\partial a_g} \text{ELL}(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{l}, \mathbf{k}) &= \sum_{i=1}^N \sum_{q_1=1}^Q \sum_{q_2=1}^Q \left[ x_{ig} - P_{x_g}(\theta_{q_1}; a_g, b_g) \right] \theta_{q_1} w_{\mathbf{x}_i, \mathbf{y}_i, \theta_{q_1}, \omega_{q_2}} \\
\frac{\partial}{\partial b_g} \text{ELL}(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{l}, \mathbf{k}) &= \sum_{i=1}^N \sum_{q_1=1}^Q \sum_{q_2=1}^Q \left[ x_{ig} - P_{x_g}(\theta_{q_1}; a_g, b_g) \right] 1 w_{\mathbf{x}_i, \mathbf{y}_i, \theta_{q_1}, \omega_{q_2}}.
\end{aligned} \tag{5.18}$$

Die Nullstellen der Gradientenvektoren (5.17) und (5.18) sind nicht explizit angebar. Ein Verfahren zur iterativen Bestimmung von Nullstellen ist der Newton-Raphson-Algorithmus. Möchte man die Nullstellen des Gradientenvektors mit dem Newton-Raphson-Algorithmus bestimmen, ist zusätzlich die Jacobi-Matrix des Gradientenvektors zu berechnen. Eine erneute Ableitung des Gradientenvektors  $\mathbf{G}_y$  nach den Itemparametern des RT-Modells führt zur Jacobi-Matrix  $\mathbf{J}_y$ . Da diese Matrix bereits eine  $3 \times 3$ -Matrix ist, sei auf eine genaue Auflistung der einzelnen Elemente verzichtet. Analog ergibt eine Ableitung des Gradientenvektors  $\mathbf{G}_x$  nach den Itemparametern des IRT-Modells die Jacobi-Matrix  $\mathbf{J}_x$ . Auch die verschiedenen Elemente dieser Matrix seien nicht einzeln aufgeführt. Der Newton-Raphson-Algorithmus basiert auf iterativen Verbesserungen einer vorläufigen Lösung. Als vorläufige Lösungen für die Itemparameter des  $g$ -ten Items kann beispielsweise im  $j$ -ten Iterationsschritt die Lösung  $[\hat{a}_{g_j}, \hat{b}_{g_j}]'$  bzw.  $[\hat{c}_{g_j}, \hat{l}_{g_j}, \hat{k}_{g_j}]'$  bestimmt worden sein. Nach Berechnung der Gradientenvektoren der Reaktionszeit  $\mathbf{G}_{y_j}$  und der Itemantworten  $\mathbf{G}_{x_j}$  an der Stelle der vorläufigen Itemparameter sowie der Jacobi-Matrizen der Reaktionszeit  $\mathbf{J}_{y_j}$  und der Itemantwort  $\mathbf{J}_{x_j}$  ergibt sich die verbesserte Lösung im  $j+1$ -ten Schritt als:

$$[\widehat{a}_{g_{j+1}}, \widehat{b}_{g_{j+1}}]' = [\widehat{a}_{g_j}, \widehat{b}_{g_j}]' - \mathbf{J}_{x_j}^{-1} \mathbf{G}_{x_j} \quad (5.19)$$

$$[\widehat{c}_{g_{j+1}}, \widehat{l}_{g_{j+1}}, \widehat{k}_{g_{j+1}}]' = [\widehat{c}_{g_j}, \widehat{l}_{g_j}, \widehat{k}_{g_j}]' - \mathbf{J}_{y_j}^{-1} \mathbf{G}_{y_j}. \quad (5.20)$$

Der Verbesserungszyklus (5.19) und (5.20) wird solange wiederholt, bis sich die Schätzwerte kaum mehr verändern. Der Newton-Raphson-Algorithmus ist ein populäres Verfahren, dessen Anwendung jedoch nicht unproblematisch ist. Der Newton-Raphson-Algorithmus konvergiert nicht zwangsläufig gegen eine Nullstelle: Die Konvergenz ist nur in einer Umgebung der Nullstelle garantiert, so dass bei weniger guten Startwerten der Algorithmus teilweise nicht zu einer Lösung kommt. Und selbst wenn es sich bei der gefundenen Lösung um eine Nullstelle der Gradientenvektoren handeln sollte, ist diese Lösung nicht notwendigerweise ein lokales Maximum der erwarteten Log-Likelihood-Funktion (5.16). Es muss daher immer überprüft werden, ob es sich bei den gefundenen Itemparametern wirklich um Werte handelt, welche die erwartete Log-Likelihood-Funktion maximieren.

Die Bestimmung des Maximums der erwarteten Log-Likelihood-Funktion beendet den M-Schritt und schließt einen Iterationszyklus des EM-Algorithmus ab. Ob ein neuer EM-Zyklus gestartet wird, hängt von der Veränderung der marginalen Log-Likelihood-Funktion ab. Fällt die Verbesserung der marginalen Log-Likelihood-Funktion für die neuen Itemparameter groß aus, wird ein weiterer EM-Zyklus mit den neuen Itemparametern begonnen. Tritt jedoch keine wesentliche Verbesserung mehr auf, wird der EM-Algorithmus abgebrochen und die im letzten Schritt berechneten Itemparameter als endgültige Lösung verwandt.

### Simulation der Itemparameterschätzung

Zur Überprüfung der Eignung des im vorherigen Abschnitt beschriebenen Schätzverfahrens wurde eine Simulation für einen Test aus fünf Items durchgeführt. Als Personenstichprobe dienten zufällig generierte Personenfähigkeiten und Residualfaktoren, welche aus einer bivariaten Standardnormalverteilung mit Korrelationskoeffizienten  $\rho = 0$  gezogen wurden. Pro Person wurden für die fünf Items exponentialverteilte Reaktionszeiten generiert, bei denen die Grundhazard-Raten  $h_{B_1}(t) = 0.67$ ,  $h_{B_2}(t) = 0.74$ ,  $h_{B_3}(t) = 0.82$ ,  $h_{B_4}(t) = 0.91$  und  $h_{B_5}(t) = 1.06$  gemäß der Beziehung  $h_{\theta, \omega}(t) = h_{B_g}(t) \exp[l_g \theta + c_g \omega]$  von den Persönlichkeitseigenschaften  $\theta$  und  $\omega$  moduliert wurden. Die Reaktionszeiten wurden dann an der Schwelle  $t_c = 1$  dichotomisiert. Die Generation der Reaktionszeiten entsprach somit dem Proportional Hazard Modell für dichotomisierte Reaktionszeiten (5.2). Itemantworten wurden nach dem zweiparametrischen Logit-Modell (5.4) erzeugt. Die bei der Simulation verwendeten Itemparameter sind für die fünf Aufgaben in Tabelle 5.1 zu finden.

Hauptkriterium bei der Auswahl der Itemparameter war es, möglichst realistische Beobachtungen zu erzeugen: Die gewählten Itemparameter führen zu relativ plausiblen marginalen Lösungswahrscheinlichkeiten, welche bei den einzelnen Items im Bereich von 0.3–0.6 lagen. Die marginalen Wahrscheinlichkeiten

Tabelle 5.1: Dichotomisierte RT mit linearer Parametrisierung: Wahre Itemparameter der Aufgaben

Item	$a_g$	$b_g$	$c_g$	$l_g$	$k_g$
1	1.50	0.66	0.50	0.20	-0.39
2	1.50	0.33	0.50	0.20	-0.29
3	1.50	0.00	0.30	0.30	-0.19
4	1.50	-0.33	0.20	0.50	-0.09
5	1.50	-0.66	0.20	0.50	0.01

einer Reaktion vor  $t_c = 1$  fielen für die einzelnen Aufgaben in den Bereich von 0.5 – 0.6. Insgesamt bedeuten die gewählten Einflussgewichte  $l_g$  nur einen relativ moderaten Einfluss der Personenfähigkeit  $\theta$  auf die Reaktionszeiten, was den empirischen Befunden über relativ schwache Beziehungen zwischen der Personenfähigkeit  $\theta$  und unterschiedlichen Speedmaßen bei Leistungstests gerecht wird – siehe hierzu die Ausführungen im ersten Kapitel. Auch waren die schwereren Items diejenigen Aufgaben, auf die langsamer reagiert wurde.

Simuliert wurden jeweils 35 Stichproben des Umfangs von 100 Versuchspersonen, 35 Stichproben des Umfangs von 1000 Versuchspersonen und 35 Stichproben des Umfangs von 10000 Versuchspersonen. Aufgrund der langen Dauer der Schätzungen (pro Datensatz bei 10000 Versuchspersonen etwa 6 Stunden) wurde auf eine größere Anzahl von Durchgängen verzichtet. Startwerte waren für alle fünf Items die Werte  $k_g = 0$ ,  $c_g = 0.3$ ,  $l_g = 0.3$ ,  $a_g = 1$  und  $b_g = 0$ . Derartige Startwerte hätte man vermutlich auch ohne Kenntnis der wahren Itemparameter gewählt. Der EM-Algorithmus wurde beendet, wenn der Zuwachs der marginalen Log-Likelihood-Funktion in einer Iteration geringer als ein bestimmter Wert ausfiel. Als Abbruchkriterium diente bei 10000 Versuchspersonen der Wert 0.5, bei 1000 Versuchspersonen der Wert 0.05 und bei 100 Versuchspersonen der Wert 0.01.<sup>3</sup> Die erwartete Log-Likelihood-Funktion wurde mit einem numerischen Quadratur-Verfahren bestimmt. Pro Dimension basierte die Quadratur auf sieben Quadraturpunkten, was ein sinnvoller Kompromiss zwischen vertretbarem Rechenaufwand und Genauigkeit darstellt.

Das Konvergenzverhalten des EM-Algorithmus war insgesamt bei den Versuchen mit größerer Fallzahl gut. Bei den Datensätzen mit 10000 und 1000 Fällen konvergierte jeder Versuch. Bei 10000 Versuchspersonen betrug die mittlere Anzahl der EM-Zyklen 11.64, die maximale Anzahl 13. Bei 1000 Versuchspersonen stieg diese jeweils auf 13.34 und 22 an. Bei kleineren Stichprobenumfängen wurde die Schätzung jedoch instabil. Bei einem Stichprobenumfang von 100 Personen konvergierte das Verfahren nur bei rund 75% der generierten Datensätze. Aus diesem Grund wurden so lange weitere Datensätze erzeugt, bis 35 Schätzversuche erfolgreich abgeschlossen werden konnten. Insgesamt stieg die notwendige Iterationszahl bei Stichproben aus 100 Personen auf durchschnitt-

<sup>3</sup>Die Log-Likelihood-Funktionen nahmen an ihrem Maximum bei 10.000 Versuchspersonen einen Wert um -60.000, bei 1000 Versuchspersonen einen Wert um -6.000 und bei 100 Versuchspersonen einen Wert um -600 an. Die gewählten Abbruchkriterien entsprachen somit in etwa derselben relativen Veränderung. Zudem zeigte sich, dass eine Verringerung des Abbruchkriteriums zu keinen starken Veränderungen der Parameterschätzer führte.

lich 20.6 an, mit einem Maximum von 54 Iterationen. Numerische Probleme, wie sie teilweise bei sehr steilen Likelihood-Funktionen auftreten können, spielen bei den gewählten Simulationsbedingungen keine Rolle. Bei größerer Itemanzahl jedoch ist eine Erhöhung der Anzahl der Quadraturpunkte sicherlich ratsam.

Die Ergebnisse der Simulation sind in den Tabellen 5.2 bis 5.4 dargestellt. Wie in Tabelle 5.2 und Tabelle 5.3 zu sehen ist, kann bei ausreichend großem Stichprobenumfang eine gute Schätzung der Parameter erreicht werden. Die Mittelwerte der Parameterschätzungen liegen immer nahe an den wahren Werten und auch die maximalen und minimalen Schätzwerte sind noch gute Schätzwerte. Weniger gut sind die Resultate bei einem Stichprobenumfang von 100 Personen, wie anhand Tabelle 5.4 festgestellt werden kann. Bei der Interpretation von Tabelle 5.4 ist zudem zu berücksichtigen, dass es sich bei den dargestellten Ergebnissen um die erfolgreichen Schätzversuche handelt, bei denen die Lösung konvergierte. Die Ergebnisse sind somit eine Auswahl günstiger Schätzversuche. Extreme Schätzwerte treten für die Trennschärfe  $a_g$  des Logit-Modells auf: Hier finden sich teilweise Werte im Bereich von 7, was ein Hinweis auf ein mögliches Konvergenzproblem ist. Aber auch die Einflussgewichte  $c_g$  und  $l_g$  werden nur schlecht geschätzt, teilweise ist sogar das Vorzeichen falsch. Auch dies deutet auf eine Instabilität des Verfahrens bei kleinen Stichprobenumfängen hin.

## 5.2 Quadratische Parametrisierung

Modelle für eine „Unfolding“-Beziehung zwischen Reaktionszeit und Personenfähigkeit  $\theta$  existieren kaum, auch wenn sich Hinweise auf eine inverse U-Beziehung finden. Ein Reaktionszeitmodell zur Modellierung einer derartigen Beziehung ist das Proportional Hazard Modell für dichotomisierte Reaktionszeiten mit der quadratischen Parametrisierung (4.23). Nach diesem Modell folgt die Hazard-Rate einer Person der Beziehung:

$$h_{\theta,\omega}(t) = h_{B_g}(t) \exp [l_g(\theta - d_g)^2 + c_g\omega]. \quad (5.21)$$

Nur die Personenfähigkeit  $\theta$  steht dabei in einer Unfolding-Beziehung zur Reaktionszeit, die zweite Persönlichkeitseigenschaft  $\omega$  beeinflusst die Hazard-Rate weiterhin monoton. Anstelle der kontinuierlichen Reaktionszeit  $t_g$  kann lediglich festgehalten werden, ob auf ein Item vor einem Zeitlimit  $t_c$  reagiert wurde ( $y_g = 1$ ) oder danach ( $y_g = 0$ ). Die für das Modell resultierende Wahrscheinlichkeit, auf Aufgabe  $g$  vor dem Zeitpunkt  $t_c$  zu reagieren, berechnet sich dabei als

$$P_{y_g}(\theta, \omega; c_g, d_g, l_g, k_g) = 1 - \exp \left[ - \exp [k_g + l_g(\theta - d_g)^2 + c_g\omega] \right], \quad (5.22)$$

wobei  $k_g = \log \left( \int_0^{t_c} h_{B_g}(z) dz \right)$  repräsentiert. Für die Wahrscheinlichkeit, bei Item  $g$  eine positive Itemantwort zu geben, wird wie bisher ein zweiparametrisches Logit-Modell (3.1) angenommen. Es gilt:

Tabelle 5.2: Itemparameterschätzung beim Proportional Hazard Modell für dichotomisierte Reaktionszeiten mit linearer Parametrisierung – Ergebnisse der 35 Simulationen mit N=10000

Parameter a					
Item	1	2	3	4	5
Wahrer Wert	1.500	1.500	1.500	1.500	1.500
Min	1.362	1.374	1.367	1.362	1.388
Max	1.601	1.594	1.583	1.620	1.624
Mean	1.485	1.497	1.483	1.497	1.511
STD	0.056	0.052	0.056	0.057	0.053
RMSE	0.057	0.053	0.059	0.057	0.054
Parameter b					
Item	1	2	3	4	5
Wahrer Wert	0.666	0.333	0.000	-0.333	-0.666
Min	0.617	0.292	-0.038	-0.372	-0.725
Max	0.723	0.372	0.039	-0.302	-0.624
Mean	0.674	0.333	-0.004	-0.329	-0.666
STD	0.025	0.021	0.019	0.019	0.026
RMSE	0.026	0.021	0.019	0.020	0.026
Parameter c					
Item	1	2	3	4	5
Wahrer Wert	0.500	0.500	0.300	0.200	0.200
Min	0.390	0.404	0.246	0.172	0.164
Max	0.542	0.504	0.408	0.273	0.289
Mean	0.470	0.466	0.317	0.224	0.224
STD	0.033	0.022	0.036	0.025	0.027
RMSE	0.044	0.041	0.039	0.035	0.036
Parameter l					
Item	1	2	3	4	5
Wahrer Wert	0.200	0.200	0.300	0.500	0.500
Min	0.158	0.148	0.265	0.466	0.464
Max	0.236	0.235	0.339	0.533	0.550
Mean	0.201	0.194	0.298	0.497	0.496
STD	0.021	0.019	0.019	0.016	0.020
RMSE	0.021	0.020	0.019	0.016	0.021
Parameter k					
Item	1	2	3	4	5
Wahrer Wert	-0.393	-0.293	-0.193	-0.093	0.006
Min	-0.432	-0.334	-0.229	-0.121	-0.035
Max	-0.347	-0.267	-0.163	-0.064	0.039
Mean	-0.394	-0.295	-0.195	-0.096	0.007
STD	0.017	0.017	0.017	0.014	0.015
RMSE	0.017	0.017	0.017	0.014	0.020

Tabelle 5.3: Itemparameterschätzung beim Proportional Hazard Modell für dichotomisierte Reaktionszeiten mit linearer Parametrisierung – Ergebnisse der 35 Simulationen mit N=1000

Parameter a					
Item	1	2	3	4	5
Wahrer Wert	1.500	1.500	1.500	1.500	1.500
Min	1.144	1.250	1.289	1.139	1.165
Max	1.819	1.901	1.788	1.788	1.840
Mean	1.518	1.512	1.503	1.496	1.502
STD	0.151	0.159	0.138	0.166	0.152
RMSE	0.152	0.160	0.138	0.166	0.152
Parameter b					
Item	1	2	3	4	5
Wahrer Wert	0.666	0.333	0.000	-0.333	-0.666
Min	0.518	0.192	-0.145	-0.493	-0.866
Max	0.780	0.539	0.137	-0.174	-0.497
Mean	0.668	0.342	0.006	-0.328	-0.671
STD	0.060	0.069	0.059	0.073	0.079
RMSE	0.060	0.069	0.059	0.073	0.079
Parameter c					
Item	1	2	3	4	5
Wahrer Wert	0.500	0.500	0.300	0.200	0.200
Min	0.249	0.260	0.130	0.061	0.049
Max	0.676	0.963	0.522	0.430	0.436
Mean	0.466	0.499	0.320	0.243	0.216
STD	0.101	0.132	0.100	0.089	0.104
RMSE	0.107	0.132	0.102	0.098	0.105
Parameter l					
Item	1	2	3	4	5
Wahrer Wert	0.200	0.200	0.300	0.500	0.500
Min	0.018	0.070	0.162	0.381	0.349
Max	0.330	0.351	0.473	0.646	0.621
Mean	0.195	0.205	0.291	0.506	0.499
STD	0.062	0.062	0.058	0.063	0.062
RMSE	0.062	0.063	0.059	0.064	0.062
Parameter k					
Item	1	2	3	4	5
Wahrer Wert	-0.393	-0.293	-0.193	-0.093	0.006
Min	-0.489	-0.395	-0.296	-0.244	-0.120
Max	-0.296	-0.191	-0.112	-0.034	0.129
Mean	-0.391	-0.292	-0.202	-0.108	0.006
STD	0.052	0.054	0.042	0.053	0.053
RMSE	0.052	0.054	0.043	0.055	0.054



Tabelle 5.4: Itemparameterschätzung beim Proportional Hazard Modell für dichotomisierte Reaktionszeiten mit linearer Parametrisierung – Ergebnisse der 35 konvergierenden Simulationen mit  $N=100$

Parameter a					
Item	1	2	3	4	5
Wahrer Wert	1.500	1.500	1.500	1.500	1.500
Min	0.554	0.753	0.697	0.888	0.573
Max	4.002	4.693	5.508	7.449	7.469
Mean	1.803	1.631	1.876	1.782	1.499
STD	0.664	0.755	0.955	1.145	1.137
RMSE	0.729	0.766	1.026	1.179	1.137
Parameter b					
Item	1	2	3	4	5
Wahrer Wert	0.666	0.333	0.000	-0.333	-0.666
Min	0.152	0.004	-0.317	-0.645	-1.573
Max	1.933	0.964	0.893	0.149	-0.166
Mean	0.642	0.395	0.023	-0.256	-0.741
STD	0.298	0.226	0.234	0.224	0.357
RMSE	0.299	0.234	0.235	0.237	0.364
Parameter c					
Item	1	2	3	4	5
Wahrer Wert	0.500	0.500	0.300	0.200	0.200
Min	-0.362	-0.402	-0.514	-0.633	-0.612
Max	1.468	1.510	0.927	0.725	0.814
Mean	0.509	0.589	0.368	0.197	0.150
STD	0.405	0.467	0.361	0.302	0.294
RMSE	0.405	0.475	0.367	0.302	0.298
Parameter l					
Item	1	2	3	4	5
Wahrer Wert	0.200	0.200	0.300	0.500	0.500
Min	-0.209	-0.276	-0.127	0.076	0.190
Max	0.722	0.638	1.049	1.276	1.222
Mean	0.253	0.243	0.312	0.604	0.616
STD	0.247	0.248	0.238	0.250	0.237
RMSE	0.252	0.251	0.239	0.264	0.270
Parameter k					
Item	1	2	3	4	5
Wahrer Wert	-0.393	-0.293	-0.193	-0.093	0.006
Min	-0.777	-0.708	-0.559	-0.509	-0.237
Max	-0.046	-0.108	-0.016	0.259	0.247
Mean	-0.418	-0.347	-0.250	-0.083	-0.001
STD	0.173	0.166	0.118	0.185	0.126
RMSE	0.175	0.174	0.131	0.186	0.127

$$P_{x_g}(\theta; a_g, d_g) = \frac{\exp[a_g(\theta - d_g)]}{1 + \exp[a_g(\theta - d_g)]}. \quad (5.23)$$

Man beachte die gegenüber Gleichung (3.1) leicht veränderte Parametrisierung. Ein wesentliches Merkmal der gemeinsamen Modellierung von Reaktionszeitverteilung und Itemantwortverteilung ist die Annahme, dass der Lokationsparameter  $d_g$  des Proportional Hazard Modells (5.21) und der Lokationsparameter  $b_g$  des zweiparametrischen Logit-Modells (3.1) identisch sind ( $b_g = d_g$ ). Die gemeinsame Wahrscheinlichkeit, bei Item  $g$  eine bestimmte Itemantwort vor oder nach dem Zeitpunkt  $t_c$  zu beobachten, lautet bei Annahme der bedingten Unabhängigkeit von Reaktionszeit und Itemantwort:

$$P(x_g, y_g | \theta, \omega; a_g, d_g, c_g, l_g, k_g) = [P_{x_g}(\theta; a_g, d_g)]^{x_g} [1 - P_{x_g}(\theta; a_g, d_g)]^{(1-x_g)} \\ [P_{y_g}(\theta, \omega; c_g, d_g, l_g, k_g)]^{y_g} [1 - P_{y_g}(\theta, \omega; c_g, d_g, l_g, k_g)]^{(1-y_g)}. \quad (5.24)$$

In Gleichung (5.24) kodiert die Zufallsvariable  $x_g$  die Art der Itemantwort, die Zufallsvariable  $y_g$  den Umstand, ob auf ein Item vor oder nach dem Zeitlimit  $t_c$  reagiert wurde. Bei Beobachtung von  $G$  Aufgaben ergibt sich aufgrund der bedingten Unabhängigkeit von Reaktionszeiten und Itemantworten die Wahrscheinlichkeit, dass eine Antwortmuster  $\mathbf{x} = [x_1, \dots, x_G]$  und  $\mathbf{y} = [y_1, \dots, y_G]$  bei einer Person mit Personenfähigkeit  $\theta$  und Residualfaktor  $\omega$  auftritt, als:

$$P(\mathbf{x}, \mathbf{y} | \theta, \omega; \mathbf{a}, \mathbf{d}, \mathbf{c}, \mathbf{l}, \mathbf{k}) = \prod_{g=1}^G P(x_g, y_g | \theta, \omega; a_g, d_g, c_g, l_g, k_g). \quad (5.25)$$

Die Vektoren  $\mathbf{a}$ ,  $\mathbf{d}$ ,  $\mathbf{c}$ ,  $\mathbf{l}$  und  $\mathbf{k}$  fassen die einzelnen Itemparameter der  $G$  Items zusammen. Möchte man die unbekannt Itemparameter anhand einer Stichprobe bestimmen, stellt sich wiederum das Problem, dass mit größer werdendem Stichprobenumfang die Anzahl der unbekannt Personenparameter ansteigt, so dass eine konsistente Schätzung der Itemparameter nicht garantiert ist.

### Schätzung der Itemparameter

Nimmt man für die Verteilung der unbeobachtbaren Personenparameter  $\theta$  und  $\omega$  eine bivariate Standardnormalverteilung mit  $\rho = 0$  an, führt eine Integration der bedingten Wahrscheinlichkeit (5.25) über die Personenparameter  $\theta$  und  $\omega$  zur marginalen Wahrscheinlichkeit eines Antwortmusters  $P(\mathbf{x}, \mathbf{y}; \mathbf{a}, \mathbf{d}, \mathbf{c}, \mathbf{l}, \mathbf{k})$ . Für die marginale Log-Likelihood-Funktion erhält man dann bei einer Stichprobe aus  $N$  Personen die Beziehung:

$$\text{MLL}(\mathbf{a}, \mathbf{d}, \mathbf{c}, \mathbf{l}, \mathbf{k}) = \log \prod_{i=1}^N P(\mathbf{x}_i, \mathbf{y}_i; \mathbf{a}, \mathbf{d}, \mathbf{c}, \mathbf{l}, \mathbf{k}). \quad (5.26)$$

Bei dem Marginal-Maximum-Likelihood-Ansatz werden die Ausprägungen der Parameter  $\mathbf{a}$ ,  $\mathbf{d}$ ,  $\mathbf{c}$ ,  $\mathbf{l}$  und  $\mathbf{k}$  gewählt, welche die marginale Log-Likelihood-Funktion maximieren. Das Maximum der marginalen Likelihood-Funktion kann über den EM-Algorithmus gefunden werden. Die einzelnen Schritte entsprechen weitgehend der bereits in Kapitel 5.1 beschriebenen Vorgehensweise. Wiederum kann der EM-Algorithmus in zwei Phasen, den E-Schritt und den M-Schritt untergliedert werden. Die genauen Details der Schätzung werden in den beiden nächsten Abschnitten beschrieben. Die Details sind eher technischer Natur und können beim ersten Lesen übersprungen werden.

### E-Schritt

Ausgangspunkt des E-Schrittes ist die Rückkehr zur Likelihood-Funktion, die sich bei Kenntnis der Ausprägungen von  $\theta$  und  $\omega$  bei den Versuchspersonen ergeben hätte. Laut der Annahmen des Modells kann die Wahrscheinlichkeit eines bestimmten Antwortmusters  $[\mathbf{x}, \mathbf{y}]$  bei einer Person mit Personenfähigkeit  $\theta$  und Residualfaktor  $\omega$  angegeben werden als:

$$\begin{aligned}
 P(\mathbf{x}, \mathbf{y} | \theta, \omega; \mathbf{a}, \mathbf{d}, \mathbf{c}, \mathbf{l}, \mathbf{k}) &= \prod_{g=1}^G P(x_g, y_g | \theta, \omega; a_g, d_g, c_g, l_g, k_g) \\
 &= \prod_{g=1}^G \left[ P_{x_g}(\theta; a_g, d_g) \right]^{x_g} \left[ 1 - P_{x_g}(\theta; a_g, d_g) \right]^{(1-x_g)} \\
 &\quad \left[ P_{y_g}(\theta, \omega; c_g, d_g, l_g, k_g) \right]^{y_g} \left[ 1 - P_{y_g}(\theta, \omega; c_g, d_g, l_g, k_g) \right]^{(1-y_g)} \quad (5.27) \\
 &= \prod_{g=1}^G \left[ \frac{\exp[a_g(\theta - d_g)]}{1 + \exp[a_g(\theta - d_g)]} \right]^{x_g} \left[ 1 - \frac{\exp[a_g(\theta - d_g)]}{1 + \exp[a_g(\theta - d_g)]} \right]^{(1-x_g)} \\
 &\quad \left[ 1 - \exp \left[ - \exp[k_g + l_g(\theta - d_g)^2 + c_g\omega] \right] \right]^{y_g} \\
 &\quad \left[ \exp \left[ - \exp[k_g + l_g(\theta - d_g)^2 + c_g\omega] \right] \right]^{(1-y_g)}.
 \end{aligned}$$

Wären die Personenparameter bekannt, könnte bei Beobachtung mehrerer Personen mit Gleichung (5.27) die Log-Likelihood-Funktion aufgestellt und die Itemparameter geschätzt werden. Das ist jedoch nicht möglich, wenn die Personenparameter unbekannt sind.

Testpersonen werden zufällig aus einer Population von potentiellen Testpersonen ausgewählt. Innerhalb dieser Population folgen die Personenfähigkeit  $\theta$  und der Residualfaktor  $\omega$  einer bestimmten Verteilung, für die in der Regel eine bivariate Standardnormalverteilung mit Korrelationskoeffizienten  $\rho = 0$  angenommen wird. Legt man diese Verteilung zugrunde, kann bei Annahme bestimmter Itemparameterausprägungen und bei Vorliegen eines bestimmten Antwortmusters  $[\mathbf{x}, \mathbf{y}]$  die bedingte Verteilung der Personenmerkmale  $\theta$  und  $\omega$  beschrieben werden mit der Funktion:

$$f(\theta, \omega | \mathbf{x}, \mathbf{y}; \mathbf{a}, \mathbf{d}, \mathbf{c}, \mathbf{l}, \mathbf{k}) = \frac{P(\mathbf{x}, \mathbf{y} | \theta, \omega; \mathbf{a}, \mathbf{d}, \mathbf{c}, \mathbf{l}, \mathbf{k}) f(\theta, \omega)}{P(\mathbf{x}, \mathbf{y}; \mathbf{a}, \mathbf{d}, \mathbf{c}, \mathbf{l}, \mathbf{k})}. \quad (5.28)$$

Diese Verteilung wird nun verwendet, um die unbekanntenen Personenparameter aus der Log-Likelihood-Funktion heraus zu integrieren. Ausgegangen wird hierbei zunächst einmal von bestimmten vorläufigen Itemparametern  $\mathbf{a}_0$ ,  $\mathbf{d}_0$ ,  $\mathbf{c}_0$ ,  $\mathbf{l}_0$  und  $\mathbf{k}_0$ . Dies können selbst gewählte Startwerte sein oder vorläufige Lösungen des vorhergehenden Iterationsschrittes. Bei Annahme der vorläufigen Itemparameter kann der bedingte Erwartungswert der Log-Likelihood-Funktion berechnet werden. Es gilt:

$$\text{ELL}(\mathbf{a}, \mathbf{d}, \mathbf{c}, \mathbf{l}, \mathbf{k}) = \sum_{i=1}^N \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \log \left[ P(\mathbf{x}_i, \mathbf{y}_i | \theta, \omega; \mathbf{a}, \mathbf{d}, \mathbf{c}, \mathbf{l}, \mathbf{k}) \right] \frac{P(\mathbf{x}_i, \mathbf{y}_i | \theta, \omega; \mathbf{a}_0, \mathbf{d}_0, \mathbf{c}_0, \mathbf{l}_0, \mathbf{k}_0) f(\theta, \omega)}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(\mathbf{x}_i, \mathbf{y}_i | \theta, \omega; \mathbf{a}_0, \mathbf{d}_0, \mathbf{c}_0, \mathbf{l}_0, \mathbf{k}_0) f(\theta, \omega) d\theta d\omega} d\theta d\omega. \quad (5.29)$$

Bei der Verteilung  $f(\theta, \omega)$  handelt es sich um die Verteilung der Persönlichkeitseigenschaften in der Population der potentiellen Testteilnehmer, die bivariate Standardnormalverteilung mit Korrelationskoeffizienten  $\rho = 0$ . Eine Lösung des Integrals kann nur über numerische Verfahren approximiert werden. Bei Verwendung einer kartesischen Quadraturregel (Stroud, 1971; Bock & Aitkin, 1981) mit den Quadraturpunkten und den Quadraturgewichten der Gauß-Hermite-Quadratur kann dieses Integral näherungsweise bestimmt werden über die Summe:

$$\text{ELL}(\mathbf{a}, \mathbf{d}, \mathbf{c}, \mathbf{l}, \mathbf{k}) = \sum_{i=1}^N \sum_{q_1=1}^{Q_1} \sum_{q_2=1}^{Q_2} \log \left[ P(\mathbf{x}_i, \mathbf{y}_i | \theta_{q_1}, \omega_{q_2}; \mathbf{a}, \mathbf{d}, \mathbf{c}, \mathbf{l}, \mathbf{k}) \right] \frac{P(\mathbf{x}_i, \mathbf{y}_i | \theta_{q_1}, \omega_{q_2}; \mathbf{a}_0, \mathbf{d}_0, \mathbf{c}_0, \mathbf{l}_0, \mathbf{k}_0)}{\sum_{q_3=1}^{Q_3} \sum_{q_4=1}^{Q_4} P(\mathbf{x}_i, \mathbf{y}_i | \theta_{q_3}, \omega_{q_4}; \mathbf{a}_0, \mathbf{d}_0, \mathbf{c}_0, \mathbf{l}_0, \mathbf{k}_0) w_{q_3} w_{q_4}} w_{q_1} w_{q_2}. \quad (5.30)$$

Bei den Größen  $w_{q_1} \dots w_{q_4}$  handelt es sich um die Gewichte der Gauss-Hermite-Quadratur, bei den Größen  $\theta_{q_1} \dots \omega_{q_4}$  um die Quadraturpunkte, die in Standardlehrbüchern zur numerischen Berechnung von Integralen tabelliert sind – siehe zum Beispiel Davis und Rabinowitz (1975). Wiederum wurden pro Dimension sieben Quadraturpunkte gewählt. Gleichung (5.30) lässt sich so deuten, als entfielen auf jedes Antwortmuster  $[\mathbf{x}_i, \mathbf{y}_i]$  Personen mit Parameterausprägung  $\theta_{q_1}$  und  $\omega_{q_2}$  in der „relativen Häufigkeit“:

$$w_{\mathbf{x}_i, \mathbf{y}_i, \theta_{q_1}, \omega_{q_2}} = \frac{P(\mathbf{x}_i, \mathbf{y}_i | \theta_{q_1}, \omega_{q_2}, \mathbf{a}_0, \mathbf{d}_0, \mathbf{c}_0, \mathbf{l}_0, \mathbf{k}_0)}{\sum_{q_3=1}^{Q_3} \sum_{q_4=1}^{Q_4} P(\mathbf{x}, \mathbf{y} | \theta_{q_3}, \omega_{q_4}, \mathbf{a}_0, \mathbf{d}_0, \mathbf{c}_0, \mathbf{l}_0, \mathbf{k}_0) w_{q_3} w_{q_4}} w_{q_1} w_{q_2}. \quad (5.31)$$

Beim M-Schritt kann folglich angenommen werden, dass ein bestimmtes Antwortmuster  $[\mathbf{x}_i, \mathbf{y}_i]$  bei Personen mit Parameterausprägung  $\theta_{q_1}$  und  $\omega_{q_1}$  mit der relativen Häufigkeit  $w_{\mathbf{x}_i, \mathbf{y}_i, \theta_{q_1}, \omega_{q_2}}$  auftrat.

**M-Schritt**

Die erwartete Log-Likelihood-Funktion ist eine Funktion der unbekanntem Itemparameter. Beim M-Schritt wird die erwartete Log-Likelihood-Funktion über die unbekanntem Itemparameter maximiert. Als vorläufige Schätzer der gesuchten Itemparameter wird die Lösung von

$$\begin{aligned} \arg \max_{\mathbf{a}, \mathbf{d}, \mathbf{c}, \mathbf{l}, \mathbf{k}} \left[ \text{ELL}(\mathbf{a}, \mathbf{d}, \mathbf{c}, \mathbf{l}, \mathbf{k}) \right] = \\ \arg \max_{\mathbf{a}, \mathbf{d}, \mathbf{c}, \mathbf{l}, \mathbf{k}} \left[ \sum_{i=1}^N \sum_{q_1=1}^{Q_1} \sum_{q_2=1}^{Q_2} \log \left[ P(\mathbf{x}_i, \mathbf{y}_i | \theta_{q_1}, \omega_{q_2}; \mathbf{a}, \mathbf{d}, \mathbf{c}, \mathbf{l}, \mathbf{k}) \right] w_{\mathbf{x}_i, \mathbf{y}_i, \theta_{q_1}, \omega_{q_2}} \right] \end{aligned} \quad (5.32)$$

gesucht. Wiederum vereinfacht sich die Maximierung dadurch, dass die erwartete Log-Likelihood-Funktion separat für die Itemparameter jedes Items maximiert werden kann. Bei der Maximierung der erwarteten Log-Likelihood-Funktion (5.32) werden die Quadraturpunkte als wahre Personenparameter angenommen, jedoch mit einem bestimmten Gewicht (5.31) versehen. Die Maximierung der erwarteten Likelihood-Funktion (5.30) kann daher analog zum Falle bekannter Personenparameter erfolgen. Allerdings ist die Maximierung der erwarteten Log-Likelihood-Funktion nicht unproblematisch. Tatsächlich besitzt diese Funktion mehrere lokale Extrempunkte. Dies erschwert die Maximierung mit Hilfe des Newton-Raphson-Ansatzes: Bei weniger guten Startwerten konvergiert das Verfahren zu einem lokalen Minimum, wodurch der gesamte EM-Algorithmus zusammenbricht. Für den ersten Maximierungsschritt wurde daher der sehr rechenintensive, jedoch wesentlich robustere Simplex-Algorithmus (Nelder & Mead, 1965; Walters, Parker, Morgan & Deming, 1991) herangezogen, der im Programmpaket OPTIM von R implementiert ist. Nach dem ersten EM-Zyklus kann die Maximierung von den verbesserten Itemparameterschätzern ausgehen, welche im ersten Schritt in Richtung des Maximums der marginalen Likelihood-Funktion gerückt sind. Da sich dadurch der Ausgangspunkt der Maximierung in größerer Nähe zum Maximum der erwarteten Log-Likelihood-Funktion befindet, ist die Tendenz des Newton-Raphson-Algorithmus geringer, zu einem alternativen Extrempunkt zu konvergieren. Dies kann ausgenutzt werden, um nach dem ersten EM-Schritt zum Newton-Raphson-Algorithmus zu wechseln, der wesentlich weniger rechenintensiv ist als der Simplex-Algorithmus. Durch diesen Wechsel kann das Verfahren beschleunigt werden. Nur der Newton-Raphson-Ansatz soll im Folgenden dargestellt werden.

Notwendige Bedingung für ein Maximum der erwarteten Log-Likelihood-Funktion ist die Auflage, dass die partiellen Ableitungen der Funktion nach den Itemparametern an der Stelle des potentiellen Maximums Null ergeben. Da die erwartete Log-Likelihood-Funktion separat für die einzelnen Items maximiert werden kann, lässt sich jedes Item gesondert betrachten. Der Gradientenvektor  $\mathbf{G}$  des  $g$ -ten Items, der die partiellen Ableitungen der erwarteten Log-Likelihood-Funktion nach den Itemparametern dieses Items enthält, ist ein Spaltenvektor mit den Elementen:

$$\begin{aligned}
\frac{\partial}{\partial a_g} \text{ELL}(\mathbf{a}, \mathbf{d}, \mathbf{c}, \mathbf{l}, \mathbf{k}) &= \sum_{i=1}^N \sum_{q_1=1}^{Q_1} \sum_{q_2=1}^{Q_2} \left[ x_{ig} - P_{x_g}(\theta_{q_1}; a_g, b_g) \right] (\theta_{q_1} - d_g) w_{\mathbf{x}_i, \mathbf{y}_i, \theta_{q_1}, \omega_{q_2}} \\
\frac{\partial}{\partial k_g} \text{ELL}(\mathbf{a}, \mathbf{d}, \mathbf{c}, \mathbf{l}, \mathbf{k}) &= \sum_{i=1}^N \sum_{q_1=1}^{Q_1} \sum_{q_2=1}^{Q_2} \left[ \frac{y_{ig} - P_{y_g}(\theta_{q_1}, \omega_{q_2}; c_g, d_g, l_g, k_g)}{P_{y_g}(\theta_{q_1}, \omega_{q_2}; c_g, d_g, l_g, k_g)} \right] \\
&\quad \times \exp [k_g + l_g(\theta_{q_1} - d_g)^2 + c_g \omega_{q_2}] w_{\mathbf{x}_i, \mathbf{y}_i, \theta_{q_1}, \omega_{q_2}} \\
\frac{\partial}{\partial l_g} \text{ELL}(\mathbf{a}, \mathbf{d}, \mathbf{c}, \mathbf{l}, \mathbf{k}) &= \sum_{i=1}^N \sum_{q_1=1}^{Q_1} \sum_{q_2=1}^{Q_2} \left[ \frac{y_{ig} - P_{y_g}(\theta_{q_1}, \omega_{q_2}; c_g, d_g, l_g, k_g)}{P_{y_g}(\theta_{q_1}, \omega_{q_2}; c_g, d_g, l_g, k_g)} \right] \\
&\quad \times \exp [k_g + l_g(\theta_{q_1} - d_g)^2 + c_g \omega_{q_2}] (\theta_{q_1} - d_g)^2 w_{\mathbf{x}_i, \mathbf{y}_i, \theta_{q_1}, \omega_{q_2}} \\
\frac{\partial}{\partial d_g} \text{ELL}(\mathbf{a}, \mathbf{d}, \mathbf{c}, \mathbf{l}, \mathbf{k}) &= \sum_{i=1}^N \sum_{q_1=1}^{Q_1} \sum_{q_2=1}^{Q_2} \left[ \left[ \frac{y_{ig} - P_{y_g}(\theta_{q_1}, \omega_{q_2}; c_g, d_g, l_g, k_g)}{P_{y_g}(\theta_{q_1}, \omega_{q_2}; c_g, d_g, l_g, k_g)} \right] \right. \\
&\quad \times \exp [k_g + l_g(\theta_{q_1} - d_g)^2 + c_g \omega_{q_2}] (-2l_g(\theta_{q_1} - d_g)) \\
&\quad \left. + \left[ x_{ig} - P_{x_g}(\theta_{q_1}; a_g, b_g) \right] (-a_g) \right] w_{\mathbf{x}_i, \mathbf{y}_i, \theta_{q_1}, \omega_{q_2}} \\
\frac{\partial}{\partial c_g} \text{ELL}(\mathbf{a}, \mathbf{d}, \mathbf{c}, \mathbf{l}, \mathbf{k}) &= \sum_{i=1}^N \sum_{q_1=1}^{Q_1} \sum_{q_2=1}^{Q_2} \left[ \frac{y_{ig} - P_{y_g}(\theta_{q_1}, \omega_{q_2}; c_g, d_g, l_g, k_g)}{P_{y_g}(\theta_{q_1}, \omega_{q_2}; c_g, d_g, l_g, k_g)} \right] \\
&\quad \times \exp [k_g + l_g(\theta_{q_1} - d_g)^2 + c_g \omega_{q_2}] \omega_{q_2} w_{\mathbf{x}_i, \mathbf{y}_i, \theta_{q_1}, \omega_{q_2}}
\end{aligned}$$

Die Nullstellen des Gradientenvektors können mit dem Newton-Raphson-Verfahren bestimmt werden. Allerdings wird hierfür noch die Jacobi-Matrix  $\mathbf{J}$  des Gradientenvektors benötigt, die man durch erneutes Ableiten des Gradientenvektors nach den Itemparametern erhält. Bei der Jacobi-Matrix  $\mathbf{J}$  handelt es sich um eine symmetrische  $5 \times 5$ -Matrix, deren Elemente hier nicht einzeln aufgelistet werden sollen. Wurde im  $j$ -ten Iterationsschritt eine vorläufige „Lösung“  $[\hat{a}_{gj}, \hat{k}_{gj}, \hat{l}_{gj}, \hat{d}_{gj}, \hat{c}_{gj}]'$  für die Itemparameter der  $g$ -ten Aufgabe gefunden, kann im  $j + 1$ -ten Schritt eine Verbesserung bestimmt werden über die Beziehung:

$$[\hat{a}_{gj+1}, \hat{k}_{gj+1}, \hat{l}_{gj+1}, \hat{d}_{gj+1}, \hat{c}_{gj+1}]' = [\hat{a}_{gj}, \hat{k}_{gj}, \hat{l}_{gj}, \hat{d}_{gj}, \hat{c}_{gj}]' - \mathbf{J}_j^{-1} \mathbf{G}_j. \quad (5.33)$$

Bei der Größe  $\mathbf{G}_j$  handelt es sich um den an der Stelle der vorläufigen Itemparameter  $[\hat{a}_{gj}, \hat{k}_{gj}, \hat{l}_{gj}, \hat{d}_{gj}, \hat{c}_{gj}]$  berechneten Gradientenvektor, bei  $\mathbf{J}_j$  um die an derselben Stelle bestimmte Jacobi-Matrix. Die schrittweise Verbesserung vorläufiger Itemparameterwerte nach Beziehung (5.33) wird solange wiederholt, bis die Lösung die gesuchte Nullstelle mit ausreichender Genauigkeit approximiert. Bei den resultierenden Itemparameterwerten handelt es sich dann um die Itemparameterwerte, welche die erwartete Log-Likelihood-Funktion des aktuellen M-Schrittes maximieren. Der M-Schritt ist damit beendet. E-Schritt und M-Schritt wechseln sich so lange ab, bis die marginale Log-Likelihood-Funktion nur noch geringfügig ansteigt.

### Simulation der Itemparameterschätzung

Zur Überprüfung des vorgeschlagenen EM-Algorithmus wurde eine Simulation durchgeführt. Für die Simulation diente das Statistikpaket R, mit welchem der EM-Algorithmus umgesetzt und die Beobachtungsdaten erzeugt wurden. Simuliert wurde die Itemparameterschätzung für einen Test aus fünf Items. Als Zufallsstichprobe von Personen wurden Werte für die Personenfähigkeit  $\theta$  und den Residualfaktor  $\omega$  aus einer bivariaten Standardnormalverteilung mit Korrelationskoeffizient  $\rho = 0$  gezogen. Auf diesen Werten basierte die anschließende Generation von Itemantworten und Reaktionszeiten für fünf Items. Als Reaktionszeitverteilung fand wiederum eine Exponentialverteilung Verwendung, deren zeitinvariante Hazard-Rate gemäß Gleichung (5.21) modifiziert wurde. Die generierten Reaktionszeiten wurden anschließend an dem Zeitlimit  $t_c = 1$  dichotomisiert. Die Generation von Itemantworten erfolgte nach dem zweiparametrischen Logit-Modell (5.23). Die für die Simulation verwendeten Itemparameter sind in Tabelle 5.5 wiedergegeben.

Tabelle 5.5: Dichotomisierte RT mit quadratischer Parametrisierung: Wahre Itemparameter der Aufgaben

Item	$a_g$	$d_g$	$c_g$	$l_g$	$k_g$
1	1.50	-0.66	0.20	0.10	0.00
2	1.50	-0.33	0.20	0.10	0.00
3	1.50	0.00	0.20	0.10	0.00
4	1.50	0.33	0.20	0.10	0.00
5	1.50	0.66	0.20	0.10	0.00

Bei der Auswahl der Itemparameter stand das Bedürfnis nach realistischen Parameterausprägungen im Vordergrund. Der Einfluss der Personenfähigkeit  $\theta$  auf die Reaktionszeitverteilung wurde bewusst schwach angesetzt, um eine Schätzbarkeit des Modells unter weniger günstigen Bedingungen zu erproben. Die Befunde in der Literatur sprechen zudem auch nur für schwache bis mittlere Beziehungen zwischen der Reaktionszeit und der Distanz von Item und Person – siehe hierzu die Übersicht im ersten Kapitel.

Es wurden 35 Datensätze im Umfang von 10000 und 1000 Personen erzeugt. Eine Simulation mit kleineren Stichprobenumfängen erschien nicht sinnvoll, da die Schätzung bereits bei 1000 Personen instabil wurde. Zur approximativen Berechnung des bedingten Erwartungswertes der Log-Likelihood-Funktion diente ein Quadraturverfahren mit sieben Quadraturpunkten pro Persönlichkeitsdimension. Diese Anzahl erwies sich als ausreichend für die Problemsituation: Bei Analyse eines Datensatzes aus 1000000 Versuchspersonen konnten die wahren Itemparameter mit großer Genauigkeit reproduziert werden. Auch sprechen die Ergebnisse in Tabelle 5.6 dafür, dass durch die Approximation keine Verzerrungen entstehen. Bei der Schätzung wurde von Startparametern ausgegangen, die man auch ohne Kenntnis der wahren Parameter hätte auswählen können:  $a_g = 1.0$ ,  $d_g = 0.0$ ,  $c_g = 0.1$ ,  $l_g = 0.0$ ,  $k_g = 0.0$ . Die Schätzung wurde abgebrochen, wenn keine bedeutende Verbesserung der marginalen Log-Likelihood

mehr erzielt werden konnte (Bei 10000 Versuchspersonen: Zuwachs kleiner als 0.1, bei 1000 Versuchspersonen: Zuwachs kleiner als 0.01<sup>4</sup>).

Das Konvergenzverhalten des Verfahrens war robust. Alle Schätzversuche konvergierten. Bei 10000 Versuchspersonen waren durchschnittlich 19.4 Iterationen erforderlich (min: 15, max: 28). Bei 1000 Versuchspersonen stieg die durchschnittliche Iterationsanzahl auf 35.7 (min: 16, max: 84) an. Dies entsprach einer Rechenzeit von etwa 6 Stunden pro Simulation. Die Ergebnisse der Schätzversuche sind in Tabelle 5.6 und Tabelle 5.7 wiedergegeben.

Wie in Tabelle 5.6 ersichtlich ist, können die Itemparameter bei 10000 Versuchspersonen relativ genau bestimmt werden. Zudem gibt es keine Anzeichen eines nennenswerten Bias. Im Gegensatz zu manchen Unfolding-Modellen aus dem IRT-Bereich ist die Schätzung auch bei weniger guten Startwerten möglich – das Vorzeichen der Itemschwierigkeit muss beispielsweise nicht vorgegeben werden. Die Schätzergebnisse sind jedoch bei kleinerem Stichprobenumfang, besonders was den Parameter  $c_g$  betrifft, unbefriedigend. Teilweise wird sogar das Vorzeichen von  $c_g$  falsch geschätzt. Aber auch der geschätzte Parameter  $\hat{a}_g$  weicht manchmal erheblich von der wahren Parameterausprägung ab. Dass  $c_g$  schlechter als  $l_g$  geschätzt wird, mag daran liegen, dass die Schätzung von  $l_g$  davon profitiert, dass die bedingte Verteilung von  $\theta$  stark von den Itemantworten beeinflusst wird, zur Bestimmung der bedingten Verteilung von  $\omega$  jedoch nur die Reaktionszeiten zur Verfügung stehen. Die bedingte Verteilung von  $\omega$  spielt beim EM-Algorithmus eine wichtige Rolle zur Schätzung von  $c_g$ , während die bedingte Verteilung von  $\theta$  wichtig für die Schätzung von  $l_g$  ist. Dies sind jedoch nur Spekulationen, die einer weiteren Überprüfung bedürfen.

### 5.3 Diskussion

Ziel des aktuellen Kapitels war es, die im ersten Abschnitt der Arbeit aufgestellten Proportional Hazard Modelle der empirischen Forschung zugänglich zu machen. Grundvoraussetzung hierfür ist die Entwicklung von Schätzmethode für die Itemparameter. Wie aus den Simulationsstudien ersichtlich ist, lassen sich die psychologisch plausiblen, zweidimensionalen Modelle für die Verteilung der Itemantworten und Reaktionszeiten bei einem Test durchaus schätzen. Jedoch sind die Anforderungen an den Stichprobenumfang und die notwendige Rechenleistung beträchtlich.

Trotz des geringen Umfangs des Tests liegt die Rechenzeit (bei einer Programmierung mit der Statistik-Software R) bei ungefähr sechs Stunden, obwohl nur sieben Quadraturpunkte pro Dimension gewählt wurden. Bei einer Vergrößerung der Anzahl der Quadraturpunkte steigt die Anzahl der Rechenschritte quadratisch an, so dass eine Erhöhung nur begrenzt möglich ist. Aber auch bei einer Verlängerung des Tests wächst der Rechenaufwand überproportional: Da der EM-Algorithmus auf den verschiedenen Antwortmustern basiert, müssten

---

<sup>4</sup>Die Log-Likelihood-Funktionen nahmen an ihrem Maximum bei 10.000 Versuchspersonen einen Wert um -60.000, bei 1000 Versuchspersonen einen Wert um -6.000 an. Die gewählten Abbruchkriterien entsprachen somit in etwa derselben relativen Veränderung. Zudem zeigte sich, dass es während der letzten Schritte des EM-Algorithmus kaum noch zu Veränderungen der Schätzwerte kam.



Tabelle 5.6: Itemparameterschätzung beim Proportional Hazard Modell für dichotomisierte Reaktionszeiten mit quadratischer Parametrisierung – Ergebnisse der 35 Simulationen mit N=10000

Parameter a					
Item	1	2	3	4	5
Wahrer Wert	1.500	1.500	1.500	1.500	1.500
Min	1.398	1.377	1.400	1.413	1.391
Max	1.620	1.586	1.617	1.612	1.593
Mean	1.486	1.484	1.515	1.494	1.502
STD	0.053	0.053	0.052	0.043	0.047
RMSE	0.056	0.055	0.054	0.043	0.047
Parameter d					
Item	1	2	3	4	5
Wahrer Wert	-0.660	-0.330	0.000	0.330	0.660
Min	-0.715	-0.371	-0.049	0.283	0.613
Max	-0.580	-0.286	0.041	0.375	0.700
Mean	-0.662	-0.330	0.001	0.332	0.656
STD	0.027	0.021	0.021	0.022	0.021
RMSE	0.027	0.021	0.021	0.022	0.021
Parameter c					
Item	1	2	3	4	5
Wahrer Wert	0.200	0.200	0.200	0.200	0.200
Min	0.110	0.102	0.114	0.107	0.126
Max	0.305	0.258	0.279	0.264	0.259
Mean	0.196	0.188	0.191	0.191	0.194
STD	0.041	0.037	0.033	0.034	0.034
RMSE	0.042	0.039	0.035	0.035	0.035
Parameter l					
Item	1	2	3	4	5
Wahrer Wert	0.100	0.100	0.100	0.100	0.100
Min	0.070	0.058	0.060	0.076	0.081
Max	0.116	0.139	0.136	0.128	0.133
Mean	0.096	0.098	0.099	0.103	0.102
STD	0.010	0.018	0.020	0.013	0.013
RMSE	0.011	0.018	0.020	0.014	0.013
Parameter k					
Item	1	2	3	4	5
Wahrer Wert	0.000	0.000	0.000	0.000	0.000
Min	-0.048	-0.041	-0.042	-0.031	-0.057
Max	0.071	0.062	0.033	0.042	0.046
Mean	0.002	0.002	0.001	0.001	-0.003
STD	0.023	0.024	0.020	0.017	0.024
RMSE	0.024	0.024	0.020	0.017	0.024

Tabelle 5.7: Itemparameterschätzung beim Proportional Hazard Modell für dichotomisierte Reaktionszeiten mit quadratischer Parametrisierung – Ergebnisse der 35 Simulationen mit N=1000

Parameter a					
Item	1	2	3	4	5
Wahrer Wert	1.500	1.500	1.500	1.500	1.500
Min	1.218	0.949	1.244	1.185	1.146
Max	1.948	2.099	1.884	2.107	1.786
Mean	1.549	1.475	1.528	1.517	1.500
STD	0.183	0.210	0.155	0.194	0.158
RMSE	0.190	0.212	0.157	0.194	0.158
Parameter d					
Item	1	2	3	4	5
Wahrer Wert	-0.660	-0.330	0.000	0.330	0.660
Min	-0.896	-0.443	-0.115	0.227	0.486
Max	-0.535	-0.167	0.095	0.484	0.855
Mean	-0.665	-0.325	-0.008	0.340	0.662
STD	0.087	0.071	0.051	0.058	0.079
RMSE	0.087	0.071	0.051	0.059	0.079
Parameter c					
Item	1	2	3	4	5
Wahrer Wert	0.200	0.200	0.200	0.200	0.200
Min	-0.436	-0.181	-0.064	-0.057	-0.108
Max	0.695	0.435	0.587	0.705	0.818
Mean	0.206	0.163	0.241	0.228	0.250
STD	0.175	0.145	0.151	0.152	0.170
RMSE	0.175	0.150	0.156	0.155	0.177
Parameter l					
Item	1	2	3	4	5
Wahrer Wert	0.100	0.100	0.100	0.100	0.100
Min	0.030	0.019	-0.011	-0.016	0.023
Max	0.280	0.181	0.252	0.252	0.183
Mean	0.112	0.106	0.117	0.130	0.111
STD	0.049	0.042	0.060	0.067	0.034
RMSE	0.051	0.043	0.062	0.073	0.036
Parameter k					
Item	1	2	3	4	5
Wahrer Wert	0.000	0.000	0.000	0.000	0.000
Min	-0.144	-0.151	-0.166	-0.178	-0.133
Max	0.110	0.095	0.165	0.143	0.128
Mean	-0.006	-0.010	-0.011	-0.033	-0.027
STD	0.063	0.060	0.075	0.075	0.061
RMSE	0.064	0.061	0.076	0.082	0.067

bei sechs Aufgaben bereits 4096 Antwortmuster anstelle der 1024 Antwortmuster bei fünf Aufgaben berücksichtigt werden (vorausgesetzt, dass alle Antwortmuster auftreten sollten). Die relativ lange Rechenzeit bei der direkten Implementierung in R könnte allerdings ohne Probleme durch eine Programmierung in einer effizienteren Programmiersprache verkürzt werden.

Für die Notwendigkeit großer Stichproben kann keine so einfache Abhilfe gefunden werden. Zwar lässt sich die Schätzung trotz der geringen Testlänge bei ausreichend großem Stichprobenumfang als unproblematisch bezeichnen: Sowohl das Modell mit linearer Parametrisierung als auch das Modell mit quadratischer Parametrisierung konvergiert bei jedem Simulationsversuch gegen das Maximum, selbst wenn von neutralen Startwerten ausgegangen wird. Diese Robustheit ist zumindest für das Modell mit quadratischer Beziehung überraschend. Die Stabilität der Schätzung beruht vermutlich darauf, dass es sich bei dem Modell ja nur teilweise um ein Unfolding-Modell handelt, da sich die Wahrscheinlichkeit einer positiven Antwort schließlich weiterhin aus einer Dominanzbeziehung zwischen Person und Aufgabe ergibt. Trotz der Robustheit des Verfahrens kann die Präzision der Schätzung erst ab einem Stichprobenumfang von 10000 Versuchspersonen als gut bezeichnet werden, zumindest bei einem Test aus fünf Aufgaben und dem Modell mit quadratischer Parametrisierung. Dass große Stichproben für die Parameterschätzung notwendig sind, ist allerdings kein Defekt der gemeinsamen Modellierung von Itemantworten und Reaktionszeiten alleine, sondern eine allgemeine Eigenschaft der IRT-Modelle, deren Itemparameter ebenfalls erst bei großen Stichprobenumfängen mit passabler Genauigkeit reproduziert werden können. Untersucht wurde jedoch nur ein relativ kurzer Test aus fünf Items. Durch eine Testverlängerung kann unter Umständen die Itemparameterschätzung verbessert werden, jedoch auf Kosten eines steigenden Rechenaufwandes.

## Kapitel 6

# Itemparameterschätzung bei Accelerated Failure Time Modellen

Die Log-Normalverteilung ist eine populäre Verteilung für die Bearbeitungszeiten bei psychologischen Tests (Thissen, 1984; Ferrando & Lorenzo-Sevas, 2007a; van der Linden, 2007). Aufgrund der Popularität und Bewährtheit dieser Verteilungsform für die Modellierung von Reaktionszeiten sollen auch für die in Kapitel 4.2.1 vorgeschlagenen Modelle für log-normalverteilte Reaktionszeiten Schätzverfahren entwickelt werden, um die Anpassung der Modelle an empirische Daten zu ermöglichen. Da es sich bei diesen Modellen um Accelerated Failure Time Modelle handelt, bilden sie eine interessante Alternative zu den bereits behandelten Proportional Hazard Modellen. Wie bereits im vorherigen Kapitel wird diesmal die gemeinsame Schätzung von IRT-Modell und Reaktionszeitmodell beschrieben.

### 6.1 Lineare Parametrisierung

In Kapitel 4.2.1 wurde für die Bearbeitungszeiten von Testitems ein Modell vorgeschlagen, das auf der Log-Normalverteilung basierte. Annahme war dabei, dass der Lokationsparameter der Log-Normalverteilung über eine lineare Beziehung auf zwei Persönlichkeitseigenschaften, die Personenfähigkeit  $\theta$  und den Residualfaktor  $\omega$ , zurückgeführt werden kann, siehe Gleichung (4.32):

$$\eta_g(\theta, \omega) = k_g + l_g\theta + c_g\omega. \quad (6.1)$$

Der Skalenparameter  $\sigma_g$  der Log-Normalverteilung gilt als itemspezifisch und variiert nicht über verschiedene Personen. Die Dichte der Reaktionszeitvertei-

lung bei Item  $g$  lautet für eine Person mit Personenfähigkeit  $\theta$  und Residualfaktor  $\omega$  in diesem Fall:<sup>1</sup>

$$f(t_g|\theta, \omega; c_g, l_g, k_g, \sigma_g) = \frac{1}{\sqrt{2\pi}\sigma_g t_g} \exp \left[ -\frac{1}{2} \frac{[\log(t_g) - \eta_g(\theta, \omega)]^2}{\sigma_g^2} \right] \quad \text{mit } t_g > 0. \quad (6.2)$$

Der Lokationsparameter  $\eta_g(\theta, \omega)$  kann dabei über Beziehung (6.1) auf die Personenfähigkeit  $\theta$  und den Residualfaktor  $\omega$  einer Person zurückgeführt werden. Als Modell für die Wahrscheinlichkeit einer positiven Antwort wurde bisher immer ein zweiparametriges Logit-Modell verwandt. Von dieser Praxis soll jedoch im aktuellen Kapitel abgewichen werden. Das zweiparametriges Logit-Modell wird im Folgenden vielmehr durch das zweiparametriges Probit-Modell ersetzt. Gerechtfertigt werden kann dieser Modellwechsel damit, dass sich Probit-Modell und Logit-Modell nur unwesentlich in den vorhergesagten Wahrscheinlichkeiten unterscheiden. Es lässt sich zeigen, dass bei Wahl eines geeigneten Skalierungsfaktors die Differenz zwischen den Itemcharakteristiken maximal 0.01 beträgt (Savalei, 2006). Trotz des geringen Unterschiedes zwischen beiden Modellen vereinfacht der Modellwechsel die Schätzung der Itemparameter jedoch erheblich.

Beim zweiparametriges Probit-Modell lässt sich die Wahrscheinlichkeit einer positiven Itemantwort berechnen als

$$P_{x_g}(\theta; a_g, b_g) = \Phi(a_g(\theta - b_g)), \quad (6.3)$$

wobei  $\Phi(x)$  die Verteilungsfunktion der Standardnormalverteilung ist (Baker & Seock-Ho, 2004). Bei Bedingung auf die Persönlichkeitseigenschaften  $\theta$  und  $\omega$  werden die Itemantworten und Reaktionszeiten einer Person als voneinander unabhängig angenommen. Für die gemeinsame Verteilung der Reaktionszeiten  $\mathbf{t} = [t_1, \dots, t_G]$  und Itemantworten  $\mathbf{x} = [x_1, \dots, x_G]$  bei  $G$  Aufgaben mit den Itemparametern  $\mathbf{a} = [a_1, \dots, a_G]$ ,  $\mathbf{b} = [b_1, \dots, b_G]$ ,  $\mathbf{c} = [c_1, \dots, c_G]$ ,  $\mathbf{l} = [l_1, \dots, l_G]$ ,  $\mathbf{k} = [k_1, \dots, k_G]$  und  $\boldsymbol{\sigma} = [\sigma_1, \dots, \sigma_G]$  gilt somit:

$$f(\mathbf{x}, \mathbf{t}|\theta, \omega; \mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{l}, \mathbf{k}, \boldsymbol{\sigma}) = \prod_{g=1}^G [P_{x_g}(\theta; a_g, b_g)]^{x_g} [1 - P_{x_g}(\theta; a_g, b_g)]^{1-x_g} \quad (6.4)$$

$$f(t_g|\theta, \omega; c_g, l_g, k_g, \sigma_g).$$

Ein ähnliches Modell wurde bereits von Thissen (1984) vorgeschlagen, dessen Modell wiederum parallelen zum Modell von Furneaux (1952) aufweist. Thissen schätzte die Parameter des Modells über einen Joint-Maximum-Likelihood-Ansatz. Eine genauere Betrachtung der Modellbestandteile (6.2), (6.1) und (6.3) legt jedoch ein anderes Schätzverfahren nahe.

<sup>1</sup>Die Darstellung der Verteilungen weicht in diesem Kapitel etwas von der Darstellung im dritten und vierten Kapitel ab. Da im aktuellen Kapitel die Parameterschätzung im Mittelpunkt steht, soll durch die explizite Nennung der Itemparameter deren Rolle herausgestellt werden.

### Schätzung der Itemparameter

Ist eine Reaktionszeit  $t_g$  log-normalverteilt, kann diese durch Logarithmierung in eine normalverteilte Zufallsvariable  $t_g^* = \log(t_g)$  transformiert werden. Werden alle Reaktionszeiten logarithmiert, sind diese nach Logarithmierung normalverteilt mit Erwartungswert  $\mu(t_g^*|\theta, \omega) = \eta_g(\theta, \omega)$  und Varianz  $\sigma_g^2$ . Bei Modell (6.1) handelt es sich dann, wie un schwer zu erkennen ist, um ein faktorenanalytisches Modell mit linearen Faktoren, für welches bewährte Schätzprozeduren existieren. Wie von Takane und de Leeuw (1987) gezeigt wurde, kann das zweiparametrische Probit-Modell (6.3) ebenfalls als faktorenanalytisches Modell mit linearen Faktoren aufgefasst werden, wenn man annimmt, dass der beobachteten, binären Itemantwort eine dichotomisierte kontinuierliche Zufallsvariable zugrunde liegt. Der Grundgedanke und die Herleitung der Itemcharakteristik ist dabei weitgehend identisch zur Begründung des zweiparametrischen Logit-Modells im dritten Kapitel, siehe dort die Randbemerkungen auf S. 26. Hierzu muss man für jedes Item eine kontinuierliche Zufallsvariable  $z_g$  annehmen, die bei einer Person mit Personenfähigkeit  $\theta$  normalverteilt ist mit der Verteilung  $z_g \sim N(a_g(\theta - b_g), 1)$ . Jede beobachtbare Itemantwort basiert auf einer Realisierung dieser unbeobachtbaren Zufallsvariablen: Eine Aufgabe wird positiv ( $x_g = 1$ ) beantwortet, wenn die Realisierung die Schwelle  $\tau = 0$  überschreitet ( $z_g > 0$ ) und negativ, wenn die Realisierung die Schwelle unterschreitet ( $z_g < 0$ ). Die Wahrscheinlichkeit, die Schwelle zu überschreiten, ergibt sich bei Normalverteilung der kontinuierlichen Variablen aus der Verteilungsfunktion der Normalverteilung und lautet  $P(\theta; a, b) = \Phi(a(\theta - b))$ . Nimmt man diesen Prozess für alle binären Itemantworten  $\mathbf{x} = [x_1, \dots, x_g]$  an, gehören zu den binären Itemantworten die latenten Zufallsvariablen  $\mathbf{z} = [z_1, \dots, z_G]$ . Wären die Ausprägungen auf den latenten Variablen  $\mathbf{z}$  bekannt, ließe sich die Korrelationsmatrix der Beobachtungen  $[\mathbf{z}, \mathbf{t}^*]$  berechnen. Diese könnte dann mit einer Faktorenanalyse ausgewertet werden. Jedoch können die unbekanntenen Korrelationen von  $\mathbf{z}$  mit  $\mathbf{z}$  und  $\mathbf{t}^*$  auch geschätzt werden, wenn nur die dichotomisierten Variablen  $\mathbf{x}$  verfügbar sind. Die Häufigkeiten der Itemantworten  $n(x_g = 1)$ ,  $n(x_h = 1)$  und  $n(x_g = 1, x_h = 1)$  bei zwei Items  $g$  und  $h$  reichen aus, um die tetrachorische Korrelation zwischen den latenten Variablen  $z_g$  und  $z_h$  zu berechnen (Drasgow, 1988). Auch die Korrelation zwischen  $t_g^*$  und  $z_h$  kann geschätzt werden, wenn man biseriale Korrelationskoeffizienten berechnet (Tate, 1955). Dadurch lassen sich die unbeobachtbaren Korrelationen in der Korrelationsmatrix von  $[\mathbf{z}, \mathbf{t}^*]$  ergänzen.<sup>2</sup>

Die Verwendung tetrachorischer Korrelationskoeffizienten und faktorenanalytischer Verfahren erleichtert die Parameterschätzung bei IRT-Modellen erheblich, umgeht dieser Ansatz doch die teilweise sehr rechenaufwendige Approximation der marginalen Log-Likelihood-Funktion. Dennoch wurde die Vorgehensweise von einzelnen Autoren kritisiert: Zum einen stellen die zur Berechnung der tetrachorischen Korrelationskoeffizienten verwendeten marginalen Häufigkeiten keine suffizienten Statistiken dar (Knol & Berger, 1991). Bei der Analy-

<sup>2</sup>Der aktuelle Schätzansatz interpretiert die binären Itemantworten als Resultat kontinuierlicher Itemantworten und versucht, die Kovarianzen zwischen den kontinuierlichen Variablen zu schätzen. Dadurch kann eine Auswertung über faktorenanalytische Methoden erfolgen. Prinzipiell kann auch der umgekehrte Weg eingeschlagen werden. Werden die Reaktionszeiten dichotomisiert, lassen sich die Itemantworten und die dichotomisierten Reaktionszeiten über Standardverfahren für Probit-Modelle auswerten. Allerdings ist in diesem Fall die spezifische Varianz der Reaktionszeiten nicht mehr schätzbar.

se einer tetrachorischen Korrelationsmatrix handelt es sich also nicht um eine „Full-Information Factor Analysis“ (Bock et al., 1988). Zum anderen ist die Matrix der tetrachorischen und biserialen Korrelationskoeffizienten möglicherweise nicht positiv-definit, so dass die meisten Faktorextraktionsverfahren ausscheiden. Beide Kritikpunkte lassen sich jedoch zumindest teilweise entschärfen. Von Devlin, Gnanadesikan und Kettenring (1975) wurde eine „Glättungsprozedur“ vorgeschlagen, mit der eine ursprünglich nicht positiv-definite Matrix durch eine positiv-definite Matrix approximiert werden kann. Auch setzt eine Faktorenanalyse nach der ULS-Methode keine positiv-definite Korrelationsmatrix voraus. Dem Kritikpunkt der suboptimalen Effizienz bei einer Analyse tetrachorischer Korrelationskoeffizienten kann entgegengehalten werden, dass Simulationsstudien für eine Eignung des Ansatzes sprechen (Knol & Berger, 1991; Parry & McArdle, 1991). Als besonders geeignet erwies sich in den Simulationsstudien von Knol und Berger (1991) die Kombination einer geglätteten tetrachorischen Korrelationsmatrix mit einer Analyse durch bestimmte Extraktionsverfahren (ULS, MINRES). Im Statistikpaket TESTFACT ist beispielsweise die Analyse tetrachorischer Korrelationen mit einer Principal Factor Analyse nach der MINRES-Procedure implementiert (Schilling & Bock, 2005). Aufgrund der Bewährung der Kombination einer „geglätteten“ tetrachorischen Korrelationsmatrix mit einer ULS-Parameterschätzung in den bekannten Simulationsstudien soll die Eignung dieser Vorgehensweise für die Parameterschätzung des aktuellen Modells erprobt werden. Insgesamt erscheint der Ansatz viel versprechender als der ursprüngliche Ansatz von Thissen (1984), da er einfacher zu implementieren und weniger rechenintensiv ist.

### Simulation der Itemparameterschätzung

Zur Überprüfung der Eignung des Schätzverfahrens wurde eine Simulation durchgeführt. Generiert wurden dabei für unterschiedlich große Stichproben die Antwortmuster eines Tests aus fünf Items. Die für die Simulation verwendeten Itemparameter finden sich in Tabelle 6.1. Wiederum stand bei der Parameterauswahl das Streben nach möglichst realitätsnahen Bedingungen im Vordergrund. Die Kommunalität der Reaktionszeiten ist relativ niedrig angesiedelt und insgesamt sind die Ladungen der Personenfähigkeit auf den Reaktionszeiten eher gering.

Tabelle 6.1: Log-normalverteilte RT mit linearer Parametrisierung: Wahre Itemparameter der Aufgaben

Item	$a_g$	$b_g$	$c_g$	$l_g$	$k_g$	$\sigma_g$
1	1.5	-1.0	0.2	0.5	1.5	1.0
2	1.5	-0.5	0.2	0.5	1.5	1.0
3	1.5	0.0	0.3	0.3	1.5	1.0
4	1.5	0.5	0.5	0.2	1.5	1.0
5	1.5	1.0	0.5	0.2	1.5	1.0

Simuliert wurden jeweils 50 Stichproben mit den Umfängen 100, 1000 und 10000 Versuchspersonen. Die Personenfähigkeiten und Residualfaktorenwerte

der einzelnen Versuchspersonen wurden durch Zufallsziehung aus einer bivariaten Standardnormalverteilung mit Korrelationskoeffizienten  $\rho = 0$  generiert. Anschließend wurden unter Verwendung der Itemparameter aus Tabelle 6.1 pro Person fünf Itemantworten und fünf Reaktionszeiten erzeugt. Hierzu diente wiederum das Programmpaket R. Nach Erzeugung eines Datensatzes wurden die tetrachorischen und biserialen Korrelationskoeffizienten mit dem Programm R über die Prozedur HETCOR berechnet. Die so erhaltene Korrelationsmatrix wurde in das Programm AMOS eingespeist und mit einer ULS-Faktorenanalyse analysiert.

Die Ergebnisse der Itemparameterschätzung sind in den Tabellen 6.2 bis 6.4 dargestellt. Bei 10000 Versuchspersonen waren alle Korrelationsmatrizen positiv definit. Bei 1000 Versuchspersonen war eine Korrelationsmatrix nicht mehr positiv definit und wurde vor einer weiteren Auswertung durch eine positiv definite Matrix approximiert. Bei einem Stichprobenumfang von 100 Personen stieg die Zahl der nicht positiv definiten Korrelationsmatrizen auf 45 an. Auch ließen sich die tetrachorische Korrelationsmatrizen bei vier Datensätzen aus 100 Versuchspersonen nicht berechnen. Das Konvergenzverhalten des ULS-Algorithmus war bei Stichproben aus 1000 und 10000 Versuchspersonen gut – bei diesem Stichprobenumfang konvergierten alle Schätzversuche. Bei den Simulationen mit Stichproben aus 100 Versuchspersonen konnte jedoch für 15 Datensätze keine Konvergenz erzielt werden, so dass die Ergebnisse in Tabelle 6.4 nur auf den konvergierenden 35 Simulationen beruhen.

Die Simulationsergebnisse in Tabelle 6.2 bis 6.4 ergänzen die Befunde zur Eignung des Schätzverfahrens. Eine präzise Itemparameterschätzung ist erst bei den beiden größeren Stichprobenumfängen möglich. Allerdings sind die Schätzwerte nur bei einer Größe von 10000 Versuchspersonen als gut zu bezeichnen. Bereits bei Stichproben aus 1000 Personen treten für die Trennschärfe  $a_g$  vereinzelt extreme Schätzwerte auf. Bei Umfängen von 100 Personen wird teilweise die Diskrepanz zwischen Schätzer und wahren Wert noch extremer. Unter Umständen ließe sich die Schätzung der Itemparameter des zweiparametrischen Probit-Modells mit einer „Full-Information-Factor“-Analyse, d.h. einer Schätzung nach der Marginal-Maximum-Likelihood-Methode verbessern. Auffällig ist, dass die Einflussgewichte  $l_g$  und  $c_g$  des Reaktionszeitmodells präziser geschätzt werden als die Trennschärfen des IRT-Modells. Dies mag daran liegen, dass die kontinuierlichen Reaktionszeiten größere Information über die Itemparameter besitzen als die binären Itemantworten.

## 6.2 Quadratische Parametrisierung

Auch mit Accelerated Failure Time Modellen kann eine „Unfolding-Beziehung“ zwischen Fähigkeitsausprägung und Reaktionszeit modelliert werden. Wenn sich auch für eine derartige Beziehung empirische Belege finden lassen (Ferrando, 2006; Akrami et al., 2007), kommen Unfolding-Modelle für Reaktionszeiten in der Literatur selten vor, einzige Ausnahme bilden die Vorschläge von Ferrando und Lorenzo-Sevas (2007a, 2007b) oder Wang (2006). Aus diesem Grund soll auch ein Verfahren entwickelt werden, um gemeinsam Modell (4.37) für log-normalverteilte Reaktionszeiten mit quadratischer Parametrisierung und ein



Tabelle 6.2: Itemparameterschätzung bei log-normalverteilten Reaktionszeiten und linearer Parametrisierung – Ergebnisse der 50 Simulationen mit N=10000

Parameter a					
Item	1	2	3	4	5
Wahrer Wert	1.500	1.500	1.500	1.500	1.500
Min	1.412	1.405	1.414	1.416	1.388
Max	1.631	1.584	1.581	1.618	1.644
Mean	1.516	1.498	1.506	1.515	1.513
STD	0.056	0.048	0.041	0.046	0.058
RMSE	0.058	0.048	0.041	0.048	0.060
Parameter b					
Item	1	2	3	4	5
Wahrer Wert	-1.000	-0.500	0.000	0.500	1.000
Min	-1.053	-0.528	-0.032	0.451	0.936
Max	-0.947	-0.457	0.041	0.549	1.044
Mean	-0.999	-0.494	-0.002	0.499	0.995
STD	0.023	0.015	0.015	0.018	0.021
RMSE	0.023	0.016	0.015	0.018	0.021
Parameter c					
Item	1	2	3	4	5
Wahrer Wert	0.200	0.200	0.300	0.500	0.500
Min	0.174	0.156	0.244	0.433	0.455
Max	0.239	0.231	0.333	0.557	0.563
Mean	0.202	0.196	0.299	0.500	0.506
STD	0.016	0.017	0.017	0.028	0.024
RMSE	0.017	0.017	0.017	0.028	0.025
Parameter l					
Item	1	2	3	4	5
Wahrer Wert	0.500	0.500	0.300	0.200	0.200
Min	0.480	0.480	0.274	0.175	0.176
Max	0.544	0.524	0.334	0.222	0.218
Mean	0.500	0.500	0.302	0.199	0.196
STD	0.013	0.011	0.012	0.011	0.011
RMSE	0.013	0.011	0.012	0.011	0.011
Parameter k					
Item	1	2	3	4	5
Wahrer Wert	1.500	1.500	1.500	1.500	1.500
Min	1.473	1.475	1.478	1.473	1.477
Max	1.525	1.521	1.529	1.527	1.520
Mean	1.502	1.500	1.499	1.501	1.501
STD	0.011	0.010	0.011	0.012	0.011
RMSE	0.011	0.010	0.011	0.012	0.011
Parameter $\sigma$					
Item	1	2	3	4	5
Wahrer Wert	1.000	1.000	1.000	1.000	1.000
Min	0.971	0.939	0.972	0.950	0.943
Max	1.043	1.035	1.035	1.052	1.049
Mean	0.997	1.003	1.000	0.999	0.992
STD	0.013	0.017	0.015	0.025	0.024
RMSE	0.014	0.017	0.015	0.052	0.025

Tabelle 6.3: Itemparameterschätzung bei log-normalverteilten Reaktionszeiten und linearer Parametrisierung – Ergebnisse der 50 Simulationen mit N=1000

Parameter a					
Item	1	2	3	4	5
Wahrer Wert	1.500	1.500	1.500	1.500	1.500
Min	1.205	1.219	1.292	1.160	1.280
Max	2.674	1.721	1.933	1.924	3.477
Mean	1.640	1.463	1.508	1.507	1.644
STD	0.332	0.113	0.141	0.151	0.369
RMSE	0.360	0.119	0.142	0.151	0.396
Parameter b					
Item	1	2	3	4	5
Wahrer Wert	-1.000	-0.500	0.000	0.500	1.000
Min	-1.113	-0.629	-0.090	0.365	0.803
Max	-0.830	-0.418	0.107	0.636	1.262
Mean	-0.980	-0.512	-0.004	0.490	0.998
STD	0.071	0.047	0.040	0.056	0.084
RMSE	0.074	0.049	0.040	0.056	0.085
Parameter c					
Item	1	2	3	4	5
Wahrer Wert	0.200	0.200	0.300	0.500	0.500
Min	0.070	0.076	0.207	0.299	0.276
Max	0.314	0.328	0.417	0.712	0.718
Mean	0.192	0.198	0.299	0.512	0.515
STD	0.056	0.061	0.050	0.077	0.083
RMSE	0.056	0.061	0.050	0.078	0.085
Parameter l					
Item	1	2	3	4	5
Wahrer Wert	0.500	0.500	0.300	0.200	0.200
Min	0.404	0.432	0.186	0.128	0.133
Max	0.572	0.584	0.366	0.321	0.289
Mean	0.501	0.507	0.296	0.198	0.206
STD	0.035	0.038	0.037	0.046	0.037
RMSE	0.035	0.039	0.037	0.046	0.037
Parameter k					
Item	1	2	3	4	5
Wahrer Wert	1.500	1.500	1.500	1.500	1.500
Min	1.415	1.452	1.418	1.403	1.423
Max	1.574	1.553	1.570	1.609	1.573
Mean	1.499	1.501	1.494	1.507	1.499
STD	0.036	0.025	0.030	0.035	0.033
RMSE	0.036	0.025	0.031	0.036	0.033
Parameter $\sigma$					
Item	1	2	3	4	5
Wahrer Wert	1.000	1.000	1.000	1.000	1.000
Min	0.881	0.911	0.862	0.688	0.707
Max	1.126	1.114	1.091	1.137	1.213
Mean	0.997	1.010	0.997	0.978	0.982
STD	0.048	0.048	0.051	0.088	0.104
RMSE	0.048	0.049	0.051	0.091	0.106

Tabelle 6.4: Itemparameterschätzung bei log-normalverteilten Reaktionszeiten und linearer Parametrisierung – Ergebnisse der 35 konvergierenden Simulationen mit  $N=100$

Parameter a					
Item	1	2	3	4	5
Wahrer Wert	1.500	1.500	1.500	1.500	1.500
Min	0.806	0.824	0.819	0.726	0.767
Max	4.608	3.016	2.777	4.455	11.071
Mean	1.971	1.646	1.496	1.789	1.258
STD	0.899	0.532	0.495	0.879	2.012
RMSE	1.015	0.552	0.495	0.925	2.285
Parameter b					
Item	1	2	3	4	5
Wahrer Wert	-1.000	-0.500	0.000	0.500	1.000
Min	-1.597	-0.798	-0.371	0.122	0.669
Max	-0.638	-0.158	0.371	0.942	1.568
Mean	-1.004	-0.481	0.013	0.525	0.980
STD	0.210	0.154	0.174	0.190	0.211
RMSE	0.210	0.156	0.175	0.191	0.212
Parameter c					
Item	1	2	3	4	5
Wahrer Wert	0.200	0.200	0.300	0.500	0.500
Min	0.007	-0.291	-0.698	-1.050	-0.913
Max	0.592	0.594	0.916	0.969	0.843
Mean	0.203	0.146	0.230	0.339	0.313
STD	0.119	0.239	0.330	0.461	0.409
RMSE	0.119	0.245	0.331	0.482	0.424
Parameter l					
Item	1	2	3	4	5
Wahrer Wert	0.500	0.500	0.300	0.200	0.200
Min	0.286	0.300	0.059	-0.043	-0.054
Max	0.727	0.785	0.549	0.491	0.488
Mean	0.507	0.541	0.299	0.195	0.194
STD	0.110	0.107	0.128	0.127	0.138
RMSE	0.111	0.114	0.128	0.127	0.138
Parameter k					
Item	1	2	3	4	5
Wahrer Wert	1.500	1.500	1.500	1.500	1.500
Min	1.287	1.250	1.266	1.228	1.271
Max	1.658	1.723	1.641	1.736	1.816
Mean	1.472	1.509	1.494	1.490	1.480
STD	0.091	0.120	0.099	0.124	0.125
RMSE	0.095	0.121	0.099	0.125	0.127
Parameter $\sigma$					
Item	1	2	3	4	5
Wahrer Wert	1.000	1.000	1.000	1.000	1.000
Min	0.741	0.646	0.477	0.323	0.446
Max	1.344	1.278	1.305	1.548	1.264
Mean	0.974	0.949	0.883	0.986	0.932
STD	0.157	0.155	0.198	0.252	0.196
RMSE	0.159	0.163	0.230	0.252	0.208

zweiparametrisches IRT-Modell an die Daten einer Stichprobe anzupassen. Kernannahme von Modell (4.37) ist die Annahme, dass die Reaktionszeit bei Aufgabe  $g$  log-normalverteilt ist und der Lokationsparameter der Log-Normalverteilung auf die Personenfähigkeit  $\theta$  und den Residualfaktor  $\omega$  einer Person zurückgeführt werden kann gemäß der Beziehung:

$$\eta_g(\theta, \omega) = k_g + l_g(\theta - d_g)^2 + c_g\omega. \quad (6.5)$$

Der Skalenparameter  $\sigma_g$  der Log-Normalverteilung ist zwar für die einzelnen Aufgaben verschieden, hängt aber nicht von der Person ab. Bei Modellierung (6.5) steht nur die Personenfähigkeit in einer Unfolding-Beziehung zur Reaktionszeit, der Residualfaktor beeinflusst die Reaktionszeit weiterhin monoton. Für eine genaue Interpretation des Modells sei auf den entsprechenden Abschnitt in Kapitel 4.2.1 verwiesen. Ein ähnliches Modell wurde bereits von Ferrando und Lorenzo-Sevas (2007a) vorgeschlagen, jedoch gingen diese Autoren nicht von einer quadratischen Beziehung aus, sondern von einer linearen Beziehung zwischen dem Lokationsparameter und der Größe  $|\theta - d_g|$ . Empirische Evidenz findet sich sowohl für den Ansatz von Ferrando als auch für die quadratische Beziehung (Ferrando, 2006; Akrami et al., 2007). Allerdings handelt es sich bei dem Modell von Ferrando und Lorenzo-Sevas (2007a) um ein etwas einfacheres Modell, da die Ladungen des zweiten Faktors über alle Items als identisch angenommen werden. Als Verteilungsdichte der Reaktionszeit  $t_g$  bei Aufgabe  $g$  resultiert bei Modell (6.5) für eine Person mit Personenfähigkeit  $\theta$  und Residualfaktor  $\omega$  die Dichtefunktion:

$$f(t_g|\theta, \omega; c_g, d_g, l_g, k_g, \sigma_g) = \frac{1}{\sqrt{2\pi}\sigma_g t_g} \exp \left[ -\frac{1}{2} \frac{[\log(t_g) - \eta_g(\theta, \omega)]^2}{\sigma_g^2} \right] \quad \text{mit } t_g > 0. \quad (6.6)$$

Der Lokationsparameter  $\eta_g(\theta, \omega)$  kann dabei über Beziehung (6.5) auf die Personenfähigkeit  $\theta$  und den Residualfaktor  $\omega$  einer Person zurückgeführt werden. Für die Wahrscheinlichkeit einer positiven Antwort bei Aufgabe  $g$  gilt das zweiparametrische Probit-Modell

$$P_{x_g}(\theta; a_g, d_g) = \Phi(a_g(\theta - d_g)), \quad (6.7)$$

wobei  $\Phi(x)$  die Verteilungsfunktion der Standardnormalverteilung ist. Der Wechsel vom zweiparametrischen Logit-Modell zum zweiparametrischen Probit-Modell beruht auf denselben Gründen, mit denen der Modellwechsel bereits im ersten Teil dieses Kapitels gerechtfertigt wurde: Obwohl zwischen den Itemcharakteristiken der beiden Modelle nur ein geringer Unterschied besteht, wird die Itemparameterschätzung durch die Annahme eines Probit-Modells erheblich vereinfacht. Man beachte, dass die Itemlokation  $d_g$  sowohl in die Verteilung der Reaktionszeit über Gleichung (6.5) als auch in die Itemcharakteristik des zweiparametrischen Probit-Modells (6.7) eingeht.

Bei Bedingung auf die Persönlichkeitseigenschaften  $\theta$  und  $\omega$  sind die Itemantworten und Reaktionszeiten voneinander unabhängig. Die gemeinsame Verteilung der Reaktionszeiten  $\mathbf{t} = [t_1, \dots, t_G]$  und Itemantworten  $\mathbf{x} = [x_1, \dots, x_G]$

bei  $G$  Aufgaben mit Itemparametern  $\mathbf{a} = [a_1, \dots, a_G]$ ,  $\mathbf{d} = [d_1, \dots, d_G]$ ,  $\mathbf{c} = [c_1, \dots, c_G]$ ,  $\mathbf{l} = [l_1, \dots, l_G]$ ,  $\mathbf{k} = [k_1, \dots, k_G]$  und  $\boldsymbol{\sigma} = [\sigma_1, \dots, \sigma_G]$  lautet bei einer Person mit Residualfaktor  $\omega$  und Personenfähigkeit  $\theta$ :

$$f(\mathbf{x}, \mathbf{t} | \theta, \omega; \mathbf{a}, \mathbf{d}, \mathbf{c}, \mathbf{l}, \mathbf{k}, \boldsymbol{\sigma}) = \prod_{g=1}^G [P_{x_g}(\theta; a_g, d_g)]^{x_g} [1 - P_{x_g}(\theta; a_g, d_g)]^{1-x_g} \quad (6.8)$$

$$f(t_g | \theta, \omega; c_g, d_g, l_g, k_g, \sigma_g).$$

Bei Beobachtung der Reaktionszeiten und Itemantworten von  $N$  Personen ergibt sich als Log-Likelihood-Funktion der Parameter die Beziehung:

$$\text{LL}(\boldsymbol{\theta}, \boldsymbol{\omega}, \mathbf{a}, \mathbf{d}, \mathbf{c}, \mathbf{l}, \mathbf{k}, \boldsymbol{\sigma}) = \log \prod_{i=1}^N f(\mathbf{x}_i, \mathbf{t}_i | \theta_i, \omega_i; \mathbf{a}, \mathbf{d}, \mathbf{c}, \mathbf{l}, \mathbf{k}, \boldsymbol{\sigma}). \quad (6.9)$$

Die Vektoren  $\boldsymbol{\theta}$  und  $\boldsymbol{\omega}$  stehen für die unbekanntenen Ausprägungen der Personenfähigkeiten und Residualfaktorwerte bei den  $N$  Versuchspersonen. Da die Anzahl der Personenparameter mit dem Stichprobenumfang ansteigt, führt eine direkte Maximierung der Log-Likelihood-Funktion nicht zwangsläufig zu konsistenten Itemparameterschätzern (Neyman & Scott, 1948).

Modell (6.5) nimmt log-normalverteilte Reaktionszeiten an, deren Lokationsparameter  $\eta_g(\theta, \omega)$  über eine quadratische Beziehung von der Personenfähigkeit  $\theta$  und über eine lineare Beziehung von dem Residualfaktor  $\omega$  abhängt. Logarithmiert man die Reaktionszeiten, handelt es sich bei Modell (6.5) um ein nonlineares, faktorenanalytisches Modell, bei dem die logarithmierten Reaktionszeiten nach Bedingung auf die Persönlichkeitseigenschaften  $\theta$  und  $\omega$  normalverteilt sind. Die Schätzung der Itemparameter wird jedoch erheblich durch den Umstand erschwert, dass die Itemlokation  $d_g$  des Reaktionszeitmodells (6.5) und des IRT-Modells (6.7) als gleich angenommen wird. Auch können zwischen den Reaktionszeiten und den Itemantworten verschiedener Items keine tetrachorische Korrelationskoeffizienten berechnet werden, da die marginale Verteilung der logarithmierten Reaktionszeiten keiner Normalverteilung gehorcht. Herkömmliche faktorenanalytische Verfahren, wie sie zur Schätzung des linearen Modells eingesetzt wurden, scheiden aus diesem Grund zur Bestimmung der Itemparameter aus.

Eine Lösung dieses Problems ist in der Betrachtung der marginalen Log-Likelihood-Funktion zu finden. Diese ergibt sich nach Ausintegration der Personenparameter. Nimmt man an, dass innerhalb einer Population von potentiellen Testteilnehmern die Personenfähigkeit  $\theta$  und der Residualfaktor  $\omega$  bivariat standardnormalverteilt sind mit Erwartungswertvektor  $\boldsymbol{\mu} = [0, 0]'$  und einer Identitätsmatrix als Varianz-Kovarianz-Matrix  $\boldsymbol{\Sigma}$ , ergibt sich folgende marginale Verteilung des Antwortmusters  $[\mathbf{x}, \mathbf{t}]$ :

$$f(\mathbf{x}, \mathbf{t}; \mathbf{a}, \mathbf{d}, \mathbf{c}, \mathbf{l}, \mathbf{k}, \boldsymbol{\sigma}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\mathbf{x}, \mathbf{t} | \theta, \omega; \mathbf{a}, \mathbf{d}, \mathbf{c}, \mathbf{l}, \mathbf{k}, \boldsymbol{\sigma}) \quad (6.10)$$

$$\frac{1}{(2\pi)\sqrt{\det(\boldsymbol{\Sigma})}} \exp \left[ -\frac{1}{2} [\theta, \omega] \boldsymbol{\Sigma}^{-1} [\theta, \omega]' \right] d\theta d\omega.$$

Als marginale Log-Likelihood-Funktion erhält man somit bei Verwendung von (6.10) für eine Stichprobe aus  $N$  Personen die Funktion:

$$\text{MLL}(\mathbf{a}, \mathbf{d}, \mathbf{c}, \mathbf{l}, \mathbf{k}, \boldsymbol{\sigma}) = \log \prod_{i=1}^N f(\mathbf{x}_i, \mathbf{t}_i; \mathbf{a}, \mathbf{d}, \mathbf{c}, \mathbf{l}, \mathbf{k}, \boldsymbol{\sigma}). \quad (6.11)$$

Es ist diese marginale Log-Likelihood-Funktion, welche beim Marginal-Maximum-Likelihood-Ansatz maximiert wird. Allerdings ist die Maximierung nicht unproblematisch. Ferrando und Lorenzo-Sevas (2007a) gingen bei der Schätzung ihres verwandten Modells so vor, dass sie zuerst die Itemparameter  $a_g$  und  $d_g$  des IRT-Modells allein aufgrund der Itemantworten bestimmten. Dies ist mit Standardsoftware für IRT-Modelle problemlos möglich. Die dadurch erhaltenen Schätzer  $\hat{a}_g$  und  $\hat{d}_g$  wurden dann bei der Schätzung der restlichen Itemparameter des Reaktionszeitmodells als Stellvertreter für die wahren Parameter  $a_g$  und  $d_g$  genutzt. Die Parameter werden somit in zwei Schritten geschätzt. Im Gegensatz zu dieser Vorgehensweise sollen im folgenden alle Itemparameter des Modells simultan geschätzt werden. Die Maximierung der marginalen Log-Likelihood-Funktion kann über einen EM-Algorithmus erfolgen. Die beiden folgenden Abschnitte beschreiben die Schritte des EM-Algorithmus und können beim ersten Lesen übersprungen werden.

### E-Schritt

Das Schätzproblem lässt sich als Missing-Data-Problem auffassen. Im Gegensatz zu den vorherigen Problemen gibt es beim aktuellen Schätzproblem jedoch zwei Arten von fehlenden Daten. Wie in den vorherigen Abschnitten stellen die Persönlichkeitseigenschaften  $\theta$  und  $\omega$  fehlende Daten dar. Die Einführung einer zweiten Art von fehlenden Daten lässt sich folgendermaßen begründen: Wie auf S. 105 dargestellt wurde, kann die beobachtete Itemantwort  $x_g$  einer Person als dichotomisierte, jedoch ursprünglich normalverteilte, latente Variable  $z_g$  aufgefasst werden, deren genaue Ausprägung allerdings unbekannt ist und bei der es sich somit um eine fehlende Beobachtung handelt. Die Einführung einer derartigen unbekanntenen, kontinuierlichen Variablen ist analog zum Prinzip der „Data Augmentation“ zu sehen, welches bei der Markov Chain Monte Carlo Schätzung von IRT-Modellen zur Anwendung kommt – als Beispiel hierfür sei die Veröffentlichung von De Knop, Rijmen und van Mechelen (2003) genannt. Könnte diese kontinuierliche, latente Variable  $z_g$  für jedes Item und jede Person beobachtet werden, würde sich die Itemparameterschätzung erheblich vereinfachen.

Bei Kenntnis der Personenparameter  $\theta$  und  $\omega$  einer Testpersonen kann die Verteilung der logarithmierten Reaktionszeiten  $\mathbf{t}^* = \log(\mathbf{t}) = [\log(t_1), \dots, \log(t_G)]$  und der latenten Variablen  $\mathbf{z} = [z_1, \dots, z_G]$  bei  $G$  Aufgaben mit der folgenden Verteilungsdichte beschrieben werden:

$$f(\mathbf{z}, \mathbf{t}^* | \theta, \omega; \mathbf{a}, \mathbf{d}, \mathbf{c}, \mathbf{l}, \mathbf{k}, \boldsymbol{\sigma}) = \prod_{g=1}^G \frac{1}{\sqrt{2\pi\sigma_g^2}} \exp \left[ -\frac{1}{2} \frac{[t_g^* - \eta_g(\theta, \omega)]^2}{\sigma_g^2} \right] \frac{1}{\sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left[ \frac{z_g - (a_g(\theta - d_g))}{1} \right]^2 \right] \quad (6.12)$$

Gleichung (6.12) ergibt sich, da die kontinuierlichen latenten Variablen  $z_{ig}$  und die logarithmierten Reaktionszeiten  $t_{ig}^*$  normalverteilt sind. Als „vollständige“ Log-Likelihood-Funktion der Itemparameter resultiert bei Beobachtung von  $N$  Personen und Kenntnis der Personenparameter die Funktion:

$$\begin{aligned} \text{LL}(\mathbf{a}, \mathbf{d}, \mathbf{c}, \mathbf{l}, \mathbf{k}, \boldsymbol{\sigma}^2) = & \sum_{i=1}^N \sum_{g=1}^G \left[ \log \left( \frac{1}{\sqrt{2\pi\sigma_g^2}} \right) - \frac{1}{2\sigma_g^2} \left( t_{ig}^* - (k_g + l_g(\theta_i - d_g) + c_g\omega_i) \right)^2 \right. \\ & \left. + \log \left( \frac{1}{\sqrt{2\pi}} \right) - \frac{1}{2} \left( z_{ig} - a_g(\theta_i - d_g) \right)^2 \right]. \end{aligned} \quad (6.13)$$

Expandiert man in Gleichung (6.13) die Klammern, führt dies zu:

$$\begin{aligned} \text{LL}(\mathbf{a}, \mathbf{d}, \mathbf{c}, \mathbf{l}, \mathbf{k}, \boldsymbol{\sigma}^2) = & \sum_{i=1}^N \sum_{g=1}^G \left[ \log \left( \frac{1}{\sqrt{2\pi\sigma_g^2}} \right) + \frac{1}{\sigma_g^2} \left( l_g t_{ig}^* \theta_i^2 - k_g c_g \omega_i + 2l_g^2 d_g \theta_i^3 - k_g l_g d_g^2 - 3l_g^2 d_g^2 \theta_i^2 \right. \right. \\ & - k_g l_g \theta_i^2 + c_g t_{ig}^* \omega_i + 2l_g^2 d_g^3 \theta_i + l_g d_g^2 t_{ig}^* - \frac{1}{2} t_{ig}^{*2} + k_g t_{ig}^* - \frac{1}{2} k_g^2 - \frac{1}{2} \theta_i^4 - \frac{1}{2} l_g^2 d_g^4 \\ & \left. - c_g^2 \omega_i^2 + 2l_g c_g d_g \theta_i \omega_i - l_g c_g \theta_i^2 \omega_i - l_g c_g d_g^2 \omega_i - 2l_g d_g t_{ig}^* \theta_i + 2k_g l_g d_g \theta_i \right) \\ & \left. + \log \left( \frac{1}{\sqrt{2\pi}} \right) - \frac{1}{2} z_{ig}^2 + a_g z_{ig} \theta_i - a_g d_g z_{ig} - \frac{1}{2} a_g^2 \theta_i^2 + a_g^2 d_g \theta_i - \frac{1}{2} a_g^2 d_g^2 \right]. \end{aligned} \quad (6.14)$$

Natürlich ist diese „vollständige“ Log-Likelihood-Funktion nicht berechenbar, da die latenten Variablen  $\theta_i$ ,  $\omega_i$  und  $\mathbf{z}_i$  bei den verschiedenen Versuchspersonen nicht beobachtet werden können und somit in Gleichung (6.14) mehrere Summanden unbestimmt bleiben. Eine Übersicht über die unbekanntenen Größen der Log-Likelihood-Funktion ist in Tabelle 6.5 wiedergegeben.

Wären diese unbekanntenen Größen bekannt, könnte die Log-Likelihood-Funktion berechnet und Schätzwerte für die gesuchten Itemparameter bestimmt werden. Der E-Schritt besteht nun darin, alle unbeobachteten Statistiken durch ihren bedingten Erwartungswert zu ersetzen. Grundannahme dabei ist, dass die

Tabelle 6.5: Übersicht über die unbeobachtbaren Größen der „vollständigen“ Log-Likelihood-Funktion (6.14)

Nr	Komponente
1	$\sum_{i=1}^N \theta_i$
2	$\sum_{i=1}^N \theta_i^2$
3	$\sum_{i=1}^N \theta_i^3$
4	$\sum_{i=1}^N \theta_i^4$
5	$\sum_{i=1}^N \omega_i$
6	$\sum_{i=1}^N \omega_i^2$
7	$\sum_{i=1}^N \theta_i \omega_i$
8	$\sum_{i=1}^N \theta_i^2 \omega_i$
9	$\sum_{i=1}^N t_{ig}^* \theta_i$
10	$\sum_{i=1}^N t_{ig}^* \theta_i^2$
11	$\sum_{i=1}^N t_{ig}^* \omega_i$
12	$\sum_{i=1}^N z_{ig}$
13	$\sum_{i=1}^N z_{ig}^2$
14	$\sum_{i=1}^N z_{ig} \theta_i$

Personen aus einer Population stammen, in der die Persönlichkeitseigenschaften  $\theta$  und  $\omega$  mit einer bestimmten Verteilung  $f(\theta, \omega)$  verteilt sind. Im Folgenden wird für diese Verteilung eine bivariate Normalverteilung mit Korrelationskoeffizienten  $\rho = 0$  angenommen. Nach Beobachtung der Itemantworten und logarithmierten Reaktionszeiten  $[\mathbf{x}, \mathbf{t}^*]$  einer Person bei den  $G$  Aufgaben kann die bedingte Verteilung der Personenparameter  $f(\theta, \omega | \mathbf{x}, \mathbf{t}^*)$  berechnet werden. Hierfür ist die Einführung einer Indikatorfunktion hilfreich: Es sei  $I(x, z)$  die Indikatorfunktion, welche für  $I(0, z < 0)$  und  $I(1, z > 0)$  den Wert Eins und ansonsten den Wert Null annimmt. Die auf die beobachteten Größen  $\mathbf{x}$  und  $\mathbf{t}^*$  bedingte Verteilung der Personenparameter  $\theta$  und  $\omega$  ergibt sich dann als:

$$\begin{aligned}
& f(\theta, \omega | \mathbf{x}, \mathbf{t}^*; \mathbf{a}, \mathbf{d}, \mathbf{c}, \mathbf{l}, \mathbf{k}, \boldsymbol{\sigma}) \\
&= \frac{\int_{-\infty}^{\infty} \prod_{g=1}^G I(x_g, z_g) f(z_g | \theta; a_g, d_g) f(t_g^* | \theta, \omega; c_g, d_g, l_g, k_g, \sigma_g) f(\theta, \omega) \, d\mathbf{z}}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{g=1}^G I(x_g, z_g) f(z_g | \theta; a_g, d_g) f(t_g^* | \theta, \omega; c_g, d_g, l_g, k_g, \sigma_g) f(\theta, \omega) \, d\mathbf{z} \, d\theta \, d\omega} \\
&= \frac{\prod_{g=1}^G P(x_g | \theta; a_g, d_g) f(t_g^* | \theta, \omega; c_g, d_g, l_g, k_g, \sigma_g) f(\theta, \omega)}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{g=1}^G P(x_g | \theta; a_g, d_g) f(t_g^* | \theta, \omega; c_g, d_g, l_g, k_g, \sigma_g) f(\theta, \omega) \, d\theta \, d\omega} \\
&= \frac{f(\mathbf{x}, \mathbf{t}^* | \theta, \omega; \mathbf{a}, \mathbf{d}, \mathbf{c}, \mathbf{l}, \mathbf{k}, \boldsymbol{\sigma}) f(\theta, \omega)}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\mathbf{x}, \mathbf{t}^* | \theta, \omega; \mathbf{a}, \mathbf{d}, \mathbf{c}, \mathbf{l}, \mathbf{k}, \boldsymbol{\sigma}) f(\theta, \omega) \, d\theta \, d\omega}.
\end{aligned} \tag{6.15}$$

Bei den Verteilungen  $f(t_g^* | \theta, \omega; c_g, d_g, l_g, k_g, \sigma_g)$  und  $f(z_g | \theta; a_g, d_g)$  handelt es sich um Normalverteilungen mit den entsprechenden Verteilungsparametern,



siehe Gleichung (6.12). Um allerdings eine konkrete bedingte Verteilung zu erhalten, ist es notwendig, für die Itemparameter in Gleichung (6.15) vorläufige Werte  $\mathbf{a}_0$ ,  $\mathbf{d}_0$ ,  $\mathbf{c}_0$ ,  $\mathbf{l}_0$ ,  $\mathbf{k}_0$  und  $\boldsymbol{\sigma}_0$  zu wählen. Dabei kann es sich entweder um frei gewählte Startwerte handeln oder um Lösungen eines früheren EM-Zyklus. Liegen vorläufige Itemparameter vor, können die bedingten Erwartungswerte der in Tabelle 6.5 aufgelisteten Komponenten bestimmt werden. Die bedingten Erwartungswerte der Komponenten 1-11 ergeben sich, wenn man die entsprechenden Komponenten über die bedingte Verteilung  $f(\theta, \omega | \mathbf{t}^*, \mathbf{x}; \mathbf{a}_0, \mathbf{d}_0, \mathbf{c}_0, \mathbf{l}_0, \mathbf{k}_0, \boldsymbol{\sigma}_0)$  integriert. Der bedingte Erwartungswert von Komponente 1 berechnet sich beispielsweise als:

$$E \left[ \sum_{i=1}^N \theta_i | \mathbf{x}_i, \mathbf{t}_i^*; \mathbf{a}_0, \mathbf{d}_0, \mathbf{c}_0, \mathbf{l}_0, \mathbf{k}_0, \boldsymbol{\sigma}_0 \right] = \sum_{i=1}^N \left[ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \theta f(\theta, \omega | \mathbf{x}_i, \mathbf{t}_i^*; \mathbf{a}_0, \mathbf{d}_0, \mathbf{c}_0, \mathbf{l}_0, \mathbf{k}_0, \boldsymbol{\sigma}_0) d\theta d\omega \right]. \quad (6.16)$$

Zwar ist dieses Integral nicht in geschlossener Form angebar, jedoch kann es über eine Gauss-Hermite-Quadratur approximiert werden. Die Anwendung einer kartesischen Quadraturregel führt zu einer gewichteten Summe von Funktionswerten an ausgewählten Stützstellen:

$$E \left[ \sum_{i=1}^N \theta_i | \mathbf{x}_i, \mathbf{t}_i^*; \mathbf{a}_0, \mathbf{d}_0, \mathbf{c}_0, \mathbf{l}_0, \mathbf{k}_0, \boldsymbol{\sigma}_0 \right] = \sum_{i=1}^N \sum_{q_1=1}^{Q_1} \sum_{q_2=1}^{Q_2} \left[ \theta_{q_1} \frac{f(\mathbf{x}_i, \mathbf{t}_i^* | \theta_{q_1}, \omega_{q_2}; \mathbf{a}_0, \mathbf{d}_0, \mathbf{c}_0, \mathbf{l}_0, \mathbf{k}_0, \boldsymbol{\sigma}_0)}{\sum_{q_3=1}^{Q_3} \sum_{q_4=1}^{Q_4} f(\mathbf{x}_i, \mathbf{t}_i^* | \theta_{q_3}, \omega_{q_4}; \mathbf{a}_0, \mathbf{d}_0, \mathbf{c}_0, \mathbf{l}_0, \mathbf{k}_0, \boldsymbol{\sigma}_0) w_{q_3} w_{q_4}} w_{q_1} w_{q_2} \right]. \quad (6.17)$$

Bei den Größen  $w_{q_1}$  bis  $w_{q_4}$  handelt es sich um die Gewichte der Gauss-Hermite-Quadratur, bei den Größen  $\theta_{q_1}$  bis  $\omega_{q_4}$  um die Stützstellen. Diese Größen sind in Standardlehrbüchern zur numerischen Integration tabelliert (Davis & Rabinowitz, 1975). Die weiteren Komponenten 2-11 in Tabelle 6.5 können analog berechnet werden. Die Bestimmung der auf die beobachtbaren Variablen  $[\mathbf{x}_i, \mathbf{t}_i^*]$  bedingten Erwartungswerte von Komponente 12-14 in Tabelle 6.5 erfordert jedoch noch einen zusätzlichen Schritt. Hierbei kommt das Gesetz der iterierten Erwartung zur Anwendung. Es gilt allgemein:

$$E[z_{ig} | \mathbf{x}_i, \mathbf{t}_i^*] = E_{\theta} [E[z_{ig} | \theta, \mathbf{x}_i, \mathbf{t}_i^*]]. \quad (6.18)$$

Bei Bedingung auf  $\theta$  sind die logarithmierten Reaktionszeiten  $\mathbf{t}_i^*$  und die latenten Variablen  $\mathbf{z}_i$  bei den  $g$  Items voneinander unabhängig, so dass der Erwartungswert  $E[z_{ig} | \theta, \mathbf{t}_i^*, \mathbf{x}_i]$  für das  $g$ -te Item einfach berechnet werden kann. Hierzu ist die Verteilung  $f(z_{ig} | \theta, \mathbf{t}_i^*, \mathbf{x}_i)$  zu betrachten. Aufgrund der Unabhängigkeit von  $\mathbf{t}_i^*$  und  $\mathbf{z}_i$  vereinfacht sich diese Verteilung zu  $f(z_{ig} | \theta, \mathbf{x}_i)$ . Bei

der bedingten Verteilung  $f(z_{ig}|\theta, \mathbf{x}_i)$  handelt es sich um eine abgeschnittene Normalverteilung, die an der Schwelle 0 gekappt wird. Die Momente der abgeschnittenen Normalverteilung sind bekannt und ergeben sich, wenn man die Transformation  $z_l = (0 - \mu(z_{ig}|\theta))/\sigma_g$  verwendet, als:

$$E[z_{ig}|x_{ig} = 1, \theta] = E[z_{ig}|z_{ig} > 0, \theta] = \mu(z_{ig}|\theta) + \sigma_g \frac{\phi(z_l)}{1 - \Phi(z_l)} \quad (6.19)$$

$$\sigma^2[z_{ig}|x_{ig} = 1, \theta] = \sigma^2[z_{ig}|z_{ig} > 0, \theta] = \sigma_g^2 \left[ 1 + z_l \frac{\phi(z_l)}{1 - \Phi(z_l)} - \left[ \frac{\phi(z_l)}{1 - \Phi(z_l)} \right]^2 \right]. \quad (6.20)$$

Analog existieren Formeln für den Erwartungswert  $E[z_{ig}|x_{ig} = 0, \theta]$  und für die Varianz  $\sigma^2[z_{ig}|x_{ig} = 0, \theta]$ . Zu beachten ist dabei, dass  $\mu(z_{ig}|\theta)$  nach der auf S. 105 beschriebenen Logik auf den Personenparameter  $\theta$  zurückgeführt werden kann über die Beziehung  $\mu(z_{ig}|\theta) = a_g(\theta - d_g)$ . Diese Vorstellung lag ja genau der Auffassung einer Itemantwort als Resultat einer dichotomisierten, latenten Variablen zu Grunde. Beziehung (6.19) und Beziehung (6.20) können zur Berechnung der Komponenten 12-14 in Tabelle 6.5 ausgenutzt werden. Sie dienen dazu, den auf die Personenfähigkeit  $\theta$  bedingten Erwartungswert der Komponenten zu berechnen, also zur Berechnung des inneren Elements in Gleichung (6.18). Den nicht auf  $\theta$  bedingten Erwartungswert erhält man dann durch erneute Erwartungswertbildung über  $\theta$ . Der auf die beobachteten Itemantworten  $\mathbf{x}_i$  und logarithmierten Reaktionszeiten  $\mathbf{t}_i^*$  bedingte Erwartungswert von Komponente 12 berechnet sich beispielsweise als

$$\begin{aligned} E \left[ \sum_{i=1}^N z_{ig} \middle| \mathbf{x}_i, \mathbf{t}_i^*; \mathbf{a}_0, \mathbf{d}_0, \mathbf{c}_0, \mathbf{l}_0, \mathbf{k}_0, \boldsymbol{\sigma}_0 \right] = \\ \sum_{i=1}^N E_{\theta} \left[ E[z_{ig}|\theta, a_{0g}, b_{0g}] \middle| \mathbf{x}_i, \mathbf{t}_i^*; \mathbf{a}_0, \mathbf{d}_0, \mathbf{c}_0, \mathbf{l}_0, \mathbf{k}_0, \boldsymbol{\sigma}_0 \right] \quad (6.21) \\ \sum_{i=1}^N \left[ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E[z_{ig}|x_{ig}, \theta] f(\theta, \omega | \mathbf{x}_i, \mathbf{t}_i^*; \mathbf{a}_0, \mathbf{d}_0, \mathbf{c}_0, \mathbf{l}_0, \mathbf{k}_0, \boldsymbol{\sigma}_0) d\theta d\omega \right], \end{aligned}$$

wobei für  $E[z_{ig}|x_{ig}, \theta]$  der Erwartungswert  $E[z_{ig}|z_{ig} > 0, \theta]$  gewählt wird, wenn  $x_{ig} = 1$  und  $E[z_{ig}|z_{ig} < 0, \theta]$ , wenn  $x_{ig} = 0$  ist. Zwar kann das Integral nicht explizit bestimmt werden, doch ist es analog zu Gleichung (6.17) über eine Gauss-Hermite-Quadratur approximierbar. Die restlichen Elemente in Tabelle 6.5 können auf die gleiche Weise bestimmt werden. Teilweise muss dabei jedoch noch die bedingte Varianz  $\sigma^2[z_{ig}|x_{ig}, \theta]$  berücksichtigt werden.

### M-Schritt

Der M-Schritt beginnt, wenn alle unbeobachteten Komponenten in der Log-Likelihood-Funktion (6.14) durch ihren bedingten Erwartungswert ersetzt wurden. Dies führt zur „erwarteten“ Log-Likelihood-Funktion. Für diese „erwartete“ Log-Likelihood-Funktion kann das Maximum bestimmt werden. Ableitung

der erwarteten Log-Likelihood-Funktion nach den unbekanntem Itemparametern und Nullsetzung der einzelnen Ableitungen ergibt ein Gleichungssystem, das nach den gesuchten Itemparametern aufzulösen ist. Auch wenn sich theoretisch explizite Lösungen für die Itemparameter angeben lassen, scheitert die direkte Berechnung dieser Lösungen jedoch aus praktischen Gründen. Eine Auflösung des Gleichungssystems nach dem Itemparameter  $\hat{b}_g$  ergab beispielsweise eine äußerst unübersichtliche Lösung: Wäre die Varianz  $\sigma_g^2$  bekannt, wäre die Lösung für  $\hat{b}_g$  bereits ein Polynom siebter Ordnung. Möchte man das Gleichungssystem weiter nach  $\sigma_g^2$  auflösen, ergibt sich eine praktisch nicht mehr handhabbare Lösungsgleichung. Die Schwierigkeit der Berechnung expliziter Lösungen weist daraufhin, dass bei der Maximierung der erwarteten Log-Likelihood-Funktion mit vielen lokalen Extremwerten zu rechnen ist. Aufgrund der damit verbundenen Komplikationen wurde für die Maximierung der Funktion der Simplex-Algorithmus von Nelder und Mead (1965) genutzt. Der Simplex-Algorithmus besitzt den Vorteil, dass er ohne Berechnung der Ableitungen auskommt und als robust gilt. Versuche, die Maximierung über einen Newton-Raphson-Algorithmus durchzuführen, waren nicht erfolgreich. Allerdings tendierte auch der Simplex-Algorithmus dazu, bei weniger guten Startwerten zu lokalen Extremwerten zu konvergieren. Die Parameterschätzung lebt somit von guten Startwerten. Aus diesem Grund wurde dem EM-Algorithmus eine vorläufige Schätzung der Itemparameter über die Momentmethoden vorgeschaltet: Die Itemparameter  $a_g$  und  $d_g$  des zweiparametrischen Probit-Modells (6.7) lassen sich einfach mit Hilfe existierender Softwarepakete bestimmen, wenn man die Reaktionszeiten zunächst ignoriert. Liegen konsistente Schätzungen von  $a_g$  und  $d_g$  vor, können die weiteren Parameter nach dem Verfahren der Momentmethode geschätzt werden. Voraussetzung ist dabei natürlich die bivariate Standardnormalverteilung der latenten Variablen  $\theta$  und  $\omega$ , von der bereits beim Marginal-Maximum-Likelihood-Ansatz ausgegangen wurde. Wie bereits in Anhang A.2.2 dargelegt, ergibt sich das erste nonzentrale Moment  $\mu_1(t_g^*)$  und das dritte zentrale Moment  $z\mu_3(t_g^*)$  der logarithmierten Reaktionszeit  $t_g^*$  bei Item  $g$  als:

$$\mu_1[t_g^*] = k_g + l_g + l_g d_g^2 \quad (6.22)$$

$$z\mu_3[t_g^*] = 8l_g^3 + 24l_g^3 d_g^2 \quad (6.23)$$

Setzt man in die Gleichungen (6.22) und (6.23) den bereits mit Hilfe der Itemantworten berechneten Schätzer  $\hat{d}_g$  ein, ist eine Auflösung des Gleichungssystems nach den unbekanntem Parametern  $l_g$  und  $k_g$  möglich. Die noch fehlenden Parameter  $\sigma_g$  und  $c_g$  sind dann über eine Faktorenanalyse bestimmbar. Für die Kovarianz der logarithmierten Reaktionszeiten zweier Items  $j$  und  $g$  gilt nach Modell (6.6) die Beziehung:

$$\sigma[t_g^*, t_j^*] = 2l_g l_j + 4l_g l_j d_g d_j + c_g c_j. \quad (6.24)$$

Für die Itemparameter  $l_g$ ,  $l_j$ ,  $d_g$  und  $d_j$  können die bereits berechneten Schätzwerte eingesetzt werden. Korrigiert man die Kovarianz der logarithmierten Reaktionszeiten zweier Items  $g$  und  $j$  um die Komponente  $2\hat{l}_g \hat{l}_j + 4\hat{l}_g \hat{l}_j \hat{d}_g \hat{d}_j$ , erhält man eine Residualkovarianz. Wird diese Korrektur für alle Items vorgenommen, ergibt sich eine Matrix mit Residualkovarianzen, die mittels einer

Faktorenanalyse ausgewertet werden kann. Allerdings ist hierfür ein Verfahren auszuwählen, das keine positiv definite Kovarianzmatrix benötigt. Bei einem ausreichend großem Datensatz sind die erhaltenen Schätzwerte bereits erstaunlich präzise.<sup>3</sup> Die so gefundenen Itemparameter können dann als Eingangsgröße für den EM-Algorithmus und als Ausgangspunkt des M-Schrittes dienen.

### Simulation der Itemparameterschätzung

Zur Überprüfung der Eignung der Schätzverfahren wurde eine Simulation durchgeführt. Generiert wurden Itemantworten und Reaktionszeiten bei einem Test aus fünf Items für unterschiedlich große Stichproben. Die bei der Simulation verwandten Itemparameter sind in Tabelle 6.6 zu finden. Ein Anliegen bei der Auswahl der Itemparameter war die Bemühung um ein realistisches Szenario: Der Einfluss der Persönlichkeitseigenschaften auf die Reaktionszeiten ist moderat. Das Einflussgewicht der Personenfähigkeit  $l_g$  ist negativ, um so eine umgekehrt U-förmige Beziehung zwischen der Reaktionszeit und der Distanz von Person und Item auf der Fähigkeitsdimension zu erreichen.

Tabelle 6.6: Log-normalverteilte RT mit quadratischer Parametrisierung: Wahre Itemparameter der Aufgaben

Item	$a_g$	$d_g$	$c_g$	$l_g$	$k_g$	$\sigma_g$
1	1.00	-1.00	0.25	-0.50	0.50	1.00
2	1.00	-0.50	0.25	-0.50	0.50	1.00
3	1.00	0.00	0.50	-0.50	0.50	1.00
4	1.00	0.50	0.50	-0.25	0.50	1.00
5	1.00	1.00	0.50	-0.25	0.50	1.00

Simuliert wurden jeweils 35 Datensätze aus 1000 und 10000 Versuchspersonen. Auf noch kleinere Stichproben wurde aufgrund der sich abzeichnenden Instabilität der Schätzung verzichtet. Für jede Versuchsperson wurden aus einer bivariaten Standardnormalverteilung mit Korrelationskoeffizient  $\rho = 0$  zufällige Eigenschaftsausprägungen gezogen. Anschließend wurden Reaktionszeiten und Itemantworten für fünf Items gemäß Modell (6.6) und Modell (6.7) erzeugt. Mit Hilfe der Momentmethode wurden vorläufige Itemparameterschätzer bestimmt. In einem ersten Schritt diente die Prozedur LTM des Programms R zur Analyse der Itemantworten mit einem zweiparametrischen Logit-Modell. Durch Reskalierung mit dem Faktor 1.702 konnte der Trennschärfeparameter  $\hat{a}_g$  des Logit-Modells in den entsprechenden Trennschärfekoeffizient des Probit-Modells transformiert werden (Savalei, 2006). Dies führte zu den vorläufigen Schätzern der Itemparameter  $\hat{a}_g$  und  $\hat{d}_g$ .

<sup>3</sup>Über die Momentmethode können übrigens auch die Itemparameter des Quasi-Likelihood-Modells mit quadratischer Parametrisierung (Kapitel 4.2.2) geschätzt werden. Bei Ableitung der Momentschätzer wurde nämlich keine Annahme über die genaue Verteilungsform der Reaktionszeiten gemacht. Als einzige zusätzliche Bedingung muss gefordert werden, dass das dritte Moment der Residualkomponente Null ist, was bei symmetrischer Verteilung jedoch immer gilt.

Eine Auflösung der Momentgleichungen (6.22) und (6.23) ergab die vorläufigen Parameterschätzer  $\widehat{k}_g$  und  $\widehat{l}_g$ . Diese Werte wurden genutzt, um die Kovarianzmatrix der logarithmierten Reaktionszeiten (6.24) um die Komponente  $2\widehat{l}_g\widehat{l}_j + 4\widehat{l}_g\widehat{l}_j\widehat{d}_g\widehat{d}_j$  zu korrigieren. Die korrigierte Kovarianzmatrix wurde mit dem Programm R über die Prozedur FACTANAL faktorisiert. Anschließend dienten diese vorläufigen Parameterschätzer als Startwerte in einem nachgeschalteten EM-Algorithmus. Die approximative Berechnung der erwarteten Log-Likelihood-Funktion basierte dabei auf einer numerischen Quadratur mit 15 Quadraturpunkten pro Dimension. Der EM-Algorithmus wurde abgebrochen, wenn sich die marginale Log-Likelihood-Funktion bei 10000 Versuchspersonen weniger als 1 bzw. bei 1000 Versuchspersonen weniger als 0.1 verbesserte.<sup>4</sup>

Für alle generierten Datensätze konnten über die Momentmethode Startwerte bestimmt werden. Probleme mit dem letzten Schritt (Faktorisierung der korrigierten Kovarianzmatrix) traten nicht auf, obwohl dies sicherlich der kritischste Abschnitt der Momentschätzung war. Die in R implementierte Variante der Faktorenanalyse basiert auf dem instabilen Verfahren der Maximum-Likelihood-Schätzung, welches eine positiv definite Kovarianzmatrix benötigt und nicht immer zu einer adäquaten Lösung konvergiert. Das Konvergenzverhalten des Momentmethoden-Schätzers ist daher eine positive Überraschung. Aber auch der EM-Algorithmus konvergierte stabil gegen eine Lösung. Bei den durchgeführten Simulationen erwies sich der Simplex-Algorithmus als relativ robust, auch wenn er nicht immer das absolute Maximum der erwarteten Log-Likelihood-Funktion fand, sondern gelegentlich in einem lokalen Maximum landete. Allerdings wurde dieser Fehler in den folgenden Zyklen des EM-Schrittes wieder ausgeglichen, so dass letztendlich immer das absolute Maximum oder zumindest ein Maximum in der Nähe der wahren Parameter gefunden wurde. Bei 10000 Versuchspersonen wurden durchschnittlich 5.3 Iterationsschritte benötigt (max: 9). Die geringe Anzahl von Iterationsschritten liegt sicherlich daran, dass mit Hilfe der Momentmethode bereits gute Startwerte gefunden werden konnten – die Momentmethode sollte eine konsistente Schätzung der Itemparameter erlauben. Bei einem Datensatz aus 1000 Versuchspersonen stieg die Anzahl der Iterationen an: Durchschnittlich wurden nun 13.1 (max: 30) Schritte benötigt.

In Tabelle 6.7 und Tabelle 6.8 sind die Ergebnisse der Schätzung aufgeführt – die vorläufigen Schätzer der Momentmethode finden sich dabei immer auf der linken Seite, die des Marginal-Maximum-Likelihood-Ansatzes sind auf der rechten Seite aufgelistet. Diese Darstellung erlaubt einen Vergleich der beiden Schätzverfahren.

Bei großem Stichprobenumfang lassen sich die Itemparameter präzise bestimmen. Steht eine Stichprobe von 10000 Versuchspersonen zur Verfügung, liefert bereits die Momentmethode gute Schätzergebnisse. Dies ist umso positiver zu bewerten, da pro Momentschätzung weniger als eine Minute Rechenzeit benötigt wurde, die Schätzung nach dem EM-Verfahren immerhin mehrere Stunden dauerte. Betrachtet man die absoluten Abweichungen, verbessert der wesentlich aufwendigere EM-Ansatz die Itemparameterschätzung nur unwesent-

<sup>4</sup>Im letzten Iterationsschritt betrug die marginale Log-Likelihood-Funktion bei 10000 Versuchspersonen etwa -100000, bei 1000 Versuchspersonen etwa -10000. Die Abbruchkriterien entsprachen somit in etwa denselben relativen Veränderungen. Auch zeigte sich, dass im letzten Iterationsschritt kaum noch Veränderungen der Itemparameter auftraten.

Tabelle 6.7: Itemparameterschätzung bei log-normalverteilten Reaktionszeiten und quadratischer Parametrisierung – Ergebnisse der 35 Simulationen mit  $N=10000$

Methode	Momentmethode					EM-Algorithmus				
<b>Parameter a</b>										
Item	1	2	3	4	5	1	2	3	4	5
Wert	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Min	0.963	0.941	0.883	0.952	0.989	0.964	0.941	0.907	0.963	0.969
Max	1.121	1.109	1.059	1.105	1.103	1.074	1.107	1.045	1.081	1.092
Mean	1.055	1.010	0.989	1.006	1.040	1.030	1.007	0.996	1.003	1.023
STD	0.035	0.038	0.037	0.033	0.030	0.026	0.033	0.029	0.028	0.027
RMSE	0.065	0.039	0.038	0.034	0.049	0.040	0.034	0.029	0.028	0.035
<b>Parameter d</b>										
Item	1	2	3	4	5	1	2	3	4	5
Wert	-1.000	-0.500	0.000	0.500	1.000	-1.000	-0.500	0.000	0.500	1.000
Min	-1.041	-0.531	-0.039	0.460	0.941	-1.041	-0.537	-0.039	0.473	0.961
Max	-0.917	-0.438	0.049	0.541	1.036	-0.941	-0.453	0.047	0.533	1.046
Mean	-0.976	-0.498	0.001	0.499	0.991	-0.996	-0.503	0.001	0.502	0.998
STD	0.028	0.021	0.021	0.019	0.024	0.020	0.018	0.015	0.017	0.023
RMSE	0.036	0.021	0.021	0.019	0.026	0.021	0.019	0.016	0.017	0.023
<b>Parameter c</b>										
Item	1	2	3	4	5	1	2	3	4	5
Wert	0.250	0.250	0.500	0.500	0.500	0.250	0.250	0.500	0.500	0.500
Min	0.159	0.182	0.440	0.419	0.417	0.197	0.234	0.455	0.462	0.437
Max	0.329	0.302	0.547	0.604	0.583	0.304	0.302	0.555	0.544	0.548
Mean	0.236	0.254	0.503	0.496	0.493	0.250	0.264	0.510	0.499	0.494
STD	0.039	0.030	0.029	0.041	0.037	0.023	0.018	0.023	0.024	0.025
RMSE	0.042	0.030	0.029	0.041	0.038	0.023	0.023	0.026	0.024	0.026
<b>Parameter l</b>										
Item	1	2	3	4	5	1	2	3	4	5
Wert	-0.500	-0.500	-0.500	-0.250	-0.250	-0.500	-0.500	-0.500	-0.250	-0.250
Min	-0.523	-0.534	-0.547	-0.303	-0.278	-0.515	-0.517	-0.523	-0.274	-0.262
Max	-0.476	-0.466	-0.467	-0.208	-0.215	-0.480	-0.470	-0.465	-0.226	-0.233
Mean	-0.499	-0.499	-0.501	-0.251	-0.247	-0.494	-0.490	-0.492	-0.249	-0.248
STD	0.010	0.016	0.018	0.021	0.012	0.008	0.012	0.014	0.011	0.007
RMSE	0.010	0.016	0.018	0.021	0.012	0.010	0.016	0.016	0.011	0.008
<b>Parameter k</b>										
Item	1	2	3	4	5	1	2	3	4	5
Wert	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500
Min	0.423	0.462	0.459	0.445	0.428	0.429	0.459	0.460	0.466	0.471
Max	0.536	0.551	0.565	0.580	0.546	0.517	0.511	0.524	0.546	0.538
Mean	0.473	0.498	0.499	0.502	0.492	0.483	0.489	0.491	0.500	0.499
STD	0.031	0.019	0.020	0.028	0.027	0.016	0.011	0.015	0.017	0.016
RMSE	0.041	0.019	0.020	0.029	0.028	0.023	0.016	0.018	0.017	0.016
<b>Parameter <math>\sigma^2</math></b>										
Item	1	2	3	4	5	1	2	3	4	5
Wert	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Min	0.914	0.889	0.954	0.958	0.885	0.987	0.960	0.953	0.962	0.952
Max	1.190	1.060	1.043	1.064	1.109	1.064	1.035	1.051	1.057	1.061
Mean	1.055	0.995	0.994	1.002	1.016	1.024	1.004	0.998	1.003	1.010
STD	0.071	0.040	0.021	0.026	0.044	0.020	0.019	0.022	0.024	0.024
RMSE	0.089	0.041	0.022	0.026	0.047	0.031	0.019	0.022	0.024	0.026

Tabelle 6.8: Itemparameterschätzung bei log-normalverteilten Reaktionszeiten und quadratischer Parametrisierung – Ergebnisse der 35 Simulationen mit  $N=1000$

Methode	Momentmethode					EM-Algorithmus				
<b>Parameter a</b>										
Item	1	2	3	4	5	1	2	3	4	5
Wert	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Min	0.820	0.860	0.795	0.804	0.876	0.779	0.899	0.829	0.803	0.888
Max	1.347	1.254	1.152	1.313	1.566	1.206	1.199	1.190	1.210	1.205
Mean	1.086	1.013	0.983	1.018	1.074	1.019	1.011	0.986	1.006	1.021
STD	0.125	0.105	0.092	0.108	0.157	0.096	0.089	0.073	0.090	0.074
RMSE	0.151	0.106	0.094	0.110	0.174	0.098	0.090	0.074	0.090	0.077
<b>Parameter d</b>										
Item	1	2	3	4	5	1	2	3	4	5
Wert	-1.000	-0.500	0.000	0.500	1.000	-1.000	-0.500	0.000	0.500	1.000
Min	-1.103	-0.638	-0.125	0.352	0.820	-1.109	-0.633	-0.094	0.348	0.826
Max	-0.844	-0.401	0.150	0.636	1.168	-0.884	-0.406	0.164	0.702	1.125
Mean	-0.957	-0.492	0.000	0.498	0.985	-1.003	-0.505	0.007	0.503	1.000
STD	0.067	0.059	0.062	0.068	0.077	0.061	0.051	0.054	0.067	0.061
RMSE	0.080	0.059	0.062	0.068	0.078	0.061	0.052	0.055	0.067	0.061
<b>Parameter c</b>										
Item	1	2	3	4	5	1	2	3	4	5
Wert	0.250	0.250	0.500	0.500	0.500	0.250	0.250	0.500	0.500	0.500
Min	0.010	0.142	0.317	0.169	0.179	0.087	0.111	0.255	0.366	0.248
Max	0.545	0.569	0.686	1.118	0.847	0.385	0.431	0.665	1.011	0.732
Mean	0.226	0.267	0.505	0.541	0.507	0.266	0.289	0.507	0.517	0.501
STD	0.125	0.103	0.084	0.186	0.138	0.062	0.070	0.079	0.121	0.101
RMSE	0.128	0.104	0.085	0.191	0.138	0.065	0.080	0.079	0.122	0.101
<b>Parameter l</b>										
Item	1	2	3	4	5	1	2	3	4	5
Wert	-0.500	-0.500	-0.500	-0.250	-0.250	-0.500	-0.500	-0.500	-0.250	-0.250
Min	-0.599	-0.639	-0.644	-0.323	-0.286	-0.564	-0.547	-0.553	-0.310	-0.305
Max	-0.420	-0.415	-0.393	-0.011	-0.154	-0.401	-0.408	-0.403	-0.198	-0.205
Mean	-0.499	-0.501	-0.493	-0.222	-0.242	-0.490	-0.486	-0.490	-0.2244	-0.250
STD	0.040	0.050	0.054	0.076	0.035	0.033	0.036	0.036	0.025	0.021
RMSE	0.040	0.050	0.055	0.081	0.036	0.035	0.038	0.037	0.025	0.021
<b>Parameter k</b>										
Item	1	2	3	4	5	1	2	3	4	5
Wert	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500
Min	0.222	0.382	0.358	0.180	0.279	0.349	0.378	0.376	0.408	0.332
Max	0.674	0.675	0.682	0.655	0.675	0.578	0.609	0.601	0.602	0.605
Mean	0.463	0.505	0.496	0.475	0.480	0.491	0.493	0.496	0.505	0.503
STD	0.093	0.067	0.061	0.112	0.092	0.056	0.041	0.048	0.043	0.051
RMSE	0.100	0.067	0.061	0.115	0.094	0.057	0.042	0.048	0.044	0.052
<b>Parameter <math>\sigma^2</math></b>										
Item	1	2	3	4	5	1	2	3	4	5
Wert	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Min	0.756	0.661	0.799	0.006	0.677	0.869	0.856	0.846	0.166	0.681
Max	1.486	1.177	1.207	1.131	1.211	1.138	1.313	1.227	1.130	1.137
Mean	1.091	0.978	1.000	0.933	0.983	1.019	0.996	1.010	0.996	0.981
STD	0.173	0.117	0.087	0.210	0.143	0.056	0.054	0.072	0.175	0.094
RMSE	0.195	0.119	0.087	0.220	0.144	0.059	0.054	0.073	0.178	0.095

lich. Pro Itemparameter sinkt die Varianz der Itemparameterschätzer nur um den Betrag von etwa 0.020 ab. Allerdings ist der Effizienzgewinn beträchtlich: In der Regel ist die Varianz der EM-Schätzer nur halb so groß wie die Varianz der Momentmethoden-Schätzer. Interessanterweise tritt dieser Effekt bei allen Itemparametern auf. Dies ist überraschend, da es sich bei der Momentmethode ja nur teilweise um eine Schätzung anhand von Stichprobenmomenten handelt: Die Itemparameter  $a_g$  und  $d_g$  werden bei diesem Ansatz ebenfalls über einen Marginal-Maximum-Likelihood-Ansatz geschätzt. Dennoch profitiert die Schätzung dieser Parameter von dem nachgeschalteten EM-Algorithmus. Eine Berücksichtigung der Reaktionszeiten verbessert somit nicht nur die Schätzung der Personenfähigkeit, sondern auch die Bestimmung der Itemparameter des IRT-Modells. Interessant ist auch der Umstand, dass beim Momentmethoden-Ansatz die Parameter  $a_g$  und  $d_g$  mit ähnlicher Präzision geschätzt werden, wie die übrigen Parameter. Dies ist deshalb bemerkenswert, da die Schätzung der übrigen Parameter einerseits auf dem mit großer Schwankung behafteten, dritten Stichprobenmoment beruht, andererseits auf den fehlerbehafteten Schätzer  $\hat{a}_g$  und  $\hat{d}_g$  basiert. Dies schlägt sich jedoch nicht wesentlich in einem großen Standardfehler nieder.

Bei einer Stichprobe von 1000 Versuchspersonen fällt die Bilanz weniger positiv aus. Zwar finden sich keine Hinweis auf einen nennenswerten Bias, weder bei der Schätzung nach der Momentmethode noch bei der Schätzung über den Marginal-Maximum-Likelihood-Ansatz. Allerdings treten vereinzelt pathologische Schätzwerte auf, beispielsweise eine Merkmalsvarianz von 0.006 (der kleinste Schätzwert bei Item 4). Dies ist ein Zeichen, dass besonders der Momentmethoden-Schätzer bei kleinen Stichproben mit Vorsicht anzuwenden ist. Dennoch kann wiederum festgestellt werden, dass durch eine gemeinsame Schätzung der Itemparameter von RT- und IRT-Modell auch die Bestimmung der Itemparameter des zweiparametrischen Probit-Modells profitiert. Dies zeigt sich im Absinken der Varianz der Schätzwerte bei gemeinsamer Schätzung über den EM-Algorithmus.

Wenn auch der EM-Algorithmus zu besseren Schätzergebnissen führt, so ist dabei zu bedenken, dass der EM-Algorithmus auf restriktiveren Annahmen beruht als der Momentmethoden-Ansatz, der auch ohne die Annahme log-normalverteilter Reaktionszeiten zu konsistenten Schätzergebnissen führt. Auch ist der EM-Ansatz mit erheblichem Rechenaufwand verbunden (bis zu 8 Stunden bei einer Berechnung mit dem Softwarepaket R). Aussagen darüber, wie sich die Schätzverfahren in der Praxis bewähren könnten, wenn die Modellannahmen nicht vollständig erfüllt sind, lassen sich allerdings von den Simulationsergebnissen nur mit Einschränkung ableiten.

### 6.3 Diskussion

Ziel des aktuellen Kapitels war es, Schätzverfahren für die im vierten Kapitel beschriebenen Accelerated Failure Time Modelle zu entwickeln. Die Verfügbarkeit präziser Itemparameterschätzer ist eine zwingende Voraussetzung für eine Nutzung der Reaktionszeitmodelle zur Diagnostik.



Wie gezeigt werden konnte, ist es möglich, die Itemparameter der einzelnen Modelle relativ präzise zu bestimmen. Zwar sind teilweise gute Startwerte notwendig, jedoch können diese über einen vorgeschalteten Momentmethoden-Schätzer auf einfache Weise gefunden werden. Außerdem sind für eine gute Schätzung der Itemparameter große Stichproben notwendig. Bereits bei Stichproben von 1000 Personen treten teilweise recht hohe Abweichungen zwischen dem wahren Wert und dem Schätzwert eines Itemparameters auf. Dies ist allerdings keine Schwäche der Reaktionszeitmodelle allein. Bei dem Momentmethoden-Schätzer des Modells mit quadratischer Parametrisierung besteht der erste Schritt in einer separaten Analyse der Itemantworten mit einem Standardschätzverfahren für zweiparametrische Logit-Modelle. Dies ist aber genau das Schätzverfahren, das man ohne Berücksichtigung der Reaktionszeiten zur Auswertung des Tests herangezogen hätte. Bei einem Vergleich der Schätzleistung dieses ersten Schrittes mit den EM-Schätzern der entsprechenden Itemparameter bei gemeinsamer Schätzung aller Modellparameter stellt man fest, dass die Parameterschätzung durch die gemeinsame Berücksichtigung der Itemantworten und Reaktionszeiten verbessert wird. Die Notwendigkeit großer Stichproben beruht also nicht alleine auf dem zusätzlichen Einbezug der Reaktionszeiten.

Es ist nicht nur die Notwendigkeit großer Stichproben, welche die Itemparameterschätzung erschwert, sondern auch die Rechenintensität der Verfahren, zumindest wenn der EM-Algorithmus zum Einsatz kommt. Beim EM-Algorithmus wird der auf die Itemantworten und Reaktionszeiten bedingte Erwartungswert der Log-Likelihood-Funktion berechnet. Im vorherigen Kapitel zur Schätzung der Reaktionszeitmodelle für dichotomisierte Reaktionszeiten vereinfachte sich die Erwartungswertbildung dadurch, dass alle Personen mit demselben Antwortmuster denselben bedingten Erwartungswert erhalten konnten. Die Zahl der zu berechnenden bedingten Erwartungswerte entspricht dabei der Anzahl der verschiedenen Antwortmuster. Diese Vereinfachung ist bei Betrachtung der kontinuierlichen Reaktionszeiten nicht mehr möglich. Hier muss der bedingte Erwartungswert für jede Person gesondert bestimmt werden. Dadurch wird das Verfahren sehr rechenintensiv. Rechenzeiten von acht Stunden sind trotz guter Startwerte bei einem Stichprobenumfang von 10000 Personen keine Seltenheit. Allerdings könnte die Rechenzeit durch eine effizientere Implementierung des Algorithmus verkürzt werden.

## Kapitel 7

# Der Beitrag von Reaktionszeiten zur Fähigkeitsschätzung

Die Idee, die bei der Bearbeitung von psychologischen Tests aufgezeichneten Reaktionszeiten in die Diagnostik einzubeziehen, ist nicht neu. Ferrando (2006, S. 145) bemerkt: „If the response time is consistently related to certain item and respondent characteristics [...], then the additional information provided by this variable could be used to obtain more accurate estimates of the item and respondent parameters“. Auch Tyron und Mulloy (1993, S. 231) regen die Verwendung von Reaktionszeiten an: „Computer measured response time provides information that is methodologically independent from item content [...]. If such information is psychologically relevant, then the reliability and validity of objective personality tests could be improved by using both the scale scores and RT“. Eine konsequente Umsetzung dieses Gedankens ist bisher allerdings ausgeblieben.

Reaktionszeiten bei Testitems sind in der psychologischen Diagnostik auf vielfältige Weise nutzbar. Beispielsweise können „Speedmaße“ als eigenständige Größen neben den Skalenwerten zur Verhaltensprognose verwendet werden (Holden & Fekken, 1993; Siem, 1996; van den Wollenberg & Creemers, 1986). Wenn man aber davon ausgeht, dass Reaktionszeiten und Itemantworten von derselben Persönlichkeitseigenschaft, der Personenfähigkeit  $\theta$ , beeinflusst werden, dann kann man die Reaktionszeiten dazu nutzen, die Schätzung der Fähigkeitsausprägung einer Person zu verbessern. Die Möglichkeit, durch die Betrachtung der Reaktionszeit die Schätzung der Personenfähigkeit zu verbessern, wurde bisher jedoch nur selten untersucht, gegenwärtig liegen nur die Arbeiten zweier Autoren vor, die Veröffentlichungen von van der Linden (2008) sowie die von Ferrando und Lorenzo-Sevas (2007a) bzw. Ferrando und Lorenzo-Sevas (2007b). Das folgende Kapitel behandelt daher die Frage, welchen Nutzen die vorgeschlagenen Reaktionszeitmodelle für die Fähigkeitsschätzung besitzen.

## 7.1 Nutzen der RT für die Schätzung der Fähigkeit

Die psychologische Diagnostik versucht, die Eigenschaften von Personen aus deren Verhalten bei standardisierten Testaufgaben zu erschließen. Grundlage vieler diagnostischer Verfahren sind Verhaltensmodelle, welche das Verhalten einer Person als Funktion bestimmter Persönlichkeitseigenschaften modellieren. Anhand dieser Modelle kann dann versucht werden, vom beobachteten Verhalten auf die zugrunde liegenden Eigenschaften zu schließen. Im Verlauf der aktuellen Arbeit wurden mehrere Modelle für Reaktionszeiten und Itemantworten vorgestellt. Bei diesen Modellen hing die gemeinsame Verteilung der Itemantworten und Reaktionszeiten von zwei Persönlichkeitseigenschaften, der Personenfähigkeit  $\theta$  und dem Residualfaktor  $\omega$ , ab. Angenommen wurde, dass die Itemantworten und Reaktionszeiten einer Person unabhängig voneinander verteilt sind, wenn auf die Persönlichkeitseigenschaften  $\theta$  und  $\omega$  als einzige systematische Ursachen bedingt wird. Beobachtet man die Itemantworten  $\mathbf{x} = [x_1, \dots, x_G]$  und die Reaktionszeiten  $\mathbf{t} = [t_1, \dots, t_G]$  einer Person bei  $G$  Aufgaben, gilt für die gemeinsame Verteilung dieser Größen somit:

$$f(\mathbf{t}, \mathbf{x}|\theta, \omega) = \left[ \prod_{g=1}^G f(t_g|\theta, \omega) \right] \left[ \prod_{g=1}^G f(x_g|\theta) \right]. \quad (7.1)$$

Für die Verteilungen  $f(t_g|\theta, \omega)$  und  $f(x_g|\theta)$  können Modelle aus dem dritten und vierten Kapitel gewählt werden. Steht eine Eichstichprobe zu Verfügung, lassen sich die Itemparameter mit den im fünften und sechsten Kapitel beschriebenen Verfahren schätzen. Fasst man die geschätzten Itemparameter als wahre Itemparameter auf, kann anhand des Antwort- und Reaktionszeitmusters einer Person deren Ausprägung auf der Fähigkeits- und Residualfaktordimension über die Maximum-Likelihood-Methode bestimmt werden. Während im dritten und vierten Kapitel die Fähigkeitsschätzung anhand der Itemantworten oder der Reaktionszeiten getrennt voneinander betrachtet wurde, werden in diesem Kapitel die Vorteile einer gemeinsamen Schätzung analysiert.

Bei der Maximum-Likelihood-Schätzung der Personenparameter werden diese als „fixe“ Verteilungsparameter aufgefasst, ungeachtet der Tatsache, dass Personenparameter eigentlich Realisierungen von Zufallsvariablen sind. Betrachtet wird dabei die auf die Personenparameter bedingte Verteilung der Beobachtungen. Die Log-Likelihood-Gleichung einer Person mit einem bestimmten Reaktionszeit- und Antwortmuster bei  $G$  Aufgaben ist:

$$\text{LL}(\theta, \omega) = \sum_{g=1}^G \log(f(t_g|\theta, \omega)) + \sum_{g=1}^G \log(f(x_g|\theta)). \quad (7.2)$$

Die Maximum-Likelihood-Schätzer der Persönlichkeitseigenschaften  $\theta$  und  $\omega$  sind diejenigen Werte  $\hat{\theta}$  und  $\hat{\omega}$ , welche die Log-Likelihood-Funktion bei den gegebenen Beobachtungsdaten  $[\mathbf{x}, \mathbf{t}]$  maximieren. Möchte man die Präzision der Eigenschaftsschätzung bestimmen, ist die Varianz der Schätzer von Interesse.

Da besonders die Personenfähigkeit  $\theta$  im Interesse des diagnostischen Prozesses steht, sei die folgende Diskussion auf die Varianz des Fähigkeitsschätzers  $\hat{\theta}$  beschränkt. Diese Varianz lässt sich mit Hilfe der Fisher-Information des Tests berechnen.

Die Fisher-Information des Tests setzt sich aus den Fisher-Informationen der Items zusammen. Diese wurden für die einzelnen Reaktionszeit- und Item-Response-Modelle im dritten und vierten Kapitel berechnet. Aufgrund der bedingten Unabhängigkeit der Itemantworten und Reaktionszeiten eines Tests können die Fisher-Informationen der Reaktionszeiten und der Itemantworten summiert werden. Da es sich um ein zweiparametriges Modell handelt, ergibt sich eine Matrix:

$$\begin{aligned} \mathbf{I}_{\mathbf{x},\mathbf{t}}(\theta, \omega) = & \sum_{g=1}^G \begin{bmatrix} -E\left[\frac{\partial^2}{\partial\theta^2} \log f(x_g|\theta)\right] & 0 \\ 0 & 0 \end{bmatrix} \\ & + \sum_{g=1}^G \begin{bmatrix} -E\left[\frac{\partial^2}{\partial\theta^2} \log f(t_g|\theta, \omega)\right] & -E\left[\frac{\partial^2}{\partial\theta\partial\omega} \log f(t_g|\theta, \omega)\right] \\ -E\left[\frac{\partial^2}{\partial\omega\partial\theta} \log f(t_g|\theta, \omega)\right] & -E\left[\frac{\partial^2}{\partial\omega^2} \log f(t_g|\theta, \omega)\right] \end{bmatrix}. \end{aligned} \quad (7.3)$$

Wie unmittelbar ersichtlich ist, handelt es sich bei der Informationsmatrix nicht um eine Diagonalmatrix, so dass die Interpretation nicht allein auf die Diagonale beschränkt bleiben kann. Im Vergleich zu einer Nichtbeachtung der Reaktionszeiten  $\mathbf{t}$  und einer Schätzung der Personenfähigkeit  $\theta$  allein aufgrund der Itemantworten  $\mathbf{x}$  kommt es aber durch die Berücksichtigung der Reaktionszeiten auf alle Fälle zu einer Reduktion der Varianz von  $\hat{\theta}$ . Dies zeigen einfache Überlegungen. Die asymptotische Kovarianzmatrix von Maximum-Likelihood-Schätzern kann über die Inverse der Informationsmatrix berechnet werden – siehe Gleichung (2.15) und Gleichung (2.16). Diese Beziehungen gelten bei Annahme bestimmter Regularitätsbedingungen auch im Falle nicht identisch verteilter Zufallsvariablen (Bradley & Gart, 1962). Die Inverse der Matrix (7.3) lässt sich symbolisch schreiben als

$$\mathbf{I}_{\mathbf{x},\mathbf{t}}^{-1}(\theta, \omega) = \begin{bmatrix} a_1 + a_2 & c \\ c & d \end{bmatrix}^{-1}, \quad (7.4)$$

wobei das Element  $a_1 = \sum_{g=1}^G -E\left[\frac{\partial^2}{\partial\theta^2} \log f(x_g|\theta)\right]$  aus dem ersten Summanden von Gleichung (7.3) stammt und die übrigen Elemente die entsprechenden Elemente des zweiten Summanden von Gleichung (7.3) sind. Die Varianz des Maximum-Likelihood-Schätzers der Personenfähigkeit  $\hat{\theta}$  findet sich in dem zu  $\theta$  gehörenden Diagonalelement  $I_{\mathbf{x},\mathbf{t}[11]}^{-1}(\theta, \omega)$  der inversen Informationsmatrix  $\mathbf{I}_{\mathbf{x},\mathbf{t}}^{-1}(\theta, \omega)$ . Eine Inversion der Informationsmatrix ergibt für dieses Element:

$$I_{\mathbf{x},\mathbf{t}[11]}^{-1}(\theta, \omega) = \frac{d}{(a_1 + a_2)d - c^2} = \frac{1}{a_1 + (a_2 - \frac{c^2}{d})}. \quad (7.5)$$

Die Größe  $1/a_1 = 1/\sum_{g=1}^G [-E\left[\frac{\partial^2}{\partial\theta^2} \log f(x_g|\theta)\right]]$  entspricht genau der Varianz eines Maximum-Likelihood-Schätzers der Personenfähigkeit  $\hat{\theta}$  bei einer

Schätzung anhand der Itemantworten alleine, also bei Schätzung ohne Beachtung der Reaktionszeiten. Es kann nun gezeigt werden, dass durch Berücksichtigung der Reaktionszeiten die Varianz des Fähigkeitsschätzers  $\hat{\theta}$  zwangsläufig abnimmt. Beim zweiten Summanden von Gleichung (7.3) handelt es sich um die Informationsmatrix der Reaktionszeiten, woraus geschlossen werden kann, dass diese Matrix positiv definit ist, mit positiven Diagonalelementen. Somit gilt nach dem Kriterium von Sylvester die Ungleichung  $a_2d - c^2 > 0$  (Searle, 1982, S. 205). Aus dieser Beziehung folgt  $a_2 > c^2/d$ , so dass  $\Delta = a_2 - c^2/d > 0$ . Dies führt zu  $1/a_1 > 1/(a_1 + \Delta)$ . Die Varianz des Maximum-Likelihood-Schätzers der Personenfähigkeit  $\hat{\theta}$  sinkt bei Berücksichtigung der Reaktionszeiten somit auf alle Fälle gegenüber einer alleinigen Schätzung aufgrund der Itemantworten, selbst wenn zusätzlich ein zweiter Personenparameter  $\omega$  beteiligt ist. Allerdings fällt aufgrund der Notwendigkeit der Schätzung von  $\omega$  die Varianzreduktion geringer aus als bei einem eindimensionalen Reaktionszeitmodell.

Tatsächlich lässt sich die durch die Berücksichtigung der Reaktionszeiten erzielte Varianzreduktion auch mit einem zweiten Argument begründen, welches eine weitere Eigenschaft der Fähigkeitsschätzung verdeutlicht. Schätzt man die Fähigkeit allein aufgrund der Itemantworten  $\mathbf{x}$  einer Person, erhält man einen ersten Schätzer  $\hat{\theta}_{\mathbf{x}}$  mit Varianz  $1/a_1$ , wobei  $a_1$  die Fisher-Information der Itemantworten aus Matrix (7.3) ist. Berücksichtigt man lediglich die Reaktionszeiten  $\mathbf{t}$  zur Fähigkeitsschätzung, erhält man einen Schätzer  $\hat{\theta}_{\mathbf{t}}$  mit Varianz  $d/(a_2d - c^2)$ . Diese Varianz ist das entsprechende Element der Inversen der Fisher-Informationsmatrix der Reaktionszeiten, siehe Gleichung (7.3). Die beiden Schätzer können mit der Linearkombination  $\hat{\theta}_p = k_x \hat{\theta}_{\mathbf{x}} + k_t \hat{\theta}_{\mathbf{t}}$  zusammengefasst werden, wobei  $k_x$  und  $k_t$  Gewichte sind mit der Restriktion  $k_x + k_t = 1$ . Handelt es sich bei  $\hat{\theta}_{\mathbf{x}}$  und  $\hat{\theta}_{\mathbf{t}}$  um konsistente Schätzer, ist der neue Schätzer  $\hat{\theta}_p$  bei einer derartigen Gewichtung ebenfalls konsistent (Greene, 2003, S. 904). Optimalerweise wählt man diejenigen Gewichte  $k_x$  und  $k_t$ , die zu minimaler Varianz von  $\hat{\theta}_p$  führen. Diese Gewichte hängen von der Varianz der Summanden  $\hat{\theta}_{\mathbf{x}}$  und  $\hat{\theta}_{\mathbf{t}}$  ab und können, wenn man die Unabhängigkeit der beiden Schätzer ausnützt, einfach berechnet werden – siehe hierfür zum Beispiel Goldberger (1998, S. 141). Beim aktuellen Beispiel sind diese optimalen Gewichte:

$$k_{\mathbf{x}} = \frac{d/(a_2d - c^2)}{1/a_1 + d/(a_2d - c^2)}$$

$$k_{\mathbf{t}} = \frac{1/a_1}{1/a_1 + d/(a_2d - c^2)}.$$

Die Komponenten  $a_1$  bis  $d$  stehen für die entsprechenden Elemente aus Matrix (7.3). Die bei dieser Gewichtung resultierende Varianz des gemittelten Schätzers  $\hat{\theta}_p$  ist dann:

$$\begin{aligned}
 \sigma^2(\widehat{\theta}_p) &= k_x^2 \sigma^2(\widehat{\theta}_x) + k_t^2 \sigma^2(\widehat{\theta}_t) \\
 &= \left[ \frac{d/(a_2 d - c^2)}{1/a_1 + d/(a_2 d - c^2)} \right]^2 \frac{1}{a_1} + \left[ \frac{1/a_1}{1/a_1 + d/(a_2 d - c^2)} \right]^2 \frac{d}{(a_2 d - c^2)} \\
 &= \frac{d}{(a_1 + a_2)d - c^2}.
 \end{aligned} \tag{7.6}$$

Dies entspricht jedoch genau dem Element  $I_{\mathbf{x}, \mathbf{t}[11]}^{-1}(\theta, \omega)$  der inversen Fisher-Informationsmatrix bei gemeinsamer Berücksichtigung der Reaktionszeiten und Itemantworten (7.5). Die resultierende Varianz (7.5) kann somit alternativ als Varianz einer optimal gewichteten Zusammenfassung zweier Schätzer aufgefasst werden. Es ist leicht verständlich, dass diese Varianz höchstens so groß sein kann wie die kleinste Einzelvarianz eines der beiden Schätzer. Dies ist ein weiterer Beleg dafür, dass durch die Berücksichtigung der Reaktionszeiten die Varianz des Fähigkeitsschätzers reduziert wird.

Feststellen lässt sich somit Folgendes: Durch die zusätzliche Berücksichtigung der Reaktionszeiten reduziert sich die Varianz bei der Schätzung der Personenfähigkeit  $\theta$  zwangsläufig. Diese Varianzreduktion ist garantiert und tritt selbst dann ein, wenn die Reaktionszeiten alleine keine präzise Schätzung der Personenfähigkeit erlauben und somit durch die Berücksichtigung der Reaktionszeiten eine sehr unsichere Datenquelle hinzugenommen wird. Allerdings fällt in diesem Fall die Varianzreduktion nur gering aus. Interessanterweise lässt sich eine Varianzreduktion trotz des Umstandes realisieren, dass durch die Betrachtung der Reaktionszeiten eine weitere Persönlichkeitseigenschaft  $\omega$  eingeführt wird und somit eine zweite Größe zu schätzen ist. Wie anhand Gleichung (7.5) zu sehen ist, verringert dies zwar den Beitrag der Reaktionszeiten zur Varianzreduktion, die Varianz des Schätzers wird dennoch reduziert. Wenn also von einer Varianzreduktion ausgegangen werden kann, stellt sich die Frage, wie groß diese bei einem Test in der Praxis ausfällt. Dies soll im nächsten Abschnitt untersucht werden.

### Analyse der tatsächlichen Varianzreduktion

Die Gleichungen (7.3) und (7.4) geben allgemein an, wie die (approximative) Varianz des Fähigkeitsschätzers  $\widehat{\theta}$  bei einem Test zu berechnen ist. Ein erster Bestandteil von Gleichung (7.3) ist die Informationsfunktion des zweiparametrischen Logit-Modells, die im dritten Kapitel in Gleichung (3.6) dargestellt wurde. Je nach gewähltem Reaktionszeitmodell müssen jedoch unterschiedliche Informationsmatrizen für den zweiten Summanden von Gleichung (7.3) eingesetzt werden. Die Informationsmatrizen der verschiedenen Reaktionszeitmodelle wurden bereits im vierten Kapitel berechnet, so dass diese Ergebnisse als bekannt vorausgesetzt werden können. Im Folgenden soll für die vier im fünften und sechsten Kapitel beschriebenen Testmodelle untersucht werden, mit welcher Varianz bei diesen Modellen die Fähigkeit geschätzt werden kann, wenn man die Reaktionszeiten berücksichtigt und mit welcher Varianz, wenn man die Reaktionszeiten ignoriert.

**PH-Modell mit linearer Parametrisierung**

In Kapitel 5.1 wurde ein Proportional Hazard Modell für dichotomisierte Zeiten mit linearer Parametrisierung betrachtet. Für dieses Modell wurde die Itemparameterschätzung bei einem Test aus fünf Items simuliert. Die Itemparameter der fünf Testaufgaben finden sich in Tabelle 7.1.

Tabelle 7.1: Dichotomisierte RT und lineare Parametrisierung: Itemparameter eines Tests aus fünf Aufgaben

Item	$a_g$	$b_g$	$c_g$	$l_g$	$k_g$
1	1.50	0.66	0.50	0.20	-0.39
2	1.50	0.33	0.50	0.20	-0.29
3	1.50	0.00	0.30	0.30	-0.19
4	1.50	-0.33	0.20	0.50	-0.09
5	1.50	-0.66	0.20	0.50	0.01

Die Varianz bei der Schätzung der Fähigkeit ohne Berücksichtigung der Reaktionszeiten erhält man über die Fisher-Information der Itemantworten des zweiparametrischen Logit-Modells gemäß Beziehung (3.7). Die Informationsmatrix der Reaktionszeiten wurde für dieses Modell im vierten Kapitel berechnet, siehe dort Gleichung (4.22). Kombiniert man die Information der Reaktionszeiten mit der Information der Itemantworten und invertiert man die Informationsmatrix, ergibt sich die Varianz des Fähigkeitsschätzers bei Verwendung der Itemantworten und der Reaktionszeiten. Die bei einem Test aus fünf Items mit Itemparametern gemäß Tabelle 7.1 resultierende Varianz des Fähigkeitsschätzers  $\hat{\theta}$  ist in Abbildung 7.1 wiedergegeben. In dieser Abbildung ist die Varianz von  $\hat{\theta}$  bei Berücksichtigung und ohne Berücksichtigung der Reaktionszeiten für die Kombination verschiedener Ausprägungen der Personenfähigkeit  $\theta$  und des Residualfaktors  $\omega$  dargestellt. Die minimale Varianz ist bei Berücksichtigung der Reaktionszeiten 0.37, ohne Berücksichtigung der Reaktionszeiten 0.40. Dies zeigt, dass die Berücksichtigung der Reaktionszeiten wenig zur Fähigkeitsschätzung beiträgt, wenn sich die zu schätzenden Ausprägungen der Eigenschaften im optimalen Diagnosebereich des Tests befinden. Wie bereits in den vorherigen Kapiteln angedeutet wurde, profitiert die Fähigkeitsschätzung von der Berücksichtigung der Reaktionszeiten hauptsächlich bei extremen Fähigkeitsausprägungen, welche sich alleine aufgrund der Itemantworten nur schlecht schätzen lassen. In diesen Bereichen ist der Nutzen der Reaktionszeiten jedoch beträchtlich.

**PH-Modell mit quadratischer Parametrisierung**

Im Einstellungsbereich sind Unfolding-Beziehungen zwischen Personen und Items plausibler als Dominanzbeziehungen. Diese Art der Abhängigkeit kann mit einem Proportional Hazard Modell für dichotomisierte Reaktionszeiten und quadratischer Parametrisierung modelliert werden. Die Schätzung eines solchen Modells wurde in Kapitel 5.2 anhand eines Tests aus fünf Items erprobt. Die dabei verwendeten Itemparameter finden sich in Tabelle 7.2.

Die Fisher-Information des Proportional Hazard Modells für dichotomisierte

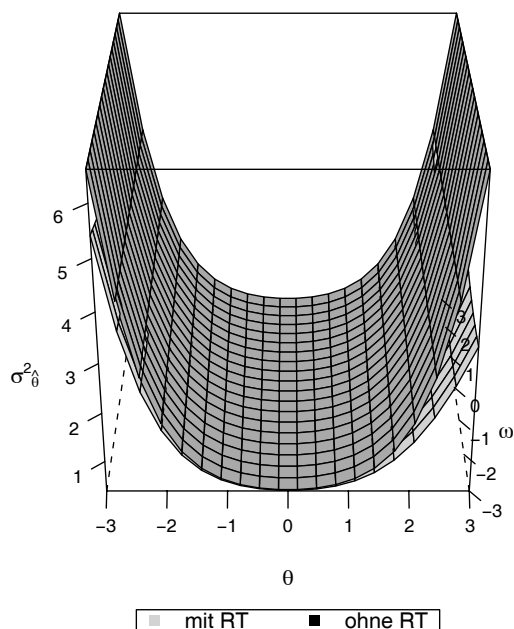


Abbildung 7.1: Dichotomisierte RT: Varianz des Fähigkeitsschätzers  $\hat{\theta}$  bei einem Test aus 5 Items mit Itemparametern gemäß Tabelle 7.1 und linearer Parametrisierung – Kontrastierung der Varianz mit Berücksichtigung (hellgraue Fläche) und ohne Berücksichtigung (dunkelgraue Fläche) der Reaktionszeiten

Reaktionszeiten und quadratischer Parametrisierung wurde im vierten Kapitel in Gleichung (4.26) angegeben. Kombiniert man die Fisher-Information der Reaktionszeiten mit der Fisher-Information des zweiparametrischen Logit-Modells, kann die Varianz des Fähigkeitsschätzers bei Schätzung anhand der Itemantworten und Reaktionszeiten berechnet werden. In Abbildung 7.2 ist die Varianz des Schätzers  $\hat{\theta}$  abgetragen, die sich für einen Test aus fünf Items mit Itemparametern gemäß Tabelle 7.2 bei Berücksichtigung der Reaktionszeiten und ohne Berücksichtigung der Reaktionszeiten ergibt. Wie der Abbildung entnommen werden kann, verbessert sich die Schätzung der Personenfähigkeit durch die Berücksichtigung der Reaktionszeiten hauptsächlich bei extremen Fähigkeitsausprägungen. Insgesamt ist der Nutzen bei diesem Modell jedoch begrenzt.

**AFT-Modell mit linearer Parametrisierung**

Als Alternative zu den Proportional Hazard Modellen bieten sich Accelerated Failure Time Modelle an. Ein spezieller Vertreter dieser Modellklasse mit lognormalverteilten Reaktionszeiten und linearer Parametrisierung wurde in Kapitel 6.1 vorgestellt. Die Anpassung eines derartigen Modells an Beobachtungsdaten wurde anhand eines Tests aus fünf Items simuliert. Die bei der Simulation angenommenen Itemparameter der fünf Items sind in Tabelle 7.3 aufgelistet.



Tabelle 7.2: Dichotomisierte RT und quadratische Parametrisierung: Itemparameter eines Tests aus fünf Aufgaben

Item	$a_g$	$d_g$	$c_g$	$l_g$	$k_g$
1	1.50	-0.66	0.20	0.10	0.00
2	1.50	-0.33	0.20	0.10	0.00
3	1.50	0.00	0.20	0.10	0.00
4	1.50	0.33	0.20	0.10	0.00
5	1.50	0.66	0.20	0.10	0.00

Tabelle 7.3: Log-normalverteilte RT und lineare Parametrisierung: Itemparameter eines Tests aus fünf Aufgaben

Item	$a_g$	$b_g$	$c_g$	$l_g$	$k_g$	$\sigma_g$
1	1.50	-1.00	0.20	0.50	1.50	1.00
2	1.50	-0.50	0.20	0.50	1.50	1.00
3	1.50	0.00	0.30	0.30	1.50	1.00
4	1.50	0.50	0.50	0.20	1.50	1.00
5	1.50	1.00	0.50	0.20	1.50	1.00

Die Berechnung der Varianz des Fähigkeitsschätzers erfordert die Kenntnis der Informationsmatrix der Reaktionszeiten. Beim aktuellen Modell entspricht diese Matrix der Matrix (4.36). Kombiniert man die Information der Reaktionszeiten mit der Information der Itemantworten, kann die Varianz von  $\hat{\theta}$  berechnet werden, die sich bei gleichzeitiger Berücksichtigung der Reaktionszeiten und der Itemantworten ergibt. Die bei einem Test aus fünf Items mit Itemparametern laut Tabelle 7.3 resultierende Varianz des Fähigkeitsschätzers  $\hat{\theta}$  ist in Abbildung 7.3 wiedergegeben. Wiederum trägt die Berücksichtigung der Reaktionszeiten vor allem zur Schätzung extremer Fähigkeiten bei. Für diese Fähigkeitsausprägungen wird jedoch die Varianz der Schätzung erheblich reduziert.

#### AFT-Modell mit quadratischer Parametrisierung

Im sechsten Kapitel wurde ein zweites Reaktionszeitmodell aus der Klasse der Accelerated Failure Time Modelle betrachtet. Bei diesem Modell stand die Personenfähigkeit in quadratischer Beziehung zum Lokationsparameter einer Log-Normalverteilung. Die Itemparameterschätzung wurde für dieses Modell anhand eines Tests aus fünf Items simuliert. Die dabei verwandten Itemparameter finden sich in Tabelle 7.4.

Die Informationsmatrix der Reaktionszeiten ergibt sich bei diesem Modell nach Beziehung (4.40). Die Kombination dieser Matrix mit der Testinformation der Itemantworten ermöglicht es, die Varianz des Fähigkeitsschätzers bei Verwendung der Itemantworten und der Reaktionszeiten zu berechnen. Die Varianz des Schätzers  $\hat{\theta}$ , die sich bei diesem Modell für einen Test aus fünf Items mit Itemparameter gemäß Tabelle 7.4 ergibt, ist für verschiedene Ausprägungen der

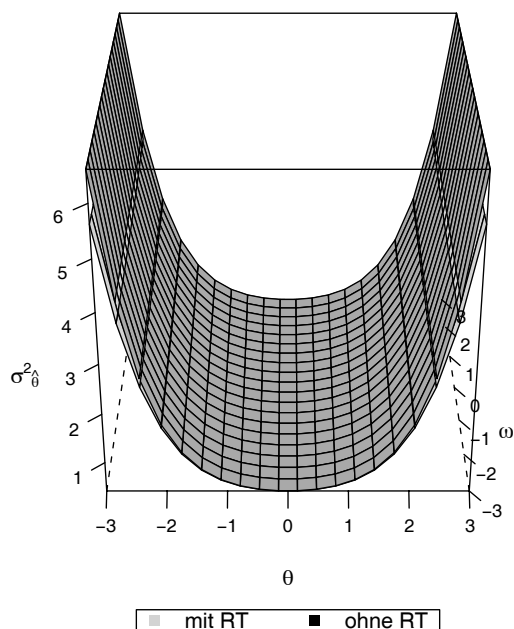


Abbildung 7.2: Dichotomisierte RT: Varianz des Fähigkeitsschätzers  $\hat{\theta}$  bei einem Test aus 5 Items mit Itemparametern gemäß Tabelle 7.2 und quadratischer Parametrisierung – Kontrastierung der Varianz mit Berücksichtigung (hellgraue Fläche) und ohne Berücksichtigung (dunkelgraue Fläche) der Reaktionszeiten

wahren Personenfähigkeit in Abbildung 7.4 wiedergegeben. Dort wird der Fall einer Schätzung mit und ohne Berücksichtigung der Reaktionszeiten unterschieden. Die Berücksichtigung der Reaktionszeiten verbessert hauptsächlich die Bestimmung extremer Fähigkeitsausprägungen. Der Nutzen der Reaktionszeiten ist jedoch erheblich: Durch den zusätzlichen Einbezug der Reaktionszeiten kann eine Obergrenze für die Varianz der Fähigkeitsschätzung garantiert werden, die für alle Fähigkeitsausprägungen gilt. Dies folgt aus dem sich ergänzenden Verhalten der Informationsmatrizen der Reaktionszeiten und Itemantworten, siehe hierzu die Ausführungen in Kapitel 4.2.1.

### Bewertung des Beitrags der RT

Die Frage nach der Varianzreduktion bei Berücksichtigung der Reaktionszeiten ist natürlich auch eine empirische Frage, die theoretisch nicht vollständig beantwortet werden kann. Der Beitrag der Reaktionszeiten hängt schließlich von den Itemparametern der Reaktionszeitmodelle ab. Dennoch können mehrere generelle Bemerkungen gemacht werden. Betrachtet man die Iteminformationsfunktionen der einzelnen Reaktionszeitmodelle im vierten Kapitel, so stellt man fest, dass sich diese Funktionen bei den meisten Reaktionszeitmodellen komple-

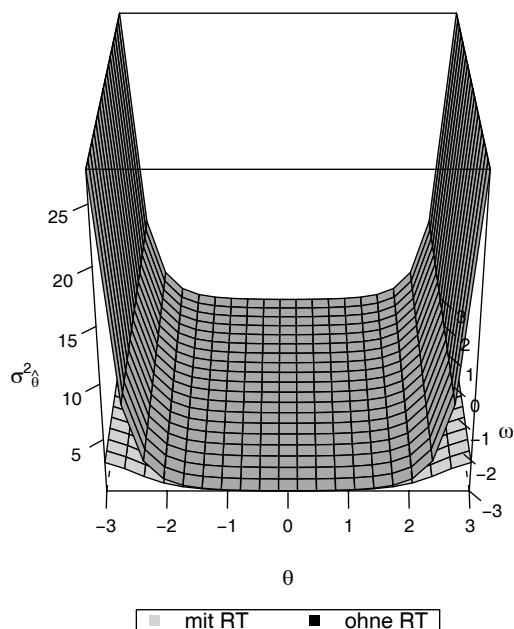


Abbildung 7.3: Log-normalverteilte RT: Varianz des Fähigkeitsschätzers  $\hat{\theta}$  bei einem Test aus 5 Items mit Itemparametern gemäß Tabelle 7.3 und linearer Parametrisierung – Kontrastierung der Varianz mit Berücksichtigung (hellgraue Fläche) und ohne Berücksichtigung (dunkelgraue Fläche) der Reaktionszeiten

mentär zur Iteminformationsfunktion des zweiparametrischen Logit-Modells verhalten: Oft sind die Informationsfunktionen der Reaktionszeitmodelle in den Wertebereichen der Personenfähigkeit groß, in denen die Iteminformation des zweiparametrischen Logit-Modells gering ausfällt. Diese Eigenschaft schlägt sich in der Varianzreduktion bei Einbezug der Reaktionszeiten nieder. Betrachtet man die Abbildungen 7.1 bis 7.4, stellt man fest, dass eine größere Varianzreduktion hauptsächlich für extreme Fähigkeitsausprägungen eintritt. Dies ist aber genau der Bereich, in denen zweiparametrische Logit-Modelle nur geringe Iteminformation und folglich auch nur geringe diagnostische Präzision besitzen. Eine Berücksichtigung der Reaktionszeiten dürfte somit hauptsächlich bei den Personen vorteilhaft sein, für die ein Test entweder zu leicht oder zu schwer ist.

Insgesamt fällt der Nutzen der Proportional Hazard Modelle für dichotomisierte Reaktionszeiten geringer aus als der Nutzen der Accelerated Failure Time Modelle. Dies liegt daran, dass bei den Proportional Hazard Modellen für dichotomisierte Reaktionszeiten aufgrund der Dichotomisierung der Reaktionszeiten ein Informationsverlust eintritt. Dieses Phänomen ist hinreichend bekannt (Schader & Schmid, 1988) und wurde bereits im vierten Kapitel angesprochen. Wie drastisch der Informationsverlust jedoch für einen konkreten Test ausfällt, wird allerdings erst anhand der Abbildungen 7.1 und 7.2 deutlich. Es ist

Tabelle 7.4: Log-normalverteilte RT und quadratische Parametrisierung: Itemparameter eines Tests aus fünf Aufgaben

Item	$a_g$	$d_g$	$c_g$	$l_g$	$k_g$	$\sigma_g$
1	1.00	-1.00	0.25	-0.50	0.50	1.00
2	1.00	-0.50	0.25	-0.50	0.50	1.00
3	1.00	0.00	0.50	-0.50	0.50	1.00
4	1.00	0.50	0.50	-0.25	0.50	1.00
5	1.00	1.00	0.50	-0.25	0.50	1.00

in der Tat eine etwas paradoxe Vorgehensweise, zuerst gegen Ende des dritten Kapitels für die Reaktionszeiten als kontinuierliches Maß der Personenfähigkeit zu werben, um dann im vierten Kapitel dieses kontinuierliche Maß zu dichotomisieren. Der Vorteil der Proportional Hazard Modelle für dichotomisierte Reaktionszeiten ist jedoch deren Flexibilität, die sogar unterschiedliche Reaktionszeitverteilungen für die verschiedenen Items des Tests zulässt. Dies ist bei anderen semiparametrischen Schätzansätzen für Proportional Hazard Modelle wie zum Beispiel dem Partial Likelihood Ansatz von Cox (1972) nicht möglich. Zudem kann der Informationsverlust begrenzt werden, wenn anstelle dichotomisierter Reaktionszeiten die Reaktionszeiten in mehrere Intervalle eingeteilt werden. Dies vergrößert etwas den Aufwand der Itemparameterschätzung, ist aber prinzipiell möglich.

## 7.2 Nutzen der RT für die Schätzung der Fähigkeit bei obliquen Faktorstruktur

Am Anfang des zweiten Teils der Arbeit wurden Überlegungen angestellt, wie die gemeinsame Verteilung der Itemantworten, Reaktionszeiten und Persönlichkeitseigenschaften über eine Population möglicher Testteilnehmer aussehen könnte. Bei Annahme der bedingten Unabhängigkeit von Reaktionszeiten und Itemantworten und bei Verwendung eines eindimensionalen IRT-Modells führten diese Überlegungen zur Beziehung:

$$f(\mathbf{t}, \mathbf{x}, \theta, \omega) = \left[ \prod_{g=1}^G f(t_g | \theta, \omega) \right] \left[ \prod_{g=1}^G f(x_g | \theta) \right] f(\theta, \omega). \quad (7.7)$$

Tatsächlich wurde bei der Itemparameterschätzung davon ausgegangen, dass die Persönlichkeitseigenschaften  $\theta$  und  $\omega$  bivariat standardnormalverteilt sind mit einer Korrelation von  $\rho = 0$ . Bei Modell (7.7) handelt es sich in diesem Fall um ein orthogonales Faktormodell, mit orthogonalen latenten Variablen  $\theta$  und  $\omega$ . Da jedoch das IRT-Modell nur auf einem Faktor lädt, können die Faktoren nicht rotiert werden, wenn die Eindimensionalität des IRT-Modells nicht aufgegeben werden soll. Die Verwendung orthogonaler Faktoren besitzt bei der gemeinsamen Modellierung von Itemantworten und Reaktionszeiten einige Popularität (Ferrando & Lorenzo-Sevas, 2007a; Thissen, 1984). Eine orthogonale

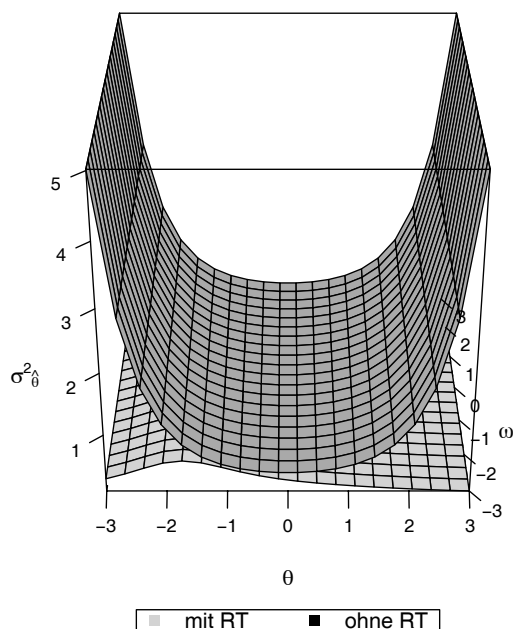


Abbildung 7.4: Log-normalverteilte RT: Varianz des Fähigkeitsschätzers  $\hat{\theta}$  bei einem Test aus 5 Items mit Itemparametern gemäß Tabelle 7.4 und quadratischer Parametrisierung – Kontrastierung der Varianz mit Berücksichtigung (hellgraue Fläche) und ohne Berücksichtigung (dunkelgraue Fläche) der Reaktionszeiten

Faktorstruktur ist jedoch nicht konkurrenzlos. Von van der Linden (2007) wurde ein obliques Faktormodell mit zwei korrelierten Faktoren vorgeschlagen, bei welchem ein Faktor nur auf den Itemantworten und der andere Faktor nur auf den Reaktionszeiten ladet. Diese Annahme führt zur alternativen Beziehung

$$f(\mathbf{t}, \mathbf{x}, \theta, \omega) = \left[ \prod_{g=1}^G f(t_g|\omega) \right] \left[ \prod_{g=1}^G f(x_g|\theta) \right] f(\theta, \omega), \quad (7.8)$$

wobei die oblique Faktorstruktur dadurch erzeugt wird, dass für  $f(\theta, \omega)$  eine bivariate Standardnormalverteilung mit einem Korrelationskoeffizienten  $\rho \neq 0$  angenommen wird. Dass neben dem Modell mit orthogonalen Faktoren auch das Modell mit obliquen Faktoren betrachtet wird, beruht auf mehreren Gründen. Zum einen finden sich tatsächlich Hinweise, die für eine korrelative Beziehung zwischen einem Speed- und einem Fähigkeitsfaktor sprechen (van der Linden, 2007; Bontempo, 2000; Scrams & Schnipke, 1997). Zum anderen spiegelt das oblique Faktormodell am direktesten die Korrelationen zwischen Fähigkeits- und Geschwindigkeitsmaßen wider, die bei Tests teilweise zu finden sind. Diese beiden Gründe sprechen dafür, auch den potentiellen Nutzen von Modell (7.8) zu

untersuchen. Tatsächlich ist das Modell von van der Linden (2007) bei linearer Parametrisierung ein Spezialfall der untersuchten, orthogonalen Faktormodelle. Transformiert man die latenten Variablen  $\theta$  und  $\omega$  nach der Regel

$$\theta = \theta \quad (7.9)$$

$$\kappa = \omega - \rho\theta, \quad (7.10)$$

führt diese Transformation zu den neuen, orthogonalen Faktoren  $\theta$  und  $\kappa$ , sofern die ursprünglichen Variablen bivariat standardnormalverteilt mit Korrelationskoeffizient  $\rho$  waren. Dies erlaubt den Ausdruck der gemeinsamen Verteilung (7.8) in Begriffen der neuen Parameter:

$$f(\mathbf{t}, \mathbf{x}, \theta, \kappa) = \left[ \prod_{g=1}^G f(t_g | \rho\theta + \kappa) \right] \left[ \prod_{g=1}^G f(x_g | \theta) \right] f(\theta, \kappa). \quad (7.11)$$

Zumindest bei den Modellen mit linearer Parametrisierung ist die Struktur von (7.11) identisch zum allgemeinen Modell (7.7) mit orthogonalen Faktoren. Allerdings schlägt sich die ursprünglich oblique Faktorstruktur in einem speziellen Ladungsmuster der Einflussgewichte der latenten Variablen  $\theta$  und  $\kappa$  nieder. Bei den Modellen mit linearer Parametrisierung wurde immer ein linearer Prädiktor  $\eta$  betrachtet, der die Verteilung der Reaktionszeiten bestimmt. Geht man wie van der Linden (2007) von einer obliquen Faktorstruktur aus, hängt der lineare Prädiktor eines Items  $g$  gemäß Gleichung (7.8) nur von dem Residualfaktor  $\omega$  über die lineare Beziehung  $\eta_g(\omega) = k_g + c_g\omega$  ab. Bei Transformation der latenten Variablen nach Beziehung (7.10) lauten die linearen Prädiktoren bei den  $G$  Items eines Tests in Begriffen der neuen Parameter:

$$\begin{aligned} \eta_1(\theta, \kappa) &= k_1 + c_1\rho\theta + c_1\kappa \\ \eta_2(\theta, \kappa) &= k_2 + c_2\rho\theta + c_2\kappa \\ &\dots \\ \eta_G(\theta, \kappa) &= k_G + c_G\rho\theta + c_G\kappa. \end{aligned} \quad (7.12)$$

Augenscheinlich gleichen diese Beziehungen den im vierten Kapitel betrachteten, zweidimensionalen Modellierungsansätzen mit linearer Parametrisierung. Jedoch besitzt die Annahme obliquen Faktoren erhebliche Konsequenzen für die Schätzung der Personenparameter. Wie bei näherer Betrachtung des Gleichungssystems (7.12) ersichtlich ist, sind die Parameter  $\theta$  und  $\kappa$  nicht identifiziert, selbst wenn die Itemparameter bekannt sind (R. Davidson & MacKinnon, 1993): Eine Veränderung von  $\theta$  kann durch eine Veränderung von  $\kappa$  ausgeglichen werden. Zwar kann die Personenfähigkeit  $\theta$  über die Itemantworten geschätzt werden. Jedoch wird diese Schätzung dann dazu benötigt, den Faktor  $\kappa$  eindeutig zu bestimmen. Tatsächlich tragen in diesem Fall die Reaktionszeiten nicht zur Schätzung der Personenfähigkeit  $\theta$  bei. Dies kann man auch daran erkennen, dass die Informationsmatrix der Reaktionszeiten nach Reparametrisierung von

$\omega$  mit  $\theta$  und  $\kappa$  nicht positiv-definit ist. Dennoch kann auch bei obliquer Faktorstruktur die Reaktionszeit zu einer Schätzung der Personenfähigkeit ausgenutzt werden. Dies erfordert jedoch ein Paradigmenwechsel, was den Status des Residualfaktors  $\omega$  betrifft. Handelt es sich bei den Persönlichkeitseigenschaften  $\theta$  und  $\omega$  um normalverteilte, oblique Faktoren, kann die Verteilung der Itemantworten, Reaktionszeiten und des Residualfaktors  $\omega$  bei Bedingung auf die Personenfähigkeit  $\theta$  geschrieben werden als:

$$f(\mathbf{t}, \mathbf{x}, \omega|\theta) = f(\mathbf{x}|\theta)f(\mathbf{t}|\omega)f(\omega|\theta). \quad (7.13)$$

Der Residualfaktor  $\omega$  ist unbeobachtbar und stellt damit eine so genannte „fehlende“ Beobachtung dar. Da der Residualfaktor  $\omega$  für die Diagnostik oft keine Rolle spielt und nicht geschätzt werden soll, kann er als Störgröße betrachtet werden. Ausintegration der in der Terminologie von Rubin (1976) unsystematisch fehlenden und daher ignorierbaren Beobachtung  $\omega$  ergibt die marginale Verteilung der Reaktionszeiten und Itemantworten

$$f(\mathbf{t}, \mathbf{x}|\theta) = f(\mathbf{x}|\theta) \int f(\mathbf{t}|\omega)f(\omega|\theta) d\omega = f(\mathbf{x}|\theta)f(\mathbf{t}|\theta), \quad (7.14)$$

wobei  $f(\mathbf{t}|\theta) = \int f(\mathbf{t}|\omega)f(\omega|\theta) d\omega$  ist. Wie in Gleichung (7.14) ersichtlich ist, sind die Reaktionszeiten  $\mathbf{t}$  auch nach Marginalisierung über  $\omega$  unabhängig von den Itemantworten  $\mathbf{x}$ , wenn auf die Personenfähigkeit  $\theta$  bedingt wird. Allerdings gilt diese Aussage nicht für die einzelnen Reaktionszeiten  $\mathbf{t} = [t_1, \dots, t_G]$  einer Person, die aufgrund der Integration über  $\omega$  ihre Unabhängigkeit verlieren. Die Abhängigkeit der Reaktionszeiten ist bei der Bestimmung der Testinformation über die Personenfähigkeit  $\theta$  zu berücksichtigen. Als Testinformationsfunktion der Reaktionszeiten und der Itemantworten über den Parameter  $\theta$  erhält man:

$$I(\theta) = I_{\mathbf{x}}(\theta) + I_{\mathbf{t}}(\theta) = \sum_{i=g}^G I_{x_g}(\theta) + I_{\mathbf{t}}(\theta). \quad (7.15)$$

Bei den Komponenten  $I_{x_g}(\theta)$  handelt es sich um die Informationsbeiträge der einzelnen Itemantworten, deren Form für das zweiparametrische Logit-Modelle im dritten Kapitel beschrieben wurde, siehe hierzu Gleichung (3.6). Aufgrund der bedingten Unabhängigkeit der Itemantworten können diese einfach aufsummiert werden. Der Informationsbeitrag der Reaktionszeiten  $I_{\mathbf{t}}(\theta)$  jedoch kann nicht mehr über eine Addition der Informationsbeiträge der einzelnen Items gebildet werden. Aufgrund der Abhängigkeit der Reaktionszeiten fällt der Informationsbeitrag der Reaktionszeiten geringer aus. Eine Berechnung der approximativen Varianz des Fähigkeitsschätzers ist über den Kehrwert der Testinformation möglich, da die Informationsfunktion (7.15) keine Matrix ist. Die Reduktion der Varianz des Fähigkeitsschätzers, die sich bei Berücksichtigung der Reaktionszeiten ergibt, hängt monoton von der Informationsfunktion der Reaktionszeiten  $I_{\mathbf{t}}(\theta)$  ab. Diese Größe soll im Folgenden näher betrachtet werden.

Zur Berechnung des Informationsbeitrages der Reaktionszeiten  $I_{\mathbf{t}}(\theta)$  ist eine Betrachtung der Score-Funktion aufschlussreich. Diese erhält man durch Ableitung der logarithmierten marginalen Reaktionszeitverteilung  $\log[f(\mathbf{t}|\theta)]$  nach der Personenfähigkeit  $\theta$ :

$$S(\theta, \mathbf{t}) = \frac{\partial}{\partial \theta} \log \left[ \int f(\mathbf{t}|\omega) f(\omega|\theta) d\omega \right]. \quad (7.16)$$

Die Score-Funktion der Reaktionszeiten kann alternativ ausgedrückt werden. Umformungen ergeben bei Vertauschbarkeit von Ableitung und Integration:

$$\begin{aligned} S(\theta, \mathbf{t}) &= \frac{\partial}{\partial \theta} \log \left[ \int f(\mathbf{t}|\omega) f(\omega|\theta) d\omega \right] \\ &= \frac{1}{\int f(\mathbf{t}|\omega) f(\omega|\theta) d\omega} \frac{\partial}{\partial \theta} \left[ \int f(\mathbf{t}|\omega) f(\omega|\theta) d\omega \right] \\ &= \frac{1}{\int f(\mathbf{t}|\omega) f(\omega|\theta) d\omega} \int f(\mathbf{t}|\omega) f(\omega|\theta) \frac{\partial}{\partial \theta} \left[ \log [f(\mathbf{t}|\omega) f(\omega|\theta)] \right] d\omega \quad (7.17) \\ &= \int \frac{f(\mathbf{t}|\omega) f(\omega|\theta)}{\int f(\mathbf{t}|\omega) f(\omega|\theta) d\omega} \frac{\partial}{\partial \theta} \left[ \log [f(\mathbf{t}|\omega) f(\omega|\theta)] \right] d\omega \\ &= \int \left[ \frac{\partial}{\partial \theta} \log f(\omega|\theta) \right] f(\omega|\mathbf{t}, \theta) d\omega. \end{aligned}$$

Die Bedeutung der Umformung erschließt sich, wenn man noch einmal die gemeinsame Verteilung der Itemantworten, Reaktionszeiten und des Residualfaktors  $f(\mathbf{x}, \mathbf{t}, \omega|\theta) = f(\mathbf{x}|\theta) f(\mathbf{t}|\omega) f(\omega|\theta)$  betrachtet. Wäre die genaue Ausprägung des Residualfaktors  $\omega$  einer Person bekannt und stünde als zusätzliche Beobachtung zur Verfügung, besäße diese Beobachtung die Score-Funktion:

$$S(\theta, \omega) = \frac{\partial}{\partial \theta} \log f(\omega|\theta). \quad (7.18)$$

Die Score-Funktion der Reaktionszeiten  $S(\theta, \mathbf{t})$  ergibt sich folglich als Erwartungswert der Score-Funktion des Residualfaktors  $S(\theta, \omega)$ , wenn dieser über die auf die beobachteten Reaktionszeiten  $\mathbf{t}$  bedingte Verteilung von  $\omega$  gebildet wird (siehe 7.17):

$$S(\theta, \mathbf{t}) = \int \left[ \frac{\partial}{\partial \theta} \log f(\omega|\theta) \right] f(\omega|\mathbf{t}, \theta) d\omega = E_{\omega} [S(\theta, \omega) | \mathbf{t}, \theta]. \quad (7.19)$$

Die Testinformation der Reaktionszeiten ist (analog zur Fisher-Information) die Varianz der Score-Funktion der Reaktionszeiten. Als Testinformation ergibt sich somit:

$$I_{\mathbf{t}}(\theta) = \sigma_{\mathbf{t}}^2 \left[ E_{\omega} [S(\theta, \omega) | \mathbf{t}, \theta] \right]. \quad (7.20)$$

Gleichung (7.20) spiegelt die Beziehung der Testinformation der Reaktionszeiten zu verschiedenen Teilkomponenten wider, zur „Information“ der Reaktionszeiten  $\mathbf{t}$  über den unbekannt Parameter  $\omega$  und zur „Information“ von  $\omega$



über  $\theta$ . Dabei können zwei Extremfälle unterschieden werden: Bei Fall (1) tragen die Reaktionszeiten nicht zur Schätzung der Personenfähigkeit bei, bei Fall (2) ergibt sich eine deterministische Beziehung.

**Zu Fall (1):**

Die Score-Funktion der Reaktionszeiten (7.19) ist genau dann Null für alle beobachtbaren Reaktionszeiten  $\mathbf{t}$ , wenn die Posteriori-Verteilung des Residualfaktors  $f(\omega|\mathbf{t}, \theta)$  der Priori-Verteilung  $f(\omega|\theta)$  entspricht. In diesem Fall gilt für die Score-Funktion nämlich unter der Voraussetzung, dass Integration und Ableitung vertauscht werden können, die Beziehung:

$$\begin{aligned}
 S(\theta, \mathbf{t}) &= \int \left[ \frac{\partial}{\partial \theta} \log f(\omega|\theta) \right] f(\omega|\mathbf{t}, \theta) d\omega \\
 &= \int \left[ \frac{\partial}{\partial \theta} \log f(\omega|\theta) \right] f(\omega|\theta) d\omega \\
 &= \int \left[ \frac{\partial}{\partial \theta} f(\omega|\theta) \right] \frac{1}{f(\omega|\theta)} f(\omega|\theta) d\omega & (7.21) \\
 &= \int \left[ \frac{\partial}{\partial \theta} f(\omega|\theta) \right] d\omega \\
 &= \frac{\partial}{\partial \theta} \left[ \int f(\omega|\theta) d\omega \right] = 0.
 \end{aligned}$$

Die Testinformation der Reaktionszeiten ist die Varianz der Score-Funktion. Diese ist aber beim oben geschilderten Fall ebenfalls Null – die Reaktionszeiten tragen nicht zur Schätzung der Personenfähigkeit bei.

**Zu Fall (2):**

Anders sieht es aus, wenn man von einer eindeutigen Bestimmbarkeit des Residualfaktors  $\omega$  einer Person durch die beobachteten Reaktionszeiten  $\mathbf{t}$  ausgeht. In diesem Fall ist die bedingte Verteilung  $f(\omega|\mathbf{t}, \theta)$  eine degenerierte Verteilung, zwischen den beobachteten Reaktionszeiten  $\mathbf{t}$  und dem Residualfaktor  $\omega$  besteht eine deterministische Beziehung der Form  $\omega = g(\mathbf{t})$ . Der auf einen bestimmten beobachteten Zeitvektor  $\mathbf{t}$  bedingte Erwartungswert der Score-Funktion vereinfacht sich in diesem Fall zu

$$S(\theta, \mathbf{t}) = E_{\omega} [S(\theta, \omega)|\mathbf{t}, \theta] = \frac{\partial}{\partial \theta} \left[ \log f(g(\mathbf{t})|\theta) \right] = \frac{\partial}{\partial \theta} \left[ \log f(\omega|\theta) \right], \quad (7.22)$$

wobei  $\omega$  der zu einem bestimmten Reaktionszeitmuster  $\mathbf{t}$  gehörende Wert des Residualfaktors ist. Die Score-Funktion der Reaktionszeiten  $\mathbf{t}$  entspräche somit der Score-Funktion des Residualfaktors  $\omega$ , wenn dieser bekannt wäre. Es ist leicht ersichtlich, dass die Testinformation der Reaktionszeiten in diesem Falle der Testinformation des Residualfaktors  $\omega$  gleicht. Dies bedeutet, dass im Falle einer eindeutigen Bestimmbarkeit von  $\omega$  über die beobachteten Reaktionszeiten  $\mathbf{t}$  die Information der Reaktionszeiten  $I_{\mathbf{t}}(\theta)$  durch die Information des Residualfaktors  $I_{\omega}(\theta)$  begrenzt wird. Dass dies generell gilt, wird im nächsten Abschnitt gezeigt.

### Dekomposition der Fisher-Information

Eine Anwendung des „Missing Information Principle“ (Mislevy & Sheehan, 1989, 1988; Orchard & Woodbury, 1972; McLachlan & Krishnan, 1997) ermöglicht es, eine Obergrenze für die Testinformation der Reaktionszeiten anzugeben. Hintergrund dabei ist die Überlegung, dass die Information als Varianz der Score-Funktion definiert ist. Die Varianz einer Funktion lässt sich aber nach dem Gesetz der iterierten Erwartung auf zwei Komponenten zurückführen (Casella & Berger, 2002): (1) Die Varianz des bedingten Erwartungswertes einer Funktion und (2) den Erwartungswert der bedingten Varianz der Funktion. In mathematischer Formulierung lautet das Gesetz der iterierten Erwartung:

$$\sigma^2[x] = E_y[\sigma^2[x|y]] + \sigma_y^2[E[x|y]]. \quad (7.23)$$

Eine einfache Umstellung ergibt:

$$\sigma_y^2[E[x|y]] = \sigma^2[x] - E_y[\sigma^2[x|y]]. \quad (7.24)$$

Dieses Prinzip kann auf die Testinformation der Reaktionszeiten angewandt werden. Hierzu muss man sich in Erinnerung rufen, dass die Testinformation der Reaktionszeiten nach Beziehung (7.20) als Varianz des bedingten Erwartungswertes der Score-Funktion des Residualfaktors  $\sigma_{\mathbf{t}}^2[E_{\omega}[S(\theta, \omega)|\mathbf{t}, \theta]]$  geschrieben werden konnte. Ersetzt man die Komponenten von (7.24) mit den entsprechenden Größen aus (7.20), lässt sich die Testinformation  $I_{\mathbf{t}}(\theta)$  der Reaktionszeiten  $\mathbf{t}$  berechnen als:

$$\begin{aligned} I_{\mathbf{t}}(\theta) &= \sigma_{\mathbf{t}}^2[E_{\omega}[S(\theta, \omega)|\mathbf{t}, \theta]] \\ &= \sigma_{\omega}^2[S(\theta, \omega)] - E_{\mathbf{t}}[\sigma_{\omega}^2[S(\theta, \omega)|\mathbf{t}, \theta]]. \end{aligned} \quad (7.25)$$

Die erste Komponente  $\sigma_{\omega}^2[S(\theta, \omega)]$  ist gerade die Information  $I_{\omega}(\theta)$  von  $\omega$  über  $\theta$ , die man erhielte, wenn  $\omega$  keine unbeobachtbare Größe wäre – also die Information einer beobachteten Zufallsvariablen mit Verteilung  $f(\omega|\theta)$ . Sie bildet somit eine Obergrenze für die in den Reaktionszeiten  $\mathbf{t}$  enthaltene Testinformation  $I_{\mathbf{t}}(\theta)$ . Dies folgt daraus, dass  $E_{\mathbf{t}}[\sigma_{\omega}^2[S(\theta, \omega)|\mathbf{t}, \theta]]$  immer positiv ist. Bemerkenswert bei dieser Aussage ist, dass dieses Maximum für eine Person nicht durch eine Vergrößerung der Anzahl der Reaktionszeitmessungen erhöht werden kann. Dies liegt daran, dass die einzelnen Reaktionszeiten  $\mathbf{t} = [t_1, \dots, t_G]$  nicht unabhängig sind, wenn über den unbeobachteten Residualfaktor  $\omega$  integriert wird. Gleichung (7.25) bedeutet folglich, dass durch die Berücksichtigung der Reaktionszeiten die Varianz des Fähigkeitsschätzers maximal um denselben Betrag reduzieren werden kann, wie durch Berücksichtigung des (unbeobachtbaren) Residualfaktors einer Person. Die Berücksichtigung der Reaktionszeiten besitzt also maximal dasselbe diagnostische Potential wie die Berücksichtigung des (unbekannten) Residualfaktors  $\omega$  einer Person.

Allerdings wird die maximal mögliche Varianzreduktion nicht erreicht: Die zweite Komponente  $E_{\mathbf{t}}[\sigma_{\omega}^2[S(\theta, \omega)|\mathbf{t}, \theta]]$  ist der Informationsverlust, der auf der

Unsicherheit des Residualfaktors  $\omega$  beruht, welcher anhand der beobachteten Reaktionszeiten  $\mathbf{t}$  nicht eindeutig erschlossen werden kann. Je „konzentrierter“ die Posteriori-Verteilung  $f(\omega|\mathbf{t})$  ausfällt, desto geringer wird mit abnehmender Varianz  $\sigma_\omega^2 \left[ \frac{\partial}{\partial \theta} \log[f(\omega|\theta)] | \theta, \mathbf{t} \right]$  der Informationsverlust (Mislevy & Sheehan, 1988) und desto größer fällt die tatsächliche Varianzreduktion durch Berücksichtigung der Reaktionszeiten aus.

### Illustratives Beispiel

Die Ausführungen des letzten Abschnittes waren teilweise recht abstrakt. Sie lassen sich jedoch anhand eines konkreten Beispiels illustrieren. Ausgangspunkt der Überlegungen zur Testinformation der Reaktionszeiten war eine Annahme über die Form der Verteilung der Reaktionszeiten  $\mathbf{t}$ , Itemantworten  $\mathbf{x}$  und des Residualfaktors  $\omega$  bei Bedingung auf die Personenfähigkeit  $\theta$  (siehe 7.8):

$$f(\mathbf{t}, \mathbf{x}, \omega | \theta) = f(\mathbf{t} | \omega) f(\mathbf{x} | \theta) f(\omega | \theta). \quad (7.26)$$

Um die Aussagen des letzten Abschnittes anhand eines Beispiels demonstrieren zu können, werden für die einzelnen Komponenten von Gleichung (7.26) konkrete Verteilungen gewählt. Da lediglich die Testinformation der Reaktionszeiten von Interesse ist, sind vorerst nur die Verteilungen  $f(\mathbf{t} | \omega)$  und  $f(\omega | \theta)$  relevant.

Häufig werden in der Testtheorie die latenten Variablen als standardnormalverteilt angenommen. Von dieser Annahme soll auch im Folgenden ausgegangen werden: In einer Population von potentiellen Testpersonen sind der Residualfaktor  $\omega$  und die Personenfähigkeit  $\theta$  bivariat standardnormalverteilt mit einem als bekannt angenommenen Korrelationskoeffizienten  $\rho$ . Der Korrelationskoeffizient  $\rho$  kann ebenso wie die Itemparameter der Aufgaben in einer Vorstudie mit ausreichender Genauigkeit geschätzt werden und gilt daher als bekannte Größe. Die gemeinsame Verteilung von  $\theta$  und  $\omega$  lautet in diesem Fall:

$$(\theta, \omega) \sim N_2 \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right). \quad (7.27)$$

In der Gleichung steht  $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  für die bivariate Normalverteilung mit Erwartungswertvektor  $\boldsymbol{\mu}$  und Varianz-Kovarianz-Matrix  $\boldsymbol{\Sigma}$ . Bei Bedingung auf die Personenfähigkeit  $\theta$  ist der Residualfaktor  $\omega$  somit univariat normalverteilt (R. Johnson & Wichern, 2002):

$$\omega | \theta \sim N(\rho\theta, 1 - \rho^2). \quad (7.28)$$

Dabei bezeichnet  $N(\mu, \sigma)$  die univariate Normalverteilung mit Lokationsparameter  $\rho\theta$  und Varianz  $1 - \rho^2$ . Als Modell für die Verteilung der Reaktionszeiten  $f(\mathbf{t} | \omega)$  kann jedes der im vierten Kapitel vorgestellten Modelle gewählt werden, wenn das Einflussgewicht der Personenfähigkeit  $l_g$  auf Null gesetzt wird. Der Effekt des Missing Information Prinzips lässt sich jedoch am besten anhand der Modelle für log-normalverteilte Reaktionszeiten demonstrieren, weshalb für

die folgende Illustration Modell (4.32) herangezogen wird. Zunächst soll nur die Reaktionszeitverteilung eines Items betrachtet werden.

Bei fehlendem Einfluss der Personenfähigkeit ( $l_g = 0$ ) folgt nach Modell (4.32) die Reaktionszeitverteilung  $f(t_g|\omega)$  einer Person beim  $g$ -ten Item eines Tests einer Log-Normalverteilung mit Skalenparameter  $\sigma_g$  und individualspezifischem Lokationsparameter  $\eta_g(\omega) = k_g + c_g\omega$ . Durch Logarithmierung kann die log-normalverteilte Reaktionszeit  $t_g$  in eine normalverteilte Zufallsvariable  $t_g^* = \log(t_g)$  transformiert werden. Bei Bedingung auf  $\omega$  gilt für die Verteilung der logarithmierten Reaktionszeit von Aufgabe  $g$  somit:

$$t_g^*|\omega \sim N(k_g + c_g\omega, \sigma_g^2). \quad (7.29)$$

Die maximal in der Reaktionszeit von Item  $g$  enthaltene Iteminformation ist die Information des Residualfaktors  $\omega$  über die Personenfähigkeit  $\theta$  – also die Information  $I_\omega(\theta)$ , die sich ergeben hätte, wenn der Residualfaktor  $\omega$  beobachtet werden könnte. Der Residualfaktor  $\omega$  ist bei Bedingung auf  $\theta$  normalverteilt mit Erwartungswert  $E(\omega|\theta) = \rho\theta$  und Varianz  $\sigma^2(\omega|\theta) = 1 - \rho^2$ . Diese Beziehung entspricht jedoch genau dem Modell eines unidimensionalen latenten Trait Modells für kontinuierliche Antworten (Mellenbergh, 1994). Die Fisher-Information von  $\omega$  über  $\theta$  ist in diesem Fall eine Funktion der Korrelation zwischen  $\omega$  und  $\theta$  und lautet:

$$I_\omega(\theta) = \frac{\rho^2}{(1 - \rho^2)}. \quad (7.30)$$

Eine genaue Herleitung dieser Beziehung findet sich in Anhang C.1. Gleichung (7.30) ist eine Obergrenze für die Testinformation der Reaktionszeiten  $\mathbf{t}$  und somit auch eine Obergrenze für die Iteminformation der Reaktionszeit  $t_g$  von Item  $g$ . Damit bestimmt die Korrelation zwischen der Personenfähigkeit  $\theta$  und dem Residualfaktor  $\omega$  die maximal mögliche Reduktion der Varianz des Fähigkeitsschätzers  $\hat{\theta}$ . Diese maximale Reduktion wird durch die Reaktionszeiten jedoch nicht erreicht, da die Testinformation der Reaktionszeiten geringer ausfällt als die Obergrenze (7.30). Wie hoch aber fällt die Iteminformation der ersten Reaktionszeit  $t_1$  aus? Bei Integration über die unbeobachtbare Größe  $\omega$  erhält man bei Bedingung auf  $\theta$  für die marginale Verteilung der logarithmierten Reaktionszeit  $t_1^*$  beim ersten Item die Verteilung  $f(t_1^*|\theta) = \int f(t_1^*|\omega)f(\omega|\theta) d\omega$ . Die Lösung des Integrals ist explizit angebar:

$$t_1^*|\theta \sim N\left(k_1 + c_1(\rho\theta), \sigma_1^2 + c_1^2(1 - \rho^2)\right). \quad (7.31)$$

Dies ergibt sich aus dem Umstand, dass die logarithmierte Reaktionszeit  $t_1^*$  und der Residualfaktor  $\omega$  bei Bedingung auf  $\theta$  bivariat normalverteilt sind. Die marginale Verteilung der logarithmierten Reaktionszeit des ersten Items besitzt daher folgende Information über die Personenfähigkeit:

$$I_{t_1^*}(\theta) = \frac{(c_1\rho)^2}{\sigma_1^2 + c_1^2(1 - \rho^2)}. \quad (7.32)$$

Für die genaue Ableitung dieser Informationsfunktion siehe Mellenbergh (1994). Wie Gleichung (7.32) entnommen werden kann, konvergiert der Informationsbeitrag des ersten Items bei abnehmender korrelativer Beziehung von  $\omega$  und  $\theta$  gegen Null. Bei perfekter Korrelation von  $\rho = 1$  erreicht die Iteminformation mit  $I_{t_1^*}(\theta) = c_1^2/\sigma_1^2$  genau die Beziehung, die ein eindimensionales Faktormodell besäße, in dem die Personenfähigkeit sowohl auf den Itemantworten als auch auf den Reaktionszeiten läßt, siehe Gleichung (4.36). Gleichung (7.32) beschreibt den Informationsbeitrag des ersten Items. Die Informationsbeiträge der weiteren Items zur Testinformation der Reaktionszeiten können jedoch aufgrund der Abhängigkeit der Reaktionszeiten nicht über die Beziehung (7.32) berechnet werden.

Da die Reaktionszeiten bei mehreren Aufgaben trotz Bedingung auf die Personenfähigkeit  $\theta$  nicht unabhängig verteilt sind, lässt sich die Testinformation der Reaktionszeiten nicht als Summe der Iteminformationen der einzelnen Reaktionszeiten berechnen. Tatsächlich kann gezeigt werden, dass bei einem Test aus  $n$  Items mit identischen Itemparametern  $c$ ,  $k$  und  $\sigma^2$  durch die Hinzunahme eines zusätzlichen Items mit denselben Itemparametern die Testinformation der Reaktionszeiten genau um folgenden Betrag ansteigt:

$$\Delta I_{t_{n+1}^*}(\theta) = \frac{\left[ c\rho - nc\rho \left[ \frac{u}{\sigma^2} - n \frac{(u)^2}{\sigma^2(\sigma^2 + nu)} \right] \right]^2}{\left[ \sigma^2 + u \right] - \left[ n \frac{(u)^2}{\sigma^2} - n^2 \frac{(u)^3}{\sigma^2(\sigma^2 + nu)} \right]}, \quad (7.33)$$

wobei in Gleichung (7.33) der Term  $u$  für  $u = c^2(1 - \rho^2)$  steht. Die Rechenschritte, die zu Gleichung (7.33) führen, sollen hier nur grob skizziert werden, eine genaue Darstellung findet sich in Anhang C.2. Gleichung (7.33) beruht darauf, dass bei abhängigen Beobachtungen die Testinformation  $I_{\mathbf{t}^*}(\theta)$  der logarithmierten Reaktionszeiten als  $I_{\mathbf{t}^*}(\theta) = I_{t_1^*}(\theta) + I_{t_2^*, t_1^*}(\theta) + \dots$  berechnet werden kann, wobei  $I_{t_2^*, t_1^*}(\theta)$  die Fisher-Information der Verteilung  $f(t_2^*|t_1^*; \theta)$  über den Parameter  $\theta$  ist (Sorensen & Gianola, 2002). Der Informationszuwachs durch das Item  $n + 1$  ergibt sich somit aus der Iteminformation der Verteilung  $f(t_{n+1}^*|t_1^*, \dots, t_n^*; \theta)$  über die Personenfähigkeit  $\theta$ . Diese bedingte Verteilung  $f(t_{n+1}^*|t_1^*, t_2^*, \dots, t_n^*; \theta)$  kann für die normalverteilten logarithmierten Reaktionszeiten im Falle von Items mit identischen Itemparametern einfach berechnet werden. Gleichung (7.33) ist die Iteminformation dieser bedingten Verteilung. Grenzwertbildungen zeigen, dass der Informationszuwachs mit zunehmendem Testumfang gegen Null konvergiert. Dies ist letztendlich auch eine Folge des Umstands, dass die Gesamtinformation der Reaktionszeiten in der maximalen Information  $I_\omega(\theta)$  eine obere Grenze besitzt.

Die Testinformation der Reaktionszeiten  $I_{\mathbf{t}^*}(\theta)$  eines Tests aus 5, 10, 25, 50 und 100 Items mit identischen Itemparametern, die sich aus der Summe der Informationsbeiträge (7.33) der einzelnen Items ergibt, ist in Abbildung 7.5 dargestellt. Als Itemparameter wurden die Ausprägungen  $c = 0.3$ ,  $k = 0$  und  $\sigma = 1$  gewählt. Abbildung 7.5 illustriert, dass trotz zunehmendem Umfang des Tests die Testinformation der Reaktionszeiten nur bis zur Obergrenze (7.30) anwächst.

Bisher wurde immer die Information der Reaktionszeiten über die Personenfähigkeit  $\theta$  betrachtet. Nicht erwogen wurden die Konsequenzen für die Varianz

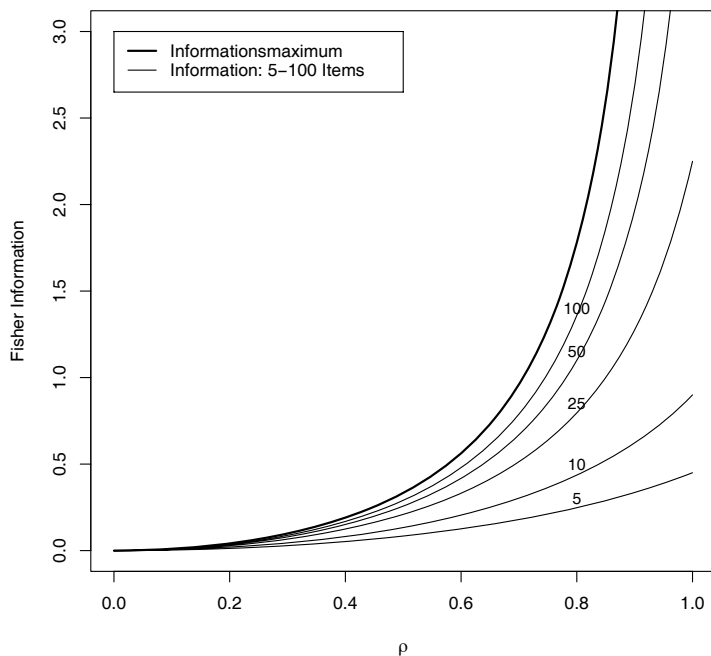


Abbildung 7.5: Testinformation der RT eines Tests aus 5, 10, 25, 50 und 100 identischen Items mit Itemparametern  $c = 0.3$ ,  $k = 0$  und  $\sigma = 1$  in Abhängigkeit von der Korrelation zwischen dem Residualfaktor und der Personenfähigkeit – Kontrastierung der realisierten Testinformation (gemäß 7.33) mit der maximalen Testinformation (7.30)

der Fähigkeitsschätzung. Allgemeine Beziehungen der Iteminformation zur Varianz von Maximum-Likelihood-Schätzern (2.9) sprechen dafür, dass es durch Einbezug eines Items nur dann zu einer bemerkbaren Varianzreduktion des Fähigkeitsschätzers kommt, wenn das Item einen großen Beitrag zur Testinformation leistet. Betrachtet man den Verlauf der Informationsfunktion in Abbildung 7.5, ist der Schluss nahe liegend, dass bei obliquen Faktorstruktur eine hohe Korrelation zwischen Personenfähigkeit und Residualfaktor notwendig ist, um die Reaktionszeiten gewinnbringend nutzen zu können. Die in Studien gefundenen Korrelationen zwischen der Personenfähigkeit und Speed-Maßen liegen im Bereich von  $|0.0|$  bis maximal  $|0.8|$  (Vigneau, Blanchet, Loranger & Pepin, 2002; Hornke, 2000). Dies bedeutet, dass unabhängig vom konkreten Reaktionszeitmodell der maximal mögliche Nutzen der Reaktionszeiten beschränkt ist und dass die Varianz des Fähigkeitsschätzers durch eine Vergrößerung der Anzahl der beobachteten Reaktionszeiten nicht beliebig reduziert werden kann. Dies ist ein fundamentaler Unterschied zwischen dem obliquen Faktormodell und dem orthogonalen Faktormodell.

Die Abhängigkeit der mit den Reaktionszeiten erzielbaren Varianzreduktion von der Korrelation zwischen der Personenfähigkeit und dem Residualfaktor soll

genauer untersucht werden. In Tabelle (7.5) sind die Itemparameter eines Tests aus fünf Items aufgeführt.

Tabelle 7.5: Oblique Faktorstruktur mit log-normalverteilten RT und linearer Parametrisierung; Itemparameter eines Tests aus fünf Aufgaben

Item	$a_g$	$b_g$	$c_g$	$k_g$	$\sigma_g$
1	1.50	0.66	1.50	0.00	1.00
2	1.50	0.33	1.50	0.00	1.00
3	1.50	0.00	1.50	0.00	1.00
4	1.50	-0.33	1.50	0.00	1.00
5	1.50	-0.66	1.50	0.00	1.00

Die gemeinsame Testinformation der Itemantworten und Reaktionszeiten kann bei Kombination eines zweiparametrischen Logit-Modells mit dem Reaktionszeitmodell (7.29) gemäß Gleichung (7.15), (3.6) und (7.33) berechnet werden. Durch Bildung des Kehrwerts der Testinformation erhält man die Varianz des Fähigkeitsschätzers. Die bei einem derartigen Test resultierende (approximative) Varianz ist für unterschiedlich starke korrelative Beziehungen zwischen  $\theta$  und  $\omega$  in Abbildung 7.6 für verschiedene Fähigkeitsniveaus wiedergegeben. In der Abbildung ist wieder erkennbar, dass Reaktionszeiten die Schätzung der Personenfähigkeit vorwiegend bei extremen Fähigkeitsausprägungen verbessern. Allerdings hängt die Verbesserung stark von der korrelativen Beziehung zwischen den obliquen Faktoren ab. Während bei moderaten Fähigkeiten nur sehr hohe Korrelationen zu einer erwähnenswerten Varianzreduktion führen, läßt sich für extreme Fähigkeitsniveaus bereits bei moderaten Korrelationen zwischen der Personenfähigkeit und dem Residualfaktor eine beachtliche Varianzreduktion realisieren.

### 7.3 Diskussion

Von einigen Forschern wird die Hoffnung geäußert, dass durch die Reaktionszeiten bei Tests die Schätzung der zugrunde liegenden Persönlichkeitseigenschaft verbessert werden könnte. Diese Hoffnung speist sich teilweise aus der Beobachtung korrelativer Beziehungen zwischen der Personenfähigkeit und unterschiedlichen Speed-Maßen der Testbearbeitung. Derartige Befunde führen manchmal sogar zu Aussagen wie die von Mayerl (2003), dass durch die Berücksichtigung der Reaktionszeiten die Information des Tests verdoppelt werden könne. Ob diese optimistische Einschätzung des Nutzens der Reaktionszeiten gerechtfertigt ist, wurde im aktuellen Kapitel untersucht. Tatsächlich ist die Reaktionszeit bereits unter schwachen Voraussetzungen zur Fähigkeitsschätzung verwendbar. Beispielsweise kann auf die genaue Angabe einer konkreten Reaktionszeitverteilung verzichtet werden. Der genaue Beitrag von Reaktionszeitmessungen zur Varianzreduktion hängt aber von der Verteilung der Reaktionszeiten und der zugrunde liegenden Modellierung ab.

Die Verteilung der Itemantworten und Reaktionszeiten könnte sich mit einem

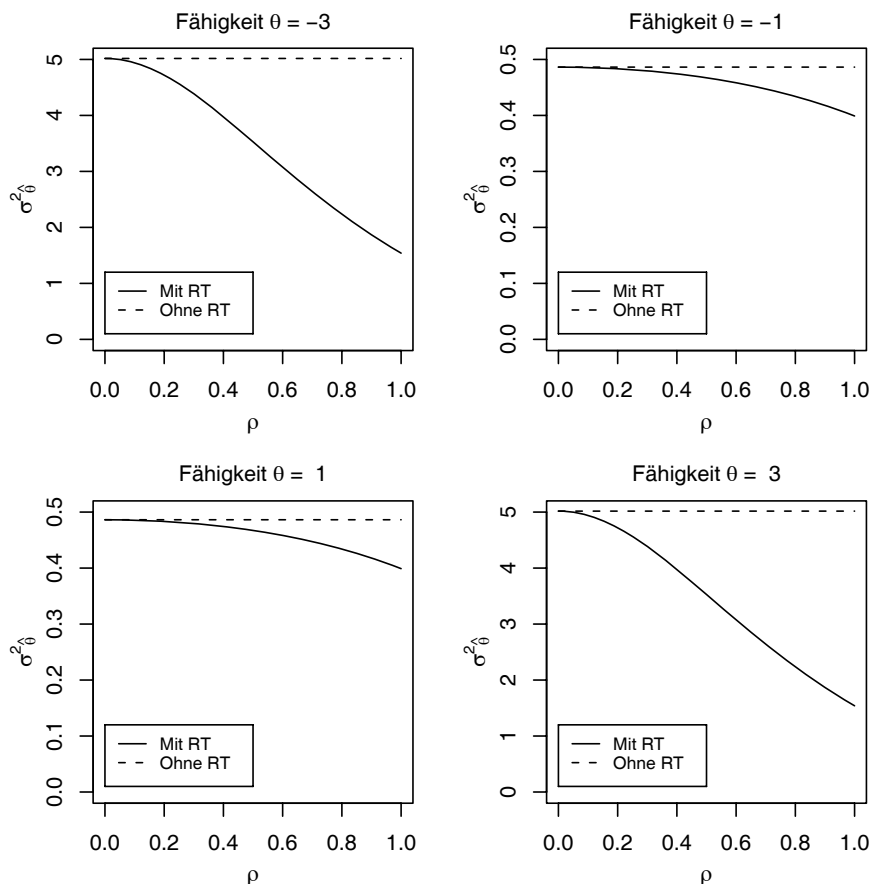


Abbildung 7.6: Log-normalverteilte RT: Varianz des Fähigkeitsschätzers  $\hat{\theta}$  bei einem Test aus 5 Items mit Itemparametern gemäß Tabelle 7.5 und linearer Parametrisierung – Kontrastierung der Varianz mit Berücksichtigung und ohne Berücksichtigung der Reaktionszeiten für unterschiedlich starke korrelative Beziehung der obliquen Faktoren

orthogonalen zweifaktoriellen Modell modellieren lassen, bei dem die Reaktionszeiten sowohl auf der Personenfähigkeit  $\theta$  als auch auf dem Residualfaktoren  $\omega$  einer Person laden. Die Implikationen dieses Falls für die Schätzung der Personenfähigkeit wurden in Kapitel 7.1 untersucht. Dort zeigte sich, dass die Berücksichtigung der Reaktionszeiten die Schätzung der Personenfähigkeit vorwiegend bei extremer Eigenschaftsausprägung verbessert. Besonders bei dem Modell mit log-normalverteilten Reaktionszeiten und der quadratischen Parametrisierung ist diese Verbesserung beachtlich. Dies liegt daran, dass die Informationsfunktionen mancher Reaktionszeitmodelle im Gegensatz zur Informationsfunktion des zweiparametrischen Logit-Modells mit extremer werdender Personenfähigkeiten nicht abnehmen. Zur Schätzung von Fähigkeiten mittlerer Ausprägung hingegen tragen die Reaktionszeiten kaum bei – hier dürften die Itemantworten zur Diagnostik in der Regel ausreichen. Der Beitrag der Reaktionszeiten zur



Varianzreduktion ist nicht beschränkt. Da die einzelnen Reaktionszeiten unabhängig voneinander sind, summieren sich ihre Informationsbeiträge. Mit einer Verlängerung des Tests steigt also der Nutzen der Reaktionszeiten unbegrenzt an. Dies ändert sich jedoch, wenn man anstatt des orthogonalen zweifaktoriellen RT-Modells ein obliques RT-Modell betrachtet.

Korrelationen zwischen der Personenfähigkeit und Speed-Maßen der Testbearbeitung mögen verschiedene Autoren dazu inspiriert haben, für die Verteilung der Reaktionszeiten und Itemantworten anstelle eines orthogonalen faktorenanalytischen Modells ein obliques Faktormodell vorzuschlagen, bei welchem ein Faktor nur die Reaktionszeiten und ein anderer Faktor nur die Itemantworten beeinflusst. Ein derartiges Modell und dessen Implikationen für die Fähigkeitsschätzung wurden in Kapitel 7.2 betrachtet. Dort zeigte sich, dass bei einer gemeinsamen Schätzung der Personenfähigkeit und des Residualfaktors die Fähigkeitsschätzung durch Berücksichtigung der Reaktionszeiten nicht verbessert werden kann. Jedoch ändert sich dies, wenn man das Sampling-Modell wechselt. Wird die Varianz des Fähigkeitsschätzers nicht mehr für eine Gruppe von Personen mit bestimmter Personenfähigkeit und bestimmtem Residualfaktor untersucht, sondern lediglich auf eine bestimmte Personenfähigkeit bedingt, können die Reaktionszeiten wiederum eingesetzt werden, um die Personenfähigkeit zu schätzen. Allerdings ist in diesem Fall der Informationsgewinn begrenzt und hängt von der Höhe der Korrelation zwischen der Personenfähigkeit und dem Residualfaktor ab. Bei kurzen Tests verbessert die Berücksichtigung der Reaktionszeiten die Schätzung extremer Fähigkeiten bereits bei moderaten Korrelationen. Bei langen Tests wächst die Information der Reaktionszeiten jedoch nicht unbegrenzt und wird im Vergleich zur Information der Itemantworten immer vernachlässigbarer. Aus diesem Grund ist das Vorliegen eines orthogonalen Faktormodells sicherlich günstiger.

Was den realen Nutzen der Reaktionszeiten bei einer praktischen Testanwendung betrifft, so müssen die gerade berichteten Ergebnisse allerdings eingeschränkt werden. Vorausgesetzt wurde bei den Ableitungen immer die Kenntnis der Itemparameter der Modelle, die jedoch erst anhand einer Eichstichprobe geschätzt werden müssen. Wenn bei kleinen Eichstichproben die Itemparameter der Reaktionszeitverteilung nur ungenau geschätzt wurden, verzerrt die falsche Ausgangslage natürlich auch die Schätzung der Personenfähigkeit  $\theta$ . Im Falle einer sehr unzuverlässigen Itemparameterschätzung bei kleinen Stichproben könnte es unter Umständen sogar besser sein, die Reaktionszeiten zu ignorieren. Gegen dieses Argument jedoch lässt sich einwenden, dass auch die Parameterschätzer des IRT-Modells fehlerbehaftet sind und zumindest bei den Simulationen im fünften und sechsten Kapitel teilweise weniger präzise bestimmt werden konnten als die Itemparameter des Reaktionszeitmodells. Ganz sicher aber fällt die theoretische Bewertung des Nutzens, den Reaktionszeiten für die Fähigkeitsschätzung besitzen, zu vorteilhaft aus, da die Itemparameterunsicherheit unberücksichtigt bleibt. Auch wurde mit dem Kehrwert der Testinformation immer nur eine theoretische Größe betrachtet, die asymptotische Varianz von Maximum-Likelihood-Schätzern, welche bei Tests geringen Umfangs nicht erreicht wird. Der durchschnittliche quadrierte Messfehler ist somit größer als die asymptotische Varianz, zumal der Fähigkeitsschätzer bei kleinen Tests einen nicht vernachlässigbaren Bias besitzen dürfte.

Auch wenn die Varianz des Fähigkeitsschätzers im Zentrum der Betrachtung stand und sicherlich eine wichtige Größe ist, spiegelt das Kriterium der Varianzreduktion den Beitrag von Beobachtungen (Itemantworten oder Reaktionszeiten) zur Schätzung der Personenfähigkeit nur unvollständig wider. Weder gibt die asymptotische Varianz Auskunft über Identifikationsprobleme von Schätzern, noch über numerische Probleme der Schätzung. Mit IRT-Modellen allein ist es beispielsweise nicht möglich, Personen mit demselben Antwortmuster hinsichtlich ihres Fähigkeitsniveaus zu unterscheiden – eine Beschränkung, die durch die Berücksichtigung der Bearbeitungszeit behoben werden kann (White, 1973). Auch existiert die Schwierigkeit, Personen eine Fähigkeitsausprägung zuzuschreiben, wenn diese alle Aufgaben gelöst oder nicht gelöst haben (Embretson & Reise, 2000, Kap. 7). Wird bei der Schätzung der Personenfähigkeit zusätzlich die Reaktionszeit berücksichtigt, treten diese Probleme nicht auf. Die Rolle der Reaktionszeit ist zudem nicht nur auf die Verbesserung der Fähigkeitsschätzung beschränkt. Ein populäres Thema in der angewandten Diagnostik ist die Behandlung und Identifikation von „schnellen Ratern“ bei Tests mit geringer persönlicher Bedeutung für die Testperson (Kong et al., 2007; Wise & Kong, 2005). Ein weiteres Problem stellen schnelle Rateantworten dar, die unter Zeitdruck gegen Ende des Tests gegeben werden (Goebel et al., 2007; Yamamoto, 1995). Bei der gemeinsamen Modellierung von Itemantwort und Reaktionszeit könnten sich unter Umständen (analog zu IRT-Modellen) Personen-Fit-Statistiken definieren lassen, mit denen atypische Reaktionen nicht nur bezüglich der Itemantwort sondern auch bezüglich der Reaktionszeit identifiziert werden könnten. Die vorgeschlagenen Reaktionszeitmodelle besitzen also weit mehr Anwendungsmöglichkeiten als eine reine Reduktion des Schätzfehlers.

# Kapitel 8

## Diskussion der Ergebnisse

Die Fähigkeit zur präzisen Messung menschlicher Eigenschaften ist eine Grundvoraussetzung für die psychologische Forschung und Diagnostik: Diagnostische Entscheidungen sind oft mit weitreichenden Konsequenzen verbunden und sollten daher auf fehlerfreien und validen Messungen entscheidungsrelevanter Persönlichkeitseigenschaften basieren. Der Nachweis von Kausalbeziehungen zwischen theoretischen Konstrukten setzt immer eine gültige Operationalisierung voraus. Es ist daher nicht verwunderlich, dass die Frage nach der Messbarkeit von Konstrukten und nach der Messgenauigkeit von Messinstrumenten eine Vielzahl von Forschungsarbeiten hervorgebracht hat. Die vermutlich wichtigste Entwicklung war die Einführung der Item-Response-Modelle, welche die Auswertung von standardisierten Tests mit Hilfe eines Messmodells ermöglichen. Allerdings wird bei den gängigen IRT-Modellen nur ein Teilaspekt des Verhaltens einer Person bei den Testaufgaben zur Diagnostik genutzt, nämlich die Itemantworten der Person. Vernachlässigt wird bisher, bis auf wenige Ausnahmen (Ferrando & Lorenzo-Sevas, 2007a; van der Linden, 2008), die potentielle Information, die in den Reaktionszeiten einer Person bei der Bearbeitung der Testaufgaben stecken könnte. Genau um diese potentielle Information der Reaktionszeiten über die Eigenschaftsausprägung einer Person geht es in der aktuellen Arbeit. Tatsächlich sprechen einige Veröffentlichungen dafür, dass nicht nur die Itemantworten, sondern auch die Reaktionszeiten mit dem zu messenden Konstrukt in Beziehung stehen. Eine Übersicht über die bisherigen Befunde, die diese Vermutung stützen, wurde im ersten Kapitel gegeben. Hervorzuheben sind zwei Beobachtungen:

1. Im Leistungsbereich finden sich Hinweise darauf, dass Maße der Personenfähigkeit mit der Reaktionszeit bei Intelligenzaufgaben korreliert sind. Hohe Korrelationen scheinen besonders bei einfachen Aufgaben aufzutreten, mit denen zwischen Personen hoher Fähigkeit nur schlecht diskriminiert werden kann.
2. Bei Selbstbeurteilungsaufgaben, wie sie in Einstellungs- oder Persönlichkeitsfragebögen vorkommen, können Urteile desto schneller abgegeben werden, je eindeutiger für eine Person die Entscheidung ist, also je stärker die Aussage von der eigenen Position abweicht.

Diese Beobachtungen waren der Ausgangspunkt für die zunächst generelle Frage, wie die Reaktionszeiten in die Schätzung von Eigenschaftsausprägungen einbezogen werden können und mit welchem Nutzen dieser Einbezug verbunden ist. Hierzu war jedoch erst eine Präzisierung der Fragestellung notwendig, um den Nutzenbegriff genauer zu definieren. Oft wird als Bewertungskriterium eines Messinstrumentes der durchschnittliche quadrierte Schätzfehler (Mean Squared Error) herangezogen. Dieses Kriterium hängt aber von der Varianz des Messergebnisses bei wiederholter Messung derselben Eigenschaftsausprägung ab. Das war der Grund dafür, unter dem Nutzen der Reaktionszeiten die durch die Berücksichtigung der Reaktionszeiten bewirkte Varianzreduktion des Eigenschaftsschätzers zu verstehen.

Die populären IRT-Modelle basieren auf der Annahme, dass eine Persönlichkeitseigenschaft, die Personenfähigkeit  $\theta$ , die Wahrscheinlichkeit einer Person beeinflusst, auf eine Testaufgabe positiv zu antworten. Die genaue Form der Beziehung zwischen der Personenfähigkeit und der Wahrscheinlichkeit einer positiven Antwort wird in der Itemcharakteristik ausgedrückt. Bei Annahme eines IRT-Modells kann anhand der Itemantworten einer Person auf deren Fähigkeitsausprägung zurück geschlossen werden. Ein populäres Schätzverfahren hierfür ist der Maximum-Likelihood-Ansatz. Die Präzision der Messung spiegelt sich in der Varianz der Fähigkeitsschätzung wider, die sich bei einer Wiederholung der Messung ergäbe. Diese Varianz lässt sich für Maximum-Likelihood-Schätzer berechnen. Diese ergibt sich aus der Fisher-Information des Tests über die unbekannte Fähigkeitsausprägung einer Person. Analog zu den IRT-Modellen für die Itemantwort lassen sich Modelle für die Reaktionszeiten bei Testaufgaben aufstellen. Grundgedanke ist dabei die Annahme, dass die Reaktionszeitverteilung ebenfalls von der Personenfähigkeit einer Person abhängt. Allerdings sprechen empirische Befunde dafür, dass noch mindestens eine weitere Persönlichkeitseigenschaft die Verteilung der Reaktionszeiten einer Person beeinflussen dürfte. Diese zweite systematische Einflussquelle wurde als Residualfaktor bezeichnet. Die genaue Beziehung zwischen der Reaktionszeitverteilung und den Persönlichkeitseigenschaften wird in einem Reaktionszeitmodell festgelegt. Geht man von einem bestimmten Reaktionszeitmodell aus, lassen sich anhand des Reaktionszeitmusters einer Person bei mehreren Testitems deren Ausprägungen auf den Persönlichkeitseigenschaften schätzen. Der Nutzen der Reaktionszeiten zeigt sich in dem Ausmaß, in dem die Varianz der Fähigkeitsschätzung durch die Berücksichtigung der Reaktionszeiten verringert wird. Im siebten Kapitel konnte gezeigt werden, dass unter relativ allgemeinen Bedingungen eine Berücksichtigung der Reaktionszeiten immer zu einer Reduktion der Varianz führt. Dieses Kapitel ist somit der eigentliche Kern der Arbeit und soll im Folgenden diskutiert werden im Hinblick auf 1) eine kritische Bewertung der Modellannahmen und 2) einer Bewertung der Ergebnisse für die Praxis.

## 8.1 Bewertung der Modellannahmen

Die Angemessenheit der aufgestellten Modelle für einen bestimmten Anwendungsbereich ist letztendlich eine empirische Frage, die nur im konkreten Anwendungsfall geklärt werden kann. Dennoch sollen die Modelle einer genaueren theoretischen Bewertung unterzogen werden, da eine kritische Analyse Hinweise

auf Verbesserungsmöglichkeiten und Perspektiven für zukünftige Forschungsarbeit ergeben kann. Betrachtet werden sollen zwei Aspekte, die Angemessenheit der einzelnen Reaktionszeitmodelle und die Plausibilität der gemeinsamen Modellierung von Reaktionszeit und Itemantwort.

### **Bewertung der Modellierung der Reaktionszeiten**

Mit welchem Verteilungsmodell Reaktionszeiten bei Tests modelliert werden können, ist eine noch ungeklärte Frage. Vorgeschlagen wurde bisher eine Vielzahl unterschiedlicher Alternativen – für eine Übersicht über die einzelnen Vorschläge sei auf die Übersicht im vierten Kapitel verwiesen. Tatsächlich ist die Frage nach der angemessenen Reaktionszeitverteilung keine theoretische Frage, sondern eine empirische, die sicher nicht allgemein geklärt werden kann: Vermutlich erfordern verschiedenartige Gegenstandsbereiche auch verschiedene Verteilungsmodelle. Dies schränkt die Generalität der bisher veröffentlichten Ansätze erheblich ein, da bei den vorgeschlagenen Reaktionszeitmodellen immer von einer bestimmten Verteilungsform ausgegangen wird. Die aktuelle Arbeit löst einen Teil dieses Problems dadurch, dass zwei verteilungsfreie Alternativen zur Modellierung der Reaktionszeiten vorgeschlagen werden, das Proportional Hazard Modell für dichotomisierte Reaktionszeiten sowie der Quasi-Likelihood-Ansatz für Accelerated Failure Time Modelle. Wenn bei diesen Ansätzen auch die genaue Verteilungsform der Reaktionszeiten keine Rolle spielt, stehen und fallen die Modelle mit der Gültigkeit der Regressionsbeziehung zwischen den Persönlichkeitseigenschaften und der Hazard Rate bzw. zwischen den Persönlichkeitseigenschaften und dem Erwartungswert der logarithmierten Reaktionszeiten. Die Gültigkeit der Regressionsbeziehung ist somit die „*Conditio sine qua non*“, ohne die eine Schätzung der Persönlichkeitseigenschaften anhand der Reaktionszeiten nicht möglich ist. Eine noch weitere Einschränkung der Annahmen scheint zum aktuellen Zeitpunkt nicht möglich zu sein.

Die in der Arbeit vorgeschlagenen Reaktionszeitmodelle wurden eingeführt, ohne die postulierten Beziehungen mit einer psychologischen Theorie zu begründen: Immer handelt es sich um Post-Hoc-Modelle, die nicht von fundierten psychologischen Modellvorstellungen über den Problemlöseprozess abgeleitet wurden. Tatsächlich ist dies auch ein Vorteil, da die Gültigkeit der Modelle somit nicht an einen konkreten Anwendungsbereich gebunden ist. In dieser Eigenschaft unterscheiden sich die Reaktionszeitmodelle nicht von den IRT-Modellen, die ebenfalls nicht auf psychologischen Problemlösemodellen basieren. Aus einer rein diagnostischen Perspektive betrachtet ist eine fundierte Begründung der verwendeten Messmodelle nicht erforderlich: Zur Temperaturmessung werden Quecksilberthermometer seit 300 Jahren eingesetzt, ohne dass zu Nutzungsbeginn der genaue thermodynamische Mechanismus verstanden wurde.

### **Bewertung der gemeinsamen Modellierung von Reaktionszeiten und Itemantworten**

Bei Annahme der bedingten Unabhängigkeit der Reaktionszeiten  $\mathbf{t}$  und der Itemantworten  $\mathbf{x}$  lässt sich die gemeinsame Verteilung der Reaktionszeiten und Itemantworten modellieren als  $f(\mathbf{x}, \mathbf{t}|\theta, \omega) = \prod_g (x_g|\theta) f(t_g|\theta, \omega)$ . Angenommen

wird somit, dass bei jedem Item die Reaktionszeit einer Person unabhängig von der Itemantwort verteilt ist und dass sich Itemantworten und Reaktionszeiten bei verschiedenen Items nicht gegenseitig beeinflussen. Die Annahme der bedingten Unabhängigkeit kann mit dem Common Cause Criterion gerechtfertigt werden, also mit der Behauptung, dass die Abhängigkeit von Phänomenen durch die Berücksichtigung der gemeinsamen Ursachen vollständig erklärt wird (Suppes & Zanotti, 1981). Diese Annahme wird bei der gemeinsamen Modellierung der Itemantworten und Reaktionszeiten häufig verwendet, siehe beispielsweise van der Linden (2007), Ferrando und Lorenzo-Sevas (2007a) oder Thissen (1984). Diese Art der Modellierung ist aber nicht zwingend. Ein besonders strittiger Punkt dürfte die Unabhängigkeit von Reaktionszeit und Itemantwort bei einem Item sein.

Nach Bloxom (1985) lässt sich die gemeinsame Verteilung der Reaktionszeit und Itemantwort bei einer Testaufgabe auf zwei Arten<sup>1</sup> modellieren: 1.) Durch die Angabe der marginalen Zeitverteilung und der auf eine bestimmte Reaktionszeit bedingten Verteilung der Itemantwort:  $f(x, t|\theta, \omega) = p(x|t, \theta)f(t|\theta, \omega)$ . Vertreter dieses Ansatzes sind beispielsweise van Breukelen und Roskam (1991), Roskam (1987), Verhelst et al. (1997) und Wang (2006). 2.) Eine alternative Modellierungsstrategie besteht in der Angabe der marginalen Verteilung der Itemantwort und der auf die Itemantwort bedingten Verteilung der Reaktionszeit:  $f(x, t|\theta, \omega) = f(t|x, \theta, \omega)p(x|\theta)$ . Dieser Ansatz wurde von Thissen (1984) angeregt und von van Breukelen (2005) in leicht veränderter Form aufgegriffen. Der Vorteil des zweiten Ansatzes besteht darin, dass die bewährten IRT-Modelle ohne Modifikationen zur Modellierung der Itemantworten herangezogen werden können. Beim ersten Ansatz hingegen besteht eine Herausforderung darin, geeignete Kombinationen der Verteilungen  $p(x|t, \theta)$  und  $f(t|\theta, \omega)$  zu finden, so dass die marginale Verteilung der Itemantwort  $p(x|\theta, \omega) = \int p(x|t, \theta)f(t|\theta, \omega) dt$  ähnliche Eigenschaften besitzt, wie die bewährten IRT-Modelle.

Tatsächlich lassen sich Hinweise darauf finden, dass die Reaktionszeitverteilung von der Art der Itemantwort abhängen könnte. Mit Ausnahme schneller Rateantworten scheinen Personen bei Leistungstests für falsche Itemantworten mehr Zeit zu benötigen als für richtige Itemantworten. Allerdings ist die Interpretation dieses Befundes oft nicht eindeutig, da in vielen Studien nicht berücksichtigt wird, dass richtige Antworten in der Regel von fähigeren und unter Umständen schnelleren Personen stammen. Zudem sind falsche Antworten bei schwierigeren Items mit längerer Bearbeitungszeit häufiger. Die Frage, ob die Reaktionszeiten bei richtigen und falschen Antworten unterschiedlich verteilt sind, lässt sich daher nicht endgültig beantworten. Auch für den Bereich der Persönlichkeitsdiagnostik gibt es Hinweise darauf, dass die Reaktionszeitverteilung von der Itemantwort abhängen könnte. Fekken, Popham und Holden konnten in einer Serie von Untersuchungen (Fekken & Holden, 1992; Popham & Holden, 1990; Holden et al., 1991; Holden & Fekken, 1993) zeigen, dass Personen mit hohen Eigenschaftsausprägungen zu eigenschaftsindikativen (d.h. positiv gepolten) Items schneller zustimmen können als Personen mit geringer Eigenschaftsausprägung. Eine Ablehnung von eigenschaftsindikativen, positiv gepolten Aussagen jedoch erfolgt bei diesen Personen langsamer als bei Per-

<sup>1</sup>Diese beiden Strategien finden sich auch in der Überlebenszeitanalyse für die gemeinsame Modellierung wiederholter Messungen und des Eintrittszeitpunktes eines Ereignisses, siehe hierfür zum Beispiel Hogan und Laird (1997).

sonen mit geringer Eigenschaftsausprägung, so dass der Befund nicht damit erklärt werden kann, dass Personen mit hoher Eigenschaftsausprägung generell schneller reagieren. Die Autoren erklären diesen Befund mit dem Einfluss von Schemata auf die Informationsverarbeitung. Auf den ersten Blick scheint das Ergebnismuster tatsächlich in Widerspruch zu einer Unabhängigkeit von Itemantwort und Reaktionszeit zu stehen. Allerdings könnte dieser Befund auch bei einer Unabhängigkeit von Itemantwort und Reaktionszeit auftreten, wenn man ein Unfolding-Modell für die Reaktionszeiten annimmt: Personen mit hoher Eigenschaftsausprägung am rechten Ende der Skala akzeptieren in der Regel alle Items am linken Skalenende. Aufgrund der großen Distanz zwischen der Person und dem Item kann die Entscheidung zudem schnell erfolgen. Items am rechten Skalenende in der Nähe der Person werden teilweise akzeptiert und teilweise abgelehnt. Die Reaktionszeit ist dabei aufgrund der geringeren Distanz von Person und Item jedoch langsamer. Mittelt man nun die Reaktionszeiten aller akzeptierten und abgelehnten Items, so tritt das beobachtete Befundmuster auf: Eine Zustimmung erfolgt bei Personen mit hoher Eigenschaftsausprägung im Durchschnitt schneller als eine Ablehnung. Die entscheidende Frage, die aber anhand der publizierten Studien alleine nicht beantwortet werden kann, lautet, ob die Distanz zwischen Person und Item oder die Wahrscheinlichkeit der von einer Person abgegebenen Itemantwort in monotoner Beziehung zur Reaktionszeit steht. Anders ausgedrückt: Wenn eine Person ein Item stark dominiert, ist dann eine Zustimmung genauso schnell wie eine Ablehnung (Voraussage des Unfolding-Modells) oder ist die Zustimmung schneller als die Ablehnung (Voraussage von Holden und Mitarbeitern). Sollte der Befund von Popham und Holden auch bei einer Kontrolle der Distanz zwischen Item und Person auftreten, wäre dies in der Tat ein Hinweis darauf, dass die Unfolding-Modelle unter Umständen nicht adäquat sind, sondern von der Vorstellung einer schemagesteuerten Itembeantwortung auszugehen ist. Es wäre in diesem Fall nicht die Distanz zwischen Person und Item, welche in monotoner Beziehung zur Reaktionszeit stünde, sondern vielmehr die Wahrscheinlichkeit der abgegebenen Itemantwort: Wahrscheinliche Itemantworten (Hohe Eigenschaftsausprägung – Zustimmung, Niedrige Eigenschaftsausprägung – Ablehnung) erfolgen dann schneller als unwahrscheinliche Itemantworten (Hohe Eigenschaftsausprägung – Ablehnung, Niedrige Eigenschaftsausprägung – Zustimmung). Psychologisch plausibel wäre auch dieses Alternativmodell, da „untypische“ Itemantworten auf zusätzlichem kognitiven Aufwand beruhen könnten. Bekannt ist, dass Entscheidungen unter Konflikt mit längerer Entscheidungszeit einhergehen (Diederich, 2003). Aber auch unter der Perspektive des Informationsgewinns besitzt die Annahme einer monotonen Beziehung zwischen der Wahrscheinlichkeit der gegebenen Antwort und der dafür benötigten Reaktionszeit ihren Charme. Immer wieder wurde darauf hingewiesen, dass eine dichotome Itemantwort weniger Information besitzt als ein kontinuierliches Antwortformat – siehe hierzu beispielsweise die Randbemerkungen im dritten Kapitel (S. 26). Wenn die Reaktionszeit als kontinuierliche Größe in Beziehung zur Wahrscheinlichkeit der abgegebenen Antwort steht, besitzt sie erheblich mehr Information über die Personenfähigkeit als die dichotome Itemantwort. Die Untersuchung eines Modells, bei welchem die Reaktionszeit zur Wahrscheinlichkeit der gegebenen Antwort in Beziehung gesetzt wird, könnte somit eine interessante Alternative zu dem gewählten Ansatz dieser Arbeit darstellen.

Der vorherige Abschnitt drehte sich um die Frage, wie die Reaktionszeit bei Bedingung auf die Itemantwort verteilt ist. Alternativ kann man auch annehmen, dass die Itemantwort von der Reaktionszeit abhängt. Teilweise wird vermutet, dass positive Itemantworten bei kurzen Reaktionszeiten weniger wahrscheinlich sind, als bei langen Itembearbeitungszeiten, siehe zum Beispiel Wang (2006) oder Roskam (1987). Dies mag bei Leistungstests plausibel sein und wird oft mit dem Speed-Accuracy-Trade-Off begründet. Wie jedoch von van der Linden (2007) festgestellt wurde, entscheiden sich Personen vor der Testbearbeitung für ein bestimmtes Geschwindigkeits- und Fähigkeitsniveau, so dass mit dem Speed-Accuracy-Trade-Off Abhängigkeiten innerhalb einer Person nicht erklärt werden können. Bei Einstellungstests hingegen kann vermutet werden, dass bei kurzen Reaktionszeiten die Wahrscheinlichkeit einer positiven Reaktion (Zustimmung) größer ausfällt. Auf Spinoza geht die These zurück, dass eine Aussage im Augenblick des Verständnisses automatisch akzeptiert wird und erst zusätzliche Reflexionen notwendig sind, damit die Aussage als falsch zurückgewiesen werden kann (Gerard, 1997). Die Frage nach der Existenz eines „Speed-Accuracy-Trade-Offs“ wird jedoch innerhalb der Einstellungsdiagnostik selten diskutiert. Ein Gedanke ist jedoch auch für Einstellungs- und Persönlichkeitstests erwägenswert. Wie im dritten Kapitel gezeigt wurde, kann bei den IRT-Modellen eine kontinuierliche, latente Variable angenommen werden, welche den binären Itemantworten zugrunde liegt. Während der Erwartungswert der latenten Variablen von der Personenfähigkeit abhängt, wird die Varianz der latenten Variablen nicht modelliert, sondern als fix angenommen. Von Haaijer et al. (2000) stammt die Idee, diese Varianz auf die Bearbeitungszeit einer Person zurück zu führen. Dies bedeutet, dass Itemantworten bei kurzer Bearbeitungszeit weniger stabil sind, als bei langer Bearbeitungszeit. Eine derartige Modellierung steht jedoch in Widerspruch zu den empirischen Befunden, dass sichere Entscheidungen mit kurzen Reaktionszeiten einhergehen. Aus diesem Grund wird die Reaktionszeit auch als Maß der Einstellungsstärke aufgefasst (Fazio, 1995). Dennoch ist der Vorschlag von Haaijer et al. (2000) eine Alternative, die in zukünftigen Forschungsarbeiten aufgegriffen werden sollte.

In der aktuellen Arbeit wurden Ansätze vorgeschlagen, wie Reaktionszeiten bei der Diagnostik berücksichtigt werden könnten. Vorgestellt wurden mehrere Reaktionszeitmodelle, mit denen ein möglichst breiter Anwendungsbereich abgedeckt werden sollte. Ein Teil der vorgeschlagenen Modelle gilt unabhängig von der genauen Verteilungsform der Reaktionszeiten, jedoch müssen restriktive Annahmen über die gemeinsame Verteilung der Itemantwort und Reaktionszeit getroffen werden. Es stellt sich die Frage, ob derartige Verteilungsannahmen nicht gänzlich vermieden werden können. In der Medizin wird teilweise der Ansatz der „Quality Adjusted Life-Years“ verfolgt. Eine Einführung in das Konzept findet sich beispielsweise bei Glasziou, Cole, Gelber, Hilden und Simes (1998). Bei diesem Ansatz wird die Zeit bis zum Tode einer Person mit dem Nutzen der Zeit gewichtet. Diese Vorgehensweise kann auf den Bereich von Leistungstests übertragen werden, wenn man den Nutzen einer Antwort von der Art der Antwort und vom Eintrittszeitpunkt abhängig macht. Eine richtige Antwort besitzt desto größeren Nutzen, je schneller die Antwort erfolgt. Durch die Modellierung des Nutzens einer Reaktion entfällt die Notwendigkeit einer gemeinsamen Modellierung von Reaktionszeit und Itemantwort. Allerdings setzt diese Vorgehensweise die oft beliebige Definition einer Nutzenfunktion voraus, welche nur



in den seltensten Fällen begründet werden kann. Bei lebensnahen Testaufgaben, wie sie teilweise in Assessment Centern eingesetzt werden und bei denen die Bearbeitungszeit teilweise einen wirklichen Wert besitzt, ist der Ansatz aber durchaus erwägenswert.

Ob die bedingte Unabhängigkeit von Reaktionszeit und Itemantwort bei einem Test gilt, ist eine empirische Frage, die theoretisch nicht geklärt werden kann. Die Übersicht über alternative Modellierungsstrategien soll daher nicht als Zweifel an der bedingten Unabhängigkeit gedeutet werden, sondern sind eher als Anregung für zukünftige Forschungsarbeiten zu verstehen. Während die Annahme der bedingten Unabhängigkeit bei Leistungstests hinterfragt werden kann, finden sich im Einstellungs- und Persönlichkeitsbereich wenige Einwände dagegen. Und es ist gerade der Einstellungs- und Persönlichkeitsbereich, in dem Reaktionszeiten besonderes diagnostisches Potential besitzen könnten, siehe hierzu die Befunde im siebten Kapitel.

## 8.2 Bewertung des Nutzens für die diagnostische Praxis

Die Frage nach der Plausibilität der vorgeschlagenen Modellierung behandelt die mögliche Einsetzbarkeit der Modelle in der Praxis. Realistische Modellannahmen sind daher eine notwendige Voraussetzung für den Erfolg einer Modellanpassung. Doch selbst wenn eines der vorgeschlagenen Reaktionszeitmodelle tatsächlich die Verteilung der Bearbeitungszeiten mit einer ausreichenden Genauigkeit beschreiben sollte, bedeutet dies nicht notwendigerweise, dass die Möglichkeit, die Reaktionszeiten modellieren zu können, mit einem praktischen Nutzen verbunden ist. Diese Frage nach dem praktischen Nutzen der Modelle soll im Folgenden erörtert werden. Zu betrachten sind dabei die Kosten und der Gewinn bei einer Berücksichtigung der Reaktionszeiten.

Um die Reaktionszeitmodelle zur Messung von Eigenschaften einsetzen zu können, ist eine vorherige Schätzung der Itemparameter notwendig. Für die Schätzung wurden in der aktuellen Arbeit verschiedene Verfahren entwickelt. Bei einer Überprüfung der Leistungsfähigkeit der Verfahren mit Simulationen zeigte sich, dass für eine akkurate Schätzung der Itemparameter große Stichproben benötigt werden. Eine passable Schätzung ist bei kurzen Tests erst ab etwa 500 Versuchspersonen möglich. Dies scheint auf den ersten Blick gegen einen zusätzlichen Einbezug der Reaktionszeiten zu sprechen. Allerdings benötigt man bereits für die alleinige Anpassung eines IRT-Modells eine große Eichstichprobe. Zudem konnte gezeigt werden, dass durch die gemeinsame Schätzung von IRT-Modell und Reaktionszeitmodell die Schätzung der Itemparameter präziser erfolgt, als bei einer alleinigen Anpassung des IRT-Modells. Durch die Modellierung der Reaktionszeiten werden also keine höheren Anforderungen an den Stichprobenumfang gestellt als bei der Anpassung eines IRT-Modells.

Unter dem Nutzen der Reaktionszeit wurde in der aktuellen Arbeit der Beitrag der Reaktionszeiten zur Reduktion der Varianz des Fähigkeitsschätzers verstanden. Dieser Aspekt ist sicherlich ein sehr wichtiger, jedoch besitzt die Berücksichtigung der Reaktionszeit weitere Vorteile, die nur am Rande erwähnt

wurden. Beispielsweise erlaubt der Einbezug der kontinuierlichen Reaktionszeit eine Unterscheidung von Personen mit denselben Itemantworten, was bei IRT-Modellen nicht möglich ist. Mit Reaktionszeiten können zudem schnelle Räteantworten identifiziert werden. Die Reduktion der Varianz ist somit nur ein Nutzenaspekt unter vielen anderen. Aufgrund seiner großen Bedeutung stand dieser Aspekt jedoch im Vordergrund der Betrachtungen. Das Ausmaß, in dem sich die Varianz des Fähigkeitsschätzers durch die Verwendung der Reaktionszeiten reduzieren lässt, kann natürlich nicht generell bestimmt werden, da der Betrag von den Parametern der Items abhängt. Dennoch lassen sich mehrere Dinge feststellen: Die Modelle für kontinuierliche Reaktionszeiten besitzen sehr günstige Iteminformationsfunktionen: Entweder sind die Iteminformationsfunktionen konstant oder sie nehmen mit extremer werdenden Eigenschaftsausprägung zu. Dadurch ergänzen sie die Informationsfunktionen zweiparametrischer IRT-Modelle optimal, bei denen die Iteminformation mit extremer werdender Personenfähigkeit absinkt. Dieses Verhalten spiegelt sich in der Varianz des Fähigkeitsschätzers wider. Vergleicht man bei diesen Modellen die Varianz des Fähigkeitsschätzers mit und ohne Berücksichtigung der Reaktionszeiten, kann man sehen, dass der wesentliche Beitrag der Reaktionszeiten zur Fähigkeitsschätzung in einer Verbesserung der Schätzung extremer Fähigkeiten besteht. Der Vorteil sich ergänzender Iteminformationsfunktionen geht bei den Modellen für dichotomisierte Reaktionszeiten verloren, da bei diesen die Iteminformation ebenfalls mit extremer werdender Eigenschaftsausprägung absinkt. Dieser ungünstige Informationsverlauf ist der Dichotomisierung der Reaktionszeiten geschuldet, die zu einem Informationsverlust führt. Dennoch verbessert der Einbezug der Reaktionszeiten auch bei den Proportional Hazard Modelle für dichotomisierte Reaktionszeiten die Schätzung extremer Fähigkeitsausprägungen, wie man anhand der Abbildungen im siebten Kapitel sieht. Zudem kann bei den Modellen für dichotomisierte Reaktionszeiten die Varianzreduktion gesteigert werden: Durch Wahl eines geeigneten Zeitlimits bei der Dichotomisierung der Reaktionszeiten kann die Lage des Informationsmaximums beeinflusst werden, was die Konstruktion maßgeschneiderter Tests erlaubt. Auch kann das Proportional Hazard Modell für dichotomisierte Reaktionszeiten erweitert werden, indem die Reaktionszeiten nicht dichotomisiert, sondern in mehrere Kategorien eingeteilt werden. Dadurch steigt zwar die Itemparameteranzahl an, da nun mehrere Konstanten geschätzt werden müssen, jedoch ändert sich das generelle Prinzip nicht. Die feinere Auflösung der Reaktionszeiten führt zu größerer Iteminformation und folglich zu größerer Varianzreduktion. Alternativ ist es möglich, bei Proportionalen Hazard Modellen die Grundhazard-Rate über Splines zu approximieren oder für Zeitabschnitte eine konstante, vom Zeitabschnitt abhängige Hazard-Rate anzunehmen. Auch diese Modellierungsansätze sollten den Informationsgewinn durch die Reaktionszeiten erhöhen und sind ein vielversprechendes Gebiet zukünftiger Forschung.

Die Testinformationsfunktion steht in Bezug zur asymptotischen Varianz des Maximum-Likelihood-Schätzers der Personenfähigkeit, wenn die wahren Itemparameter bekannt sind und der Testumfang unbegrenzt anwächst. Über das Verhalten des Fähigkeitsschätzers bei kleinen Stichproben kann somit keine Aussage gemacht werden, so lange nicht die Eigenschaften der Schätzer bei kleinen Stichproben untersucht werden, wie dies bereits für die IRT-Modelle unternommen wurde (Lord, 1983). Unbekannt ist auch, welchen Einfluss die fehlerhafte

Schätzung der Itemparameter auf die Fähigkeitsmessung ausübt. Tatsächlich gilt die Beziehung der Testinformation zur Varianz des Fähigkeitsschätzers nur bei Kenntnis der wahren Itemparameter. Welchen Effekt die zusätzliche Itemparameterschätzung auf die Fähigkeitsschätzung besitzt, muss in weiteren Forschungsarbeiten geklärt werden. Aus dem IRT-Bereich ist bekannt, dass die zusätzliche Schätzung der Itemparameter bei kleinen Eichstichproben zu einer beträchtlichen Vergrößerung des mittleren quadrierten Fehlers führen kann (Tsutakawa & Johnson, 1990).

In der aktuellen Arbeit ging es darum, die Varianz bei der Schätzung der Personenfähigkeit zu reduzieren. Die Verwendung der Reaktionszeiten ist hierfür zwar nahe liegend, jedoch nicht konkurrenzlos. Wenn soziodemographische Variablen in Beziehung zur Personenfähigkeit stehen, kann eine Fehlerreduktion auch dadurch erreicht werden, dass diese Variablen in die Fähigkeitsschätzung einbezogen werden (Adams, Wilson & Wu, 1997; Mislevy, 1987). Die Frage nach dem Nutzen der Reaktionszeit ist somit immer relativ: Lohnt sich der Aufwand einer expliziten Modellierung der Reaktionszeiten im Vergleich zum Einbezug von zusätzlichen, soziodemographischen Variablen? Auch diese Frage kann natürlich nicht generell beantwortet werden und ist gesondert zu untersuchen. Ein Vorteil der Verwendung von Reaktionszeiten gegenüber der Verwendung von soziodemographischen Variablen dürfte aber die möglicherweise größere Akzeptanz sein, da die Fähigkeitsschätzung bei Nutzung der Reaktionszeiten nur vom eigenen Verhalten abhängt und nicht von der unbeeinflussbaren Zugehörigkeit zu soziodemographischen Gruppen.

### 8.3 Ausblick

Ganz am Anfang dieser Arbeit wurde die Frage nach dem diagnostischen Nutzen von Reaktionszeiten gestellt. Um diese Frage zu beantworten, wurden verschiedene Reaktionszeitmodelle auf ihren potentiellen Nutzen hin untersucht. Welchen Nutzen die vorgeschlagenen Reaktionszeitmodelle in der Praxis besitzen, kann durch die rein theoretische Untersuchung nicht beantwortet werden – hierzu ist empirische Forschungsarbeit notwendig. Doch selbst wenn unter dem Aspekt des praktischen Nutzens die Verwendung von Reaktionszeiten im Zusammenhang mit der Schätzung von Persönlichkeitseigenschaften nicht abschließend geklärt werden kann, so zeigt die Arbeit doch das Potential, welches die Reaktionszeit als komplementäre Information zur Itemantwort besitzt. Dass bisher nur wenige Versuche unternommen wurden, diese Potential auch tatsächlich zu nutzen, mag vermutlich auch an der Schwierigkeit gelegen haben, die Reaktionszeiten bei der Bearbeitung von Tests aufzuzeichnen. Durch die Verfügbarkeit computeradaptiver Tests wird die Aufzeichnung von Reaktionszeiten jedoch erheblich erleichtert. Es ist daher zu hoffen, dass in der nahen Zukunft Reaktionszeiten in der Diagnostik eine wichtige Rolle spielen werden.

Teil III

Anhang

# Anhang A

## Momentmethoden-Schätzer ausgewählter RT-Modelle

### A.1 Modelle für exponentialverteilte Reaktionszeiten

#### A.1.1 Lineare Parametrisierung

Nach Modell (4.7) sind die Reaktionszeiten einer Person mit Personenfähigkeit  $\theta$  und Residualfaktor  $\omega$  bei Item  $g$  exponentialverteilt mit der Verteilungsdichte:

$$f(t_g|\theta, \omega) = \lambda_g(\theta, \omega) \exp[-\lambda_g(\theta, \omega)t_g] \quad \text{mit} \quad \log[\lambda_g(\theta, \omega)] = k_g + l_g\theta + c_g\omega. \quad (\text{A.1})$$

Die Schätzung der Einflussgewichte  $l_g$  und  $c_g$  sowie der Konstanten  $k_g$  kann über die Methode der Momente erfolgen, wenn man davon ausgeht, dass die Testpersonen aus einer Population gezogen werden, in welcher die Persönlichkeitseigenschaften  $\theta$  und  $\omega$  bivariat standardnormalverteilt sind mit Korrelation  $\rho = 0$ . Hierzu benötigt man die ersten beiden marginalen nonzentralen Momente. Diese lassen sich nach dem Gesetz der iterierten Erwartungswertbildung berechnen (Casella & Berger, 2002). Dieses Gesetz besagt, dass:

$$E[t_g] = E_{\theta, \omega}[E[t_g|\theta, \omega]]. \quad (\text{A.2})$$

Bei Bedingung auf die Persönlichkeitseigenschaften  $\theta$  und  $\omega$  ist die Reaktionszeit exponentialverteilt mit Verteilungsparameter  $\lambda_g(\theta, \omega)$ . Der bedingte Erwartungswert dieser Exponentialverteilung ist eine Funktion des Verteilungsparameters  $\lambda_g(\theta, \omega)$ . Es gilt:  $E[t_g|\theta, \omega] = 1/\lambda_g(\theta, \omega)$ . Der marginale Erwartungswert der Reaktionszeit von Item  $g$  bei bivariat standardnormalverteilten Persönlichkeitseigenschaften  $\theta$  und  $\omega$  lässt sich nach (A.2) berechnen als:

$$E[t_g] = E_{\theta, \omega} [E[t_g | \theta, \omega]] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\lambda_g(\theta, \omega)} \frac{1}{2\pi} \exp\left[-\frac{1}{2}\theta^2\right] \exp\left[-\frac{1}{2}\omega^2\right] d\theta d\omega. \quad (\text{A.3})$$

Bei Verwendung der Beziehung  $\lambda_g(\theta, \omega) = \exp[k_g + l_g\theta + c_g\omega]$  ergeben Umformungen:

$$\begin{aligned} E[t_g] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\exp[k_g + l_g\theta + c_g\omega]} \frac{1}{2\pi} \exp\left[-\frac{1}{2}\theta^2\right] \exp\left[-\frac{1}{2}\omega^2\right] d\theta d\omega \\ &= \exp[-k_g] \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi} \exp\left[-\frac{1}{2}(\theta^2 + 2l_g\theta)\right] \exp\left[-\frac{1}{2}(\omega^2 + 2c_g\omega)\right] d\theta d\omega \\ &= \exp\left[-k_g + \frac{1}{2}l_g^2 + \frac{1}{2}c_g^2\right] \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi} \exp\left[-\frac{1}{2}(\theta + l_g)^2\right] \exp\left[-\frac{1}{2}(\omega + c_g)^2\right] d\theta d\omega \\ &= \exp\left[-k_g + \frac{1}{2}l_g^2 + \frac{1}{2}c_g^2\right]. \end{aligned} \quad (\text{A.4})$$

Auch das zweite marginale nonzentrale Moment lässt sich mit der Methode der iterierten Erwartungswertbildung berechnen. Bei Bedingung auf die Persönlichkeitseigenschaften  $\theta$  und  $\omega$  besitzt die bedingte Verteilung der Reaktionszeit  $t_g$  das zweite nonzentrale Moment  $E[t_g^2 | \theta, \omega] = 2/\lambda_g^2(\theta, \omega)$ . Das zweite marginale nonzentrale Moment ergibt sich dann als:

$$E[t_g^2] = E_{\theta, \omega} [E[t_g^2 | \theta, \omega]] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{2}{\lambda_g^2(\theta, \omega)} \frac{1}{2\pi} \exp\left[-\frac{1}{2}\theta^2\right] \exp\left[-\frac{1}{2}\omega^2\right] d\theta d\omega. \quad (\text{A.5})$$

Eine quadratische Ergänzung von (A.5), deren einzelnen Schritte analog zur Vorgehensweise bei (A.4) sind, führt zur Lösung:

$$E[t_g^2] = 2 \exp[-2k_g + 2l_g^2 + 2c_g^2]. \quad (\text{A.6})$$

Das zweite nonzentrale Moment  $E(t_g t_j)$  kann nach demselben Prinzip berechnet werden. Bei Anwendung des Gesetzes der iterierten Erwartungswertbildung erhält man:

$$E[t_g t_j] = \exp\left[-k_g - k_j + \frac{1}{2}(l_g + l_j)^2 + \frac{1}{2}(c_g + c_j)^2\right]. \quad (\text{A.7})$$

Betrachtet man zwei Items  $g$  und  $j$ , so ergibt sich durch die Momentglei-

chungen (A.4) bis (A.7) folgendes Gleichungssystem:

$$\begin{aligned}
 \log [E[t_g]] &= -k_g + \frac{1}{2}l_g^2 + \frac{1}{2}c_g^2 \\
 \log [E[t_j]] &= -k_j + \frac{1}{2}l_j^2 + \frac{1}{2}c_j^2 \\
 \log [E[t_g^2]] &= \log(2) - 2k_g + 2l_g^2 + 2c_g^2 \\
 \log [E[t_j^2]] &= \log(2) - 2k_j + 2l_j^2 + 2c_j^2 \\
 \log [E[t_g t_j]] &= -k_g - k_j + \frac{1}{2}(l_g + l_j)^2 + \frac{1}{2}(c_g + c_j)^2.
 \end{aligned} \tag{A.8}$$

Umformungen von (A.8) führen zur Beziehung:

$$\begin{aligned}
 \log [E[t_g^2]] - \log(2) - 2 \log [E[t_g]] &= l_g^2 + c_g^2 \\
 \log [E[t_j^2]] - \log(2) - 2 \log [E[t_j]] &= l_j^2 + c_j^2 \\
 \log [E[t_g t_j]] - \log [E[t_g]] - \log [E[t_j]] &= l_g l_j + c_g c_j.
 \end{aligned} \tag{A.9}$$

In Gleichungssystem (A.9) können die theoretischen Momente durch die empirischen Momente ersetzt werden. Tatsächlich besitzt das Gleichungssystem (A.9) die Struktur eines faktorenanalytischen Modells, jedoch ohne Term für die spezifische Varianz. Dies bedeutet, dass die durch Gleichungssystem (A.9) gebildete Matrix bei mehreren Items über eine Hauptkomponentenanalyse faktorisiert werden kann (R. Johnson & Wichern, 2002, S. 484). Die gesuchten Itemparameter  $\mathbf{l} = [l_1, \dots, l_g]$  und  $\mathbf{c} = [c_1, \dots, c_g]$  entsprechen dann den ersten beiden Eigenvektoren. Die auf diese Weise geschätzten Itemparameter sind jedoch nicht eindeutig, da die Dimensionen der latenten Variablen rotiert werden können.

## A.1.2 Quadratische Parametrisierung

Laut Modell (4.11) sind die Reaktionszeiten einer Person bei einem Item  $g$  exponentialverteilt mit Verteilungsdichte:

$$f(t_g | \theta, \omega) = \lambda_g(\theta, \omega) \exp [-\lambda_g(\theta, \omega)t_g] \text{ mit } \log [\lambda_g(\theta, \omega)] = k_g + l_g(\theta - d_g)^2 + c_g \omega \tag{A.10}$$

Unter Anwendung des Gesetzes der iterierten Erwartung (Casella & Berger, 2002) können die ersten vier marginalen nonzentralen Momente der Reaktionszeit bei einem Item berechnet werden. Die Berechnung basiert dabei auf der Annahme, dass Testpersonen zufällig aus einer Population potentieller Testteilnehmer gezogen wurden, in welcher die Persönlichkeitseigenschaften  $\theta$  und  $\omega$  unabhängig standardnormalverteilt sind. Auf eine genaue Darstellung der einzelnen Rechenschritte bei der Bestimmung der Momente soll jedoch verzichtet werden – sie sind analog zu der bereits in Abschnitt A.1.1 beschriebenen Vorgehensweise. Allerdings muss bei der Bestimmung der Momente vorausgesetzt werden, dass das Einflussgewicht  $l_g$  positiv ist. Dies ist allerdings keine gravierende Einschränkung, da Modell (4.11) auch nur für positive Einflussgewichte empirisch sinnvoll ist. Die ersten vier nonzentralen Momente lauten:

$$E[t_g] = \frac{1}{\sqrt{1+2l_g}} \exp \left[ -k_g - \frac{l_g d_g^2}{1+2l_g} + \frac{1}{2} c_g^2 \right] \quad (\text{A.11})$$

$$E[t_g^2] = \frac{2}{\sqrt{1+4l_g}} \exp \left[ -2k_g - \frac{2l_g d_g^2}{1+4l_g} + 2c_g^2 \right] \quad (\text{A.12})$$

$$E[t_g^3] = \frac{6}{\sqrt{1+6l_g}} \exp \left[ -3k_g - \frac{3l_g d_g^2}{1+6l_g} + \frac{9}{2} c_g^2 \right] \quad (\text{A.13})$$

$$E[t_g^4] = \frac{24}{\sqrt{1+8l_g}} \exp \left[ -4k_g - \frac{4l_g d_g^2}{1+8l_g} + 8c_g^2 \right]. \quad (\text{A.14})$$

Die theoretischen Erwartungswerte  $E[t_g]$  bis  $E[t_g^4]$  können durch die empirischen Momente  $m_1[t_g]$  bis  $m_4[t_g]$  konsistent geschätzt werden. Werden die theoretischen Erwartungswerte durch die empirischen Momente ersetzt, kann das Gleichungssystem nach den unbekanntem Itemparametern aufgelöst werden. Das Einflussgewicht  $\hat{l}_g$  erhält man als Lösung des Gleichungssystems:

$$\begin{aligned} & \left[ \log [m_3[t_g]] - 3 \log [m_2[t_g]] + 3 \log [m_1[t_g]] \right] \frac{96}{1+8\hat{l}} \\ & - \left[ \log(6) - \frac{1}{2} \log [1+6\hat{l}_g] + \frac{3}{2} \log [1+4\hat{l}_g] - \frac{3}{2} \log [1+2\hat{l}_g] \right] \frac{96}{1+8\hat{l}_g} \\ & - \left[ \log [m_4[t_g]] - 6 \log [m_2[t_g]] + 8 \log [m_1[t_g]] \right] \frac{24}{1+6\hat{l}_g} \\ & + \left[ \log(12) - \frac{1}{2} \log [1+8\hat{l}_g] + 3 \log [1+4\hat{l}_g] - 4 \log [1+2\hat{l}_g] \right] \frac{24}{1+6\hat{l}_g} = 0. \end{aligned} \quad (\text{A.15})$$

Gleichung (A.15) ist nicht explizit lösbar – zudem besitzt die Gleichung teilweise zwei Nullstellen, so dass für  $\hat{l}_g$  zwei Lösungen existieren. Allerdings scheidet oft eine der Lösungen für  $\hat{l}_g$  aus, da in diesem Fall die geschätzte Itemlokation  $\hat{d}_g$  eine komplexe Zahl wäre. Es gilt nämlich für  $\hat{d}_g$  die Beschränkung:

$$\begin{aligned} d_g^2 = & \left[ - \left[ \log [m_3[t_g]] - 3 \log [m_2[t_g]] + 3 \log [m_1[t_g]] \right] \right. \\ & \left. + \left[ \log(6) - \frac{1}{2} \log [1+6\hat{l}_g] + \frac{3}{2} \log [1+4\hat{l}_g] - \frac{3}{2} \log [1+2\hat{l}_g] \right] \right] \quad (\text{A.16}) \\ & \left[ \frac{(1+6\hat{l}_g)(1+4\hat{l}_g)(1+2\hat{l}_g)}{24\hat{l}_g^3} \right] > 0. \end{aligned}$$

Allerdings ist der Schätzer der Itemlokation  $\hat{d}_g$  nicht eindeutig identifiziert, da das Vorzeichen durch die Momentgleichungen unbestimmt bleibt –  $\hat{d}_g$  geht



immer als quadratischer Term in die Momentgleichungen (A.11) bis (A.14) ein. Auch das Vorzeichen von  $\widehat{c}_g$  ist nicht eindeutig definiert. Der Grund der fehlenden Eindeutigkeit liegt an der Unbeobachtbarkeit der Persönlichkeitseigenschaften  $\theta$  und  $\omega$ . Definiert man die latenten Skalen der Persönlichkeitseigenschaften alleine über die Festlegung der Varianz, ist eine Skalenreflexion immer möglich. Nach Festlegung des Vorzeichens der Itemlokation  $d_g$  und des Einflussgewichts  $c_g$  eines Items sind jedoch die Itemparameter der anderen Items eindeutig definiert. Hierzu ist jedoch zusätzlich die Kovarianz zweier Items zu berücksichtigen. Das erwartete Produkt der Reaktionszeiten zweier Aufgaben  $g$  und  $j$  lautet

$$E[t_g t_j] = \frac{1}{\sqrt{1 + 2l_g + 2l_j}} \exp \left[ -(k_g + k_j) - (l_g d_g^2 + l_j d_j^2) + \frac{2(l_g d_g + l_j d_j)^2}{1 + 2l_g + 2l_j} + \frac{(c_g + c_j)^2}{2} \right], \quad (\text{A.17})$$

so dass bei Festlegung des Vorzeichens einer Itemlokation die Vorzeichen der anderen Itemlokationen eindeutig bestimmt sind. Diese Festlegung einer Itemlokation ist analog zur Notwendigkeit der Skalenverankerung bei IRT-Modellen zu sehen, bei denen die Skala der latenten Variablen ebenfalls reflektiert werden kann.

## A.2 Modelle für log-normalverteilte Reaktionszeiten

### A.2.1 Lineare Parametrisierung

Nach Modell (4.32) ist die logarithmierte Reaktionszeit  $t_g^* = \log(t_g)$  von Aufgabe  $g$  bei Bedingung auf die Persönlichkeitseigenschaften  $\theta$  und  $\omega$  normalverteilt mit Erwartungswert  $E[t_g^* | \theta, \omega] = \eta_g(\theta, \omega) = k_g + l_g \theta + c_g \omega$  und Standardabweichung  $\sigma_g$ . Es gilt daher:

$$t_g^* = k_g + l_g \theta + c_g \omega + e_g \quad \text{mit} \quad e_g \sim N(0, \sigma_g^2). \quad (\text{A.18})$$

Versuchspersonen werden aus einer Population von potentiellen Testpersonen gezogen. Geht man davon aus, dass in der Population der potentiellen Testpersonen die Personenfähigkeit  $\theta$  und der Residualfaktor  $\omega$  unabhängig standardnormalverteilt sind, folgt auch die Marginalverteilung der logarithmierten Reaktionszeit  $f(t_g^*) = \int f(t_g^* | \theta, \omega) f(\theta, \omega) d\theta d\omega$  einer Normalverteilung. Dies ist eine Folge des Umstandes, dass die Summe normalverteilter Zufallsvariablen ebenfalls normalverteilt ist. Für die Herleitung der folgenden Verteilungseigenschaften der marginalen Reaktionszeitverteilung ist jedoch die genaue Verteilungsform unwesentlich. Entscheidend ist vielmehr die Annahme, dass die Persönlichkeitseigenschaften  $\theta$  und  $\omega$  unabhängig von der Residualkomponente  $e_g$  verteilt sind und dass Residualkomponenten verschiedener Items voneinander

unabhängig sind. In diesem Fall kann die Varianz der marginalen Reaktionszeitverteilung als Summe der Varianzen der einzelnen Komponenten berechnet werden:

$$\sigma^2(t_g^*) = \sigma^2(k_g + l_g\theta + c_g\omega + e_g) = l_g^2 + c_g^2 + \sigma_g^2. \quad (\text{A.19})$$

Auch die marginale Kovarianz zweier logarithmierter Reaktionszeiten  $t_g^*$  und  $t_j^*$  ist bei Ausnutzung der Unabhängigkeit von Persönlichkeitseigenschaften und Residualkomponenten bestimmbar:

$$\sigma(t_g^*, t_j^*) = \sigma((k_g + l_g\theta + c_g\omega + e_g), (k_j + l_j\theta + c_j\omega + e_j)) = l_g l_j + c_g c_j \quad (\text{A.20})$$

Diese Rückführung der Kovarianz auf die Multiplikation zweier Einflussgewichte entspricht dabei genau dem Modell der Faktorenanalyse (R. Johnson & Wichern, 2002). Somit stehen alle faktorenanalytischen Verfahren zur Schätzung der Itemparameter zur Verfügung.

## A.2.2 Quadratische Parametrisierung

Nach Modell (4.37) folgt die Verteilung der Reaktionszeit einer Person mit den Persönlichkeitseigenschaften  $\theta$  und  $\omega$  bei Aufgabe  $g$  einer Log-Normalverteilung mit Skalenparameter  $\sigma_g$  und Lokationsparameter  $\eta_g(\theta, \omega) = k_g + l_g(\theta - d_g)^2 + c_g\omega$ . Eine Reparametrisierung führt zum äquivalenten Modell:

$$\begin{aligned} \eta_g(\theta, \omega) &= k_g + l_g(\theta - d_g)^2 + c_g\omega \\ &= (k_g + l_g d_g^2) + l_g\theta^2 - 2l_g d_g\theta + c_g\omega \\ &= \beta_{g0} + \beta_{g1}\theta^2 + \beta_{g2}\theta + \beta_{g3}\omega \end{aligned} \quad (\text{A.21})$$

Die Transformationen  $\beta_{g0} = (k_g + l_g d_g^2)$ ,  $\beta_{g1} = l_g$ ,  $\beta_{g2} = -2l_g d_g$  und  $\beta_{g3} = c_g$  können jederzeit wieder rückgängig gemacht werden und stellen keine Parameterrestriktionen dar. Nach Logarithmieren der Reaktionszeit von Item  $g$  gilt:

$$t_g^* = \beta_{g0} + \beta_{g1}\theta^2 + \beta_{g2}\theta + \beta_{g3}\omega + e_g \text{ mit } e_g \sim N(0, \sigma_g^2) \quad (\text{A.22})$$

Bei Modell (A.22) handelt es sich um das Modell einer nonlinearen Faktorenanalyse, für welches verschiedene Schätzverfahren existieren (Yalcin & Amezmyia, 2001). Geht man davon aus, dass eine Stichprobe von Versuchspersonen aus einer Population potentieller Testteilnehmer gezogen wird, in welcher die Persönlichkeitseigenschaften  $\theta$  und  $\omega$  unabhängig standardnormalverteilt sind, können die Itemparameter jedoch einfach über die Stichprobenmomente geschätzt werden. Der Erwartungswert  $\mu[t_g^*]$  und das zweite zentrale Moment  $z\mu_2[t_g^*]$  bis fünfte zentrale Moment  $z\mu_5[t_g^*]$  der logarithmierten Reaktionszeiten lauten in diesem Fall:

$$\begin{aligned}
 \mu[t_g^*] &= \beta_{g0} + \beta_{g1} \\
 z\mu_2[t_g^*] &= 2\beta_{g1}^2 + \beta_{g2}^2 + \beta_{g3}^2 + \sigma_g^2 \\
 z\mu_3[t_g^*] &= 8\beta_{g1}^3 + 6\beta_{g1}\beta_{g2}^2 \\
 z\mu_4[t_g^*] &= 60\beta_{g1}^4 + 3\beta_{g2}^4 + 60\beta_{g1}^2\beta_{g2}^2 + (12\beta_{g1}^2 + 6\beta_{g2}^2)\beta_{g3}^2 \\
 &\quad + (12\beta_{g1}^2 + 6\beta_{g2}^2 + 6\beta_{g3}^2 + 3\sigma_g^2)\sigma_g^2 \\
 z\mu_5[t_g^*] &= 544\beta_{g1}^5 + 60\beta_{g1}^3\beta_{g2}^2 + 60\beta_{g1}\beta_{g2}^4 \\
 &\quad + (80\beta_{g1}^3 + 60\beta_{g1}\beta_{g2}^2)\beta_{g3}^2 + (80\beta_{g1}^3 + 60\beta_{g1}\beta_{g2}^2)\sigma_g^2
 \end{aligned} \tag{A.23}$$

Ersetzt man in den Momentgleichungen (A.23) die theoretischen Momente  $\mu[t_g^*]$  bis  $z\mu_5[t_g^*]$  mit den korrespondierenden Stichprobenmomenten  $m_1[t_g^*]$  bis  $zm_5[t_g^*]$ , dann kann das Gleichungssystem nach den unbekanntem Parametern aufgelöst werden. Umformungen ergeben für den Schätzer des Parameters  $\hat{\beta}_{g1}$  bzw. des Einflussgewichtes  $\hat{l}_g$  die Bedingung:

$$-1536 \hat{\beta}_{g1}^6 + 480 z m_3[t_g^*] \hat{\beta}_{g1}^3 + (60 z m_2[t_g^*] z m_3[t_g^*] - 6 z m_5[t_g^*]) \hat{\beta}_{g1} = 0 \tag{A.24}$$

Durch Bestimmung der Nullstellen von (A.24) kann  $\hat{\beta}_{g1}$  gefunden werden. Die Funktion (A.24) besitzt unter Umständen drei non-triviale Nullstellen, von denen aber maximal zwei positiv sind. Eine der Lösungen scheidet jedoch oft aus, da für diese Lösung nach der Momentbedingung  $z m_3[t_g^*] = 8\hat{\beta}_{g1}^3 + 6\hat{\beta}_{g1}\hat{\beta}_{g2}^2$  der Schätzer  $\hat{\beta}_{g2}^2$  einen negativen Wert besitzen müsste. Die übrigen Itemparameter lassen sich bei Kenntnis von  $\hat{\beta}_{g1}$  berechnen, wenn man die restlichen Momentbedingungen ausnützt. In Gleichungssystem (A.23) gehen die Größen  $\beta_{g2}$  und  $\beta_{g3}$  immer als Quadrate in die einzelnen Gleichungen ein. Dies bedeutet, dass das Vorzeichen dieser Komponenten nicht bestimmt ist, sondern festgelegt werden muss. Ursache hierfür ist die Unbestimmtheit der Skalen der latenten Variablen  $\theta$  und  $\omega$ , welche bei alleiniger Festlegung der Varianz immer noch reflektiert werden können. Legt man die Vorzeichen jedoch beim ersten Item fest, sind die Skalen auch für die weiteren Items verankert. Hierzu muss jedoch die Kovarianzmatrix der Items berücksichtigt werden. Schätzer für die ursprünglichen Itemparameter  $\hat{l}_g$ ,  $\hat{c}_g$ ,  $\hat{d}_g$  und  $\hat{k}_g$  erhält man durch Retransformation der geschätzten Itemparameter  $\hat{\beta}_{g0}$ ,  $\hat{\beta}_{g1}$ ,  $\hat{\beta}_{g2}$  und  $\hat{\beta}_{g3}$ .

Anzumerken ist, dass die Schätzbarkeit der Parameter über die Momentmethode zwar die (asymptotische) Identifikation des Modells belegt. Da die Stichprobenmomente ferner konsistent geschätzt werden können, sind die Momentmethoden-Schätzer ebenfalls konsistent. Jedoch ist die Momentmethode aufgrund der starken Stichprobenschwankung besonders des fünften Momentes nicht die optimale Schätzmethode. Simulationen sprechen in der Tat für eine große Varianz der Momentmethoden-Schätzer. In den Simulationen ergab sich zudem das Problem, dass bei kleineren Stichprobenumfängen der Schätzer des Einflussgewichtes  $\hat{l}_g$  keine reelle Lösung besaß. Wie in Abschnitt 6.2 gezeigt wird, können die Momentmethoden-Schätzer als Startwerte für eine Schätzung nach dem Marginal-Maximum-Likelihood-Ansatz dienen.

# Anhang B

## Iteminformationsfunktionen der RT-Modelle

### B.1 Modelle für exponentialverteilte Reaktionszeiten

#### B.1.1 Lineare Parametrisierung

Der Beitrag der Reaktionszeiten zur Reduktion der Varianz des Fähigkeitschätzers hängt von der Fisher-Information der Reaktionszeiten ab. Die Fisher-Information entspricht der Varianz der Score-Funktion  $S(\theta, \omega, t)$  der Reaktionszeitverteilung (siehe 2.11). Alternativ lässt sie sich bei Gültigkeit bestimmter Regularitätsbedingungen als negativer Erwartungswert der zweiten Ableitungen der Log-Likelihood-Funktion nach den Persönlichkeitseigenschaften bestimmen. Laut Modell (4.7) sind die Reaktionszeiten bei einer Aufgabe exponentialverteilt mit personenspezifischem Intensitätsparameter  $\log [\lambda_g(\theta, \omega)] = k_g + l_g\theta + c_g\omega$ . Die zweite Ableitung der logarithmierten Reaktionszeitverteilung nach dem Personenparameter  $\theta$  lautet bei dem Modell:

$$\begin{aligned} \frac{\partial^2}{\partial \theta^2} \log [f(t_g|\theta, \omega)] &= \frac{\partial^2}{\partial \theta^2} \left[ \log \left[ \lambda_g(\theta, \omega) \exp(-\lambda_g(\theta, \omega)t_g) \right] \right] \\ &= \frac{\partial^2}{\partial \theta^2} \left[ \log \left[ \exp(k_g + l_g\theta + c_g\omega) \exp(-\exp(k_g + l_g\theta + c_g\omega)t_g) \right] \right] \\ &= \frac{\partial^2}{\partial \theta^2} \left[ (k_g + l_g\theta + c_g\omega) - \exp(k_g + l_g\theta + c_g\omega)t_g \right] \\ &= -l_g^2 \exp(k_g + l_g\theta + c_g\omega)t_g. \end{aligned} \tag{B.1}$$

Analoge Rechenschritte für die weiteren Ableitungen führen zur Hesse-Matrix:

$$\mathbf{H}(\theta, \omega, t_g) = -\exp(k_g + l_g\theta + c_g\omega)t_g \begin{bmatrix} l_g^2 & l_gc_g \\ l_gc_g & c_g^2 \end{bmatrix}. \quad (\text{B.2})$$

Der Erwartungswert einer mit Verteilungsparameter  $\lambda_g(\theta, \omega)$  exponentialverteilten Zufallsvariablen  $t_g$  entspricht  $1/\lambda_g(\theta, \omega) = \exp(-k_g - l_g\theta - c_g\omega)$ . Einsetzen dieses Wertes in Gleichung (B.2) und Veränderung des Vorzeichens ergibt die Fisher-Information (4.10):

$$\mathbf{I}_g(\theta, \omega) = -E \left[ -\exp(k_g + l_g\theta + c_g\omega)t_g \begin{bmatrix} l_g^2 & l_gc_g \\ l_gc_g & c_g^2 \end{bmatrix} \right] = \begin{bmatrix} l_g^2 & l_gc_g \\ l_gc_g & c_g^2 \end{bmatrix}. \quad (\text{B.3})$$

### B.1.2 Quadratische Parametrisierung

Modell (4.11) führt den Verteilungsparameter  $\lambda_g(\theta, \omega)$  der Exponentialverteilung über eine quadratische Funktion auf die Personenfähigkeit  $\theta$  und über eine lineare Funktion auf den Residualfaktor  $\omega$  zurück:  $\log[\lambda_g(\theta, \omega)] = k_g + l_g(\theta - d_g)^2 + c_g\omega$ . Die zweite Ableitung der Log-Likelihood-Funktion nach  $\theta$  führt zu:

$$\begin{aligned} \frac{\partial^2}{\partial \theta^2} \log [f(t_g|\theta, \omega)] &= \frac{\partial^2}{\partial \theta^2} \left[ \log \left[ \lambda_g(\theta, \omega) \exp(-\lambda_g(\theta, \omega)t_g) \right] \right] \\ &= \frac{\partial^2}{\partial \theta^2} \left[ \log \left[ \exp[k_g + l_g(\theta - d_g)^2 + c_g\omega] \exp[-\exp[k_g + l_g(\theta - d_g)^2 + c_g\omega]t_g] \right] \right] \\ &= 2l_g - \left[ [2l_g + 4l_g^2(\theta - d_g)^2] \exp[k_g + l_g(\theta - d_g)^2 + c_g\omega] \right] t_g \\ &= 2l_g - [2l_g + 4l_g^2(\theta - d_g)^2] \lambda_g(\theta, \omega)t_g. \end{aligned} \quad (\text{B.4})$$

Bildung der restlichen Ableitungen ergibt die Hesse-Matrix:

$$\mathbf{H}(\theta, \omega, t_g) = \begin{bmatrix} 2l_g & 0 \\ 0 & 0 \end{bmatrix} - \lambda_g(\theta, \omega)t_g \begin{bmatrix} 2l_g + 4l_g^2(\theta - d_g)^2 & 2l_gc_g(\theta - d_g) \\ 2l_gc_g(\theta - d_g) & c_g^2 \end{bmatrix}. \quad (\text{B.5})$$

Der Erwartungswert von  $t_g$  ist bei einer exponentialverteilten Zufallsvariablen genau  $1/\lambda_g(\theta, \omega)$ . Es gilt somit für die Informationsfunktion, bei welcher der Erwartungswert der negativen Hesse-Matrix der Log-Likelihood-Funktion gebildet wird, die im vierten Kapitel aufgeführte Beziehung (4.14):

$$\mathbf{I}_g(\theta, \omega) = \begin{bmatrix} 4l_g(\theta - d_g)^2 & 2l_gc_g(\theta - d_g) \\ 2l_gc_g(\theta - d_g) & c_g^2 \end{bmatrix}. \quad (\text{B.6})$$

## B.2 Modelle für dichotomisierte Reaktionszeiten

### B.2.1 Lineare Parametrisierung

Nach dem Proportional Hazard Modell für dichotomisierte Reaktionszeiten mit linearer Parametrisierung (4.18) ergibt sich die Wahrscheinlichkeit einer Reaktion vor einem Zeitlimit ( $y_g = 1$ ) beziehungsweise nach einem Zeitlimit ( $y_g = 0$ ) bei Item  $g$  als:

$$P(y_g|\theta, \omega) = \left[1 - \exp[-\exp[k_g + l_g\theta + c_g\omega]]\right]^{y_g} \left[\exp[-\exp[k_g + l_g\theta + c_g\omega]]\right]^{1-y_g}. \quad (\text{B.7})$$

Die Schritte zur Bestimmung der Informationsfunktion sind offensichtlich. Logarithmieren und zweimaliges Ableiten nach der Personenfähigkeit  $\theta$  führt zu:

$$\begin{aligned} \frac{\partial^2}{\partial \theta^2} \log [P(y_g|\theta, \omega)] &= -y_g \frac{l_g^2 [\exp[k_g + l_g\theta + c_g\omega]]^2 [\exp[\exp[k_g + l_g\theta + c_g\omega]]]}{[\exp[\exp[k_g + l_g\theta + c_g\omega]] - 1]^2} \\ &+ y_g \frac{l_g^2 [\exp[k_g + l_g\theta + c_g\omega]]}{[\exp[\exp[k_g + l_g\theta + c_g\omega]] - 1]} - (1 - y_g) l_g^2 [\exp[k_g + l_g\theta + c_g\omega]] \end{aligned} \quad (\text{B.8})$$

Bildung des Erwartungswertes über  $y_g$  und Veränderung des Vorzeichens ergibt das erste Element der Informationsmatrix:

$$I_{g[11]}(\theta, \omega) = l_g^2 \frac{[\exp[k_g + l_g\theta + c_g\omega]]^2}{\exp[\exp[k_g + l_g\theta + c_g\omega]] - 1} \quad (\text{B.9})$$

Analoge Rechenschritte für die übrigen Elemente der Informationsmatrix führen zu Beziehung (4.22):

$$\mathbf{I}_g(\theta, \omega) = \frac{[\exp[k_g + l_g\theta + c_g\omega]]^2}{\exp[\exp[k_g + l_g\theta + c_g\omega]] - 1} \begin{bmatrix} l_g^2 & l_g c_g \\ l_g c_g & c_g^2 \end{bmatrix}. \quad (\text{B.10})$$

### B.2.2 Quadratische Parametrisierung

Modell (4.23) nimmt eine Unfolding-Beziehung zwischen der Personenfähigkeit und der Reaktionszeit bei einer Aufgabe an. Die Wahrscheinlichkeit einer Reaktion vor einem Zeitlimit ( $y_g = 1$ ) beziehungsweise nach einem Zeitlimit ( $y_g = 0$ ) bei Item  $g$  folgte dabei der Beziehung:

$$P(y_g|\theta, \omega) = \left[1 - \exp\left[-\exp[k_g + l_g(\theta - d_g)^2 + c_g\omega]\right]\right]^{y_g} \left[\exp\left[-\exp[k_g + l_g(\theta - d_g)^2 + c_g\omega]\right]\right]^{1-y_g}. \quad (\text{B.11})$$

Zweimaliges Ableiten von  $\log [P(y_g|\theta, \omega)]$  nach der Personenfähigkeit  $\theta$  ergibt:

$$\begin{aligned} \frac{\partial^2}{\partial \theta^2} \log [P(y_g|\theta, \omega)] = & \\ & -y_g \frac{4l_g^2(\theta - d_g)^2 [\exp[k_g + l_g(\theta - d_g)^2 + c_g\omega]]^2 [\exp[\exp[k_g + l_g(\theta - d_g)^2 + c_g\omega]]]}{\left[\exp[\exp[k_g + l_g(\theta - d_g)^2 + c_g\omega]] - 1\right]^2} \\ & + y_g \frac{(2l_g + 4l_g^2(\theta - d_g)^2) [\exp[k_g + l_g(\theta - d_g)^2 + c_g\omega]]}{\left[\exp[\exp[k_g + l_g(\theta - d_g)^2 + c_g\omega]] - 1\right]} \\ & - (1 - y_g)(2l_g + 4l_g^2(\theta - d_g)^2) [\exp[k_g + l_g(\theta - d_g)^2 + c_g\omega]]. \end{aligned} \quad (\text{B.12})$$

Das zur Personenfähigkeit  $\theta$  gehörende Element der Informationsmatrix  $I_{g[11]}(\theta, \omega)$  erhält man nach Bildung des Erwartungswertes über  $y_g$  und Änderung des Vorzeichens:

$$I_{g[11]}(\theta, \omega) = \frac{[\exp(k_g + l_g(\theta - d_g)^2 + c_g\omega)]^2}{\exp[\exp[k_g + l_g(\theta - d_g)^2 + c_g\omega]] - 1} 4l_g^2(\theta - d_g)^2. \quad (\text{B.13})$$

Analoge Berechnungen führen zur Informationsmatrix (4.26):

$$\mathbf{I}_g(\theta, \omega) = \frac{[\exp(k_g + l_g(\theta - d_g)^2 + c_g\omega)]^2}{\exp[\exp[k_g + l_g(\theta - d_g)^2 + c_g\omega]] - 1} \begin{bmatrix} 4l_g^2(\theta - d_g)^2 & 2l_g c_g(\theta - d_g) \\ 2l_g c_g(\theta - d_g) & c_g^2 \end{bmatrix}. \quad (\text{B.14})$$

## B.3 Modelle für log-normalverteilte Reaktionszeiten

### B.3.1 Lineare Parametrisierung

Modell (4.32) geht von log-normalverteilten Reaktionszeiten aus, bei denen der Lokationsparameter der Verteilung in linearer Beziehung zur den Persönlichkeitseigenschaften  $\theta$  und  $\omega$  steht:  $\eta_g(\theta, \omega) = k_g + l_g\theta + c_g\omega$ . Die log-normalverteilte Reaktionszeit  $t_g$  kann durch Logarithmierung in eine normalverteilte Zufallsvariable  $t_g^*$  transformiert werden. Der Erwartungswert der normalverteilten Zufallsvariablen ist dann der Lokationsparameter der Log-Normalverteilung  $\eta_g(\theta, \omega)$ ,

die Varianz der Skalenparameter  $\sigma_g$ . Nach Transformation der Reaktionszeit  $t_g$  bei Item  $g$  in  $t_g^* = \log(t_g)$  gilt für die Verteilung von  $t_g^*$  somit:

$$f(t_g^*|\theta, \omega) = \frac{1}{\sqrt{2\pi}\sigma_g} \exp \left[ -\frac{1}{2\sigma_g^2} [t_g^* - (k_g + l_g\theta + c_g\omega)]^2 \right]. \quad (\text{B.15})$$

Zweimaliges Ableiten der logarithmierten Verteilungsdichte nach der Personenfähigkeit  $\theta$  ergibt die Funktion:

$$\begin{aligned} \frac{\partial^2}{\partial \theta^2} \log [f(t_g^*|\theta, \omega)] &= \frac{\partial^2}{\partial \theta^2} \left[ \log \left[ \frac{1}{\sqrt{2\pi}\sigma_g} \right] - \frac{1}{2\sigma_g^2} [t_g^* - (k_g + l_g\theta + c_g\omega)]^2 \right] \\ &= -\frac{l_g^2}{\sigma_g^2}. \end{aligned} \quad (\text{B.16})$$

Analoge Berechnungen für die weiteren Ableitungen und Veränderung des Vorzeichens führen zur Informationsmatrix (4.36):

$$\mathbf{I}_g(\theta, \omega) = \frac{1}{\sigma_g^2} \begin{bmatrix} l_g^2 & l_g c_g \\ l_g c_g & c_g^2 \end{bmatrix}. \quad (\text{B.17})$$

### B.3.2 Quadratische Parametrisierung

Eine andere Informationsfunktion besitzt Modell (4.37), das von einem quadratischen Einfluss der Personenfähigkeit  $\theta$  auf den Lokationsparameter der Log-Normalverteilung ausgeht:  $\eta_g(\theta, \omega) = k_g + l_g(\theta - d_g)^2 + c_g\omega$ . Wiederum ergeben sich Vereinfachungen, wenn man die Reaktionszeiten logarithmiert. Durch Logarithmieren wird die ursprünglich log-normalverteilte Reaktionszeit  $t_g$  eines Items in eine normalverteilte Zufallsvariable  $t_g^* = \log(t_g)$  transformiert. Die Verteilung der logarithmierten Reaktionszeit folgt dann der Beziehung:

$$f(t_g^*|\theta, \omega) = \frac{1}{\sqrt{2\pi}\sigma_g} \exp \left[ -\frac{1}{2\sigma_g^2} [t_g^* - (k_g + l_g(\theta - d_g)^2 + c_g\omega)]^2 \right]. \quad (\text{B.18})$$

Zweimaliges Ableiten der logarithmierten Verteilungsdichte resultiert in:

$$\begin{aligned} \frac{\partial^2}{\partial \theta^2} \log [f(t_g^*|\theta, \omega)] &= \frac{\partial^2}{\partial \theta^2} \left[ \log \left[ \frac{1}{\sqrt{2\pi}\sigma_g} \right] - \frac{1}{2\sigma_g^2} [t_g^* - (k_g + l_g(\theta - d_g)^2 + c_g\omega)]^2 \right] \\ &= \frac{2l_g}{\sigma_g^2} [t_g^* - (k_g + l_g(\theta - d_g)^2 + c_g\omega)] - \frac{4l_g^2}{\sigma_g^2} (\theta - d_g)^2. \end{aligned} \quad (\text{B.19})$$



Bildung der weiteren Ableitungen führt zur Hesse-Matrix:

$$\mathbf{H}(\theta, \omega, t_g^*) = \frac{1}{\sigma_g^2} \begin{bmatrix} 2l_g [t_g^* - [k_g + l_g(\theta - d_g)^2 + c_g\omega]] - 4l_g^2(\theta - d_g)^2 & -2l_g c_g(\theta - d_g) \\ -2l_g c_g(\theta - d_g) & -c_g^2 \end{bmatrix}. \quad (\text{B.20})$$

Nach Bildung des Erwartungswertes und Veränderung des Vorzeichens ergibt sich die Informationsmatrix (4.40):

$$\mathbf{I}_g(\theta, \omega) = \frac{1}{\sigma_g^2} \begin{bmatrix} 4l_g^2(\theta - d_g)^2 & 2l_g c_g(\theta - d_g) \\ 2l_g c_g(\theta - d_g) & c_g^2 \end{bmatrix}. \quad (\text{B.21})$$

## Anhang C

# Informationsgewinn durch Betrachtung der RT

### C.1 Fisher-Information bei Kenntnis des Residualfaktors

Sind der Residualfaktor  $\omega$  und die Personenfähigkeit  $\theta$  in einer Population potentieller Testteilnehmer bivariat standardnormalverteilt mit dem bekannten Korrelationskoeffizienten  $\rho$ , folgt die bedingte Verteilung  $f(\omega|\theta)$  einer Normalverteilung mit Verteilungsdichte:

$$f(\omega|\theta) = \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp \left[ -\frac{1}{2(1-\rho^2)} (\omega - \rho\theta)^2 \right]. \quad (\text{C.1})$$

Ist der Residualfaktor  $\omega$  einer Person bekannt, kann dieser als zusätzliche Beobachtung zur Schätzung der Personenfähigkeit verwendet werden. Die Fisher-Information dieser zusätzlichen Beobachtung ist der negative Erwartungswert der zweifachen Ableitung von  $\log [f(\omega|\theta)]$  nach der Personenfähigkeit  $\theta$ . Eine Logarithmierung der Verteilungsdichte (C.1) und zweimaliges Ableiten nach der Personenfähigkeit  $\theta$  führen zu:

$$\frac{\partial^2}{\partial \theta^2} \log [f(\omega|\theta)] = -\frac{\rho^2}{(1-\rho^2)}. \quad (\text{C.2})$$

Veränderung des Vorzeichens und Bildung des Erwartungswertes über  $\omega$  ergibt die Informationsfunktion (7.30):

$$I_\omega(\theta) = \frac{\rho^2}{(1-\rho^2)}. \quad (\text{C.3})$$

## C.2 Testinformation bei Korrelation von Personenfähigkeit und Residualfaktor

In Kapitel 7.2 blieb die Herleitung des Informationsgewinns durch die Beobachtung einer zusätzlichen Reaktionszeit, Formel (7.33), ausgespart. Dies soll nun nachgeholt werden. Ausgangsbedingung der Überlegungen war die Annahme von Modell (7.29). Nach diesem Modell ist die logarithmierte Reaktionszeit  $t_g^* = \log(t_g)$  von Item  $g$  bei Bedingung auf den Residualfaktor  $\omega$  normalverteilt mit Erwartungswert  $E[t_g^*|\omega] = \eta_g(\omega) = k_g + c_g\omega$  und Varianz  $\sigma_g^2$ :

$$t_g^*|\omega \sim N(k_g + c_g\omega, \sigma_g^2). \quad (\text{C.4})$$

Bei den weiteren Ausführungen wird vorausgesetzt, dass der Residualfaktor  $\omega$  und die Personenfähigkeit  $\theta$  in einer Population potentieller Testteilnehmer bivariat standardnormalverteilt sind mit der als bekannt angenommenen Korrelation  $\rho$ . Es seien  $\mathbf{t}^* = [t_1^*, \dots, t_n^*] = [\log(t_1), \dots, \log(t_n)]$  die logarithmierten Reaktionszeiten bei  $n$  verschiedenen Items. Bei Bedingung auf die Personenfähigkeit  $\theta$  gilt für die gemeinsame Verteilung des Residualfaktors  $\omega$  und der logarithmierten Reaktionszeiten  $\mathbf{t}^*$  einer Person bei einem Test aus  $n$  Items die Beziehung:

$$f(\mathbf{t}^*, \omega|\theta) = f(\mathbf{t}^*|\omega)f(\omega|\theta). \quad (\text{C.5})$$

In dieser Gleichung kann für die Verteilung  $f(\mathbf{t}^*|\omega)$  die Verteilung (C.4) eingesetzt werden, für die Verteilung  $f(\omega|\theta)$  sei auf die Verteilung (C.1) verwiesen. Werden mehrere Items mit gleichen Itemparametern  $c$  und  $k$  und  $\sigma$  beobachtet, ist die marginale Reaktionszeitverteilung  $f(\mathbf{t}^*|\theta)$  bei Integration über die bedingte Verteilung  $f(\omega|\theta)$  eine multivariate Normalverteilung der Form:

$$\mathbf{t}^*|\theta \sim N_n \left( \begin{bmatrix} c\rho\theta + k \\ c\rho\theta + k \\ \vdots \end{bmatrix}, \begin{bmatrix} c^2(1-\rho^2) + \sigma^2 & c^2(1-\rho^2) & \dots \\ c^2(1-\rho^2) & c^2(1-\rho^2) + \sigma^2 & \dots \\ \vdots & \vdots & \ddots \end{bmatrix} \right). \quad (\text{C.6})$$

Dies leitet sich daraus ab, dass bei Bedingung auf die Personenfähigkeit  $\theta$  der Residualfaktor  $\omega$  normalverteilt ist mit Erwartungswert  $E[\omega|\theta] = \rho\theta$  und Varianz  $\sigma^2(\omega|\theta) = 1 - \rho^2$ . Die Annahme identischer Itemparameter ist für die Ableitung der bedingten Verteilung nicht zwingend. Werden jedoch identische Itemparameter angenommen, handelt es sich bei der marginalen Varianz-Kovarianz-Matrix der logarithmierten Reaktionszeiten um eine „Equi-Correlation-Matrix“. Dieser Umstand wird bei späteren Berechnungen noch ausgenutzt werden. Die gemeinsame marginale Verteilung (C.6) der logarithmierten Reaktionszeiten  $\mathbf{t}^* = [t_1^*, \dots, t_n^*]$  eines Tests aus  $n$  Items kann in eine Serie von bedingten Verteilungen zerlegt werden:

$$f(\mathbf{t}^*|\theta) = f(t_1^*|\theta) f(t_2^*|\theta, t_1^*) \dots \quad (\text{C.7})$$

Bei abhängigen Beobachtungen ist die Fisher-Information aller Beobachtungen nicht mehr die Summe der Informationsbeiträge der einzelnen Beobachtungen. Kann die gemeinsame Verteilung aber als eine Serie bedingter Verteilungen (C.7) geschrieben werden, lässt sich die Fisher-Information aller Beobachtungen dennoch als Summe einzelner Informationsbeiträge berechnen. In diesem Fall aber gilt die Beziehung

$$I_{\mathbf{t}^*}(\theta) = I_{t_1^*}(\theta) + I_{t_2^*.t_1^*}(\theta) + \dots, \quad (\text{C.8})$$

wobei  $I_{t_2^*.t_1^*}(\theta)$  für die Fisher-Information der Verteilung der zweiten Reaktionszeit  $t_2^*$  bei Bedingung auf die erste Reaktionszeit  $t_1^*$  steht, d.h. für die Fisher-Information der Verteilung  $f(t_2^*|\theta, t_1^*)$  über den unbekanntenen Personenparameter  $\theta$  (Sorensen & Gianola, 2002). Um den Informationsbeitrag der  $n+1$ -ten Reaktionszeit zu bestimmen, muss somit die Verteilung dieser Reaktionszeit bei Bedingung auf die übrigen Reaktionszeiten, also die Verteilung  $f(t_{n+1}^*|\theta, t_1^*, \dots, t_n^*)$ , betrachtet werden. Da die logarithmierten Reaktionszeiten multivariat normalverteilt sind, ist es möglich, diese bedingte Verteilung explizit anzugeben. Man kann dabei die Eigenschaft ausnutzen, dass bei multivariat normalverteilten Zufallsvariablen die bedingten Verteilungen ebenfalls normalverteilt sind. Partitioniert man multivariat normalverteilte Zufallsvariablen  $\mathbf{x}$  in zwei Blöcke  $\mathbf{x}' = [\mathbf{x}_1, \mathbf{x}_2]$ , so sind die Variablen in Block  $\mathbf{x}_1$  bei Bedingung auf bestimmte Variablenausprägungen der Variablen in Block  $\mathbf{x}_2$  multivariat normalverteilt mit den Verteilungskennwerten (R. Johnson & Wichern, 2002):

$$\boldsymbol{\mu}_{1|2} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2) \quad (\text{C.9})$$

$$\boldsymbol{\Sigma}_{11|2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}. \quad (\text{C.10})$$

Bei den Matrizen  $\boldsymbol{\Sigma}_{22}$ ,  $\boldsymbol{\Sigma}_{12}$  ... handelt es sich um die Varianz-Kovarianz-Matrizen der entsprechenden Partitionen. Die Berechnung der inversen Matrix  $\boldsymbol{\Sigma}_{22}^{-1}$  vereinfacht sich im Falle equi-korrelierter Variablen (Ni & Kedem, 2000). Lässt sich eine Matrix  $\boldsymbol{\Sigma}$  als  $\boldsymbol{\Sigma} = a\mathbf{I}_n + b\mathbf{E}_n$  schreiben, wobei  $\mathbf{I}_n$  die Identitätsmatrix und  $\mathbf{E}_n$  eine Matrix mit dem Element 1 in jeder Zelle ist, so lautet die Inverse dieser speziellen Matrix  $\boldsymbol{\Sigma}^{-1} = (1/a)\mathbf{I}_n - (b/(a^2 + nab))\mathbf{E}_n$ . In genau diese Form aber kann die Varianz-Kovarianzmatrix der logarithmierten Reaktionszeiten (C.6) gebracht werden. Eine Anwendung der Gesetzmäßigkeit von Ni und Kedem (2000) auf das aktuelle Problem ergibt für die bedingte Varianz (C.10) der  $n+1$ -ten logarithmierten Reaktionszeit  $t_{n+1}^*$ , wenn auf die übrigen logarithmierten Reaktionszeiten  $t_1^*, t_2^*, \dots, t_n^*$  bedingt wird, die Varianz:

$$\sigma^2(t_{n+1}^*|\theta, t_1^*, \dots, t_n^*) = \sigma^2 + c^2(1-\rho^2) - n \frac{(c^2(1-\rho^2))^2}{\sigma^2} - n^2 \frac{(c^2(1-\rho^2))^3}{(\sigma^4 + nc^2\sigma^2(1-\rho^2))}. \quad (\text{C.11})$$

Die bedingte Varianz ist somit nicht von den Realisationen der beobachteten, bedingenden Reaktionszeiten  $t_1^*, t_2^*, \dots, t_n^*$  abhängig – eine bekannte Eigenschaft

der multivariaten Normalverteilung. Der Erwartungswert der  $n+1$ -ten logarithmierten Reaktionszeit  $t_{n+1}^*$  bei Bedingung auf die beobachteten Reaktionszeiten  $t_1^*, t_2^*, \dots, t_n^*$  ergibt sich nach (C.9) als:

$$E(t_{n+1}^*|\theta, t_1^*, \dots, t_n^*) = [k + c\rho\theta] + \sum_{j=1}^n \left[ \frac{c^2(1-\rho^2)}{\sigma^2} - n \frac{[c^2(1-\rho^2)]^2}{\sigma^2[\sigma^2 + nc^2(1-\rho^2)]} \right] [t_j^* - (k + c\rho\theta)]. \quad (\text{C.12})$$

Bei Bedingung auf die vorherigen logarithmierten Reaktionszeiten  $t_1^*$  bis  $t_n^*$  ist die zusätzliche Reaktionszeit  $t_{n+1}^*$  nach Logarithmierung somit normalverteilt mit den oben berechneten Verteilungskennwerten  $E(t_{n+1}^*|\theta, t_1^*, \dots, t_n^*)$  und  $\sigma^2(t_{n+1}^*|\theta, t_1^*, \dots, t_n^*)$ . Bildung der zweiten Ableitung der bedingten logarithmierten Verteilung  $\log [f(t_{n+1}^*|\theta, t_1^*, \dots, t_n^*)]$  von  $t_{n+1}^*$  nach der Personenfähigkeit  $\theta$  und Integration über die Zeit ergibt die Iteminformation der  $n+1$ -ten Reaktionszeit, Formel (7.33):

$$I_{t_{n+1}^*}^*(\theta) = - \frac{\left[ c\rho - nc\rho \left[ \frac{c^2(1-\rho^2)}{\sigma^2} - n \frac{(c^2(1-\rho^2))^2}{\sigma^2(\sigma^2 + nc^2(1-\rho^2))} \right] \right]^2}{[\sigma^2 + c^2(1-\rho^2)] - \left[ n \frac{(c^2(1-\rho^2))^2}{\sigma^2} - n^2 \frac{(c^2(1-\rho^2))^3}{\sigma^2(\sigma^2 + n(c^2(1-\rho^2)))} \right]}. \quad (\text{C.13})$$

# Literaturverzeichnis

- Adams, R., Wilson, M. & Wu, M. (1997). Multilevel item response models: An approach to errors in variables regression. *Journal of Educational and Behavioral Statistics*, 22, 47–76.
- Akrami, N., Hedlund, L. & Ekehammar, B. (2007). Personality scale response latencies as self-schema indicators: The inverted-U effect revisited. *Personality and Individual Differences*, 43, 611–618.
- Andrich, D. & Luo, G. (1993). A hyperbolic cosine latent trait model for unfolding dichotomous single-stimulus responses. *Applied Psychological Measurement*, 17, 253–276.
- Ashby, F. & Maddox, W. (1994). A response time theory of separability and integrality in speeded classification. *Journal of Mathematical Psychology*, 38, 423–466.
- Baker, F. & Seock-Ho, K. (2004). *Item response theory: Parameter estimation techniques*. New York: Marcel Dekker.
- Bassili, J. (1995). Response latency and the accessibility of voting intentions: What contributes to accessibility and how it affects vote choice. *Personality and Social Psychology Bulletin*, 21, 686–695.
- Bassili, J. (1996). The how and why of response latency measurement in telephone surveys. In N. Schwarz & S. Sudman (Hg.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research* (S. 319–346). San Francisco: Jossey-Bass.
- Böckenholt, U., Albert, D., Aschenbrenner, M. & Schmalhofer, F. (1991). The effects of attractiveness, dominance, and attribute differences on information acquisition in multiattribute binary choice. *Organizational Behavior and Human Decision Processes*, 49, 258–281.
- Berger, M. (1982). The “scientific approach“ to intelligence: An overview of its history with special reference to mental speed. In H. Eysenck (Hg.), *A model for intelligence* (S. 13–43). New York: Springer.
- Bickel, P., Buyske, S., Chang, H. & Ying, Z. (2001). On maximizing item information and matching difficulty with ability. *Psychometrika*, 66, 69–77.
- Birnbaum, A. (1968). Classification by ability levels. In F. Lord & M. Novick (Hg.), *Statistical theories of mental test scores* (S. 436–452). Reading: Addison-Wesley.

- Bloxom, B. (1985). Considerations in psychometric modeling of response time. *Psychometrika*, *50*, 383–397.
- Bock, R. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, *37*, 29–51.
- Bock, R. & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, *46*, 443–459.
- Bock, R., Gibbons, R. & Muraki, E. (1988). Full-information item factor analysis. *Applied Psychological Measurement*, *12*, 261–280.
- Bontempo, B. (2000). *Assessing speededness using probabilistic models*. Unpublished doctoral dissertation, University of Chicago, Chicago.
- Bos, C. (2002). *A comparison of marginal likelihood computation methods* (Tingbergen Institute Discussion Paper Nr. TI2002-084/4). Amsterdam: Vrije Universiteit.
- Bradburn, M., Clark, T., Love, S. & Altman, D. (2003). Survival analysis part II: Multivariate data analysis – an introduction to concepts and methods. *British Journal of Cancer*, *89*, 431–436.
- Bradley, R. & Gart, J. (1962). The asymptotic properties of ML estimators when sampling from associated populations. *Biometrika*, *49*, 205–214.
- Callegaro, M., Yang, Y., Bhola, D. & Dillman, D. (2005). *Response latency as an indicator of optimization. a study comparing job applicants' and job incumbents' response time on a web survey* (Survey Research and Methodology Nr. 11). Lincoln: University of Nebraska.
- Casella, G. & Berger, R. (2002). *Statistical inference*. Pacific Grove: Duxbury.
- Casey, M. & Tryon, W. (2001). Validating a double-press method for computer administration of personality inventory items. *Psychological Assessment*, *13*, 521–530.
- Catanzaro, J. (1997). *The effect of response latencies on the validity of personality scales*. Unpublished doctoral dissertation, Fordham University, New York.
- Clarkson, D. & Zhan, Y. (2002). Using spherical-radial quadrature to fit generalized linear mixed effects models. *Journal of Computational and Graphical Statistics*, *11*, 639–659.
- Cliff, N. & Keats, J. (2003). *Ordinal measurement in the behavioral sciences*. Mahwah: Lawrence Erlbaum.
- Cohen, A., Wollack, J., Bolt, D. & Mroch, A. (2002, April). *A mixture Rasch model analysis of test speededness*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans. Abstract retrieved December 12, 2008, from <http://testing.wisc.edu/conference%20papers.html>.
- Cox, D. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society, B*, *34*, 187–220.
- Cox, D. (1975). Partial likelihood. *Biometrika*, *62*, 269–276.

- Crow, E. & Shimizu, K. (1988). *Lognormal distributions: Theory and applications*. New York: Marcel Dekker.
- Cunningham, W., Preacher, K. & Banaji, M. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science*, *12*, 163–170.
- Danthiir, V., Wilhelm, O. & Schacht, A. (2005). Decision speed in intelligence tests: Correctly an ability. *Psychology Science*, *47*, 200–229.
- Davidson, M. (1977). On a metric, unidimensional unfolding model for attitudinal and developmental data. *Psychometrika*, *42*, 523–548.
- Davidson, R. & MacKinnon, J. (1993). *Estimation and inference in econometrics*. New York: Oxford University Press.
- Davis, P. & Rabinowitz, P. (1975). *Methods of numerical integration*. New York: Academic Press.
- De Knop, S., Rijmen, F. & van Mechelen, I. (2003). *Bayesian estimation of basic IRT models* (IAP Statistics Network Technical Report Nr. 0351). Leuven: IAP Statistics Network.
- Dempster, A., Laird, N. & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, B*, *39*, 1–38.
- Devlin, S., Gnanadesikan, R. & Kettenring, J. (1975). Robust estimation and outlier detection with correlation coefficients. *Biometrika*, *62*, 531–545.
- Diederich, A. (2003). Decision making under conflict: Decision time as a measure of conflict strength. *Psychonomic Bulletin and Review*, *10*, 167–176.
- Dodd, B. & Koch, W. (1987). Effects of variations in item step values on item and test information in the partial credit model. *Applied Psychological Measurement*, *11*, 371–384.
- Dodd, B. & Koch, W. (1994). Item and scale information functions for the successive intervals Rasch model. *Educational and Psychological Measurement*, *54*, 873–885.
- Doksum, K. (1987). An extension of partial likelihood methods for proportional hazard models to general transformation models. *The Annals of Statistics*, *15*, 325–345.
- Doksum, K. & Gasko, M. (1990). On a correspondence between models in binary regression analysis and in survival analysis. *International Statistical Review*, *58*, 243–252.
- Dolan, C., van der Maas, H. & Molenaar, P. (2002). A framework for ML estimation of parameters of (mixtures of) common reaction time distributions given optional truncation or censoring. *Behavior Research Methods, Instruments and Computers*, *34*, 304–323.
- Donoghue, J. (1994). An empirical examination of the IRT information of polytomously scored reading items under the generalized partial credit model. *Journal of Educational Measurement*, *31*, 295–311.
- Dragow, F. (1988). Polychoric and polyserial correlations. In L. Kotz &



- N. Johnson (Hg.), *Encyclopedia of statistical sciences* (Bd. 7, S. 69–74). New York: Wiley.
- Dunn, T., Lushene, R. & O'Neil, H. (1972). Complete automation of the MMPI and a study of its response latencies. *Journal of Consulting and Clinical Psychology, 39*, 381–387.
- Ebel, R. (1953). The use of response time measurements in the construction of educational achievement tests. *Educational and Psychological Measurement, 13*, 391–401.
- Embretson, S. & Reise, S. (2000). *Item response theory for psychologists*. Mahwah: Lawrence Erlbaum.
- Espinoza-Varas, B. & Watson, C. (1994). Effects of decision criterion on response latencies of binary decisions. *Perception and Psychophysics, 55*, 190–203.
- Evans, M. & Swartz, T. (1995). Methods for approximating integrals in statistics with special emphasis on Bayesian integration problems. *Statistical Science, 10*, 254–272.
- Eysenck, H. (1967). Intelligence assessment: A theoretical and experimental approach. *British Journal of Educational Psychology, 37*, 81–98.
- Fahrmeir, L. & Tutz, G. (1994). *Multivariate statistical modelling based on generalized linear models*. Berlin: Springer.
- Fazio, R. (1995). Attitudes as object-evaluation associations: Determinants, consequences, and correlates of attitude accessibility. In R. Petty & J. Krosnick (Hg.), *Attitude strength: Antecedents and consequences* (S. 247–282). Mahwah: Lawrence Erlbaum.
- Fazio, R. & Olson, M. (2003). Implicit measures in social cognition: Their meaning and use. *Annual Review of Psychology, 54*, 297–327.
- Fekken, G. & Holden, R. (1992). Response latency evidence for viewing personality traits as schema indicators. *Journal of Research in Personality, 26*, 103–120.
- Ferrando, P. (2006). Person-item distance and response time: An empirical study in personality measurement. *Psicológica, 27*, 137–148.
- Ferrando, P. & Demestre, J. (2007). Características del ítem que predicen el tiempo de respuesta en medidas de personalidad [Itemeigenschaften, welche die Reaktionszeit bei Persönlichkeitsfragebögen vorhersagen]. *Psicothema, 19*, 337–342.
- Ferrando, P. & Lorenzo-Sevas, U. (2007a). An item response theory model for incorporating response time data in binary personality items. *Applied Psychological Measurement, 31*, 525–543.
- Ferrando, P. & Lorenzo-Sevas, U. (2007b). A measurement model for likert responses that incorporates response time. *Multivariate Behavioral Research, 42*, 675–706.
- Fischer, G. & Molenaar, I. (1995). *Rasch models – foundations, recent developments, and applications*. Berlin: Springer.
- Furneaux, W. (1952). Some speed, error and difficulty relationships within a

- problem-solving situation. *Nature*, *170*, 37–38.
- Gerard, B. (1997). Psychic reality and unconscious belief: A reconsideration. *International Journal of Psycho-Analysis*, *78*, 327–333.
- Glasziou, P., Cole, B., Gelber, R., Hilden, J. & Simes, R. (1998). Quality adjusted survival analysis with repeated quality of life measures. *Statistics in Medicine*, *17*, 1215–1229.
- Goegebeur, Y., De Boeck, P., Wollack, J. & Cohen, A. (2007). *A speeded item response model with gradual process change* (IAP Statistics Network Technical Report Nr. 0689). Leuven: IAP Statistics Network.
- Goldberger, A. (1998). *Introductory econometrics*. Cambridge: Harvard University Press.
- Gonzalez, J., Tuerlinckx, F., De Boeck, P. & Cools, R. (2006). Numerical integration in logistic-normal models. *Computational Statistics and Data Analysis*, *51*, 1535–1548.
- Greene, W. (2003). *Econometric analysis*. New Jersey: Prentice Hall.
- Haaijer, R., Kamakura, W. & Wedel, M. (2000). Response latencies in the analysis of conjoint choice experiments. *Journal of Marketing Research*, *37*, 376–382.
- Hanley, C. (1962). The ‘difficulty’ of a personality inventory item. *Educational and Psychological Measurement*, *22*, 577–584.
- Harrison, A., Hwalek, M., Raney, D. & Fritz, J. (1978). Cues to deception in an interview situation. *Social Psychology*, *41*, 156–161.
- Heathcote, A., Brown, S. & Mewhort, D. (2002). Quantile maximum likelihood estimation of response time distributions. *Psychonomic Bulletin and Review*, *9*, 394–401.
- Hogan, J. & Laird, N. (1997). Mixture models for the joint distribution of repeated measures and event times. *Statistics in Medicine*, *16*, 239–257.
- Holden, R. & Fekken, G. (1993). Can personality test item response latencies have construct validity? Issues of reliability and convergent and discriminant validity. *Personality and Individual Differences*, *15*, 243–248.
- Holden, R., Fekken, G. & Cotton, D. (1991). Assessing psychopathology using structured test-item response latencies. *Psychological Assessment*, *3*, 111–118.
- Holden, R. & Kroner, D. (1992). Relative efficacy of differential response latencies for detecting faking on a self-report measure of psychopathology. *Psychological Assessment*, *4*, 170–173.
- Hornke, L. (2000). Item response times in computerized adaptive testing. *Psicológica*, *21*, 175–189.
- Hougaard, P. (1999). Fundamentals of survival data. *Biometrics*, *55*, 13–22.
- Hsu, L., Santelli, J. & Hsu, J. (1989). Faking detection validity and incremental validity of response latencies to MMPI subtle and obvious items. *Journal of Personality Assessment*, *53*, 278–295.

- Jannarone, R. (1996). Measuring quickness and correctness concurrently: A conjunctive IRT approach. In G. Engelhard & M. Wilson (Hg.), *Objective measurement: Theory into practice* (Bd. 3, S. 387–402). Norwood: Ablex.
- Jansen, M. (1997). Rasch's model for reading speed with manifest explanatory variables. *Psychometrika*, *62*, 393–409.
- Jensen, A. (1982). Reaction time and psychometric g. In H. Eysenck (Hg.), *A model for intelligence* (S. 93-132). New York: Springer.
- Jin, Z. (2005, May). *Estimation methods for accelerated failure time models*. Paper presented at the International Symposium on Applied Stochastic Models and Data Analysis, Brest. Abstract retrieved December 12, 2008, from conferences.telecom-bretagne.eu/asmda2005/IMG/pdf/proceedings/1249.pdf.
- Johnson, N., Kotz, S. & Balakrishnan, N. (1994). *Continuous univariate distributions, vol 1*. New York: Wiley.
- Johnson, R. & Wichern, D. (2002). *Applied multivariate statistical analysis*. New Jersey: Prentice Hall.
- Kalbfleisch, J. & Prentice, R. (1980). *The statistical analysis of failure time data*. New York: Wiley.
- Kiefer, J. & Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Annals of Mathematical Statistics*, *27*, 887–906.
- Klein, J. & Moeschberger, M. (1997). *Survival analysis - techniques for censored and truncated data*. Berlin: Springer.
- Knol, D. & Berger, M. (1991). Empirical comparison between factor analysis and multidimensional item response models. *Multivariate Behavioral Research*, *26*, 457–477.
- Kong, X., Wise, S. & Bhola, D. (2007). Setting the response time threshold parameter to differentiate solution behavior from rapid-guessing behavior. *Educational and Psychological Measurement*, *67*, 606–619.
- Kraut, R. (1978). Verbal and nonverbal cues in the perception of lying. *Journal of Personality and Social Psychology*, *36*, 380-391.
- Kuiper, N. (1981). Convergent evidence for the self as a prototype: The „Inverted-U RT Effect“ for self and other judgments. *Personality and Social Psychology Bulletin*, *7*, 438–443.
- Kuncel, R. (1973). Response processes and relative location of subject and item. *Educational and Psychological Measurement*, *33*, 545–563.
- Lachman, S. (1993). The relationship between rate and quality of performance on achievement tests. *Educational and Psychological Measurement*, *53*, 815–819.
- Lavergne, C. & Vigneau, F. (1997). Response speed on aptitude tests as an index of intellectual performance: A developmental perspective. *Personality and Individual Differences*, *23*, 283–290.
- Lehmann, E. (1983). *Theory of point estimation*. New York: Wiley.

- Lessafre, E. & Spiessens, B. (2001). On the effect of the number of quadrature points in a logistic random-effects model: An example. *Applied Statistics*, *50*, 325–335.
- Limpert, E., Stahel, W. & Abbt, M. (2001). Log-normal distributions across the sciences: Keys and clues. *BioScience*, *51*, 341–352.
- Lohman, D. (1989). Individual differences in errors and latencies on cognitive tasks. *Learning and Individual Differences*, *1*, 179–202.
- Lord, F. (1983). Unbiased estimators of ability parameters, of their variance, and of their parallel-forms reliability. *Psychometrika*, *48*, 233–245.
- Lord, F. & Novick, M. (1968). *Statistical theories of mental test scores*. Reading: Addison-Wesley.
- Luce, R. (1991). *Response times: Their role in inferring elementary mental organization*. New York: Oxford University Press.
- Luo, G. (1998). A general formulation for unidimensional unfolding and pairwise preference models: Making explicit the latitude of acceptance. *Journal of Mathematical Psychology*, *42*, 400–417.
- Luo, G. & Andrich, D. (2005). Information functions for the general dichotomous unfolding model. In S. Alagumalai, D. Curtis & N. Hongi (Hg.), *Applied rasch measurement: A book of exemplars* (S. 309–328). Berlin: Springer.
- MacLennan, R., Jackson, D. & Bellantino, N. (1988). Response latencies and the computerized assessment of intelligence. *Personality and Individual Differences*, *9*, 811–816.
- Maddox, W., Ashby, F. & Gottlob, L. (1998). Response time distributions in multidimensional perceptual categorization. *Perception and Psychophysics*, *60*, 620–637.
- Maraun, M. & Rossi, N. (2001). The extra-factor phenomenon revisited: Unidimensional unfolding as a quadratic factor analysis. *Applied Psychological Measurement*, *25*, 77–87.
- Maris, E. (1993). Additive and multiplicative models for gamma distributed random variables, and their application as psychometric models for response times. *Psychometrika*, *58*, 445–469.
- Marubini, E. & Valsecchi, M. (1995). *Analysing survival data from clinical trials and observational studies*. Chichester: Wiley.
- Mayerl, J. (2003). *Können Nonattitudes durch die Messung von Antwortreaktionszeiten ermittelt werden? Eine empirische Analyse computergestützter Telefoninterviews*. (Bericht Nr. 2/2003). Stuttgart: Institut für Sozialwissenschaften. (Schriftenreihe des Instituts für Sozialwissenschaften SISS)
- McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society, B*, *42*, 109–142.
- McCullagh, P. & Nelder, J. (1989). *Generalized linear models*. New York: Chapman and Hall.
- McCulloch, C. & Searle, S. (2001). *Generalized, linear, and mixed models*. New

- York: Wiley.
- McFarland, R. (1928). The role of speed in mental ability. *Psychological Bulletin*, *25*, 595–612.
- McGilchrist, C. (1993). REML estimation for survival models with frailty. *Biometrics*, *49*, 221–225.
- McLachlan, G. & Krishnan, T. (1997). *The EM algorithm and extensions*. New York: Wiley.
- Mellenbergh, G. (1994). A unidimensional latent trait model for continuous item responses. *Multivariate Behavioral Research*, *29*, 223–236.
- Mislevy, R. (1987). Exploiting auxiliary information about examinees in the estimation of item parameters. *Applied Psychological Measurement*, *11*, 81–91.
- Mislevy, R. & Sheehan, K. (1988). *The information matrix in latent-variable models* (Bericht Nr. ETS-RR-88-24-ONR). New Jersey: Educational Testing Service.
- Mislevy, R. & Sheehan, K. (1989). The role of collateral information about examinees in item parameter estimation. *Psychometrika*, *54*, 661–679.
- Molenaar, I. (1995). Some background for item response theory and the Rasch model. In G. Fischer & I. Molenaar (Hg.), *Rasch-models: Foundations, recent developments, and applications* (S. 3–14). Berlin: Springer.
- Mulligan, K., Grant, J., Mockabee, S. & Monson, J. (2003). Response latency methodology for survey research: Measurement and modeling strategies. *Political Analysis*, *11*, 289–301.
- Mummendey, H. (1995). *Die Fragebogen-Methode*. Göttingen: Hogrefe.
- Murdock, B. (1985). An analysis of the strength-latency relationship. *Memory and Cognition*, *13*, 511–521.
- Nelder, J. & Mead, R. (1965). A simplex method for function minimization. *The Computer Journal*, *7*, 308–313.
- Neubauer, A. (1990). Speed of information processing in the hick paradigm and response latencies in a psychometric intelligence test. *Personality and Individual Differences*, *11*, 147–152.
- Neyman, J. & Scott, E. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, *16*, 1–32.
- Ni, Z. & Kedem, B. (2000). Normal orthant probabilities in the equicorrelation case. *Journal of Mathematical Analysis and Applications*, *246*, 280–295.
- Orchard, T. & Woodbury, M. (1972). A missing information principle: Theory and applications. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, *1*, 697–715.
- Osawa, F. (2005). New item response model with parameters which explains learning process and the application. *Systems and Computers in Japan*, *36*, 99–106.
- Oseas, L. (1966). Give the first natural answer. *Journal of Counseling Psycho-*

- logy, 13, 454–458.
- Parry, C. & McArdle, J. (1991). An applied comparison of methods for least-squares factor analysis of dichotomous variables. *Applied Psychological Measurement*, 15, 35–46.
- Pashley, P. (1987). *The analysis of latency data using the inverse Gaussian distribution*. Unpublished doctoral dissertation, McGill University, Montreal.
- Penfield, R. (2005). Unique properties of Rasch model item information functions. *Journal of Applied Measurement*, 6, 355–365.
- Pieters, J. & van der Ven, A. (1982). Precision, speed, and distraction in time-limit tests. *Applied Psychological Measurement*, 6, 93–109.
- Popham, S. & Holden, R. (1990). Assessing MMPI constructs through the measurement of response latencies. *Journal of Personality Assessment*, 54, 469–478.
- Rabe-Hesketh, S., Skrondal, A. & Pickles, A. (2001). Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal*, 2, 1–21.
- Rabe-Hesketh, S., Skrondal, A. & Pickles, A. (2005). Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics*, 128, 301–323.
- Rafaeli, S. & Tractinsky, N. (1991). Time in computerized tests: A multi-trait, multimethod investigation of general-knowledge and mathematical-reasoning on-line examinations. *Computers in Human Behavior*, 7, 215–225.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests. studies in mathematical psychology I*. Copenhagen: Danmarks pædagogiske Institut.
- Rogers, T. (1973). Toward a definition of the difficulty of a personality item. *Psychological Reports*, 33, 159–166.
- Roskam, E. (1987). Towards a psychometric theory of intelligence. In E. Roskam & R. Suck (Hg.), *Progress in mathematical psychology* (S. 151–174). Amsterdam: North-Holland.
- Roskam, E. (1997). Models for speed and time-limit tests. In W. van der Linden & R. Hambleton (Hg.), *Handbook of modern item response theory* (S. 187–208). New York: Springer.
- Rubin, D. (1976). Inference and missing data. *Biometrika*, 63, 581–592.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph*, 17, 1–100.
- Samejima, F. (1973). A comment on Birnbaum's three parameter logistic model in the latent trait theory. *Psychometrika*, 38, 221–233.
- Samejima, F. (1983). Some methods and approaches of estimating the operating characteristics of discrete item responses. In H. Wainer & S. Messick (Hg.), *Principles of modern psychological measurement: A festschrift for Frederic*

- M. Lord (S. 159–182). Mahwah: Lawrence Erlbaum.
- Savalei, V. (2006). Logistic approximation to the normal: The KL rationale. *Psychometrika*, *71*, 763–767.
- Schader, M. & Schmid, F. (1988). Maximum-Likelihood-Schätzung aus gruppierten Daten - eine Übersicht. *OR Spektrum*, *10*, 1–12.
- Schaeffer, G., Reese, C., Steffen, M., McKinley, R. & Mills, C. (1993). *Field test of a computer-based GRE general test* (Bericht Nr. ETS-RR-93-07). Princeton: Educational Testing Service.
- Scheiblechner, H. (1979). Specifically objective stochastic latency mechanisms. *Journal of Mathematical Psychology*, *19*, 19–38.
- Schilling, S. & Bock, R. (2005). High-dimensional maximum marginal likelihood item factor analysis by adaptive quadrature. *Psychometrika*, *70*, 533–555.
- Schneider, J. & Hübner, R. (1980). Einfluss von Verfälschungsinstruktionen auf die Bearbeitungszeit von Persönlichkeitsfragebögen. *Zeitschrift für experimentelle und angewandte Psychologie*, *27*, 565–579.
- Schnipke, D. (1999). *The influence of speededness on item-parameter estimation* (Bericht Nr. LSAC-R-96-07). Princeton: Law School Admission Council.
- Schnipke, D. & Scrams, D. (1997). Modeling item response times with a two-state mixture model: A new method of measuring speededness. *Journal of Educational Measurement*, *34*, 213–232.
- Scrams, D. & Schnipke, D. (1997, April). *Making use of response times in standardized tests: Are accuracy and speed measuring the same thing?* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans. Abstract retrieved December 13, 2008, from <http://eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED409357>.
- Searle, S. (1982). *Matrix algebra useful for statistics*. New York: Wiley.
- Siem, F. (1996). The use of response latencies to enhance self-report personality measures. *Military Psychology*, *8*, 15–27.
- Skrondal, A. & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal and structural equation models*. Boca-Raton: Chapman-Hall.
- Sorensen, D. & Gianola, D. (2002). *Likelihood, Bayesian, and MCMC methods in quantitative genetics*. Berlin: Springer.
- Stricker, L. & Alderton, D. (1999). Using response latency measures for a biographical inventory. *Military Psychology*, *11*, 169–188.
- Stroud, A. (1971). *Approximate calculation of multiple integrals*. Englewood Cliffs: Prentice-Hall.
- Suppes, P. & Zanotti, M. (1981). When are probabilistic explanations possible? *Synthese*, *48*, 191–199.
- Swygert, K. (2003). *The relationship of item-level response times with test-taker and item variables in an operational CAT environment* (Bericht Nr.

- LSAC-R-98-10). Princeton: Law School Admission Council.
- Takane, Y. & de Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, *53*, 393–408.
- Takane, Y. & Sergent, J. (1983). Multidimensional scaling models for reaction times and same-different judgments. *Psychometrika*, *48*, 393–423.
- Tate, R. (1955). The theory of correlation between two continuous variables when one is dichotomized. *Biometrika*, *42*, 205–216.
- Tatsuoka, K. & Tatsuoka, M. (1978). *Time-score analysis in criterion-referenced tests* (Bericht Nr. 1978-02-00). Urbana: Illinois University.
- Therneau, T. & Grambsch, P. (2000). *Modeling survival data: Extending the Cox model*. New York: Springer.
- Thissen, D. (1976). Information in wrong responses to the Raven Progressive Matrices. *Journal of Educational Measurement*, *13*, 201–214.
- Thissen, D. (1984). Timed testing: An approach using item response theory. In D. Weiss (Hg.), *New horizons in testing: Latent trait test theory and computerized adaptive testing* (S. 179–203). New York: Academic Press.
- Thorndike, E., Bregman, E., Cobb, M. & Woodyard, E. (1927). *The measurement of intelligence*. New York: Teachers College Press.
- Tsutakawa, R. & Johnson, J. (1990). The effect of uncertainty of item parameter estimation on ability estimates. *Psychometrika*, *55*, 371–390.
- Tuerlinckx, F., Rijmen, F., Verbeke, G. & De Boeck, P. (2006). Statistical inference in generalized linear mixed models: A review. *British Journal of Mathematical and Statistical Psychology*, *59*, 225–255.
- Tyron, W. & Mulloy, J. (1993). Further validation of computer-assessed response time to emotionally evocative stimuli. *Journal of Personality Assessment*, *61*, 231–236.
- Ulrich, R. & Miller, J. (1993). Information processing models generating lognormally distributed reaction times. *Journal of Mathematical Psychology*, *37*, 513–525.
- van Breukelen, G. (1995). Parallel information processing models compatible with lognormally distributed response times. *Journal of Mathematical Psychology*, *39*, 396–399.
- van Breukelen, G. (2005). Psychometric modeling of response speed and accuracy with mixed and conditional regression. *Psychometrika*, *70*, 359–376.
- van Breukelen, G. & Roskam, E. (1991). A Rasch model for the speed-accuracy tradeoff in time-limited tests. In J. Doignon & J. Falmagne (Hg.), *Mathematical psychology: Current developments* (S. 251–271). New York: Springer.
- van den Wollenberg, A. & Creemers, P. (1986). On the predictive validity of a new scoring procedure for time-limit intelligence tests. *Tijdschrift voor Onderwijsresearch*, *11*, 53–60.



- van der Linden, W. (2006). A lognormal model for response times on test items. *Journal of Educational and Behavioral Statistics*, *31*, 181–204.
- van der Linden, W. (2007). A hierarchical framework for modeling speed and accuracy on test items. *Psychometrika*, *72*, 287–308.
- van der Linden, W. (2008). Using response times for item selection in adaptive tests. *Journal of Educational and Behavioral Statistics*, *31*, 5–20.
- van der Linden, W. & Hambleton, R. (1997). *Handbook of modern item response theory*. Berlin: Springer.
- van der Linden, W., Scrams, D. & Schnipke, D. (1999). Using response-time constraints to control for differential speededness in computerized adaptive testing. *Applied Psychological Measurement*, *23*, 195–210.
- van der Linden, W. & van Krimpen-Stoop, E. (2003). Using response times to detect aberrant responses in computerized adaptive testing. *Psychometrika*, *68*, 251–265.
- Verhelst, N., Verstralen, H. & Jansen, M. (1997). A logistic model for time-limit tests. In W. van der Linden & R. Hambleton (Hg.), *Handbook of modern item response theory* (S. 169–185). New York: Springer.
- Verkuilen, J. (2006, June). *The Fisher information function in unfolding response models*. Paper presented at the Annual Meeting of the Psychometric Society IMPS, Montreal. Abstract retrieved December 13, 2008, from [https://netfiles.uiuc.edu/jayv/www/ideal\\_point\\_IIF\\_06.20.06.pdf](https://netfiles.uiuc.edu/jayv/www/ideal_point_IIF_06.20.06.pdf).
- Vigil, A. & Codorniu, M. (2004). Aggression and inhibition deficits, the role of functional and dysfunctional impulsivity. *Personality and Individual Differences*, *37*, 1431–1440.
- Vigneau, F., Blanchet, L., Loranger, M. & Pepin, M. (2002). Response latencies measured on IQ tests: dimensionality of speed indices and the relationship between speed and level. *Personality and Individual Differences*, *33*, 165–182.
- Walters, F., Parker, L., Morgan, S. & Deming, S. (1991). *Sequential simplex optimization*. Boca Raton: CRC Press.
- Wang, T. (2006). *A model for the joint distribution of item response and response time using a one-parameter Weibull distribution* (Bericht Nr. 20). Iowa: Center for Advanced Studies in Measurement and Assessment.
- Wedderburn, R. (1974). Quasi-likelihood functions, generalized linear models and the Gauss-Newton method. *Biometrika*, *61*, 439–447.
- White, P. (1973). Individual differences in speed, accuracy, and persistence: A mathematical model for problem solving. In H. Eysenck (Hg.), *The measurement of intelligence* (S. 246–260). Lancaster: MTP Press.
- Wise, S. & DeMars, C. (2006). An application of item response time: The effort-moderated IRT model. *Educational Measurement*, *43*, 19–38.
- Wise, S. & Kong, X. (2005). Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education*, *18*, 163–183.

- Wittenbrink, B. & Schwarz, N. (2007). *Implicite measures of attitude*. London: Guilford Press.
- Yalcin, I. & Amemyia, Y. (2001). Nonlinear factor analysis as a statistical method. *Statistical Science*, *16*, 275–294.
- Yamamoto, K. (1995). *Estimating the effects of test length and test time on parameter estimation using the HYBRID model* (Bericht Nr. ETS-TR-95-2). Princeton: Educational Testing Service.

# Erklärung

Ich erkläre: Ich habe die vorgelegte Dissertation selbständig und nur mit den Hilfen angefertigt, die ich in der Dissertation angegeben habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten oder nicht veröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht.

Gießen, den 5.1.2009

Unterschrift: \_\_\_\_\_