

# Messen von Fähigkeiten und Einstellungen mittels IRT-Methoden

Diplomarbeit

im Nachdiplomstudium in angewandter Statistik  
am Institut für mathematische Statistik und Versicherungslehre der Universität Bern

vorgelegt von

**Paul Ruppen**

von Saas-Grund

2010

Leiter der Arbeit: Dr. Daniel Dietrich

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung und Problemstellung</b>	<b>1</b>
<b>2</b>	<b>IRT für dichotome Variablen</b>	<b>2</b>
2.1	Charakteristische Item-Kurve . . . . .	2
2.2	Parameter-Schätzmethoden . . . . .	5
2.2.1	Schätzen der Itemparameter bei gegebenen Fähigkeiten . . . . .	5
2.2.2	Schätzen der Fähigkeiten bei gegebenen Item-Parametern . . . . .	10
2.2.3	Gemeinsames Schätzen der Fähigkeiten und der Item-Parameter . . . . .	12
2.2.4	Marginal Maximum Likelihood-Schätzmethode . . . . .	14
2.3	Diagnostik von Modellen für zweiwertige Items . . . . .	15
2.3.1	Überprüfung der Daten . . . . .	15
2.3.2	Untersuchungen auf Güte der Anpassung . . . . .	16
2.3.3	Informationsfunktionen . . . . .	17
<b>3</b>	<b>IRT für ordinalskalierte polytome Variablen</b>	<b>19</b>
3.1	Kumulative Charakteristische Item-Kategorie-Kurven . . . . .	19
3.2	Parameter-Schätzmethoden . . . . .	20
3.2.1	Schätzen der Item-Parameter bei gegebenen Einstellungsstärken . . . . .	20
3.2.2	Schätzen der Einstellungsstärken bei gegebenen Item-Parametern . . . . .	24
3.2.3	Schätzen der Item- und der Einstellungsstärkeparameter . . . . .	25
3.3	Diagnostik von Modellen für mehrwertige ordinalskalierte Items . . . . .	26
3.3.1	Überprüfung der Daten . . . . .	26
3.3.2	Untersuchung auf Angemessenheit der Items . . . . .	26
3.3.3	Informationsfunktionen . . . . .	27
<b>4</b>	<b>Anwendung der GIRT auf Beispieldaten</b>	<b>28</b>
<b>5</b>	<b>Schlusswort</b>	<b>36</b>
<b>A</b>	<b>Anhang</b>	<b>37</b>
A.1	Skalen . . . . .	37
A.2	Beweise einiger Eigenschaften der logistischen ICC . . . . .	40
A.3	Partielle Ableitungen der Funktion $\pi$ nach $a$ und nach $b$ . . . . .	42
A.4	Herleitung der Kovarianzmatrix der Itemparameter . . . . .	43
A.5	Ableitung von $p$ nach dem Fähigkeitsparameter $\theta_i$ . . . . .	44
A.6	Ableitungen der CIRCCs nach den Itemparametern . . . . .	44
A.7	Erwartungswerte der zweiten Ableitungen der logistischen CIRCCs . . . . .	48
A.8	Ableitungen der CIRCCs nach den Fähigkeiten . . . . .	50
	<b>Index</b>	<b>53</b>
	<b>Glossar</b>	<b>55</b>
	<b>Literatur</b>	<b>57</b>

# 1 Einleitung und Problemstellung

Ziel der Item Response Theorie (IRT) ist die Messung von verborgenen Eigenschaften wie Intelligenz, Fähigkeiten, oder Einstellungen von Personen auf einem möglichst hohen Skalenniveau (Baker, 2001). Gegenüber der traditionellen Testtheorie hat die IRT den Vorteil, dass das Schätzen der Parameter von den spezifischen Items oder von den untersuchten Personen abgesehen von allfälligen Auswirkungen auf die Varianz der Schätzer unabhängig ist, sofern die Items dieselbe verborgene Eigenschaft messen und die Personen zur selben Untersuchungsgruppe gehören. Entsprechend erübrigt sich die traditionelle, aufwendige Kalibrierung von Items. Zudem kann man die Parameter verschiedener Tests auf eine identische Skala abbilden, wenn die Tests mindestens ein gemeinsames Item enthalten. Damit kann man z.B. die Leistung von Schülern einer bestimmten Altersklasse über die Zeit hinweg vergleichbar machen, aber auch die Entwicklung einer Kohorte von Schülern messen. Während in der klassischen Testtheorie bei der Verfolgung einer Kohorte die Tests genau gleich schwer sein mussten, ist dies bei der IRT nicht der Fall. Man kann später einen schwierigeren Test anwenden und die Resultate trotzdem vergleichen. Dies ist von Vorteil, weil die Trennschärfe der Tests bezüglich gegebener Fähigkeitsniveaus nicht überall dieselbe ist.

Es gibt verschiedene IRT-Modelle:

- Zweiparameter-Modelle: Sie weisen einen Schwierigkeits- und Trennschärfeparameter auf.
- Rasch-Modelle oder Einparameter-Modelle: Diese weisen einen Schwierigkeitsparameter auf. Der Trennschärfeparameter wird für alle Items auf 1 gesetzt (Rasch, 1980, Kapitel 5).
- Dreiparameter-Modelle: Neben einem Schwierigkeits- und einem Trennschärfeparameter wird noch ein Rate-Parameter eingefügt, der die Wahrscheinlichkeit zufälligerweise richtiger Antworten ausdrückt.

Jedes dieser Modelle weist Unterarten auf, z.B. je nach dem Skalenniveau der Daten, die analysiert werden (für eine Darstellung der Fülle von Modellen s. van der Linden und Hambleton (1997)).

Für die Anwendung der IRT braucht man mehrere Items, die dieselbe verborgene Eigenschaft messen. Entsprechend kann man die IRT auch als Variante der verschiedenen Methoden wiederholter Messung auffassen (s. für eine entsprechende Darstellung bezüglich des Raschmodells Perline, Wright und Wainer (1979) und für eine systematische Darstellung derartiger Verfahren Agresti (2002, Kapitel 11 und 12)).

Allgemeiner können die entwickelten Methoden dazu verwendet werden, um dichotom oder ordinalskalierte Variablen, welche dieselbe verborgene Eigenschaft messen, unter bestimmten Bedingungen auf eine intervallskalierte Variable umzurechnen (s. für Erläuterungen zu den verschiedenen Skalenniveaus Anhang A.1 S. 37). Diese Umrechnung von Variablen auf eine intervallskalierte Variable hat den Vorteil, dass erstens Verfahren, welche so skalierte Variablen voraussetzen, auf inhaltliche Bereiche angewendet werden können, die traditionell solchen Verfahren nicht zugänglich waren (Leistungsmessung in der Schule; Stärke der Einstellung zu politischen Themen oder zu Produkten, etc.). So kann man z.B. die klassische Regression auf entsprechend transformierte Daten anwenden. Zweitens kann man mit diesem Vorgehen eine Variablenreduktion vornehmen. Korrelierte oder gruppenweise korrelierte Variablen können

in eine oder mehrere unkorrelierte Variablen verdichtet werden, was für die Anwendung von statistischen Verfahren von Vorteil sein kann (Rizopoulos, 2006, S.1). Diesbezüglich kann man die IRT als eine Faktorenanalyse betrachten, die für nominal, dichotom oder polytom ordinal skalierte Variablen geeignet ist.

Ziel der vorliegenden Arbeit ist es, ein einfaches Zweiparameter-Modell für dichotome und ordinalskalierte Variablen darzulegen, die Berechnung der Parameter herzuleiten und das Modell für ordinalskalierte mehrwertige Variablen mit Hilfe einer geeigneten Software auf Beispieldaten anzuwenden. Das dichotome Modell wird inhaltlich am Beispiel der Messung von Fähigkeiten eingeführt, das polytome am Beispiel der Einstellungsstärke. Bei der Anwendung auf Beispieldaten soll die intervallskalierte Messung der Freude von 15-jährigen Personen an Naturwissenschaften durchgeführt werden, wie sie in der PISA-Studie 2006 (OECD, 2007, S. 388) angewendet wurde (PISA = Programm for International Student Assessment).

## 2 IRT für dichotome Variablen

### 2.1 Charakteristische Item-Kurve

Die *Item Response Theorie* geht von der Analyse einzelner Items aus. Items sind gewöhnlich Fragen, welche z.B. in einem Fragebogen auftauchen und die von den Untersuchungsobjekten beantwortet werden. In den einfacheren Modellen werden Fragen richtig oder falsch beantwortet, d.h. durch das Item wird eine dichotome statistische Variable festgelegt. Bezüglich solcher Items sind zwei Eigenschaften von Bedeutung: die Schwierigkeit der Items und deren Trennschärfe.

Die *Schwierigkeit* eines Items kann bezüglich der Fähigkeiten von Personen definiert werden: das gleiche Item kann schwierig sein für Personen mit wenig Fähigkeit, während es für Personen mit mehr Fähigkeit leicht sein kann. Allgemein kann man die Schwierigkeit eines Items bezüglich der Fähigkeit einer Person als die Wahrscheinlichkeit betrachten, dass die Person das Item richtig beantwortet. Dabei gibt es verschiedene Auffassungen, wie die entsprechenden Wahrscheinlichkeiten zu verstehen sind (Fischer & Molenaar, 1995, Seite 5 ff.):

1. als Modell für die relativen Häufigkeit, mit der Personen mit der Fähigkeit der untersuchten Person die Frage richtig beantworten oder
2. als Modell für die relative Häufigkeit der richtigen Antworten auf mehrere Fragen derselben Schwierigkeit durch eine Person oder
3. als Modell für die relative Häufigkeit der richtigen Antworten auf dieselbe Frage durch dieselbe Person zu verschiedenen Zeitpunkten oder Zuständen.

Entsprechende Sichtweisen spielen bei den zu wählenden Parameter-Schätzmethoden eine Rolle.

Ordnet man die Wahrscheinlichkeit, das Item korrekt zu beantworten, den Fähigkeiten der Personen zu, so müssen die Funktionswerte mit den Fähigkeiten steigen. Dabei setzen wir voraus, dass die Fähigkeiten auf einer Intervallskala messbar sind. Solche strikt monoton steigende Funktionen nennt man Charakteristische Item-Kurve (Item Characteristic Curve; ICC).

**Definition 1.** Eine stetige, strikt monoton steigende Funktion

$$ICC_j : \mathbb{R} \rightarrow [0, 1],$$

welche die Wahrscheinlichkeit korrekten Antwortens auf ein Item  $j$  in Abhängigkeit von den möglichen Fähigkeiten  $\theta$  von Personen ausdrückt, heisst „Charakteristische Item-Kurve“ (ICC).

Die ICC hat gemäss dieser Definition die Form von strikt monoton steigenden Verteilungsfunktionen stetig verteilter Zufallsvariablen (s. für ein graphisches Beispiel die Abbildung 1 S. 4).

Die ICC's verschiedener Items können sich in der horizontalen Lage unterscheiden. Offenbar können Items schwieriger oder leichter sein als andere. Sie können dies in spezifischen Intervallen oder überall sein.

**Definition 2.** Ein Item 1 ist im Intervall  $I$  schwieriger als ein Item 2 genau dann, wenn  $ICC_1(\theta) < ICC_2(\theta)$  für  $\theta \in I$ . Ist  $ICC_1(\theta) < ICC_2(\theta)$  für  $\theta \in \mathbb{R}$ , nennen wir Item 1 schwieriger als Item 2.

Es wäre nützlich, die Schwierigkeit von Items durch eine Zahl und diese in derselben Skala wie die Fähigkeiten ausdrücken zu können. Es ist naheliegend, die Schwierigkeit eines Items als die reelle Zahl  $\theta$  festzulegen, bei der die Wahrscheinlichkeit 0.5 beträgt, dass eine Person mit der Fähigkeit  $\theta$  die korrekte Antwort gibt. Diese Überlegung führt zur folgenden Definition:

**Definition 3.**  $s_j \in \mathbb{R}$  ist die Schwierigkeit des Items  $j$  genau dann, wenn  $s_j = \theta$  in  $ICC_j(\theta) = 0.5$ .

Es ist zu beachten, dass die Schwierigkeiten verschiedener Items 1 und 2 nur dann mittels der Schwierigkeitsparameter  $s_1$  und  $s_2$  vergleichbar sind, wenn sich die ICC's der beiden Items nicht schneiden.

Unter der *Trennschärfe* eines Items versteht man seine Eignung, gut oder schlecht zwischen Personen mit unterschiedlichen Fähigkeiten unterscheiden zu können. Verändern sich die Wahrscheinlichkeiten der korrekten Beantwortung von Fragen in Abhängigkeit von den Fähigkeiten stark, so unterscheidet das Item in diesem Intervall die unterschiedlichen Fähigkeiten gut. Da die ICC zwischen 0 und 1 liegt und strikt monoton steigend ist, wird eine starke Trennschärfe in einem Intervall durch eine schwache in anderen Intervallen erkauft.

**Definition 4.** Sei die  $ICC_j$  des Items  $j$  differenzierbar in  $\theta$ . Die *Trennschärfe* des Items  $j$  in  $\theta$  ist die Steigung der  $ICC_j$  in  $\theta$ .

Sei die  $ICC_j$  des Items  $j$  differenzierbar in  $ICC_j(\theta) = 0.5$ . Die *Trennschärfe* des Items  $j$  ist die Steigung der  $ICC_j$  in  $\theta$  mit  $ICC_j(\theta) = 0.5$ .

In der Folge geht es darum, geeignete Funktionen zu finden, mit denen wir ICC's modellieren können. Entsprechende Funktionen sollten mittels Variation der Parameter flexibel anpassbar sein, eine sinnvolle Interpretation der Parameter erlauben und eine möglichst einfache Berechnung der Parameter gewährleisten. Ein oft gewähltes Modell ist die logistische Verteilungsfunktion (s. für eine graphische Darstellung einer solchen Funktion Abbildung 1, S. 4)

$$p : \mathbb{R} \rightarrow ]0, 1[ \tag{1}$$

$$p(\theta) := \frac{\exp(\alpha(\theta - \beta))}{1 + \exp(\alpha(\theta - \beta))}.$$

Eine logistisch verteilte Zufallsvariable, deren Verteilung durch die Verteilungsfunktion (1) bestimmt ist, weist den Erwartungswert  $\beta$  und die Standardabweichung  $\frac{\pi}{\alpha\sqrt{3}}$  auf (Müller, 1991, S. 227).

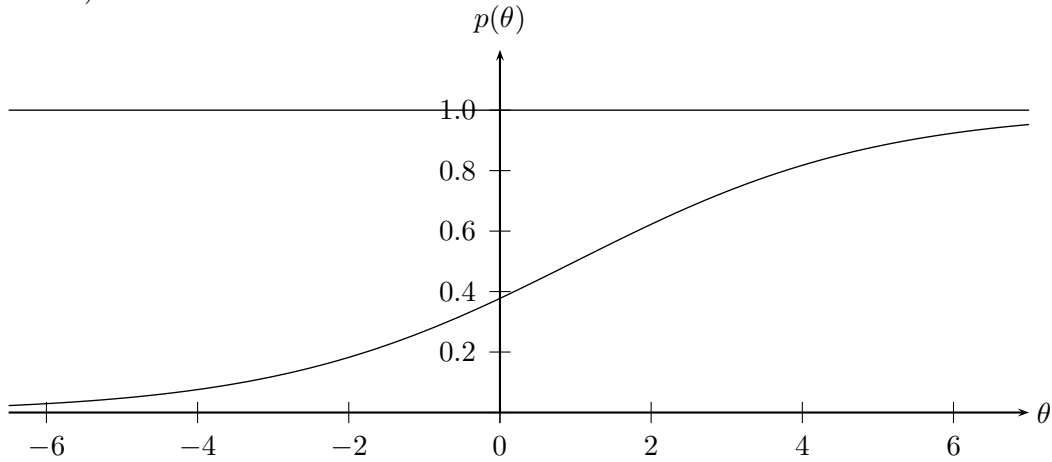


Abbildung 1: Logistische ICC mit den Parametern  $\alpha = 0.5$  und  $\beta = 1$

Die logistische Verteilungsfunktion ist strikt monoton steigend für  $\alpha > 0$ . Sie nimmt die grösste Steigung dort an, wo  $p(\theta) = 0.5$  ist und zwar in  $\theta = \beta$ . Der Parameter  $\beta$  drückt also die Schwierigkeit des Items aus. Im Punkt  $\beta$  hat die Kurve die Steigung  $\frac{\alpha}{4}$ . Damit drückt  $\alpha$  die Trennschärfe des Items aus. Die logistische ICC weist entsprechend die höchste Trennschärfe in ihrem Schwierigkeitsparameter auf. Die Kurve ist punktsymmetrisch in  $(\beta, 0.5)$ . Es gilt  $\lim_{\theta \rightarrow \infty} p(\theta) = 1$  und  $\lim_{\theta \rightarrow -\infty} p(\theta) = 0$ . Zuletzt gilt

$$\ln \left( \frac{p(\theta)}{1 - p(\theta)} \right) = \alpha (\theta - \beta), \quad (2)$$

d.h. die logarithmierte Chance (Odds), eine korrekte Antwort zu liefern, ist eine affine Funktion der Fähigkeiten (für die Beweise einiger dieser Aussagen s. Seite 40). Die logarithmierte Chance wird auch Log-Odds oder Logit genannt. Eine ICC, die mit der logistischen Verteilungsfunktion (1) modelliert wird, wird künftig „logistische ICC“ genannt.

**Bemerkung 5.** Zu Beginn der IRT-Entwicklung wurde die Verteilungsfunktion der Normalverteilung zur Modellierung der ICC verwendet. Bei dieser werden die Antwortwahrscheinlichkeiten durch  $\varphi(\theta) := \Phi(a(\theta - \beta))$  ausgedrückt, wobei  $\Phi$  die Verteilungsfunktion der Standardnormalverteilung ist.  $\Phi$  weist den Nachteil auf, nicht in geschlossener Form vorzuliegen. Zudem lässt sich in dem der Gleichung (2) entsprechenden Ausdruck

$$\Phi^{-1}(\varphi(x)) = \alpha (\theta - \beta)$$

für den linken Term keine so direkte Interpretation wie für (2) finden (s. für eine entsprechende Diskussion Baker und Kim (2004, S. 6 ff.)). Das logistische Modell wird für den Spezialfall des Raschmodells ( $\alpha = 1$ ) inhaltlich ausführlich diskutiert in Fischer und Molenaar (1995, S. 15 ff.).

Bei der Einführung der ICC's wurde vorausgesetzt, dass die Fähigkeiten von Personen auf einer Intervallskala messbar sind. Damit entspricht die logistische ICC einer logistischen Regressionskurve. Im Gegensatz zur logistischen Regression wird man aber in der Anwendung

die intervallskalierten Fähigkeiten von Personen nicht kennen. Es wird in der Folge darum gehen, Methoden zu entwickeln, um die Schwierigkeits- und Trennschärfeparameter der Items zusammen mit den Fähigkeitsparametern von Personen zu schätzen. Dazu starten wir mit den einfacheren Problemen der Schätzung der Itemparameter bei gegebenen Fähigkeitsparametern und der Schätzung der Fähigkeitsparameter bei gegebenen Itemparametern. Anschliessend wird eine Schätzmethode eingeführt, die beide Parametertypen gemeinsam schätzt: die *Joint Maximum Likelihood-Schätzmethode* (JMLE - „E“ für „Estimation“). Zusätzlich wird die *Marginal Maximum Likelihood-Schätzmethode* (MMLE) skizziert, welche die Itemparameter unter der Voraussetzung einer spezifischen Verteilung der Fähigkeitsparameter schätzt. Es gibt weitere Schätzmethoden, die hier nicht behandelt werden (u.a. Conditional Maximum Likelihood Estimation; Marginal Bayesian Parameter Estimation (Baker & Kim, 2004)).

## 2.2 Parameter-Schätzmethoden

### 2.2.1 Schätzen der Itemparameter bei gegebenen Fähigkeiten

Sind die Fähigkeiten von Personen bekannt und als auf einer Intervallskala messbar vorausgesetzt, so kann man die Parameter  $\alpha$  und  $\beta$  der logistischen ICC mit den bekannten Maximum-Likelihood-Methoden (ML-Methoden) für die logistische Regression schätzen (s. für entsprechende Methoden z.B. Agresti (1990, S. 112 ff.) und Baker und Kim (2004, S. 38 ff.)). Setzt man nämlich

$$\begin{aligned} a &:= \alpha \\ b &:= -\alpha\beta, \end{aligned} \tag{3}$$

so wird aus

$$\begin{aligned} \alpha(\theta - \beta) &= \alpha\theta - \alpha\beta \\ &= a\theta + b, \end{aligned}$$

wobei die logistischen Regressionsfunktion durch

$$\begin{aligned} \pi &: \mathbb{R} \rightarrow ]0, 1[ \\ \pi(\theta) &:= \frac{\exp(a\theta + b)}{1 + \exp(a\theta + b)} \end{aligned} \tag{4}$$

gegeben ist. Als Ausgangspunkt für das bei ML-Methoden benötigte Wahrscheinlichkeitsmodell betrachtet man relative Häufigkeiten, mit denen Personen mit spezifischen Fähigkeiten  $\theta_i$  die Frage richtig beantworten -  $i \in \mathbb{N}_m^*$  mit der Anzahl  $m$  von Klassen von Personen mit Fähigkeitsparameter  $\theta_i$ . Geeignet ist entsprechend die Betrachtung der binomialverteilten Statistik  $X_i$ , welche die Anzahl  $x_i$  der korrekten Antworten pro Fähigkeitsklasse  $\theta_i$  als Werte annimmt. Die Parameter der Verteilung sind  $n_i$  (Anzahl der Personen mit dem Fähigkeitsparameter  $\theta_i$ ) und  $\pi(\theta_i)$  (Wahrscheinlichkeit, dass eine Person mit dem Fähigkeitsparameter  $\theta_i$  die korrekte Antwort gibt). Die Wahrscheinlichkeitsfunktion  $f_w$  der gemeinsamen Verteilung der  $m$  binomialverteilten Statistiken  $X_i$  ist dann unter der Voraussetzung der Unabhängigkeiten dieser Statistiken

$$\begin{aligned} f_w &: \prod_{i=1}^m \mathbb{N}_{n_i} \rightarrow [0, 1] \\ f_w(x_1, \dots, x_m) &:= \prod_{i=1}^m \binom{n_i}{x_i} \pi(\theta_i)^{x_i} (1 - \pi(\theta_i))^{n_i - x_i}. \end{aligned}$$

Dies ergibt mit (4) die folgende Likelihood-Funktion  $f$  in den Variablen  $a$  und  $b$ , wobei die Binomialkoeffizienten  $\binom{n_i}{x_i}$  weggelassen werden. Sie sind bei gegebenen  $n_i$  und  $x_i$  konstant und fallen deshalb bei der üblichen Suche nach Maxima (Logarithmieren und Ableiten) weg.

$$\begin{aligned} f : \mathbb{R}^2 &\rightarrow ]0, 1] \\ f(a, b) &:= \prod_{i=1}^m \pi(a, b)^{x_i} (1 - \pi(a, b))^{n_i - x_i} \\ &= \prod_{i=1}^m \left( \frac{\exp(a\theta_i + b)}{1 + \exp(a\theta_i + b)} \right)^{x_i} \left( 1 - \frac{\exp(a\theta_i + b)}{1 + \exp(a\theta_i + b)} \right)^{n_i - x_i} \end{aligned}$$

Durch Logarithmieren erhält man die Log-Likelihood-Funktion (LL-Funktion):

$$\begin{aligned} \ln f : \mathbb{R}^2 &\rightarrow [0, -\infty[ \\ \ln f(a, b) &:= \sum_{i=1}^m x_i \ln \pi(a, b) + (n_i - x_i) \ln(1 - \pi(a, b)) \\ &= \sum_{i=1}^m x_i \ln \frac{\exp(a\theta_i + b)}{1 + \exp(a\theta_i + b)} + \sum_{i=1}^m (n_i - x_i) \ln \left( 1 - \frac{\exp(a\theta_i + b)}{1 + \exp(a\theta_i + b)} \right) \end{aligned}$$

Zu suchen sind die Parameter  $\hat{a}$  und  $\hat{b}$ , die diese Funktion bei gegebenen  $x_i$  und  $n_i$  maximieren. Es ergeben sich die geschätzten Werte für die Parameter  $a$  und  $b$  in

$$\ln \left( \frac{\pi(\theta)}{1 - \pi(\theta)} \right) = a\theta + b.$$

Das Maximum der LL-Funktion wird nicht wie sonst bei ableitbaren Funktionen üblich mittels Ableitung und Nullsetzung berechnet, sondern direkt mit Hilfe der sogenannten Newton-Raphson-Methode oder des Fisher-Scorings. Es handelt sich um iterative Verfahren. Zuerst wird die Newton-Raphson-Methode vorgestellt, die im Falle der logistischen Regression allerdings mit dem Fisher-Scoring zusammenfällt. Die Newton-Raphson-Methode setzt voraus, dass (i) die Funktion in einer Umgebung  $U$  der Lage  $(a, b)$  des Maximums konkav ist und (ii) der Startpunkt  $(a_1, b_1)$  des Iterationsverfahrens sowie die Zwischenergebnisse  $(a_i, b_i)$  in  $U$  liegen. Konkret wird ein Startpunkt  $(a_1, b_1)$  gewählt, dessen LL-Funktionswert in der Nähe des Maximums liegt, z.B. mittels Berechnung der klassischen linearen Regression durch die empirischen Logits

$$\ln \left( \frac{x_i + 0.5}{n_i - x_i + 0.5} \right).$$

In diesem Punkt wird dann das Taylor-Polynom 2. Grades  $g$  berechnet:

$$\begin{aligned} g : \mathbb{R}^2 &\rightarrow \mathbb{R} \\ g(\mathbf{a}) &:= \ln f(\mathbf{a}_1) + \mathbf{q}_{\mathbf{a}_1}^t (\mathbf{a} - \mathbf{a}_1) + \frac{1}{2} (\mathbf{a} - \mathbf{a}_1)^t \mathbf{H}_{\mathbf{a}_1} (\mathbf{a} - \mathbf{a}_1) \end{aligned}$$

mit

$$\begin{aligned} \mathbf{a} &= \begin{pmatrix} a \\ b \end{pmatrix}, \\ \mathbf{a}_1 &= \begin{pmatrix} a_1 \\ b_1 \end{pmatrix} \quad \text{und} \end{aligned}$$



$$\mathbf{q} = \begin{pmatrix} \frac{\partial \ln f(a,b)}{\partial a} \\ \frac{\partial \ln f(a,b)}{\partial b} \end{pmatrix},$$

dem Vektor der ersten partiellen Ableitungen der logistischen LL-Funktion nach  $a$  und nach  $b$ .

$$\mathbf{H} = \begin{pmatrix} \frac{\partial^2 \ln f(a,b)}{\partial a \partial a} & \frac{\partial^2 \ln f(a,b)}{\partial b \partial a} \\ \frac{\partial^2 \ln f(a,b)}{\partial a \partial b} & \frac{\partial^2 \ln f(a,b)}{\partial b \partial b} \end{pmatrix}$$

ist die Hesse-Matrix, d.h. der Matrix der zweiten partiellen Ableitungen der logistischen LL-Funktion nach  $a$  und  $b$ .

Der Vektor

$$\mathbf{q}_{\mathbf{a}_1}$$

entsteht aus  $\mathbf{q}$  durch Ersetzen von  $a$  und  $b$  durch  $a_1$  und  $b_1$ . Dies gilt auch für die Matrix

$$\mathbf{H}_{\mathbf{a}_1},$$

die aus  $\mathbf{H}$  durch Ersetzen von  $a$  und  $b$  durch  $a_1$  und  $b_1$  entsteht.

Das Taylor-Polynom  $g$  hat in  $\mathbf{a}_1$  dieselbe maximale Steigung wie die LL-Funktion  $\ln f$  und zwar in dieselbe Richtung sowie dasselbe (konkave) Krümmungsverhalten wie  $\ln f$ . Es hat entsprechend das Maximum von  $\mathbf{a}$  ausgesehen ungefähr oder genau in dieselbe Richtung wie  $\ln f$ .  $\mathbf{a}_2$  sei die Lage dieses Maximums des Taylorpolynoms. Im nächsten Schritt wird in  $\mathbf{a}_2$  das Taylor-Polynom 2. Grades berechnet, wiederum die Lage des Maximums bestimmt etc. Im Allgemeinen konvergiert das Verfahren schnell, so dass gewöhnlich nur wenige Durchgänge nötig sind, um ein gegebenes Präzisionsniveau zu erreichen - definiert durch die Differenz zweier aufeinanderfolgender Näherungswerte, die eine vorgegebene Grenze unterschreitet.

Durch Ableiten von  $g$  und Nullsetzen erhält man

$$g'(\mathbf{a}) = \mathbf{q}_{\mathbf{a}_1} + \mathbf{H}_{\mathbf{a}_1} (\mathbf{a} - \mathbf{a}_1) = 0$$

und durch Auflösen nach  $\mathbf{a}$  unter der Voraussetzung der Regularität von  $\mathbf{H}_{\mathbf{a}_1}$  die Lage des Maximums von  $g$

$$\mathbf{a}_2 := \mathbf{a} = \mathbf{a}_1 - \mathbf{H}_{\mathbf{a}_1}^{-1} \mathbf{q}_{\mathbf{a}_1}.$$

Für die  $t + 1$ -te Iteration gilt dabei

$$\mathbf{a}_{t+1} = \mathbf{a}_t - \mathbf{H}_{\mathbf{a}_t}^{-1} \mathbf{q}_{\mathbf{a}_t}.$$

Diese Methoden werden nun auf die logistische Regression angewendet. Für

$$\pi : \mathbb{R}^2 \rightarrow ]0, 1[$$

$$\pi(a, b) := \frac{\exp(a\theta_i + b)}{1 + \exp(a\theta_i + b)}$$

gilt (s. Seite 42)

$$\frac{\partial}{\partial a} (\pi(a, b)) = \theta_i \pi(a, b) (1 - \pi(a, b)) \text{ und}$$

$$\frac{\partial}{\partial b} (\pi(a, b)) = \pi(a, b) (1 - \pi(a, b)).$$

Damit ergeben sich mit

$$\frac{\partial}{\partial a} \ln(\pi(a, b)) = \frac{\theta_i \pi(a, b) (1 - \pi(a, b))}{\pi(a, b)} = \theta_i (1 - \pi(a, b)),$$

$$\frac{\partial}{\partial b} \ln(\pi(a, b)) = \frac{\pi(a, b) (1 - \pi(a, b))}{\pi(a, b)} = 1 - \pi(a, b),$$

$$\frac{\partial}{\partial a} \ln(1 - \pi(a, b)) = \frac{-\theta_i \pi(a, b) (1 - \pi(a, b))}{1 - \pi(a, b)} = -\theta_i \pi(a, b)$$

und

$$\frac{\partial}{\partial b} \ln(1 - \pi(a, b)) = \frac{-\pi(a, b) (1 - \pi(a, b))}{1 - \pi(a, b)} = -\pi(a, b)$$

die Ableitungen

$$\begin{aligned} \frac{\partial}{\partial a} \ln f(a, b) &= \frac{\partial}{\partial a} \sum_{i=1}^m x_i \ln \pi(a, b) + (n_i - x_i) \ln(1 - \pi(a, b)) \\ &= \sum_{i=1}^m x_i \theta_i (1 - \pi(a, b)) - (n_i - x_i) \theta_i \pi(a, b) \\ &= \sum_{i=1}^m \theta_i (x_i - n_i \pi(a, b)) \end{aligned}$$

und

$$\begin{aligned} \frac{\partial}{\partial b} \ln f(a, b) &= \frac{\partial}{\partial a} \sum_{i=1}^m x_i \ln \pi(a, b) + (n_i - x_i) \ln(1 - \pi(a, b)) \\ &= \sum_{i=1}^m x_i (1 - \pi(a, b)) - (n_i - x_i) \pi(a, b) \\ &= \sum_{i=1}^m (x_i - n_i \pi(a, b)). \end{aligned}$$

Zudem erhält man

$$\begin{aligned} \frac{\partial^2}{\partial a \partial a} \ln f(a, b) &= \frac{\partial}{\partial a} \sum_{i=1}^m (\theta_i x_i - \theta_i n_i \pi(a, b)) \\ &= - \sum_{i=1}^m \theta_i^2 n_i \pi(a, b) (1 - \pi(a, b)), \end{aligned}$$

$$\begin{aligned} \frac{\partial^2}{\partial a \partial b} \ln f(a, b) &= \frac{\partial}{\partial b} \sum_{i=1}^m (\theta_i x_i - \theta_i n_i \pi(a, b)) \\ &= - \sum_{i=1}^m \theta_i n_i \pi(a, b) (1 - \pi(a, b)), \end{aligned}$$

$$\begin{aligned}\frac{\partial^2}{\partial b \partial a} \ln f(a, b) &= \frac{\partial}{\partial a} \sum_{i=1}^m (x_i - n_i \pi(a, b)) \\ &= - \sum_{i=1}^m \theta_i n_i \pi(a, b) (1 - \pi(a, b))\end{aligned}$$

und

$$\begin{aligned}\frac{\partial^2}{\partial b \partial b} \ln f(a, b) &= \frac{\partial}{\partial b} \sum_{i=1}^m (x_i - n_i \pi(a, b)) \\ &= - \sum_{i=1}^m n_i \pi(a, b) (1 - \pi(a, b)),\end{aligned}$$

womit alle Grössen für den Iterationsprozess hergeleitet sind.

Zu beachten ist, dass bei der zweiten Ableitung die Werte  $x_i$  der Statistiken  $X_i$  verschwinden. Entsprechend fallen die Erwartungswerte  $E(l(X_i))$  der Komponenten  $l(X_i)$  der Hesse-Matrix mit diesen Komponenten zusammen. Die Matrix der Negation dieser Erwartungswerte der Hesse-Matrix nennt man „Informationsmatrix“. Die Inverse der Informationsmatrix ist die Kovarianzmatrix der Schätzstatistiken der Parameter (Agresti, 2002, S. 10). Die geschätzten Schwierigkeits- und Trennschärfeparameter der logistischen ICC erhält man mit den Festlegungen (3) von Seite 5:

$$\begin{aligned}\hat{\alpha} &= \hat{a} \\ \hat{\beta} &= -\frac{\hat{b}}{\hat{a}}.\end{aligned}$$

Entsprechend erhält man für die Schätzstatistiken  $A$  und  $B$ , welche die Werte  $\hat{a}$  und  $\hat{b}$ , sowie die Schätzstatistiken  $\mathfrak{A}$  und  $\mathfrak{B}$ , welche die Werte  $\hat{\alpha}$  und  $\hat{\beta}$  annehmen, mit

$$\bar{\theta} := \frac{\sum_{i=1}^m n_i \theta_i \pi(a, b) (1 - \pi(a, b))}{\sum_{i=1}^m n_i \pi(a, b) (1 - \pi(a, b))}$$

die folgenden Varianzen (die Kovarianzmatrix wird in A.4 auf Seite 43 hergeleitet; für die genau Form der unten stehen Varianzformen s. Baker und Kim (2004, S. 41), die Maxwell (1959) zitieren):

$$V(A) = \frac{1}{\sum_{i=1}^m n_i \pi(A, B) (1 - \pi(A, B)) (\theta_i - \bar{\theta})^2} = V(\mathfrak{A})$$

$$V(B) = \frac{1}{A^2} \frac{1}{\sum_{i=1}^m n_i \pi(A, B) (1 - \pi(\hat{a}, \hat{b}))} + \bar{\theta}^2 V(A)$$

und

$$V(\mathfrak{B}) = \frac{1}{A^2} \frac{1}{\sum_{i=1}^m (n_i) \pi(A, B) (1 - \pi(A, B))} + V(A) (\mathfrak{B} - \bar{\theta})^2.$$

Das Fisher-Scoring besteht darin, in der Formel

$$\mathbf{a}_{t+1} := \mathbf{a} = \mathbf{a}_t - \mathbf{H}_{\mathbf{a}_t}^{-1} \mathbf{q}_{\mathbf{a}_t}$$

die Hesse-Matrix durch die Negation der Informationsmatrix zu ersetzen. Da im Falle der logistischen Regression die beiden Matrizen identisch sind, fallen die beiden Näherungsmethoden für diesen Fall zusammen.

Die Parameter sind nicht schätzbar, wenn ein Item nur korrekte oder nur falsche Antworten erhält. Solche Items sind aus den Daten zu streichen. Dies gilt auch für die in der Folge eingeführten Methoden bezüglich dichotomer Items.

### 2.2.2 Schätzen der Fähigkeiten bei gegebenen Item-Parametern

Unter der Voraussetzung, dass die Item-Parameter von  $j \in \mathbb{N}_n^*$  Items mit logistischen ICCs bekannt sind, möchten man nun die Fähigkeitsparameter von  $i \in \mathbb{N}_N^*$  Personen auf dem Hintergrund deren Antwortenverhaltens bezüglich der  $n$  Items schätzen. Die Antworten ergeben für die Person  $i$  einen Antwortvektor  $(x_{i1}, \dots, x_{in})$ , mit  $x_{ij} \in \{0, 1\}$  und 0 für „falsch“ sowie 1 für „wahr“. Die Ergebnisse  $x_{ij}$  betrachten wir dabei als Werte von bernoulli-verteilten Zufallsvariablen  $X_{ij}$ . Die Fähigkeitsparameter  $\theta_i$  einer jeden Person  $i$  werden separat geschätzt - unter der Voraussetzung, dass das Antwortverhalten der verschiedenen Personen voneinander unabhängig ist. Die Wahrscheinlichkeit, dass die Person  $i$  auf das Item  $j$  bei einer Fähigkeit von  $\theta_i$  die Antwort  $x_{ij}$  gibt, ist

$$P(X_{ij} = x_{ij}) = (p_j(\theta_i))^{x_{ij}} (1 - p_j(\theta_i))^{1-x_{ij}}$$

mit

$$p_j(\theta_i) = \frac{\exp(\alpha_j(\theta_i - \beta_j))}{1 + \exp(\alpha_j(\theta_i - \beta_j))}$$

(s. Seite 4, Formel (1)). Die gemeinsame Verteilung unter der Voraussetzung, dass das Antwortverhalten der Person  $i$  auf die  $n$  verschiedenen Items unabhängig ist, ist dann:

$$P((X_{i1}, \dots, X_{in}) = (x_{i1}, \dots, x_{in})) = \prod_{j=1}^n (p_j(\theta_i))^{x_{ij}} (1 - p_j(\theta_i))^{1-x_{ij}}$$

Damit erhält man die Likelihood-Funktion bezüglich des Parameters  $\theta_i$  :

$$f(\theta_i) = \prod_{j=1}^n (p_j(\theta_i))^{x_{ij}} (1 - p_j(\theta_i))^{1-x_{ij}}$$

und die LL-Funktion

$$\ln f(\theta_i) = \sum_{j=1}^n [x_{ij} \ln(p_j(\theta_i)) + (1 - x_{ij}) \ln(1 - p_j(\theta_i))].$$

Das näherungsweise Maximum der LL-Funktion wird wiederum mit dem Newton-Raphson-Verfahren berechnet. Mit

$$\frac{dp_j(\theta_i)}{d\theta_i} = \alpha_j p_j(\theta_i) (1 - p_j(\theta_i))$$

(s. Seite 44) und mit

$$\begin{aligned}\frac{d \ln (p_j(\theta_i))}{d \theta_i} &= \frac{\alpha_j p_j(\theta_i) (1 - p_j(\theta_i))}{p_j(\theta_i)} = \alpha_j (1 - p_j(\theta_i)) \\ \frac{d \ln (1 - p_j(\theta_i))}{d \theta_i} &= \frac{-\alpha_j p_j(\theta_i) (1 - p_j(\theta_i))}{1 - p_j(\theta_i)} = -\alpha_j p_j(\theta_i)\end{aligned}$$

erhält man

$$\begin{aligned}\frac{d \ln f(\theta_i)}{d \theta_i} &= \sum_{j=1}^n [x_{ij} \alpha_j (1 - p_j(\theta_i)) - (1 - x_{ij}) \alpha_j p_j(\theta_i)] \\ &= \sum_{j=1}^n [x_{ij} \alpha_j - x_{ij} \alpha_j p_j(\theta_i) - \alpha_j p_j(\theta_i) + x_{ij} \alpha_j p_j(\theta_i)] \\ &= \sum_{j=1}^n [x_{ij} \alpha_j - \alpha_j p_j(\theta_i)].\end{aligned}$$

Für die zweite Ableitung nach  $\theta_i$  erhält man:

$$\frac{d}{d \theta_i} \left( \sum_{j=1}^n x_{ij} \alpha_j - \alpha_j p_j(\theta_i) \right) = - \sum_{j=1}^n \alpha_j^2 p_j(\theta_i) (1 - p_j(\theta_i))$$

Auch hier kann man feststellen, dass in der zweite Ableitung die Werte der Zufallsvariablen  $X_{ij}$  verschwinden. Mit  $\theta_1$  als Startwert erhält man für das Taylorpolynom 2. Grades

$$\begin{aligned}g: \mathbb{R} &\rightarrow \mathbb{R} \\ g(\theta) &:= \ln f(\theta_1) + q_{\theta_1}(\theta - \theta_1) + \frac{1}{2} h_{\theta_1}(\theta - \theta_1)^2,\end{aligned}$$

wobei  $q_{\theta_1}$  die erste Ableitung von  $\ln f(\theta)$  nach  $\theta$  ist, in die  $\theta_1$  eingesetzt wurde und  $h_{\theta_1}$  die zweite Ableitung von  $\ln f(\theta)$  nach  $\theta$  ist, in die  $\theta_1$  eingesetzt wurde.

Ableiten und Nullsetzen ergibt

$$\frac{d}{dx} g(\theta) = q_{\theta_1} + h_{\theta_1}(\theta - \theta_1) = 0,$$

woraus sich

$$\theta_2 := \theta = \theta_1 - \frac{q_{\theta_1}}{h_{\theta_1}}$$

ergibt. Allgemein gilt damit für den  $t + 1$ -ten Schritt

$$\theta_{t+1} := \theta_t - \frac{q_{\theta_t}}{h_{\theta_t}}.$$

Den geschätzten Standardfehler der Schätzstatistik  $\Theta_i$ , welche als Schätzwert  $\hat{\theta}_i$  für die  $i$ -te Person oder die  $i$ -te Personengruppe annimmt, erhält man aus der zweiten Ableitung der LL-Funktion (Kehrwert der Negation des Erwartungswertes der 2. Ableitung der LL-Funktion)

$$V(\Theta_i) = \frac{1}{\sum_{j=1}^n \alpha_j^2 p_j(\Theta_i) (1 - p_j(\Theta_i))}.$$

An der Formel kann man ablesen, wo die Varianz am kleinsten ist.  $p_j(\theta_i)(1 - p_j(\theta_i))$  ist maximal, wenn  $p_j(\theta_i) = 0.5$ , d.h. wenn  $\theta_i = \beta_j$ . Eine kleine Varianz ergibt sich dort zudem, wenn die Schwierigkeitsparameter  $\beta_j$  der verschiedenen Items  $j$  möglichst nahe beieinander liegen und die  $a_j$  möglichst gross sind.

Ähnlich wie bei der Schätzung der Item-Parameter ist zu beachten, dass Personen, die alle Fragen richtig oder alle Fragen falsch beantworten, aus dem Datensatz gestrichen werden müssen.

### 2.2.3 Gemeinsames Schätzen der Fähigkeiten und der Item-Parameter

Gewöhnlich sind weder die Item- noch die Fähigkeitsparameter bekannt. Sie müssen entsprechend alle geschätzt werden. Es gibt dazu verschiedene Verfahren, von denen in der Folge zwei kurz behandelt werden. Es wird zuerst die Joint Maximum Likelihood-Schätzmethode (JM-LE) eingeführt. Es liegen  $i \in \mathbb{N}_N^*$  Personen und  $j \in \mathbb{N}_n^*$  Items mit den Antwortmöglichkeiten  $x_{ij} \in \{0, 1\}$  vor, wobei die  $x_{ij}$  als Werte von  $N \cdot n$  bernoulli-verteilten Zufallsvariablen  $X_{ij}$  zu betrachten sind. Die gemeinsame Verteilung unter der Voraussetzung der Unabhängigkeit dieser Zufallsvariablen ist

$$P((X_{11}, \dots, X_{Nn}) = (x_{ij}, \dots, x_{Nn})) = \prod_{i=1}^N \prod_{j=1}^n (p_j(\theta_i))^{x_{ij}} (1 - p_j(\theta_i))^{1-x_{ij}},$$

wobei  $p_j(\theta_i)$  die Wahrscheinlichkeit einer korrekten Antwort der Person  $i$  beim Vorliegen der logistischen ICC des Items  $j$  ist. Daraus erhält man die LL-Funktion in  $2n + N$  Parametern

$$\ln f(\alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_n, \theta_1, \dots, \theta_N) = \sum_{i=1}^N \sum_{j=1}^n x_{ij} \ln p_j(\alpha_j, \beta_j, \theta_i) + (1 - x_{ij}) \ln (1 - p_j(\alpha_j, \beta_j, \theta_i)).$$

Um das Maximum der LL-Funktion mit Hilfe der Newton-Raphson-Methode zu berechnen, wird man auf die folgende Gleichung geführt (s. Seite 10)

$$\mathbf{a}_{t+1} = \mathbf{a}_t - \mathbf{H}_{\mathbf{a}_t}^{-1} \mathbf{q}_{\mathbf{a}_t},$$

wobei in diesem Falle  $\mathbf{a}_t$ ,  $\mathbf{a}_{t+1}$  und  $\mathbf{q}_{\mathbf{a}_t}$  Vektoren der Länge  $2n + N$  sind,  $\mathbf{q}_{\mathbf{a}_t}$  der Vektor der partiellen Ableitungen nach den Parametern  $\alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_n, \theta_1, \dots, \theta_N$ , in welche die Komponenten von  $\mathbf{a}_t$  für die Komponenten von  $\mathbf{a}$  eingesetzt wurden und  $\mathbf{H}$  die  $(2n + N, 2n + N)$ -Matrix der zweifachen partiellen Ableitungen nach  $\alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_n, \theta_1, \dots, \theta_N$  ist, in welche die Komponenten von  $\mathbf{a}_t$  für die Komponenten von  $\mathbf{a}$  eingesetzt wurden.

Um dieses Gleichungssysteme zu vereinfachen, geht man von folgenden Voraussetzungen aus:

- Die Zufallsvariablen  $\Theta_i$ , welche die Werte  $\hat{\theta}_i$  annehmen, sind unabhängig. Entsprechend werden die zweiten partiellen Ableitungen  $\frac{\partial^2 f}{\partial \theta_i \partial \theta_k}$  für  $j \neq k$  Null.
- Die Zufallsvariablen  $\mathfrak{A}_j$  und  $\mathfrak{B}_j$ , welche die Werte  $\hat{\alpha}_j$  und  $\hat{\beta}_j$  annehmen, sind unabhängig von den Zufallsvariablen  $\Theta_i$ , wodurch die zweiten partiellen Ableitungen  $\frac{\partial^2 f}{\partial \alpha_i \partial \theta_k}$  und  $\frac{\partial^2 f}{\partial \beta_i \partial \theta_k}$  Null werden.

- Die Zufallsvektoren  $(\mathfrak{A}_i, \mathfrak{B}_i)$  und  $(\mathfrak{A}_l, \mathfrak{B}_l)$  sind für  $i \neq l$  unabhängig, wodurch die zweiten partiellen Ableitungen  $\frac{\partial^2 f}{\partial \alpha_i \partial \beta_l}$ ,  $\frac{\partial^2 f}{\partial \beta_i \partial \alpha_l}$ ,  $\frac{\partial^2 f}{\partial \alpha_i \partial \alpha_l}$  und  $\frac{\partial^2 f}{\partial \beta_i \partial \beta_l}$  Null werden.

Es bleibt eine Matrix zurück, welche auf der Diagonalen zuerst  $n$   $(2, 2)$ -Matrizen mit den zweiten partiellen Ableitungen  $\frac{\partial^2 f}{\partial \alpha_i \partial \alpha_i}$ ,  $\frac{\partial^2 f}{\partial \alpha_i \partial \beta_i}$ ,  $\frac{\partial^2 f}{\partial \beta_i \partial \alpha_i}$  und  $\frac{\partial^2 f}{\partial \beta_i \partial \beta_i}$  aufweist und dann  $N$  zweite partielle Ableitungen  $\frac{\partial^2 f}{\partial \theta_i \partial \theta_i}$ . Auf Grund der Orthogonalität dieser Matrix, kann man die Fähigkeits-Parameter bei bekannten Item-Parametern einzeln schätzen. Ebenso kann man die Item-Parameter bei bekannten Fähigkeits-Parametern pro Item schätzen. Diese Überlegungen führen zum folgenden, von Birnbaum vorgeschlagenen iterativen Verfahren, um die Parameter zu schätzen (Birnbaum, 1968, S. 420).

Man startet mit einem grob geschätzten, standardisierten Fähigkeitsvektor der untersuchten Personen, berechnet die Itemparameter, dann wiederum die Fähigkeitsparameter, welche standardisiert werden. Mit Hilfe der Fähigkeitsparameter werden erneut die Itemparameter berechnet, etc., bis sich die Änderungen bei den Iterationen höchstens um einen festzulegenden Betrag unterscheiden.

Es stellt sich das Problem, wie ein geeigneter Start-Fähigkeitsvektor zu bestimmen ist. Man könnte z.B. für jede Person  $i$  die Summe

$$F_i := \sum_{j=1}^n x_{ij}$$

der erreichten Punkte  $x_{ij}$  der  $n$  Items berechnen und diese Summen standardisieren. Eine weitere Möglichkeit besteht darin, mit einer durch eine grobe Schätzung der Item-Schwierigkeiten gewichteten Summe von Punkten zu rechnen: Sei  $h_j$  die relative Häufigkeit, mit der das Item  $j$  korrekt beantwortet wird. Dann kann man folgende grobe Schätzung der Fähigkeiten einer Person  $i$  vornehmen:

$$F'_i := \sum_{j=1}^n x_{ij} (1 - h_j)$$

Wir gewichten also die Antwort auf das  $j$ -te Item der Person  $i$  mit  $1 - h_j$ . Je einfacher das Item, desto grösser wird  $h_j$  sein und desto weniger trägt entsprechend eine korrekte Antwort zur geschätzten Fähigkeit der Person bei. Anschliessend werden die  $F'_i$  für alle Personen standardisiert:

$$\theta'_i := \frac{F'_i - \overline{F'}}{s_{F'}}$$

wobei  $\overline{F'}$  der Mittelwert der  $F'_i$  ist und  $s_{F'}$  die Standardabweichung der  $F'_i$ . Mit Hilfe der  $\theta'_i$  werden dann die Itemparameter gemäss obigen Methoden geschätzt.

Am Schluss des Iterationsverfahrens werden die Fähigkeitsparameter standardisiert und die Itemparameter auf die derart gebildete Skala umgerechnet (Baker & Kim, 2004, S. 91) durch

$$\hat{\beta}'_j = \frac{\hat{\beta}_j - \overline{\hat{\theta}}}{\hat{s}_\theta}$$

und

$$\hat{\alpha}'_j = \hat{s}_\theta \hat{\alpha}_j.$$

Die JMLE-Methode führt nicht zu einem konsistenten Schätzverfahren, solange nicht alle möglichen  $2^n$  Antwortvektoren in einer Stichprobe vorkommen. Durch das Wachsen der

Stichprobengröße können in diesem Falle neu zu schätzende Fähigkeitsparameter hinzukommen. Dieses Wachstum der Anzahl der Fähigkeitsparameter wirkt auf die Schätzung der Itemparameter zurück (s. für eine entsprechende Diskussion, die auf eine Arbeit von Neyman und Scott (1948) zurückgeht, Baker und Kim (2004, S. 100)). Die in der Folge skizzierte MMLE-Schätzmethode kann als ein Lösungsversuch dieses Problems verstanden werden.

#### 2.2.4 Marginal Maximum Likelihood-Schätzmethode

Bei der *Marginal Maximum Likelihood-Schätzmethode (MMLE)* geht man - wie der Name sagt - von einer Randverteilung aus. Der Fähigkeitsparameter  $\theta_i$  der  $i$ -ten Person wird als Wert einer Zufallsvariable  $\Theta_i$  betrachtet, deren Verteilung durch eine Dichte  $g$  gegeben ist. Diese unbekannte Verteilung ist für alle Personen identisch und die entsprechenden Zufallsvariablen sind unabhängig. Die Hinzunahme von Personen in die Stichprobe hat unter diesen Voraussetzungen keinen Einfluss auf die Konsistenz der ML-Schätzung der Itemparameter. Oft wird für die Dichte von  $g$  die Standardnormalverteilung gewählt. Man betrachtet nun die Verteilung

$$P((X_{i1}, \dots, X_{in}) = (x_{i1}, \dots, x_{in}) \mid \Theta_i = \theta_i)$$

des Zufallsvektors  $(X_{i1}, \dots, X_{in})$ , unter der Bedingung, dass die Zufallsvariable  $\Theta_i$  den spezifischen Wert  $\theta_i$  annimmt. Bei gegebenem  $\theta_i$  ist der Zufallsvektor  $(X_{i1}, \dots, X_{in})$  von der Zufallsvariable  $\Theta_i$  unabhängig. Dies wird lokale Unabhängigkeit genannt: die Antwortvektoren verschiedener Personen mit derselben Fähigkeit sind unabhängig. Entsprechend ist in  $\theta_i$  die gemeinsame Verteilung von  $(X_{i1}, \dots, X_{in})$  und  $\Theta_i$  bestimmt durch das Produkt von  $g$  und der bedingten Dichtefunktion, die durch

$$P((X_{i1}, \dots, X_{in}) = (x_{i1}, \dots, x_{in}) \mid \Theta_i = \theta_i)$$

gegeben ist. Damit gilt

$$P((X_{i1}, \dots, X_{in}) = (x_{i1}, \dots, x_{in})) = \int_{-\infty}^{+\infty} P((X_{i1}, \dots, X_{in}) = (x_{i1}, \dots, x_{in}) \mid \Theta_i = \theta_i) g(\theta_i) d\theta_i,$$

denn

$$\begin{aligned} & \int_{-\infty}^{+\infty} P((X_{i1}, \dots, X_{in}) = (x_{i1}, \dots, x_{in}) \mid \Theta_i = \theta_i) g(\theta_i) d\theta_i \\ &= \int_{-\infty}^{+\infty} \frac{P((X_{i1}, \dots, X_{in}) = (x_{i1}, \dots, x_{in}), \Theta_i = \theta_i)}{g(\theta_i)} g(\theta_i) d\theta_i \\ &= \int_{-\infty}^{+\infty} P((X_{i1}, \dots, X_{in}) = (x_{i1}, \dots, x_{in}), \Theta_i = \theta_i) d\theta_i. \end{aligned}$$

Damit ist  $P((X_{i1}, \dots, X_{in}) = (x_{i1}, \dots, x_{in}))$  die Randverteilung der gemeinsamen Verteilung, die durch  $P((X_{i1}, \dots, X_{in}) = (x_{i1}, \dots, x_{in}) \mid \Theta_i = \theta_i) g(\theta_i)$  gegeben ist. Aus

$$P(X_{i1}, \dots, X_{in} = x_{i1}, \dots, x_{in}) = \int_{-\infty}^{+\infty} \prod_{j=1}^n (p_j(\theta_i))^{x_{ij}} (1 - p_j(\theta_i))^{1-x_{ij}} g(\theta_i) d\theta_i$$



erhält man mit

$$\begin{aligned}\alpha &:= (\alpha_1, \dots, \alpha_n) \\ \beta &:= (\beta_1, \dots, \beta_n)\end{aligned}$$

und

$$P_i(\alpha, \beta, \theta_i) := \prod_{j=1}^n (p_j(\alpha_j, \beta_j, \theta_i))^{x_{ij}} (1 - p_j(\alpha_j, \beta_j, \theta_i))^{1-x_{ij}}$$

die folgende Likelihood-Funktion in den Parametervektoren  $\alpha$  und  $\beta$

$$f_i(\alpha, \beta) = \int_{-\infty}^{+\infty} P_i(\alpha, \beta, \theta_i) g(\theta_i) d\theta_i$$

und damit für die gemeinsame LL-Funktion aller Personen

$$\ln \prod_{i=1}^N f_i(\alpha, \beta) = \sum_{i=1}^N \ln \int_{-\infty}^{+\infty} P_i(\alpha, \beta, \theta_i) g(\theta_i) d\theta_i.$$

Es gilt nun wiederum, die Parameter so zu bestimmen, dass die LL-Funktion maximal wird. Es gibt dazu verschiedene Verfahren und Zusammensetzungen von solchen, die hier nicht diskutiert werden (s. z.B. Baker und Kim (2004, S. 157, ff.), Bock und Lieberman (1970) sowie Bock und Aitkin (1981)). Nach der Bestimmung der Item-Parameter können die Fähigkeitsparameter z.B. mit den oben eingeführten Methoden von Unterkapitel 2.2.2 berechnet werden.

Der Schwachpunkt der MMLE-Methode besteht in einer begründbaren Wahl von  $g$ : die Schätzung einer angemessenen Verteilung der Fähigkeitsparameter setzt voraus, dass man die Häufigkeitsverteilung der Fähigkeitsparameter kennt und damit eine rechtfertigbare Festlegung der Abstände zwischen diesen *vor* deren Schätzung möglich ist. Die MMLE-Methode ist im R-Paket ltm implementiert (Rizopoulos, 2006).

## 2.3 Diagnostik von Modellen für zweiwertige Items

Vor der Berechnung der Parameter sind die Daten auf ihre Plausibilität bezüglich der Anwendung der IRT zu untersuchen. Nach deren Berechnung ist zu untersuchen, ob die geschätzten ICCs zu den Daten passen und wie präzise die Schätzer in welchen Bereichen sind. Es werden ein paar diesbezüglich Methoden vorgestellt (für weiterführende Literatur, s. van der Linden und Hambleton (1997)).

### 2.3.1 Überprüfung der Daten

Ein erstes Bild der Daten kann man sich verschaffen, indem man die Anzahl der Antwortvektoren ermittelt, die nur falsche oder die nur richtige Antworten enthalten. Zudem bestimmt man die Anzahl der Personen, die nur korrekte oder nur falsche Antworten geben. Man kann die absolute Häufigkeitsverteilung der korrekten Antworten betrachten. Dies kann durch eine Analyse der relativen Häufigkeiten ergänzt werden. Mit diesen kann man zudem das Logit pro Item berechnen, definiert als

$$\text{Logit}(\text{Item}_j) = \ln(h_j / (1 - h_j))$$

oder alternativ durch Verwendung der empirischen Logits, wie sie auf Seite 6 definiert sind.

Die IRT setzt voraus, dass die Items dieselbe verborgene Eigenschaft messen. Entsprechend sollten diese korreliert sein. Rizopoulos (2006, S. 6) schlägt diesbezüglich die Messung der paarweisen Abhängigkeit via  $\chi^2$ -Tests für die  $2 \times 2$ -Tabellen vor, die durch Kreuzen der Itemvariablen erzeugt werden. Nichtsignifikante Ergebnisse deuten auf mangelnde Abhängigkeit hin. Allerdings gibt der  $\chi^2$ -Test nicht die Richtung der Abweichung an. Besser wären deshalb wohl der rechtsseitige exakte Test von Fisher oder die Angabe der Odds-Ratio samt Ergebnissen entsprechender Tests. Die Odds-Ratios müssten bezüglich der korrekten Antworten signifikant grösser als 1 sein.

Zudem sollten die Itemvariablen mit der verborgenen Variable (Fähigkeiten der Probanden) korrelieren. Als erste Näherung an diese Fähigkeiten kann man die Summe der korrekten Antworten betrachten. Rizopoulos (2006, S. 6) schlägt deshalb vor, die Pearson-Korrelation zwischen jedem der Itemvariablen und der Variable „Summe der korrekten Antworten pro Person“ zu berechnen. Geeigneter als die Pearson-Korrelation wären jedoch wohl Kendalls  $\tau_b$  oder Gamma (s. für Kendalls  $\tau_b$  Agresti (1990, S. 34) und für Gamma Agresti (1990, S. 22 f.)), da die Summen höchstens als ordinalskaliert aufgefasst werden können und wenige unterschiedliche Summen möglich sind (viele Bindungen).

### 2.3.2 Untersuchungen auf Güte der Anpassung

Die Güte der Anpassung der geschätzten ICCs an die Daten kann graphisch überprüft werden. Man kann jede geschätzte Kurve mit den empirische Log-Odds (s. Seite 6) vergleichen, sofern pro Fähigkeitskategorie genügend viele Personen vorkommen. Zusätzlich kann man die unterschiedlichen Arten von Residuen untersuchen, die im Rahmen der logistischen Regression von Bedeutung sind (z.B. Devianz-Residuen, Pearson-Residuen, etc.).

Für jedes Item kann man für die logistische Regressionsfunktion mit den Parametern  $a$  und  $b$  die üblichen Tests auf Güte der Anpassung durchführen ( $\chi^2$ -Anpassungstest; Likelihood-Ratio-Test- $\chi^2$ -Test; Testen der residualen Devianz („Scaled Deviance“)). Da in der IRT auch die Fähigkeitsparameter geschätzt werden, ist die Anzahl der Freiheitsgrade in diesen Tests entsprechend anzupassen. Für eine ausführlichere und flächendeckendere Behandlung der verschiedenen Möglichkeiten, die Güte der Anpassung zu überprüfen, sei auf Agresti (1990, S. 83) oder Hosmer und Lemeshow (2000, Kapitel 5) verwiesen.

Für eine Anpassungsprüfung jeweils zweier Itemvariablen schlägt Rizopoulos (2006, S. 7) die Berechnung der einzelnen Summanden des  $\chi^2$ -Koeffizienten vor, die durch den paarweisen Vergleich der Itemvariablen mit den gemäss Modell zu erwartenden Häufigkeiten entstehen. Es wird für die  $2 \times 2$ -Tabelle der Wert

$$\frac{\left(k_{ij} - k_{ij}^e\right)^2}{k_{ij}^e}$$

berechnet ( $k_{ij}$  faktische Häufigkeit,  $k_{ij}^e$  erwartete Häufigkeit der Zelle  $(i, j)$ ), der für jede Zelle  $(i, j)$ ,  $i, j \in \{1, 2\}$ , identisch ist. Es ergeben sich  $\binom{n}{2}$  solcher Werte ( $n =$  Anzahl der Items). Die Anpassung wird als gut beurteilt, wenn kein solcher Wert grösser als 3.5 ist. Die Wahl von 3.5 wird nicht weiter begründet (Ist das standardisierte Residuum 3.5, so ist der Chi-Quadrat-Koeffizient  $4 \cdot 3.5 = 14$ . Der  $p$ -Wert mit einem Freiheitsgrad ist damit 0.000182811). Die Abweichung wird zudem durch einen  $\chi^2$ -Test beurteilt. Schliesslich kann man diese Überprüfung auch für jeweils drei gekreuzte Itemvariablen durchführen: man vergleicht in

den  $2 \times 2 \times 2$ -Tabellen die faktischen mit den erwarteten Häufigkeiten.  $\binom{n}{3}$  solcher Werte werden berechnet.

Für eine Überprüfung der Güte des gesamten Modells verwendet Rizopoulos (2006, S. 7) einen parametrischen Bootstrap-Pearson- $\chi^2$ -Test. Unter der Annahme der berechneten Parameter und des berechneten Modells werden Daten simuliert. Die ausgezählten korrekten Antworten pro Fähigkeitsklasse werden dann mit den faktischen mittels eines Pearson- $\chi^2$ -Tests verglichen.

Zu beachten ist, dass im IRT-Zusammenhang die verschiedenen Methoden nicht nur - wie dies bei der logistischen Regression der Fall ist - die Anpassung des Modells bei gegebenen  $\theta_i$ -Werten zu begutachten erlauben. Gemeinsam mit den Item-Parametern wird auch die Güte der Abstände zwischen den Fähigkeitsparametern überprüft, denn die Item-Parameter und Fähigkeitsparameter werden so geschätzt, dass *insgesamt* eine optimale *gegenseitige* Anpassung der Parameter unter der vorausgesetzten allgemeinen Form der ICC gefunden wird.

### 2.3.3 Informationsfunktionen

Um die Varianz der Fähigkeitsparameter in den verschiedenen Bereichen zu beurteilen, wird die Informationsfunktion begutachtet. Die Information eines Schätzers ist umgekehrt proportional zu seiner Varianz: Je grösser die Streuung des Schätzers, desto uninformativer ist er. Entsprechend wird die Information eines Schätzers  $\mathfrak{D}$  definiert als

$$I(\mathfrak{D}) := \frac{1}{V(\mathfrak{D})}.$$

Die Varianz des Schätzers  $\Theta_i$ , der die Fähigkeitsparameter  $\theta_i$  von Personen als Werte annimmt, ist (s. Seite 12)

$$V(\Theta_i) = \frac{1}{\sum_{j=1}^n \hat{\alpha}_j^2 p_j(\Theta_i) (1 - p_j(\Theta_i))}, \quad (5)$$

wobei  $n$  die Anzahl der Items ist. Damit ist die Information von  $\Theta_i$  gegeben durch

$$I(\Theta_i) = \sum_{j=1}^n \hat{\alpha}_j^2 p_j(\Theta_i) (1 - p_j(\Theta_i)). \quad (6)$$

Man kann die Information als Funktion von  $\theta$  darstellen. So erhält man für eine Itematterie eine Funktion

$$T_I : \mathbb{R} \rightarrow \mathbb{R}^+$$

$$T_I(\theta) = \sum_{j=1}^n \hat{\alpha}_j^2 p_j(\theta) (1 - p_j(\theta)),$$

die *Testinformationsfunktion* genannt wird (s. Abbildung 2 für ein Beispiel). Die Testinformationsfunktion drückt aus, in welchen Fähigkeitsbereichen die Information einer Itematterie und damit Präzision des Schätzers grösser oder seine Varianz kleiner ist.

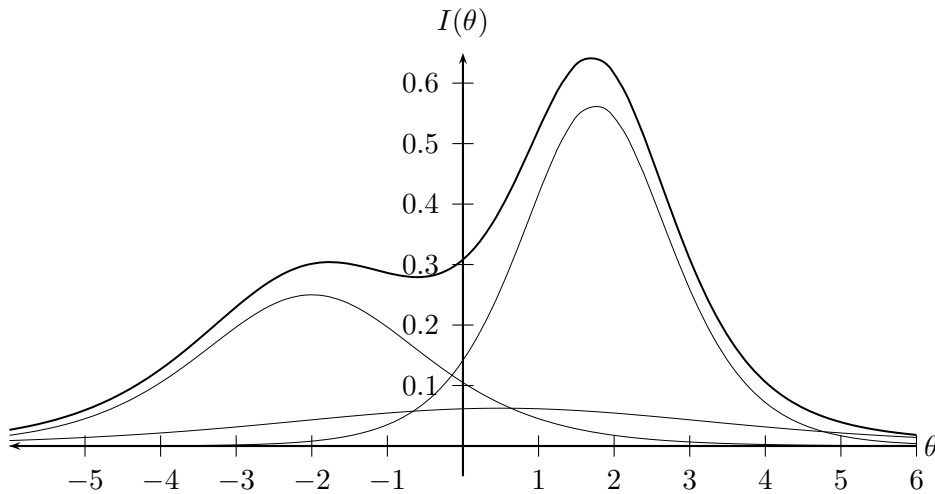


Abbildung 2: Testinformationsfunktion (fette Kurve) und die Iteminformationsfunktionen für drei mit logistischen ICCs modellierte Items mit den Parametern  $\alpha_1 = 1$ ,  $\beta_1 = -2$ ;  $\alpha_2 = 0.5$ ,  $\beta_2 = 0.5$ ;  $\alpha_3 = 1.5$ ,  $\beta_3 = 1.75$

An den Formeln (5) und (6) ersieht man, dass die verschiedenen Items zur Verringerung der Varianz des Schätzers  $\Theta_i$  und damit zu dessen Information beitragen. Ein Item  $j$  trägt dort am meisten zur Information bei, wo  $\hat{\alpha}_j^2 p_j(\theta)(1 - p_j(\theta))$  maximal ist, d.h. dort, wo der Schwierigkeitsparameter des Items liegt. Den Beitrag der einzelnen Items zur Information kann man überprüfen, indem man die Funktionen

$$I_j : \mathbb{R} \rightarrow \mathbb{R}^+$$

$$I_j(\theta) := \hat{\alpha}_j^2 p_j(\theta)(1 - p_j(\theta))$$

der  $n$  Items einzeln analysiert (s. Abbildung 2). Diese Funktionen werden *Iteminformationsfunktionen* genannt. Man erhält in einem Intervall eine flache Testinformationsfunktion, wenn die Schwierigkeitsparameter der Items über das Intervall regelmässig streuen und die Trennschärfeparameter nicht zu gross sind. Umgekehrt erhält man eine Testinformationsfunktion mit einer scharfen Spitze, wenn die Schwierigkeitsparameter der Items nahe beieinander liegen und die Trennschärfen gross sind. Mit Hilfe der Informationsfunktion kann man entsprechend überprüfen, ob eine Itembatterie für ein spezifisches Ziel geeignet ist oder nicht. Will man in einem bestimmten, kleinen Bereich unterschiedlichen Fähigkeiten möglichst genau messen (z.B. bei einem Aufnahmeexamen), so sollte die Testinformationsfunktion in diesem Bereich möglichst grosse Werte annehmen. Will man aber in einem breiten Bereich die unterschiedlichen Fähigkeiten möglichst gleich präzise messen (z.B. Leistungsmessung in der Schule), so sollte die Testinformationsfunktion in diesem Bereich möglichst konstant sein.

Diese Überlegungen können für bereits durchgeführte Tests angestellt werden, um zu überprüfen, wie gut sie in einem spezifischen Bereich die Fähigkeiten messen, oder man kann damit Tests angesichts einer Testbatterie mit bekannten Itemparametern spezifischen Zielen anzupassen: Kennt man die Itemparameter von Items, kann man einen Test zusammenstellen, der im anvisierten Bereich optimal ist.

### 3 IRT für ordinalskalierte polytome Variablen

Es gibt verschiedene Modelle für ordinalskalierte Variablen mit mehr als zwei Ausprägungen. Hier wird das sogenannte *Graded Item Response-Modell* (GIRT) behandelt, das auf Fukimo Samejima (1969) zurückgeht. Dabei wird dem Aufbau des vorherigen Kapitels gefolgt. Zuerst wird für polytome Variablen ein der ICC analoges Modell entwickelt. Anschliessend werden Parameterschätzmethoden und die Diagnostik diskutiert.

#### 3.1 Kumulative Charakteristische Item–Kategorie-Kurven

Items sind auch im Falle von mit GIRT-Modellen behandelten Variablen im Allgemeinen Fragen, welche z.B. in einem Fragebogen auftauchen und die von Untersuchungsobjekten beantwortet werden. Pro Item werden mehrere Antwortmöglichkeiten offeriert, wobei auf Personen, welche das Item unterschiedlich beantworten, eine *schwache Ordnung* bestimmt wird. Eine schwache Ordnung wird hier definiert als eine mengentheoretische Relation, welche irreflexiv, asymmetrisch und transitiv ist sowie auf die Elemente verschiedener Äquivalenzklassen einer Zerlegung definiert ist (s. auch Anhang A.1, S. 37).

**Beispiel 6.** *Durch die statistische Variable „Zufriedenheit“ mit den Werten „sehr zufrieden“, „zufrieden“, „mässig zufrieden“, „unzufrieden“ (Likert-Skala) wird eine Zerlegung der Menge der antwortenden Personen in vier Äquivalenzklassen erzeugt: die Äquivalenzklassen der sehr zufriedenen, der zufriedenen, der mässig zufriedenen und der unzufriedenen Menschen. Zwischen den Elementen unterschiedlicher Äquivalenzklassen kann in eine der beiden Richtungen die Relation „ist zufriedener als“ definiert werden, welche irreflexiv, asymmetrisch und transitiv ist. Durch diese Relation werden die Personen unterschiedlicher Äquivalenzklassen gemäss der Stärke Ihrer Zufriedenheit geordnet.*

Die Antwortmöglichkeiten auf Fragen, die einem Item entsprechen, werden „Itemkategorien“ genannt. Die Itemkategorien drücken die unterschiedliche Stärke der Einstellung einer Person aus. Bezüglich der Items sind wiederum zwei Eigenschaften von Bedeutung: das Anspruchsniveau des Items und die Trennschärfe. Das Anspruchsniveau des Items entspricht dabei der Schwierigkeit von Items im dichotomen Fall.

Das *Anspruchsniveau* eines Items kann bezüglich der Einstellungsstärke von Personen definiert werden: ein Item kann anspruchsvoll sein für Personen mit tiefer Einstellungsstärke und anspruchlos sein für Personen mit hoher Einstellungsstärke. Allgemein kann das Anspruchsniveau eines Items bezüglich einer Person als die Wahrscheinlichkeit betrachtet werden, dass ihre Antwort in hohe Itemkategorien fällt oder in tiefe: fällt die Antwort der Person mit hoher Wahrscheinlichkeit in tiefe Itemkategorien, so ist das Item bezüglich ihrer Einstellungsstärke anspruchsvoll. Diese Grundideen sind nun zu präzisieren.

Wir gehen davon aus, dass das Item  $K > 2$  Kategorien enthält und dass die Wahl von Kategorien mit grösserem  $k$  durch eine Person umso wahrscheinlicher ist, je stärker ihre Einstellung ist. Zudem wird vorausgesetzt, dass die Einstellungsstärken  $\theta$  intervallskaliert sind. Für jede Itemkategorie  $k < K$  erhält man eine strikt monoton steigende Funktion  $\mathfrak{C}_k$  zwischen 0 und 1, wenn man die Wahrscheinlichkeit in Abhängigkeit von  $\theta$  betrachtet, dass von Personen mit  $\theta$  eine Itemkategorie  $l$  mit  $l > k$  gewählt wird. Auf diese Art erhält man  $K - 1$  Funktionen, die hier *Kumulative Charakteristische Item–Kategorie-Kurven* genannt werden. (Cumulative Item-Response Characteristic Curves; CIRCC) Diese sollten sich nicht schneiden. Setzen wir für  $k = K$

$$\mathfrak{C}_K := 0$$

und für  $k = 0$

$$\mathfrak{C}_0 := 1,$$

so gilt für die Wahrscheinlichkeit  $p_k$ ,  $k \in \mathbb{N}_K^*$ , die  $k$ -te Kategorie zu wählen,

$$p_k(\theta) = \mathfrak{C}_{k-1}(\theta) - \mathfrak{C}_k(\theta). \quad (7)$$

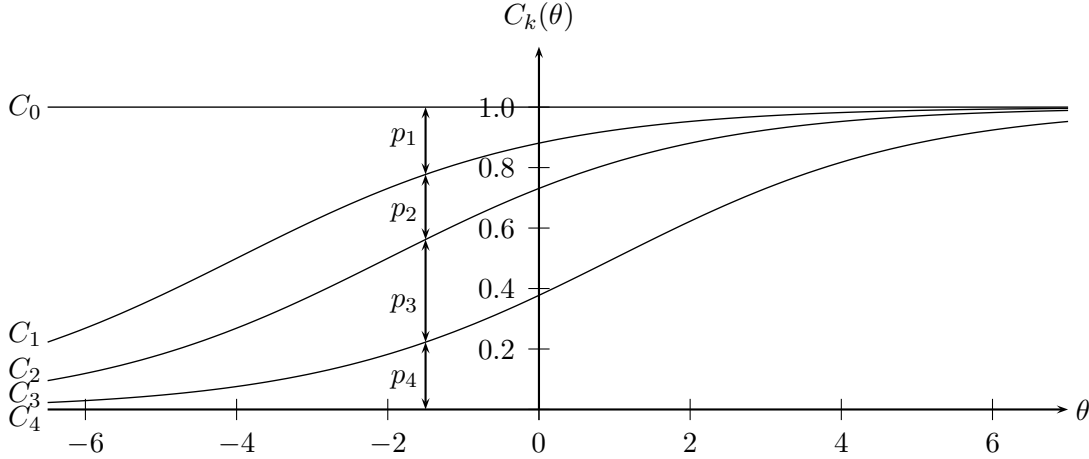


Abbildung 3: Logistische CIRCCs für eine polytome Variable mit 4 Kategorien und  $C_0 = 1$  sowie  $C_4 = 0$ .  $p_i$  ist die Wahrscheinlichkeit, bei einer Einstellungsstärke von  $\theta = -1.5$  eine Antwort in der  $i$ -te Kategorie zu geben ( $\alpha = 0.5; b_1 = 2; b_2 = 1; b_3 = -0.5; \beta_1 = -4; \beta_2 = -2; \beta_3 = 1$ ).

Wählt man für die CIRCCs und für  $k < K$  das logistische Modell

$$C_k : \mathbb{R} \rightarrow ]0, 1[$$

$$C_k(\theta) := \frac{\exp(\alpha(\theta - \beta_k))}{1 + \exp(\alpha(\theta - \beta_k))},$$

schneiden sich die Kurven nicht, wenn der Trennschärfeparameter  $\alpha$  für alle CIRCCs des Items identisch ist und für die Parameter  $\beta_k$  gilt:  $\beta_{k+1} > \beta_k$  für  $k + 1 < K$ . Wir nennen die CIRCCs bei dieser Wahl „logistisch“. Zu beachten ist, dass für CIRCCs gilt:

$$\mathfrak{C}_k(\theta) > \mathfrak{C}_{k+1}(\theta)$$

für  $0 \leq k < K$  (s. Abbildung 3).

Als nächstens geht es darum, Parameterschätzmethoden einzuführen. Es wird dabei wie bei den dichotomen Variablen vorgegangen: für die Entwicklung der Grundideen ist die JMLE-Methode geeignet. Dazu wird vorgängig wiederum die Schätzung der Item-Parameter auf der Grundlage von Einstellungsstärken und die Schätzung der Einstellungsstärken auf der Grundlage der Item-Parameter behandelt.

## 3.2 Parameter-Schätzmethoden

### 3.2.1 Schätzen der Item-Parameter bei gegebenen Einstellungsstärken

Die Item-Parameter werden mit ML-Methoden geschätzt. Entsprechend ist zuerst ein passendes Wahrscheinlichkeitsmodell für die Verteilung der Antworten auf die Kategorien eines Items

einzuführen. Ausgangspunkt für das bei ML-Methoden benötigte Wahrscheinlichkeitsmodell betrachtet man relative Häufigkeiten, mit denen Personen mit spezifischer Einstellungsstärke  $\theta_i$  eine spezifische Item-Kategorie  $k$  wählen -  $i \in \mathbb{N}_m^*$  mit der Anzahl  $m$  von Klassen von Personen mit unterschiedlichen Einstellungsstärken  $\theta_i$ . Geeignet ist entsprechend die Betrachtung der multinomialverteilten Statistik  $(X_{i1}, \dots, X_{iK})$ , welche die Anzahlen  $(x_{i1}, \dots, x_{iK})$  der Antworten in den  $K$  Itemkategorien durch Personen mit der Einstellungsstärke  $\theta_i$  als Werte annimmt. Die Parameter der Verteilung sind  $n_i$  - die Anzahl der Personen mit der Einstellungsstärke  $\theta_i$  - und  $p_k$ , die Wahrscheinlichkeit, dass von Personen mit  $\theta_i$  die  $k$ -te Itemkategorie gewählt wird. Es gilt

$$P((X_{i1}, \dots, X_{iK}) = (x_{i1}, \dots, x_{iK})) = \frac{n_i!}{\prod_{k=1}^K x_{ik}!} \prod_{k=1}^K p_k^{x_{ik}}(\theta_i)$$

mit

$$\sum_{k=1}^K p_k(\theta_i) = 1.$$

Bei logistischen CIRCCs und mit den bereits bekannten Vereinfachungen  $a := \alpha$  und  $b_k := -\alpha\beta_k$  (s. Seite 5, Definitionsgleichungen (3)) ergibt sich mit (7) von Seite 20 für die Wahrscheinlichkeiten  $p_k(\theta_i)$

$$\begin{aligned} p_1(\theta_i) &= 1 - \pi_1(\theta_i) = 1 - \frac{\exp(a\theta + b_1)}{1 + \exp(a\theta + b_1)} \\ p_k(\theta_i) &= \pi_{k-1}(\theta_i) - \pi_k(\theta_i) \\ &= \frac{\exp(a\theta + b_{k-1})}{1 + \exp(a\theta + b_{k-1})} - \frac{\exp(a\theta + b_k)}{1 + \exp(a\theta + b_k)} \text{ für } k \in \mathbb{N}_{K-1}^* \setminus \{1\} \\ p_K(\theta_i) &= \frac{\exp(a\theta + b_{K-1})}{1 + \exp(a\theta + b_{K-1})}. \end{aligned}$$

Lässt man den für die Parameterschätzmethoden überflüssigen Term

$$\frac{n_i!}{\prod_{k=1}^K x_{ik}!}$$

weg, ergibt sich beim Vorliegen von  $m$  Klassen von Personen mit unterschiedlichen Einstellungsstärken  $\theta_i$  die Likelihood-Funktion in den Parametern  $a, b_1, \dots, b_{K-1}$

$$f(a, b_1, \dots, b_{K-1}) = \prod_{i=1}^m \prod_{k=1}^K p_{ik}^{x_{ik}}(a, b_1, \dots, b_{K-1})$$

und damit die LL-Funktion

$$\begin{aligned}
\ln f(a, b_1, \dots, b_{K-1}) &= \sum_{i=1}^m \sum_{k=1}^K x_{ik} \ln p_{ik}(a, b_1, \dots, b_{K-1}) \\
&= \sum_{i=1}^m \left( x_{i1} \ln \left( 1 - \frac{\exp(a\theta + b_1)}{1 + \exp(a\theta + b_1)} \right) + \right. \\
&\quad \sum_{k=2}^{K-1} x_{ik} \ln \left( \frac{\exp(a\theta + b_{k-1})}{1 + \exp(a\theta + b_{k-1})} - \frac{\exp(a\theta + b_k)}{1 + \exp(a\theta + b_k)} \right) \\
&\quad \left. + x_{iK} \ln \frac{\exp(a\theta + b_{K-1})}{1 + \exp(a\theta + b_{K-1})} \right).
\end{aligned} \tag{8}$$

Als nächstes gilt es nun, die Parameter  $a, b_1, \dots, b_{K-1}$  zu schätzen. Man kann diese mit dem Newton-Raphson-Verfahren oder mit dem Fisher-Scoring berechnen. Für beide müssen die ersten und zweiten Ableitungen nach den Parametern bestimmt werden. Während man für das Newton-Raphson-Verfahren unmittelbar die zweiten Ableitungen verwendet, müssen für das Fisher-Scoring die Erwartungswerte der Komponenten der Hesse-Matrix bezüglich den Zufallsvariablen  $X_{ik}$  ermittelt werden.

Es gilt (s. Beweise S. 44) mit

$$\begin{aligned}
P_{ik}^* &:= \frac{\exp(a\theta_i + b_k)}{1 + \exp(a\theta_i + b_k)} \\
Q_{ik}^* &:= 1 - P_{ik}^* \\
P_{i0}^* &:= 1 \\
P_{iK}^* &:= 0 \\
\ln f &:= \ln f(a, b_1, \dots, b_{K-1})
\end{aligned}$$

und

$$p_{ik} := p_{ik}(a, b_1, \dots, b_{K-1})$$

$$1. \quad \frac{\partial}{\partial b_k} \ln f = \sum_{i=1}^m \left( -\frac{x_{ik} P_{ik}^* Q_{ik}^*}{P_{i,k-1}^* - P_{ik}^*} + \frac{x_{i,k+1} P_{ik}^* Q_{ik}^*}{P_{ik}^* - P_{i,k+1}^*} \right) =: L_k$$

$$2. \quad \frac{\partial}{\partial a} \ln f = \sum_{i=1}^m \theta_i \sum_{k=1}^K \frac{x_{ik} (P_{i,k-1}^* Q_{i,k-1}^* - P_{ik}^* Q_{ik}^*)}{P_{i,k-1}^* - P_{ik}^*} := L_a$$

$$3. \quad \frac{\partial^2}{\partial b_k \partial b_k} \ln f =$$

$$\begin{aligned}
&\sum_{i=1}^m \left( \left( -x_{ik} \frac{P_{ik}^* Q_{ik}^* (1 - 2P_{ik}^*) p_{ik} - P_{ik}^* Q_{ik}^* P_{ik}^* Q_{ik}^*}{p_{ik}^2} \right) + \right. \\
&\quad \left. \left( x_{i,k+1} \frac{P_{ik}^* Q_{ik}^* (1 - 2P_{ik}^*) p_{i,k+1} - P_{ik}^* Q_{ik}^* P_{ik}^* Q_{ik}^*}{p_{i,k+1}^2} \right) \right)
\end{aligned}$$

$$4. \quad \frac{\partial^2}{\partial b_k \partial b_{k-1}} \ln f = \sum_{i=1}^m \left( x_{ik} \frac{P_{ik}^* Q_{ik}^* P_{i,k-1}^* Q_{i,k-1}^*}{p_{ik}^2} \right)$$



$$5. \frac{\partial^2}{\partial b_k \partial b_{k+1}} \ln f = \sum_{i=1}^m \left( x_{ik+1} \frac{P_{ik}^* Q_{ik}^* P_{ik+1}^* Q_{ik+1}^*}{p_{ik+1}^2} \right)$$

$$6. \frac{\partial^2}{\partial a \partial a} \ln f =$$

$$\sum_{i=1}^m \sum_{k=1}^K x_{ik} \left\{ \frac{\theta_i^2 \left[ P_{i,k-1}^* Q_{i,k-1}^* (1 - 2P_{i,k-1}^*) - P_{ik}^* Q_{ik}^* (1 - 2P_{ik}^*) \right]}{p_{ik}} - \frac{\theta_i^2}{p_{ik}^2} (P_{i,k-1}^* Q_{i,k-1}^* - P_{ik}^* Q_{ik}^*)^2 \right\}$$

$$7. \frac{\partial^2 \ln f}{\partial b_k \partial a} \ln f =$$

$$\sum_{i=1}^m \left( -x_{ik} \frac{p_{ik} \frac{\partial}{\partial a} (P_{ik}^* Q_{ik}^*) - P_{ik}^* Q_{ik}^* \frac{\partial}{\partial a} p_{ik}}{p_{ik}^2} + x_{ik+1} \frac{p_{ik+1} \frac{\partial}{\partial a} (P_{ik}^* Q_{ik}^*) - P_{ik}^* Q_{ik}^* \frac{\partial}{\partial a} p_{ik+1}}{p_{ik+1}^2} \right)$$

mit

$$\begin{aligned} \frac{\partial}{\partial a} (P_{ik}^* Q_{ik}^*) &= \theta_i P_{ik}^* Q_{ik}^* (1 - 2P_{ik}^*) \\ \frac{\partial}{\partial a} p_{ik} &= \theta_i (P_{i,k-1}^* Q_{i,k-1}^* - P_{ik}^* Q_{ik}^*). \end{aligned}$$

Die zweiten partiellen Ableitungen  $\frac{\partial^2 \ln f}{\partial b_k \partial b_j}$  mit  $|k - j| > 1$  werden zu Null, da in  $\frac{\partial \ln f}{\partial b_k}$  der Ausdruck  $b_j$  nicht vorkommt. Für das Fisher-Scoring braucht man nun noch die Erwartungswerte der Komponenten der Hesse-Matrix bezüglich der Zufallsvariablen  $X_{ik}$ . Es gilt mit

$$n_i := \sum_{k=1}^K x_{ik}$$

und

$$E(X_{ik}) = p_{ik} n_i$$

(für Beweise s. Seite 48):

$$1. E \left( \frac{\partial^2 \ln f}{\partial b_k \partial b_k} \right) = - \sum_{i=1}^m n_i (P_{ik}^* Q_{ik}^*)^2 \left( \frac{1}{p_{ik}} + \frac{1}{p_{ik+1}} \right) =: \Lambda_{kk}$$

$$2. E \left( \frac{\partial^2 \ln f}{\partial b_k \partial b_{k-1}} \right) = \sum_{i=1}^m n_i P_{ik}^* Q_{ik}^* P_{i,k-1}^* Q_{i,k-1}^* \left( \frac{1}{p_{ik}} \right) =: \Lambda_{kk-1}$$

$$3. E \left( \frac{\partial^2 \ln f}{\partial b_k \partial b_{k+1}} \right) = \sum_{i=1}^m n_i P_{ik}^* Q_{ik}^* P_{i,k+1}^* Q_{i,k+1}^* \left( \frac{1}{p_{ik+1}} \right) =: \Lambda_{kk+1}$$

$$4. E \left( \frac{\partial^2 \ln f}{\partial a \partial a} \right) = - \sum_{i=1}^m n_i \theta_i^2 \sum_{k=1}^K (P_{i,k-1}^* Q_{i,k-1}^* - P_{ik}^* Q_{ik}^*)^2 \left( \frac{1}{p_{ik}} \right) =: \Lambda_{aa}$$

$$5. E \left( \frac{\partial^2 \ln f}{\partial b_k \partial a} \right) = \sum_{i=1}^m P_{ik}^* Q_{ik}^* n_i \theta_i \left( \frac{(P_{i,k-1}^* Q_{i,k-1}^* - P_{ik}^* Q_{ik}^*)}{p_{ik}} + \frac{(P_{ik}^* Q_{ik}^* - P_{i,k+1}^* Q_{i,k+1}^*)}{p_{ik+1}} \right) =: \Lambda_{ka}$$

Damit sind alle Grössen für das Fisher-Scoring hergeleitet. Der Iterationsprozess wird durch die folgende Gleichung beschrieben:

$$\begin{pmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{b}_3 \\ \vdots \\ \hat{b}_{K-1} \\ \hat{a} \end{pmatrix}_{t+1} = \begin{pmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{b}_3 \\ \vdots \\ \hat{b}_{K-1} \\ \hat{a} \end{pmatrix}_t - \begin{pmatrix} \Lambda_{11} & \Lambda_{12} & 0 & \cdots & 0 & \Lambda_{1a} \\ \Lambda_{21} & \Lambda_{22} & \Lambda_{23} & \cdots & 0 & \Lambda_{2a} \\ 0 & \Lambda_{32} & \Lambda_{33} & \cdots & 0 & \Lambda_{3a} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \Lambda_{K-1K-1} & \Lambda_{K-1a} \\ \Lambda_{a1} & \Lambda_{a2} & \Lambda_{a3} & \cdots & \Lambda_{aK-1} & \Lambda_{aa} \end{pmatrix}_t^{-1} \begin{pmatrix} L_1 \\ L_2 \\ L_3 \\ \vdots \\ L_{K-1} \\ L_a \end{pmatrix}$$

Nach der Berechnung der Parameter  $a, b_1, \dots, b_{K-1}$  werden daraus die Parameter  $\alpha, \beta_1, \dots, \beta_{K-1}$  berechnet. Diese bestimmen die CIRCCs, welche die Grenzen der Intervalle festlegen, die  $p_k$  definieren. Das Anspruchsniveau einer Itemkategorie wird besser durch eine Funktion ausgedrückt, die durch die Mitte dieser Intervalle verläuft. Entsprechend werden die Parameter  $\beta_k$  umgerechnet in

$$\begin{aligned} \beta_1^* &:= \beta_1 \\ \beta_k^* &:= \frac{\beta_{k-1} - \beta_k}{2} \text{ für } 1 < k < K \\ \beta_K^* &:= \beta_{K-1}, \end{aligned}$$

wodurch man für jede Itemkategorie einen Parameter erhält.

### 3.2.2 Schätzen der Einstellungsstärken bei gegebenen Item-Parametern

Unter der Voraussetzung, dass die Item-Parameter der  $j \in \mathbb{N}_n^*$  Items mit  $K_j$  Kategorien bekannt sind, möchte man die Fähigkeitsparameter von  $i \in \mathbb{N}_N^*$  Personen bestimmen. Da auf einmal nur eine Person betrachtet wird, wird der Index  $i$  gewöhnlich weggelassen. Die Antworten ergeben für eine Person  $i$  einen Antwortvektor  $x_i = (x_{1k_1}, \dots, x_{nk_n})$  mit  $x_{jk_j} \in \{0, 1\}$ , wobei  $x_{jk_j} = 1$ , wenn die Antwort für das Item  $j$  in die Kategorie  $k$  fällt. Mit

$$jk := jk_j$$

gilt bei Unabhängigkeit

$$P((X_{1k}, \dots, X_{nk}) = (x_{1k}, \dots, x_{nk})) = \prod_{j=1}^n \prod_{k=1}^{K_j} p_{jk}^{x_{jk}}.$$

Die LL-Funktion ist

$$\ln f(\theta_i) = \sum_{j=1}^n \sum_{k=1}^{K_j} x_{jk} \ln p_{jk}.$$

Bei der Verwendung der logistischen CIRCCs - und den Abkürzungen des Kapitels 3.2.1 (S. 22) ergeben sich die folgenden Ableitungen (Beweise A.8, S. 50):

$$\frac{\partial \ln f}{\partial \theta_i} = \sum_{j=1}^n \sum_{k=1}^{K_j} x_{jk} a_j \frac{P_{jk-1}^* Q_{jk-1}^* - P_{jk}^* Q_{jk}^*}{p_{jk}}$$

Die zweite Ableitung ist

$$\frac{\partial^2 \ln f}{\partial \theta_i \partial \theta_i} = \sum_{j=1}^n \sum_{k=1}^{K_j} x_{jk} \left\{ \frac{a_j^2 \left[ P_{i,k-1}^* Q_{ik-1}^* (1 - 2P_{i,k-1}^*) - P_{ik}^* Q_{ik}^* (1 - 2P_{ik}^*) \right]}{p_{ik}} - \frac{a_j^2}{p_{jk}^2} (P_{jk-1}^* Q_{jk-1}^* - P_{jk}^* Q_{jk}^*)^2 \right\}.$$

Für den Erwartungswert der zweiten Ableitung bezüglich  $X_{ik}$  erhält man:

$$E \left( \frac{\partial^2 \ln f}{\partial \theta_i \partial \theta_i} \right) = - \sum_{j=1}^n \sum_{k=1}^{K_j} \frac{n_{jk} a_j^2}{p_{jk}} (P_{jk-1}^* Q_{jk-1}^* - P_{jk}^* Q_{jk}^*)^2.$$

Die Varianz von  $\Theta_i$  ist damit

$$V(\Theta_i) = \frac{1}{\sum_{j=1}^n \sum_{k=1}^{K_j} \frac{n_{jk} a_j^2}{p_{jk}} (P_{jk-1}^* Q_{jk-1}^* - P_{jk}^* Q_{jk}^*)^2}.$$

Die Iterationsgleichung für das Fisher-Scoring ist entsprechend:

$$(\theta_i)_{t+1} = (\theta_i)_t + \frac{\sum_{j=1}^n \sum_{k=1}^{K_j} x_{jk} a_j \frac{P_{jk-1}^* Q_{jk-1}^* - P_{jk}^* Q_{jk}^*}{p_{jk}}}{\sum_{j=1}^n \sum_{k=1}^{K_j} \frac{n_{jk} a_j^2}{p_{jk}} (P_{jk-1}^* Q_{jk-1}^* - P_{jk}^* Q_{jk}^*)^2}_t$$

Für jede Personengruppe mit demselben Antwortvektor ist eine Berechnung nötig.

### 3.2.3 Schätzen der Item- und der Einstellungsstärkeparameter

Für das Schätzen sowohl der Item- als auch der Einstellungsparameter gibt es wiederum verschiedene Verfahren. Auf Grund der oben gemachten Überlegungen (s. Seite 12) gilt auch bezüglich der GIRT, dass Itemparameter pro Item und Einstellungsstärken pro Person geschätzt werden können. Entsprechend kann man die Parameter mit Hilfe des JMLE-Birnbaum-Verfahrens in einem Iterationsprozess bestimmen. Man startet mit grob geschätzten Einstellungsstärken. Diese kann man für eine Person als die Summe der gewichteten Antworten auf die Items bestimmen. Die Antworten werden aufsteigend durchnummeriert und diese Nummerierung wird als Gewichtung verwendet. Die Einstellungsstärken werden standardisiert. Anschliessend werden die Itemparameter nach der oben dargestellten Methode geschätzt, dann die Einstellungsstärken für jede Person, etc., bis eine vorgegebene Präzision erreicht ist.

Neben der JMLE-Methode gibt es unter anderem die MMLE-Methode, die im R-Paket ltm (Rizopoulos, 2006) implementiert ist.

### 3.3 Diagnostik von Modellen für mehrwertige ordinalskalierte Items

#### 3.3.1 Überprüfung der Daten

Um zu kontrollieren, ob die Itemvariablen zusammenhängen und damit geeignet sind, dieselbe verborgene Eigenschaft zu messen, wird für jedes Paar von Itemvariablen eine Korrelation berechnet. Da die Daten ordinalskaliert sind, erfolgt dies mit Kendalls  $\tau_b$  oder mit Gamma. Für jeden Korrelationskoeffizienten kann zudem ein p-Wert berechnet werden, um zu überprüfen, ob allfällige Korrelationen gut oder schlecht durch Zufall erklärbar sind.

Oft wird für Daten der hier vorausgesetzten Art Cronbachs  $\alpha$  berechnet (Cronbach, 1951). Allerdings setzt diese Kennzahl metrisch skalierte Daten voraus - es werden unter anderem Varianzen berechnet, die auf quadrierten Differenzen ruhen. Distanzen spielen entsprechend eine Rolle. Die Verwendung von  $\alpha$  ist somit nicht über alle Zweifel erhaben. Als grobes, mit Vorsicht zu geniessendes Mass kann es aber wohl trotzdem zur Anwendung gelangen.

#### 3.3.2 Untersuchung auf Angemessenheit der Items

Im Prinzip können die Methoden auf Überprüfung der Güte der Anpassung für den dichotomen Fall auf jede der CIRCCs angewendet werden. Sie werden hier nicht wiederholt. Zudem ergibt sich die Möglichkeit die Angemessenheit eines Items bezüglich der Befragten zu überprüfen, indem man die Charakteristischen Item-Kategorie-Kurven (Item-Response-Category Characteristic Curve; IRCCC) für jede Kategorie eines Items graphisch darstellt: Es wird die Wahrscheinlichkeit, dass die  $k$ -te Kategorie gewählt wird, in Abhängigkeit von  $\theta$  betrachtet. Die IRCCC für die erste Kategorie ist dabei eine strikt monoton sinkende Kurve: die Wahrscheinlichkeit, dass sie gewählt wird, sinkt in Abhängigkeit vom Willen, eine hohe Kategorie zu wählen. Mittlere Kategorien weisen eine Glockenkurve auf, während die höchste Kategorie einer strikt monoton steigende Kurve entspricht (s. Abbildung 4). Es ist offenbar

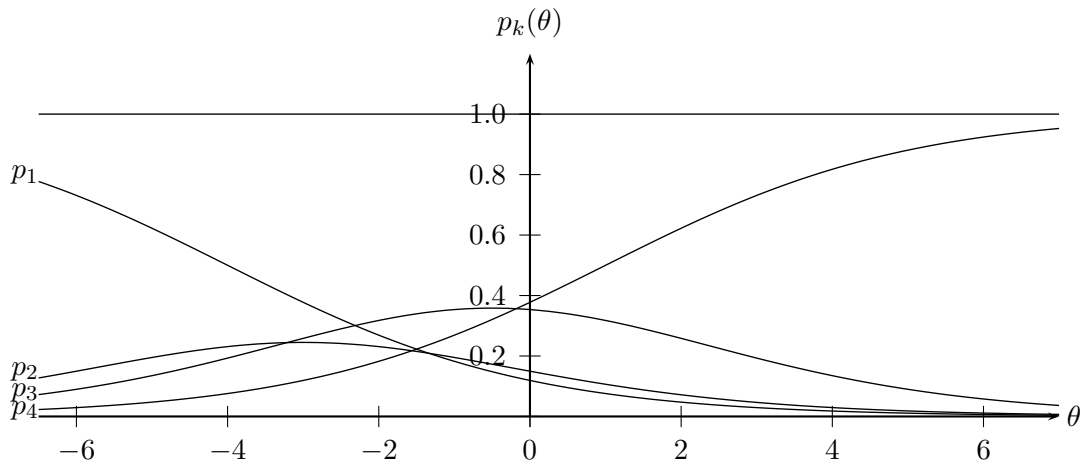


Abbildung 4: Logistische IRCCCs der CIRCCs von Abbildung 3 (Seite 20)

günstig, wenn die zu messenden Einstellungsparameter im Bereich liegen, wo *alle* Kurven möglichst stark von 0 oder von 1 verschieden sind. Umgekehrt sollten Items so konstruiert sein, dass ein solcher Bereich für sie existiert.

### 3.3.3 Informationsfunktionen

Die Varianz der Schätzstatistik  $\Theta_i$  ist wie oben gezeigt

$$V(\Theta_i) = \frac{1}{\sum_{j=1}^n \sum_{k=1}^{K_j} \frac{n_{jk} a_j^2}{p_{jk}} (P_{jk-1}^* Q_{jk-1}^* - P_{jk}^* Q_{jk}^*)^2}.$$

Je mehr Personen, desto kleiner die Varianz. Um den Beitrag der Items oder der Itemkategorien zur Testinformation zu betrachten, setzt man die Anzahl der Personen auf 1 und man erhält die Testinformationsfunktion

$$T_I(\theta) = \sum_{j=1}^n \sum_{k=1}^{K_j} \frac{a_j^2}{p_{jk}} (P_{jk-1}^* Q_{jk-1}^* - P_{jk}^* Q_{jk}^*)^2.$$

Diese ist zusammengesetzt aus den Informationsfunktionen  $I_{jk}$  der einzelnen Kategorien  $k$  der verschiedenen Items  $j$ , die wir Itemkategorie-Informationsfunktionen nennen:

$$I_{jk}(\theta) = \frac{a_j^2}{p_{jk}} (P_{jk-1}^* Q_{jk-1}^* - P_{jk}^* Q_{jk}^*)^2$$

Um eine Übersicht über bereichsweise Präzision der Itemschätzer und des Tests zu erhalten, zeichnet man für jedes Item in eine Graphik die Iteminformationsfunktionsfunktion

$$I_j(\theta) = \sum_{k=1}^{K_j} \frac{a_j^2}{p_{jk}} (P_{jk-1}^* Q_{jk-1}^* - P_{jk}^* Q_{jk}^*)^2$$

samt den Itemkategorie-Informationsfunktionen. Zusätzlich zeichnet man alle Iteminformationsfunktionen und die Testinformationsfunktion in eine weitere Abbildung. Die Graphiken entsprechen der von Abbildung 2 auf Seite 18.

## 4 Anwendung der GIRT auf Beispieldaten

Als nächstes geht es darum, als Beispiel die IRT für Variablen umzusetzen, die in der PISA-Studie (OECD, 2007) mit dieser Methode behandelt wurden. Es wird dabei mit fiktiven Daten gearbeitet, da die Daten der PISA-2006-Studie strengen Datenschutzbestimmungen unterliegen. Die selbst erzeugten Daten weisen dieselbe Struktur wie die PISA-2006-Daten auf (Anzahl Variablen, Anzahl Ausprägungen). Die IRT-Methode wurde in der PISA-Studie 2006 auf viele Variablensätze angewendet (z.B. Teilindices für die Berechnung des sozio-ökonomischen Status, s. OECD (2009, S. 346) und OECD (2007, S. 385), Interesse an Naturwissenschaften, Zukunftsorientierte Motivation in Naturwissenschaften, Selbstkonzept in Naturwissenschaften, Grad der Vertrautheit mit Umweltthemen, etc. (s. OECD (2007, S. 388 ff.)). Hier wird als Beispiel die Berechnung des *Indexes der Freude an Naturwissenschaften* gewählt. Er beruht „darauf, inwieweit die Schülerinnen und Schüler folgenden Aussagen zustimmen:

- a Im Allgemeinen macht es mir Spaß, mich mit naturwissenschaftlichen Themen zu befassen,
- b ich lese gerne etwas über Naturwissenschaften,
- c ich beschäftige mich gerne mit naturwissenschaftlichen Problemen,
- d ich eigne mir gerne neues Wissen in den Naturwissenschaften an und
- e ich bin interessiert, Neues in den Naturwissenschaften zu lernen.

Es wurde eine Vierpunkteskala mit folgenden Antwortkategorien verwendet: „stimme ganz zu“, „stimme eher zu“, „stimme eher nicht zu“ und „stimme gar nicht zu“ (OECD, 2007, S. 388).

Für die Berechnungen wird das R-Paket ltm eingesetzt (für R s. R Development Core Team (2006), für das ltm-Paket s. Rizopoulos (2006); im Artikel wird das Paket detailliert beschrieben). Das ltm-Paket arbeitet mit MMLE-Schätzmethoden. Gemäss einem von Rizopoulos im ltm-Paket gelieferten Beispiel (Environment - Attitude to the Environment), werden die Variablen als Faktoren eingegeben. Das Programm behandelt aber auch ein Dataframe mit geordneten Faktoren (ordered factor) korrekt. Die hier verwendeten Daten sind im R-Workspace „irt\_pisa.RData“ unter „pisa1“ zu finden (s. CD am Schluss dieser Arbeit). Für die Variablen wurden die folgenden Abkürzungen verwendet:

- „spass“ für „im Allgemeinen macht es mir Spaß, mich mit naturwissenschaftlichen Themen zu befassen“,
- „LeseGern“ für „ich lese gerne etwas über Naturwissenschaften“,
- „BeschaeftGerne“ für „ich beschäftige mich gerne mit naturwissenschaftlichen Problemen“,
- „NeuesWissen“ für „ich eigne mir gerne neues Wissen in den Naturwissenschaften an“ und
- „NeuesLernen“ für „ich bin interessiert, Neues in den Naturwissenschaften zu lernen“.

Für die identischen Ausprägungen der Itemvariablen wurden die Abkürzungen „SGarNichtZu“ für „stimme gar nicht zu“, „SEherNichtZu“ für „stimme eher nicht zu“, „SEherZu“ für „stimme eher zu“ und „SGanz“ für „stimme ganz zu“ gewählt.

Selber können solche Daten erzeugt werden mit dem im R-Workspace „irt\_pisa.RData“ gespeicherten Befehl „pisa()“. Der Befehl erlaubt die Variation der Anzahl Daten pro Kategorie der ersten Itemvariable. Die Häufigkeitsverteilung der anderen Item-Variablen werden nach dem jeweils selben Algorithmus mit Hilfe von Zufallszahlen berechnet. Der Befehl kann nützlich sein, um die Varianz der Schätzer in Abhängigkeit von der Stichprobengröße zu untersuchen.

Das ltm-Paket liefert einige beschreibende Statistiken:

---

```
> descript(pisa1)
```

Descriptive statistics for the 'pisa1' data-set

Sample:

5 items and 1450 sample units; 0 missing values

Proportions for each level of response:

	SGarNichtZu	SEherNichtZu	SEherZu	SGanzZu
spass	0.1724	0.3448	0.3793	0.1034
LeseGern	0.0566	0.2959	0.5759	0.0717
BeschaeftGerne	0.0841	0.3641	0.5228	0.029
NeuesWissen	0.1379	0.3766	0.4483	0.0372
NeuesLernen	0.1524	0.411	0.4131	0.0234

Frequencies of total scores:

	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Freq	4	22	72	84	96	134	152	155	157	222	181	83	62	20	5	1

Cronbach's alpha:

	value
All Items	0.8175
Excluding spass	0.7031
Excluding LeseGern	0.8026
Excluding BeschaeftGerne	0.7988
Excluding NeuesWissen	0.7939
Excluding NeuesLernen	0.7909

---

Wären die Itemvariablen metrisch skaliert, würde Cronbachs  $\alpha$  zeigen, dass die Items gut gewählt sind: es liegt in einem Bereich, der in der Literatur oft als günstig betrachtet wird. Vom ltm-Befehl „descript()“ werden zusätzlich die Pearson-Korrelationen pro Itemvariablenpaar berechnet. Angesichts des Skalenniveaus macht dies jedoch wenig Sinn, weshalb

die Berechnung von Gamma oder Kendalls  $\tau_b$  angemessener ist. Diese Kennzahlen sollten nur berechnet werden, wenn die gemeinsame Häufigkeitsverteilung des jeweiligen Variablenpaars auf der Haupt- oder der Nebendiagonale massiert ist. Im vorliegenden Fall sollten die Häufigkeiten allerdings auf der Hauptdiagonalen liegen, da eine positive Korrelation der Itemvariablen für die IRT-Berechnungen bestehen sollte. Beispielshalber wird eine entsprechende Kreuztabelle geliefert:

---

```
>table(pisa1$LeseGern,pisa1$NeuesWissen)
```

	SGarNichtZu	SEherNichtZu	SEherZu	SGanzZu
SGarNichtZu	30	36	16	0
SEherNichtZu	100	199	128	2
SEherZu	69	288	439	39
SGanzZu	1	23	67	13

---

Die Häufigkeiten sind um die Hauptdiagonale massiert. Solche Kreuztabellen sollten für alle möglichen Paare begutachtet werden. Bei den vorliegenden Daten ist die Konzentration auf der Hauptdiagonalen für alle Variablenpaare gegeben, wobei die Kreuztabellen hier nicht wiedergegeben werden.

Mit dem ltm-Befehl „`rcor.test(pisa1,method="kendall")`“ können alle Itemvariablen des Dataframes paarweise auf Korrelation geprüft werden. Man erhält das Resultat:

---

```
>rcor.test(pisa1,method="kendall")
```

	spass	LeseGern	BeschaeftGerne	NeuesWissen	NeuesLernen
spass	*****	0.536	0.551	0.557	0.56
LeseGern	<0.001	*****	0.317	0.325	0.347
BeschaeftGerne	<0.001	<0.001	*****	0.341	0.355
NeuesWissen	<0.001	<0.001	<0.001	*****	0.354
NeuesLernen	<0.001	<0.001	<0.001	<0.001	*****

upper diagonal part contains correlation coefficient estimates

lower diagonal part contains corresponding p-values

---

Es zeigt sich, dass alle Itemvariablen paarweise signifikant positiv korrelieren, wobei die Korrelation nicht überall hoch ist. Die kleinen p-Werte sind durch die grosse Stichprobe zu erklären, die selbst bei eher kleinen Abweichungen von 0 zu signifikanten Ergebnissen führen wird.

Als nächstes wird ein Modell berechnet. ltm offeriert dabei die Möglichkeit, ein Modell zu schätzen, das für alle Items denselben Trennschärfeparameter annimmt. Als Resultat erhält man:



---

```
> pisafit1=grm(pisa1,constrained=T)
> pisafit1
```

Call:

```
grm(data = pisa1, constrained = T)
```

Coefficients:

	Extrmt1	Extrmt2	Extrmt3	Dscrmn
spass	-1.289	0.063	1.715	2.042
LeseGern	-2.073	-0.505	1.951	2.042
BeschaeftGerne	-1.811	-0.179	2.541	2.042
NeuesWissen	-1.424	0.045	2.382	2.042
NeuesLernen	-1.353	0.205	2.668	2.042

Log.Lik: -6967.346

---

Unter „Dscrmn“ (für „discrimination“) wird der für alle Items identisch geschätzte Trennschärfeparameter angegeben. Unter „Extrmtx“ (für „extremity“) werden die Parameter für das Anspruchsniveau der x-ten CIRCC angegeben.

Weitere Angaben wie das Bayessche Informationskriterium (BIC), das Akaike's Informationskriterium (AIC) oder Informationen zum Schätzprozess (Präzision und Anzahl verwendete Punkte bei der Gauss-Hermitschen-Integrationsregel) erhält man mit dem Befehl „summary()“.

**Bemerkung 7.** *Das BIC ist definiert durch  $-2L(\mu, y) + k \ln(n)$  mit  $L(\mu, y) =$  Maximale Likelihood (= Log.Lik. im obigen Ausdruck);  $k =$  Anzahl geschätzter Parameter;  $n =$  Stichprobengröße. Das AIC wird definiert durch  $-2L(\mu, y) + 2k$ .*

Wir überprüfen die Anpassung der Daten ans Modell mit Hilfe der „Margins-Methode“:

---

```
> margins(pisafit1)
```

Call:

```
grm(data = pisa1, constrained = T)
```

Fit on the Two-Way Margins

	spass	LeseGern	BeschaeftGerne	NeuesWissen	NeuesLernen
spass	-	39.92	50.41	52.46	47.53
LeseGern		-	44.69	40	41.36
BeschaeftGerne			-	32.32	24.35
NeuesWissen				-	30.58
NeuesLernen					-

---

Diese Überprüfung wird noch mittels der „margins“-Methode mit dreifacher Kreuzung durchgeführt:

---

```
> margins(pisafit1,type="three")
Call:
grm(data = pisa1, constrained = T)
```

Fit on the Three-Way Margins

	Item i	Item j	Item k	$(O - E)^2 / E$	
1	1	2	3	207.1	
2	1	2	4	208.3	
3	1	2	5	180.34	
4	1	3	4	220.77	
5	1	3	5	188.07	
6	1	4	5	193.37	
7	2	3	4	109.55	
8	2	3	5	143.69	
9	2	4	5	111.44	
10	3	4	5	227.66	***

\*\*\* denotes triplets of items with lack-of-fit

---

In einem Fall liegt eine Abweichung der Daten vom Modell vor. Für die GIRT offeriert das ltm-Paket keinen Bootstrap-Test für eine Gesamtüberprüfung des Modells.

Als nächstes gilt es zu überprüfen, ob ein Modell ohne die Erzwingung identischer Trennschärfeparameter besser passt:

---

```
> pisafit2=grm(pisa1)
> pisafit2
Call:
grm(data = pisa1)
```

Coefficients:

	Extrmt1	Extrmt2	Extrmt3	Dscrmn
spass	-1.142	0.097	1.442	4.028
LeseGern	-2.173	-0.463	2.276	1.692
BeschaeftGerne	-1.97	-0.213	3.108	1.48
NeuesWissen	-1.55	0.157	2.811	1.617
NeuesLernen	-1.443	0.301	3.01	1.707

Log.Lik: -6829.196

---

Man sieht, dass die Trennschärfe des ersten Items stark von der der übrigen Items abweicht. Es ergibt sich mit der  $2 \times 2$ -Margins-Methode eine Abweichung, die nicht zum Modell passt. Im  $2 \times 2 \times 2$ -Fall ergibt sich hingegen keine derartige Abweichung. Das ltm-Paket bietet den Likelihood-Ratio-Test an, um Modelle zu vergleichen:

---

```
> anova(pisafit1,pisafit2)
```

Likelihood Ratio Table

	AIC	BIC	log.Lik	LRT	df	p.value
pisafit1	13966.69	14051.16	-6967.35			
pisafit2	13698.39	13803.98	-6829.2	276.3	4	<0.001

---

Das zweite Modell mit variablen Trennschärfeparametern passt besser. Das BIC und das AIC werden reduziert, obwohl es im zweiten Modell mehr Parameter hat.

Mit dem ltm-Paket kann man die Information des Tests in spezifischen Intervallen berechnen. Angegeben wird zusätzlich die Gesamtinformation und der Anteil der Information im Intervall an der Gesamtinformation:

---

```
> information(pisafit2,c(-4,4))
```

```
Call: grm(data = pisa1)
```

```
Total Information = 29.05
```

```
Information in (-4, 4) = 27.99 (96.33%)
```

```
Based on all the items
```

---

Diese Berechnung kann man auch für einzelne Items oder Itemgruppen durchführen:

---

```
> information(pisafit2,c(-4,4),items=1)
```

```
Call: grm(data = pisa1)
```

```
Total Information = 11.7
```

```
Information in (-4, 4) = 11.7 (100%)
```

```
Based on items 1
```

---

Man kann daraus berechnen, dass das Item 1 für

$$\frac{11.7}{29.05}100 = 40.275\%$$

der Information des Tests im spezifizierten Intervall verantwortlich ist. Will man die Information für Itemgruppen in einem Bereich berechnen lassen, muss z.B. für die Gruppe mit den ersten drei Items „items=c(1,2,3)“ eingegeben werden.

Zuletzt sind noch die IRCCCs und die Informationskurven zu analysieren: Die IRCCCs erhält man mit

```
>par(mfrow=c(3,2))
```

```
>plot(pisafit2,cx="bottomright",lwd=1,cex.main=1.5,cex.lab=1.3,cex=1.1, lty=c(1,2,3,4))
```

(s. Abbildung 5). Als Iteminformationskurven und Testinformationskurven erhält man die unten stehenden Abbildungen 6 und 7 mit

```
> plot(pisafit2,type="IIC",annot=F, lwd=3,cex.main=1.5,cex.lab=1.3)
```

```
> plot(pisafit2,type="IIC",items=0,annot=F, lwd=3,cex.main=1.5,cex.lab=1.3)
```

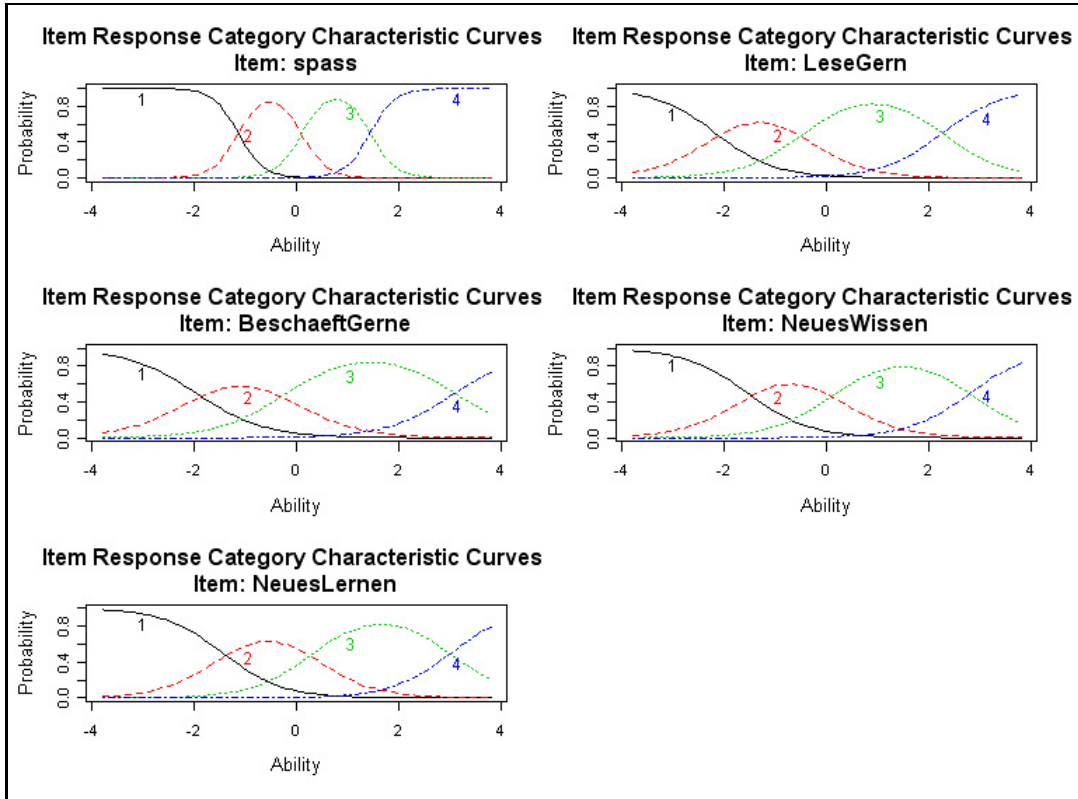


Abbildung 5: ltm-IRCCs des Beispiels

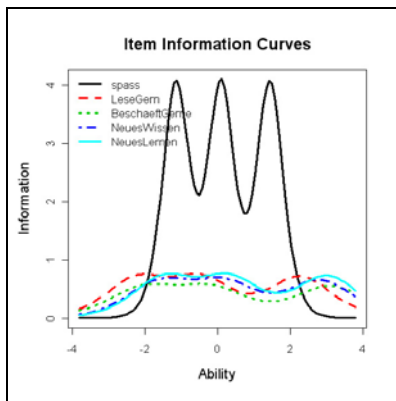


Abbildung 6: ltm-Iteminformationsfunktionen des Beispiels

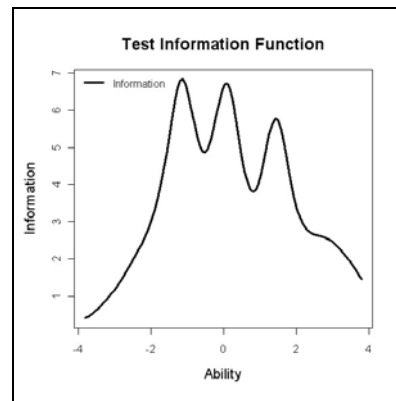


Abbildung 7: ltm-Testinformationsfunktion des Beispiels

Abbildung 8: ltm-IRCCs des Beispiels

Nachdem diese Analysen unter Umständen gezeigt haben, dass die Güte der Anpassung gut ist und die Items an angemessener Stelle liegen, um die Einstellungsstärken zu messen, gilt es noch die geschätzten Einstellungsstärken pro Antwortvektor zu berechnen. Es folgt die ltm-Ausgabe der ersten und die letzten fünf Zeilen einer Ausgabe von 213 Zeilen. Zu beachten ist, dass nicht alle möglichen Antwortvektoren vorkommen ( $4^5 = 1024$ ).

---

```
> factor.scores(pisafit2)
```

```
Call:
```

```
grm(data = pisa1)
```

```
Scoring Method: Empirical Bayes
```

```
Factor-Scores for observed response patterns:
```

	spass	LeseG	BeschaeftG	NeuesW	NeuesL	Obs	Exp	z1	se.z1
1	1	1	1	1	1	4	11.985	-2.225	0.541
2	1	1	1	1	2	4	3.717	-1.798	0.468
3	1	1	1	2	1	2	4.474	-1.822	0.471
4	1	1	1	2	2	4	2.526	-1.508	0.404
5	1	1	1	3	1	2	0.523	-1.67	0.472
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
209	4	4	4	3	3	4	3.103	2.053	0.483
210	4	4	4	3	4	2	0.988	2.48	0.524
211	4	4	4	4	2	1	0.06	2.085	0.521
212	4	4	4	4	3	2	1.429	2.456	0.523
213	4	4	4	4	4	1	0.896	2.924	0.535

---

Unter den Variablennamen werden die Ausprägungen der Variablen angegeben: pro Zeile stehen entsprechend die vorkommenden Antwortvektoren. Unter „Obs“ stehen die Anzahl Antwortvektoren des jeweiligen Typs, die im Datensatz vorkommen. „Exp“ steht für die Anzahl der erwarteten Antwortvektoren. Bei „z1“ stehen die geschätzten Einstellungsparameter und unter „se.z1“ deren geschätzter Standardfehler. Man sieht, dass die Standardabweichungen der Einstellungsparameter trotz der keineswegs kleinen Stichprobe ziemlich gross sind. In der Nähe von Null sind sie etwas kleiner, am Rand der Verteilung der z-Werte etwas grösser (maximaler Standardfehler 0.5409294, minimaler 0.3448174, mittlerer Standardfehler 0.406255). Die z-Werte weisen eine Spannweite von 5.148894 auf. Betrachten wir 0.95-Vertrauensintervalle, ergibt sich bei Normalverteilung der Schätzer ein Anteil von  $\frac{0.406255 \cdot 1.96}{5.148894} = 0.15465$  des Vertrauensintervalls an der Spannweite, womit ungefähr  $0.15465 \cdot 1450 = 224.24$  Personen innerhalb eines Vertrauensintervalls liegen und damit bezüglich Einstellungsstärke statistisch nicht unterscheidbar sind. Durch die Vergrößerung der Stichprobe auf 21500 Probanden - mit dem oben erwähnten R-Skript erstellt, s. R-Dataframe „pisa2006“ auf der beiliegenden CD - ergab sich keine Verringerung der Standardfehler (min: 0.3478278, max: 0.6336044, Mittel: 0.4097857). In der Praxis scheint sich die Konsistenz der Schätzer der Einstellungsstärken bei der Verwendung der Methoden des ltm-Pakets nicht zu realisieren.

## 5 Schlusswort

Ziel der Arbeit war es, einen Einstieg in die IRT-Modelle zu finden und diese selber in der Praxis anwenden zu können. Es erwies sich, dass das Gebiet der IRT-Modelle und der Parameter-Schätzmethoden recht gross ist und dass es kaum möglich ist, in Rahmen der für die Diplomarbeit vorgesehenen Frist das ganze Feld bis ins Detail durchzuarbeiten - eine Erkenntnis, die wohl für viele, noch so spezialisierte Bereiche der Statistik gilt. In der Arbeit wurden entsprechend Grundideen an den einfachsten Modellen nachvollzogen.

Es wurde das geschichtlich gesehen zuerst entwickelte Modell für dichotome Items dargestellt - allerdings nicht mit Hilfe der ursprünglich verwendeten Verteilungsfunktion der Standardnormalverteilung, sondern der später gebrauchten logistischen Funktion. Anschliessend wurden die zuerst entwickelten Parameter-Schätzmethoden eingeführt: das gemeinsame Schätzen von Item- und Fähigkeitsparametern (JMLE). Dieses gemeinsame Schätzen besteht allerdings in einem iterativen Verfahren (Birnbäum-Verfahren) und es weist den Nachteil auf, nicht zu konsistenten Schätzern zu führen. Heute wird für das Schätzen der Item-Parameter oft die Marginal Maximum Likelihood-Schätzmethode (MMLE) verwendet, welche eine spezifische Verteilung der Fähigkeitsparameter voraussetzt. Auf diesem - eventuell diskutablen - Hintergrund ist es möglich, die Item-Parameter konsistent zu schätzen. Diese Methode ist im verwendeten R-Paket ltm implementiert. Sie konnte in der Arbeit nur kurz skizziert werden, ohne auf technische Details einzugehen. Die im Paket verwendete Schätzmethode für die Fähigkeitsparameter konnte hier nicht behandelt werden.

Es wäre vermutlich nützlich, jeweils mit allen vorhandenen Methoden zu arbeiten. Liefern sie ähnliche Resultate, könnte dies als Gütezeichen verstanden werden. Liefern Sie unterschiedliche Resultate, wäre eine eingehendere Analyse der Daten erforderlich, um die Gründe für die Unterschiede zu verstehen.

Auch für die ordinalen, polytomen Items wurde das zuerst entwickelte Modell eingeführt: das von Fukimo Samejima (1969) ausgearbeitete Graded Item Response-Modell. Es wurde wiederum die Birnbäum-Methode dargelegt, da diese eine gute Einsicht in die Grundideen erlaubt.

Für den Fall der dichotomen Modelle wurde nachgewiesen, dass die IRT zu intervallskalierten Daten führt (s. Anhang A.1). Dies ist ein bemerkenswertes Ergebnis. Lange glaubte man, dass Intervallskalen eine empirische Operation des Zusammenfügens voraussetzen wie z.B. das Aneinanderfügen von Stäben auf einer Geraden oder das Aufeinanderschichten von Gewichten auf einer Wage (s. für eine ausführliche Darlegung der Voraussetzungen des metrischen Messens Stegmüller (1970, S. 19-109)). Diese Operation muss durch die Skala  $f$  mit Hilfe der Addition in den rationalen Zahlen gespiegelt werden können, d.h. es muss gelten  $f(x) + f(y) = f(x * y)$  - wobei  $x * y$  das durch die Operation des Zusammenfügens aus den Objekten  $x$  und  $y$  entstandene neue Objekt aus der Menge der zu messenden Objekte ist. Durch die IRT ist es möglich, mittels mehrerer Variablen, welche dieselbe Einstellung oder Fähigkeit messen, für Objektbereiche Intervallskalen zu verwenden, welche bisher dieser Art von Messung nicht zugänglich waren.

Es gilt es allerdings zu beachten, dass es sich bei den berechneten Werten um Schätzungen mit einem - relativ zur Spannweite der Daten gesehen - doch recht grossen Standardfehler handelt. Die dadurch entstehenden Messfehler sind anderer Natur als die beim klassischen Messen. Es ergibt sich in diesem Zusammenhang das Problem, wie sich dieser Umstand auf statistische Verfahren wie z.B. der linearen Regression auswirkt, die auf so erzeugte Daten angewendet werden.

# A Anhang

## A.1 Skalen

**Definition 1.** Sei eine Menge  $B$  von zu messenden Objekten gegeben und eine Zerlegung  $P$  von  $B$  mit den Elementen  $A_i$  (= Äquivalenzklassen). Dann wird eine Abbildung  $f : B \rightarrow \mathbb{R}$  (Mess)skala (bezüglich  $P$ ) genannt genau dann, wenn:  $f(x) = f(y)$  genau dann, wenn  $x, y \in A_i$ .

Erfüllt eine Skala nur die Bedingungen der Definition 1, wird sie „Nominalskala“ genannt. Nominalskalen sind bestimmt bis auf Transformationen  $h$ , die injektiv sind: Wenn  $f$  eine Nominalskala bezüglich  $P$  ist, so ist jede Funktion  $g : B \rightarrow \mathbb{R}$  mit

- $g = h \circ f$  und
- $h \circ f(x) = h \circ f(y)$  genau dann, wenn  $f(x) = f(y)$

eine Nominalskala bezüglich  $P$  ( $\circ$  ist die Verkettung von Funktionen).

**Definition 2.** Sei eine Menge  $B$  von zu messenden Objekten gegeben und eine Zerlegung  $P$  von  $B$  mit den Elementen  $A_i$ . Sei  $f$  eine Skala.  $f$  wird „Ordinalskala“ (bezüglich  $P$  und  $T$ ) genannt genau dann, wenn

- es eine asymmetrische, irreflexive und transitive Relation  $T$  auf  $B$  gibt, so dass  $(x, y) \in T$  oder  $(y, x) \in T$  genau dann wenn  $x$  und  $y$  zu verschiedenen Äquivalenzklassen  $A_i$  gehören, und
- wenn  $(x, y) \in T$  mit  $x \in A_i$  und  $y \in A_j$  dann ist  $(x', y') \in T$  für alle  $x' \in A_i$  und  $y' \in A_j$
- es gilt  $f(x) > f(y)$  genau dann, wenn  $(x, y) \in T$ .

Ordinalskalen sind bestimmt bis auf strikt monoton steigende Transformationen  $h$ . Wenn  $f$  eine Ordinalskala bezüglich  $P$  und  $T$  ist, so ist jede Funktion  $g : B \rightarrow \mathbb{R}$  mit

- $g = h \circ f$  und
- $h$  ist strikt monoton steigend

eine Ordinalskala bezüglich  $P$  und  $T$ . Zu beachten ist, dass durch solche Transformationen Abstandsverhältnisse zwischen Skalenwerten nicht erhalten bleiben. Gibt es z.B. drei Äquivalenzklassen  $A_i$  und sei  $f$  mit  $f(x_1) = 1$ ,  $f(x_2) = 2$  und  $f(x_3) = 3$ ,  $x_i \in A_i$ , eine Ordinalskala. Dann ist auch  $g$  mit  $g(x_1) = 1$ ,  $g(x_2) = 10$  und  $g(x_3) = 1000$  eine Ordinalskala, wobei man  $g$  durch die Verkettung  $h \circ f$  erhält und  $h : \{1, 2, 3\} \rightarrow \mathbb{R}$  eine strikt monoton steigende Funktion ist mit  $h(1) = 1$ ,  $h(2) = 10$ ,  $h(3) = 1000$ . Es gilt

$$\frac{f(x_2) - f(x_1)}{f(x_3) - f(x_2)} = 1 \neq \frac{g(x_2) - g(x_1)}{g(x_3) - g(x_2)} = \frac{9}{990}.$$

**Definition 3.** Sei eine Menge  $B$  von zu messenden Objekten gegeben und eine Zerlegung  $P$  von  $B$  mit den Elementen  $A_i$ . Sei  $f$  eine Ordinalskala bezüglich  $P$  und  $T$ . Dann ist  $f$  eine Intervallskala (bezüglich  $P$  und  $T$ ) genau dann, wenn  $f$  durch den Messvorgang bis auf affine Transformationen  $h(f(x)) = af(x) + b$  mit  $b \in \mathbb{R}$ ,  $a \in \mathbb{R}^+ \setminus \{0\}$  bestimmt ist.

Intervallskalen sind etwa dann gegeben, wenn es eine empirische Verknüpfungen  $*$  gibt, so dass  $f(x * y) = f(x) + f(y)$ . für  $x, y, x * y \in B$ . Beispiele für solche Verknüpfungen sind das Aneinanderfügen von Stäben auf einer Geraden oder das Aufeinanderlegen von Gewichten in einer Wagschale. Präziser gilt:

**Satz 4.** *Sei eine Menge  $B$  von zu messenden Objekten gegeben und eine Zerlegung  $P$  von  $B$  mit den Elementen  $A_i$ . Sei  $f$  eine Ordinalskala bezüglich  $P$  und  $T$ . Wenn es eine empirische Verknüpfungsrelation  $*$  gibt, so dass  $f(x * y) = f(x) + f(y)$  und wenn es für alle  $r \in \mathbb{Q}$  ein  $x \in B$  gibt, so dass  $f(x) = r$ , dann ist  $f$  durch den Messvorgang bis auf affine Transformationen  $h(f(x)) = af(x) + b$  mit  $b \in \mathbb{R}$ ,  $a \in \mathbb{R}^+ \setminus \{0\}$  bestimmt.*

*Beweis.* Seien zwei Ordinalskalen  $f$  und  $g$  bezüglich  $P$  und  $T$  gegeben, so dass jede die Voraussetzungen des Satzes erfüllt. Zu zeigen ist:

Wenn  $g(x) = h(f(x))$ , dann ist  $h(f(x)) = af(x) + b$  mit  $a > 0$  und  $a, b \in \mathbb{R}$ .

Gemäss Voraussetzung gibt es ein  $x_0 \in B$ , so dass  $f(x_0) = 0$  und ein  $y_0 \in B$ , so dass  $g(y_0) = 0$ .

OBdA sei  $(y_0, x_0) \in T$ . Es gibt damit ein  $z \in B$ , so dass

$$x_0 * z = y_0$$

und folglich

$$g(x_0 * z) = g(x_0) + g(z) = g(y_0) = 0.$$

Damit ist

$$b := g(x_0) = -g(z)$$

und, da  $g(x) = h(f(x))$  und  $f(x_0) = 0$ , gilt

$$\begin{aligned} g(x) + b &= g(x) + g(x_0) \\ &= g(x + x_0) \\ &= h(f(x + x_0)) \\ &= h(f(x) + f(x_0)) \\ &= h(f(x)) \end{aligned}$$

und entsprechend

$$g(x) = h(f(x)) - b.$$

Setzt man

$$h'(f(x)) := h(f(x)) - b$$

gilt mit

$$\begin{aligned} g(x + y) &= h(f(x + y)) - b \\ g(x) &= h(f(x)) - b \\ g(y) &= h(f(y)) - b \end{aligned}$$

$$h'(x + y) = h'(x) + h'(y).$$



Es handelt sich um die Funktionalgleichung von Cauchy, welche in  $\mathbb{Q}$  die Lösung

$$h'(r) = ar$$

aufweist. Damit ist  $ar = h(r) - b$  und  $h(r) = ar + b$ . Um  $g(x) > g(y)$  für  $(x, y) \in T$  zu bewahren, muss  $a > 0$  sein.  $\square$

Lange glaubte man, für Intervallskalen sei die Existenz von empirischen Operationen der beschriebenen Art nötig. Der Beweis des folgenden Satzes zeigt, dass dem nicht so ist und dass durch die IRT-Verfahren intervallskalierte Daten erzeugt werden. Der Beweis folgt Fischer und Molenaar (1995, S. 20 f.):

**Satz 5.** *Wenn es für jeden Parameter  $\beta$  mit  $\beta \in \mathbb{R}$  eine stetige, strikt monoton steigende Funktion  $f : \mathbb{R} \rightarrow ]0, 1[$  mit  $f(\theta - \beta)$  gibt, dann sind  $\beta$  und die Parameter  $\theta_i$  bis auf affine Transformationen  $h(\beta) = af(\beta) + b_1$  und  $h(\theta_1) = af(\theta_1) + b_2$  mit  $b_i \in \mathbb{R}$ ,  $a \in \mathbb{R}^+ \setminus \{0\}$  bestimmt.*

*Beweis.* Sei

$$f(\theta_i - \beta) = m(k(\theta_i) - l(\beta)),$$

so dass  $k$  und  $l$  strikt monoton steigende Funktionen sind. Dann gilt mit  $h(\theta_i - \beta) := m^{-1} \circ f(\theta_i - \beta)$

$$\begin{aligned} k(\theta_i) - l(\beta) &= m^{-1} \circ f(\theta_i - \beta) \\ &= h(\theta_i - \beta). \end{aligned} \tag{9}$$

Für  $\theta_i = 0$  gilt

$$\begin{aligned} k(0) - l(\beta) &= h(0 - \beta) \iff \\ l(\beta) &= -h(-\beta) + k(0). \end{aligned} \tag{10}$$

Für  $\beta = 0$  gilt

$$\begin{aligned} k(\theta_i) - l(0) &= h(\theta_i - 0) \iff \\ k(\theta_i) &= h(\theta_i) + l(0). \end{aligned} \tag{11}$$

Damit erhält man mit (9)

$$h(\theta_i) + l(0) + h(-\beta) - k(0) = h(\theta_i - \beta) \tag{12}$$

und für  $\theta_i = 0 = \beta$

$$\begin{aligned} h(0) + l(0) + h(0) - k(0) &= h(0) \iff \\ h(0) &= k(0) - l(0), \end{aligned}$$

womit aus (12) folgt

$$\begin{aligned} h(\theta_i) + l(0) + h(-\beta) - k(0) &= h(\theta_i - \beta) \iff \\ h(\theta_i) - h(0) + h(-\beta) &= h(\theta_i - \beta) \iff \\ h(\theta_i) - h(0) + h(-\beta) - h(0) &= h(\theta_i - \beta) - h(0). \end{aligned}$$

Mit

$$h^*(x) := h(x) - h(0)$$

und  $\beta' := -\beta$  ergibt sich daraus

$$h^*(\theta_i) + h^*(\beta') = h^*(\theta_i + \beta').$$

Dies ist die Funktionalgleichung von Cauchy, die in  $\mathbb{R}$  unter der Voraussetzung  $a > 0$  die Lösung  $h^*(x) = ax$  aufweist. Damit ergibt sich  $ax = h(x) - h(0)$  und  $h(x) = ax + h(0)$ . Aus (10) und (11) folgt

$$\begin{aligned} l(\beta) &= a\beta - h(0) + k(0) \\ k(\theta_i) &= a\theta_i + h(0) + l(0), \end{aligned}$$

denn mit

$$\begin{aligned} h(\beta') &= h(-\beta) \\ &= a(-\beta) + h(0) \end{aligned}$$

gilt

$$\begin{aligned} l(\beta) &= -h(-\beta) + k(0) \\ &= a\beta - h(0) + k(0). \end{aligned}$$

□

Für Intervallskalen gilt, dass Abstandsverhältnisse bei verschiedenen Skalierungen erhalten bleiben. Es gilt mit  $g(x) = af(x) + b$ :

$$\frac{f(x) - f(y)}{f(z) - f(r)} = \frac{af(x) + b - af(y) - b}{af(z) + b - af(r) - b} = \frac{g(x) - g(y)}{g(z) - g(r)}$$

## A.2 Beweise einiger Eigenschaften der logistischen ICC (Seite 4)

- $p$  ist strikt monoton steigend für  $\alpha > 0$ .

$$\begin{aligned} & \frac{d}{d\theta} \frac{\exp(\alpha(\theta - \beta))}{1 + \exp(\alpha(\theta - \beta))} \\ &= \frac{\alpha \exp(\alpha(\theta - \beta)) (1 + \exp(\alpha(\theta - \beta))) - \exp(\alpha(\theta - \beta)) \alpha \exp(\alpha(\theta - \beta))}{(1 + \exp(\alpha(\theta - \beta)))^2} \\ &= \frac{\alpha \exp(\alpha(\theta - \beta)) [1 + \exp(\alpha(\theta - \beta)) - \exp(\alpha(\theta - \beta))]}{(1 + \exp(\alpha(\theta - \beta)))^2} \\ &= \frac{\alpha \exp(\alpha(\theta - \beta))}{(1 + \exp(\alpha(\theta - \beta)))^2} \end{aligned}$$

Da

$$\frac{\exp(\alpha(\theta - \beta))}{(1 + \exp(\alpha(\theta - \beta)))^2} > 0,$$

hängt die Steigung nur vom Vorzeichen von  $\alpha$  ab.

- Steigung  $\alpha$  ist maximal in  $\theta$  mit  $p(\theta) = 0.5$  und zwar in  $\theta = \beta$ .  
Mit

$$c := c(\theta) := \exp(\alpha(\theta - \beta))$$

gilt

$$\frac{d}{d\theta}c = \alpha c$$

und damit

$$\begin{aligned} \frac{d}{d\theta} \left( \alpha \frac{\exp(\alpha(\theta - \beta))}{(1 + \exp(\alpha(\theta - \beta)))^2} \right) &= \frac{d}{d\theta} \left( \alpha \frac{c}{(1 + c)^2} \right) \\ &= \alpha \frac{\alpha c (1 + c)^2 - c 2(1 + c) \alpha c}{(1 + c)^4} \\ &= \alpha \frac{\alpha c (1 + c) [1 + c - 2c]}{(1 + c)^4} \\ &= \alpha^2 \frac{c(1 + c) [1 - c]}{(1 + c)^4} \\ &= \alpha^2 \frac{c(1 - c)}{(1 + c)^3} \end{aligned}$$

Mit Nullsetzen:

$$\begin{aligned} \alpha^2 \frac{c(1 - c)}{(1 + c)^3} = 0 &\iff \\ 1 - c = 0 \end{aligned}$$

und damit

$$\begin{aligned} 1 - \exp(\alpha(\theta - \beta)) = 0 &\iff \\ \exp(\alpha(\theta - \beta)) = 1 &\iff \\ \alpha(\theta - \beta) = 0 &\iff \\ \theta - \beta = 0 &\iff \\ \theta = \beta \end{aligned}$$

Eingesetzt ergibt sich

$$p(\beta) = \frac{\exp(\alpha(\beta - \beta))}{1 + \exp(\alpha(\beta - \beta))} = \frac{1}{2}$$

Damit liegt dort ein Extremum. Die zweite Ableitung ist

$$\begin{aligned}
\frac{d}{d\theta} \alpha^2 \frac{c(1-c)}{(1+c)^3} &= \alpha^2 \frac{[ac(1-c) + c(-ac)](1+c)^3 - c(1-c)3(1+c)^2 ac}{(1+c)^6} \\
&= \alpha^2 \frac{[ac(1-c) + c(-ac)](1+c) - c(1-c)3ac}{(1+c)^4} \\
&= \alpha^2 \frac{[ac - ac^2 - ac^2](1+c) - c(1-c)3ac}{(1+c)^4} \\
&= \alpha^3 c \frac{[1-2c](1+c) - c(1-c)3}{(1+c)^4} \\
&= \alpha^3 c \frac{1-2c^2 - c - 3c + 3c^2}{(1+c)^4} \\
&= \alpha^3 c \frac{1-4c+c^2}{(1+c)^4}
\end{aligned}$$

Mit  $c(0.5) = \exp(\alpha(0.5 - 0.5)) = 1$  gilt

$$\alpha^3 \frac{1-4+1}{(1+1)^4} = \alpha^3 \frac{-2}{2^4} < 0$$

für  $\alpha > 0$ .

- Im Punkt  $\beta$  hat die Kurve die Steigung  $\frac{\alpha}{4}$ .

$$p'(\beta) = \alpha \frac{\exp(\alpha(\beta - \beta))}{(1 + \exp(\alpha(\beta - \beta)))^2} = \alpha \frac{1}{(1+1)^2} = \frac{\alpha}{4}$$

- Die Log-Odds der logistischen ICC sind eine affine Funktion der Parameter  $\alpha$  und  $\beta$ .

$$\begin{aligned}
\ln \left( \frac{p(\theta)}{1-p(\theta)} \right) &= \ln \frac{\frac{\exp(\alpha(\theta-\beta))}{1+\exp(\alpha(\theta-\beta))}}{1 - \frac{\exp(\alpha(\theta-\beta))}{1+\exp(\alpha(\theta-\beta))}} \\
&= \ln \frac{\frac{\exp(\alpha(\theta-\beta))}{1+\exp(\alpha(\theta-\beta))}}{\frac{1+\exp(\alpha(\theta-\beta)) - \exp(\alpha(\theta-\beta))}{1+\exp(\alpha(\theta-\beta))}} \\
&= \ln \exp(\alpha(\theta - \beta)) \\
&= \alpha(\theta - \beta)
\end{aligned}$$

### A.3 Partielle Ableitungen der Funktion $\pi$ nach $a$ und nach $b$ (Seite 7)

Es gilt

$$\begin{aligned}
1 - \pi(a, b) &= 1 - \frac{\exp(a\theta_i + b)}{1 + \exp(a\theta_i + b)} \\
&= \frac{1 + \exp(a\theta_i + b) - \exp(a\theta_i + b)}{1 + \exp(a\theta_i + b)} \\
&= \frac{1}{1 + \exp(a\theta_i + b)}
\end{aligned}$$

und mit

$$\frac{\partial}{\partial b} \exp(a\theta_i + b) = \exp(a\theta_i + b) \frac{\partial}{\partial b} (a\theta_i + b) = \exp(a\theta_i + b)$$

$$\begin{aligned} \frac{\partial}{\partial b} \pi(a, b) &= \frac{\partial}{\partial b} \frac{\exp(a\theta_i + b)}{1 + \exp(a\theta_i + b)} \\ &= \frac{\exp(a\theta_i + b) (1 + \exp(a\theta_i + b)) - \exp(a\theta_i + b) \exp(a\theta_i + b)}{(1 + \exp(a\theta_i + b))^2} \\ &= \frac{\exp(a\theta_i + b) (1 + \exp(a\theta_i + b) - \exp(a\theta_i + b))}{(1 + \exp(a\theta_i + b))^2} \\ &= \frac{\exp(a\theta_i + b)}{(1 + \exp(a\theta_i + b))^2} \\ &= \frac{\exp(a\theta_i + b)}{1 + \exp(a\theta_i + b)} \frac{1}{1 + \exp(a\theta_i + b)} \\ &= \pi(a, b) (1 - \pi(a, b)) \end{aligned}$$

und mit

$$\frac{\partial}{\partial a} \exp(a\theta_i + b) = \exp(a\theta_i + b) \frac{\partial}{\partial a} (a\theta_i + b) = \theta_i \exp(a\theta_i + b)$$

$$\begin{aligned} \frac{\partial}{\partial a} \pi(a, b) &= \frac{\partial}{\partial a} \frac{\exp(a\theta_i + b)}{1 + \exp(a\theta_i + b)} \\ &= \frac{\theta_i \exp(a\theta_i + b) (1 + \exp(a\theta_i + b)) - \theta_i \exp(a\theta_i + b) \exp(a\theta_i + b)}{(1 + \exp(a\theta_i + b))^2} \\ &= \frac{\theta_i \exp(a\theta_i + b) (1 + \exp(a\theta_i + b) - \exp(a\theta_i + b))}{(1 + \exp(a\theta_i + b))^2} \\ &= \frac{\theta_i \exp(a\theta_i + b)}{(1 + \exp(a\theta_i + b))^2} \\ &= \frac{\theta_i \exp(a\theta_i + b)}{1 + \exp(a\theta_i + b)} \frac{1}{1 + \exp(a\theta_i + b)} \\ &= \theta_i \pi(a, b) (1 - \pi(a, b)). \end{aligned}$$

#### A.4 Herleitung der Kovarianzmatrix der Itemparameter (Seite 9)

Die Informationsmatrix der logistischen ICC, bei der in allen Zellen

$$E \left( \frac{\partial^2 \ln f}{\partial x \partial y} \right) = \frac{\partial^2 \ln f}{\partial x \partial y}$$

gilt, ist die Negation von

$$\begin{pmatrix} a & c \\ c & d \end{pmatrix} := \begin{pmatrix} \frac{\partial^2 \ln f}{\partial a \partial a} & \frac{\partial^2 \ln f}{\partial a \partial b} \\ \frac{\partial^2 \ln f}{\partial b \partial a} & \frac{\partial^2 \ln f}{\partial b \partial b} \end{pmatrix}$$

Die Inverse dieser Matrix kann wie folgt berechnet werden:

$$\begin{pmatrix} \frac{1}{a} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & c & 1 & 0 \\ c & d & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & \frac{1}{a}c & \frac{1}{a} & 0 \\ c & d & 0 & 1 \end{pmatrix}$$

$$\begin{aligned}
& \begin{pmatrix} 1 & 0 \\ -c & 1 \end{pmatrix} \begin{pmatrix} 1 & \frac{1}{a}c & \frac{1}{a} & 0 \\ c & d & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & \frac{1}{a}c & \frac{1}{a} & 0 \\ 0 & d - \frac{1}{a}c^2 & -\frac{1}{a}c & 1 \end{pmatrix} \\
& \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{d - \frac{1}{a}c^2} \end{pmatrix} \begin{pmatrix} 1 & \frac{1}{a}c & \frac{1}{a} & 0 \\ 0 & d - \frac{1}{a}c^2 & -\frac{1}{a}c & 1 \end{pmatrix} = \begin{pmatrix} 1 & \frac{1}{a}c & \frac{1}{a} & 0 \\ 0 & 1 & -\frac{1}{a} \frac{c}{d - \frac{1}{a}c^2} & \frac{1}{d - \frac{1}{a}c^2} \end{pmatrix} \\
& \begin{pmatrix} 1 & -\frac{c}{a} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & \frac{1}{a}c & \frac{1}{a} & 0 \\ 0 & 1 & -\frac{1}{a} \frac{c}{d - \frac{1}{a}c^2} & \frac{1}{d - \frac{1}{a}c^2} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \frac{1}{a} + \frac{1}{a^2} \frac{c^2}{d - \frac{1}{a}c^2} & -\frac{1}{a} \frac{c}{d - \frac{1}{a}c^2} \\ 0 & 1 & -\frac{1}{a} \frac{c}{d - \frac{1}{a}c^2} & \frac{1}{d - \frac{1}{a}c^2} \end{pmatrix} \\
& \begin{pmatrix} \frac{1}{a} + \frac{1}{a^2} \frac{c^2}{d - \frac{1}{a}c^2} & -\frac{1}{a} \frac{c}{d - \frac{1}{a}c^2} \\ -\frac{1}{a} \frac{c}{d - \frac{1}{a}c^2} & \frac{1}{d - \frac{1}{a}c^2} \end{pmatrix} = \begin{pmatrix} \frac{1}{a} + \frac{1}{a} \frac{c^2}{ad - c^2} & -\frac{c}{ad - c^2} \\ -\frac{c}{ad - c^2} & \frac{a}{ad - c^2} \end{pmatrix} \\
& = \begin{pmatrix} \frac{ad - c^2 + c^2}{a(ad - c^2)} & -\frac{c}{ad - c^2} \\ -\frac{c}{ad - c^2} & \frac{a}{ad - c^2} \end{pmatrix} = \begin{pmatrix} \frac{d}{ad - c^2} & -\frac{c}{ad - c^2} \\ -\frac{c}{ad - c^2} & \frac{a}{ad - c^2} \end{pmatrix}.
\end{aligned}$$

### A.5 Ableitung der logistischen ICC nach dem Fähigkeitsparameter $\theta_i$ von Seite 11

$$\begin{aligned}
\frac{dp_j(\theta_i)}{d\theta_i} &= \frac{d}{d\theta_i} \frac{\exp(\alpha_j(\theta_i - \beta_j))}{1 + \exp(\alpha_j(\theta_i - \beta_j))} \\
&= \frac{\alpha_j \exp(\alpha_j(\theta_i - \beta_j)) (1 + \exp(\alpha_j(\theta_i - \beta_j))) - \alpha_j \exp(\alpha_j(\theta_i - \beta_j)) \exp(\alpha_j(\theta_i - \beta_j))}{(1 + \exp(\alpha_j(\theta_i - \beta_j)))^2} \\
&= \frac{\alpha_j \exp(\alpha_j(\theta_i - \beta_j))}{1 + \exp(\alpha_j(\theta_i - \beta_j))} \frac{1}{1 + \exp(\alpha_j(\theta_i - \beta_j))} \\
&= \alpha_j p_j(\theta_i) (1 - p_j(\theta_i))
\end{aligned}$$

### A.6 Ableitungen der logistischen CIRCCs nach den Itemparametern (Seite 22)

Mit den Abkürzungen und der Gleichung 8 von Seite 22 gilt

$$\begin{aligned}
\ln f &= \sum_{i=1}^m \sum_{k=1}^K x_{ik} \ln p_{ik} \\
&= \sum_{i=1}^m \left( x_{i1} \ln(1 - P_{i1}^*) + \sum_{k=2}^{K-1} x_{ik} \ln(P_{ik-1}^* - P_{ik}^*) + x_{iK} \ln P_{iK}^* - 0 \right) \\
&= \sum_{i=1}^m \left( \sum_{k=1}^K x_{ik} \ln(P_{ik-1}^* - P_{ik}^*) \right).
\end{aligned}$$

Zudem gilt mit den Ableitungen von Anhang A.3, S. 42, für  $1 \leq k < K$ :

$$\frac{\partial}{\partial b_k} p_{ik+1} = \frac{\partial}{\partial b_k} (P_{ik}^* - P_{ik+1}^*) = \frac{\partial}{\partial b_k} P_{ik}^* = P_{ik}^* (1 - P_{ik}^*) = P_{ik}^* Q_{ik}^*$$

Für  $k = K - 1$  ist  $P_{ik+1}^* = P_{iK}^* = 0$ . Damit ist  $\frac{\partial}{\partial b_k} P_{iK}^* = 0 = P_{iK}^* (1 - P_{iK}^*)$ .

$$\frac{\partial}{\partial b_k} p_{ik} = \frac{\partial}{\partial b_k} (P_{ik-1}^* - P_{ik}^*) = -\frac{\partial}{\partial b_k} P_{ik}^* = -P_{ik}^* Q_{ik}^*$$

Für  $k = 1$  ist  $P_{ik-1}^* = P_{i0}^* = 1$ . Damit ist  $\frac{\partial}{\partial b_k} P_{i0}^* = 0 = P_{i0}^* (1 - P_{i0}^*)$ .

$$\begin{aligned}\frac{\partial}{\partial a} p_{ik+1} &= \frac{\partial}{\partial a} (P_{ik}^* - P_{ik+1}^*) = \theta_i P_{ik}^* (1 - P_{ik}^*) - \theta_i P_{ik+1}^* (1 - P_{ik+1}^*) \\ &= \theta_i P_{ik}^* Q_{ik}^* - \theta_i P_{ik+1}^* Q_{ik+1}^*\end{aligned}$$

Für  $k = K - 1$  ist  $P_{ik+1}^* = P_{iK}^* = 0$ . Damit ist  $\frac{\partial}{\partial a} P_{iK}^* = 0 = \theta_i P_{iK}^* (1 - P_{iK}^*)$ .

$$\frac{\partial}{\partial a} p_{ik} = \frac{\partial}{\partial a} (P_{ik-1}^* - P_{ik}^*) = \theta_i P_{ik-1}^* Q_{ik-1}^* - \theta_i P_{ik}^* Q_{ik}^*.$$

Für  $k = 1$  ist  $P_{ik-1}^* = P_{i0}^* = 1$ . Damit ist  $\frac{\partial}{\partial a} P_{i0}^* = 0 = \theta_i P_{i0}^* (1 - P_{i0}^*)$ .

$$\begin{aligned}\frac{\partial}{\partial b_k} P_{ik}^* Q_{ik}^* &= \frac{\partial}{\partial b_k} P_{ik}^* (1 - P_{ik}^*) \\ &= P_{ik}^* (1 - P_{ik}^*) (1 - P_{ik}^*) - P_{ik}^* P_{ik}^* (1 - P_{ik}^*) \\ &= P_{ik}^* (1 - P_{ik}^*) [1 - P_{ik}^* - P_{ik}^*] \\ &= P_{ik}^* (1 - P_{ik}^*) (1 - 2P_{ik}^*) \\ &= P_{ik}^* Q_{ik}^* (1 - 2P_{ik}^*)\end{aligned}$$

$$\begin{aligned}\frac{\partial}{\partial a} P_{ik}^* Q_{ik}^* &= \frac{\partial}{\partial a} P_{ik}^* (1 - P_{ik}^*) \\ &= \theta_i P_{ik}^* (1 - P_{ik}^*) (1 - P_{ik}^*) - \theta_i P_{ik}^* P_{ik}^* (1 - P_{ik}^*) \\ &= \theta_i P_{ik}^* Q_{ik}^* (1 - 2P_{ik}^*).\end{aligned}$$

Damit erhält man:

1)

$$\begin{aligned}\frac{\partial}{\partial b_k} \ln f &= \sum_{i=1}^m \left( \frac{\partial}{\partial b_k} (x_{ik} \ln (P_{ik-1}^* - P_{ik}^*) + x_{ik+1} \ln (P_{ik}^* - P_{ik+1}^*)) \right) \\ &= \sum_{i=1}^m \left( x_{ik} \frac{\frac{\partial}{\partial b_k} (P_{ik-1}^* - P_{ik}^*)}{P_{ik-1}^* - P_{ik}^*} + x_{ik+1} \frac{\frac{\partial}{\partial b_k} (P_{ik}^* - P_{ik+1}^*)}{P_{ik}^* - P_{ik+1}^*} \right) \\ &= \sum_{i=1}^m \left( -\frac{x_{ik} P_{ik}^* Q_{ik}^*}{P_{ik-1}^* - P_{ik}^*} + \frac{x_{ik+1} P_{ik}^* Q_{ik}^*}{P_{ik}^* - P_{ik+1}^*} \right) \\ &= \sum_{i=1}^m \left( -\frac{x_{ik} P_{ik}^* Q_{ik}^*}{p_{ik}} + \frac{x_{ik+1} P_{ik}^* Q_{ik}^*}{p_{ik+1}} \right)\end{aligned}$$

2)

$$\begin{aligned}
\frac{\partial}{\partial a} \ln f &= \sum_{i=1}^m \sum_{k=1}^K \frac{\partial}{\partial a} \ln p_{ik}^{x_{ik}} \\
&= \sum_{i=1}^m \sum_{k=1}^K x_{ik} \frac{\frac{\partial}{\partial a} p_{ik}}{p_{ik}} \\
&= \sum_{i=1}^m \theta_i \sum_{k=1}^K \frac{x_{ik} (P_{ik-1}^* Q_{ik-1}^* - P_{ik}^* Q_{ik}^*)}{p_{ik}} \\
&= \sum_{i=1}^m \theta_i \sum_{k=1}^K \frac{x_{ik} (P_{ik-1}^* Q_{ik-1}^* - P_{ik}^* Q_{ik}^*)}{P_{ik-1}^* - P_{ik}^*}
\end{aligned}$$

3)

$$\begin{aligned}
&\frac{\partial^2}{\partial b_k \partial b_k} \ln f \\
&= \frac{\partial}{\partial b_k} \sum_{i=1}^m \left( -\frac{x_{ik} P_{ik}^* Q_{ik}^*}{p_{ik}} + \frac{x_{ik+1} P_{ik}^* Q_{ik}^*}{p_{ik+1}} \right) \\
&= \sum_{i=1}^m \left( \left( -x_{ik} \frac{P_{ik}^* Q_{ik}^* (1 - 2P_{ik}^*) p_{ik} - P_{ik}^* Q_{ik}^* P_{ik}^* Q_{ik}^*}{p_{ik}^2} \right) + \right. \\
&\quad \left. + \left( x_{ik+1} \frac{P_{ik}^* Q_{ik}^* (1 - 2P_{ik}^*) p_{ik+1} - P_{ik}^* Q_{ik}^* P_{ik}^* Q_{ik}^*}{p_{ik+1}^2} \right) \right)
\end{aligned}$$

4)

$$\begin{aligned}
&\frac{\partial^2}{\partial b_k \partial b_{k-1}} \ln f \\
&= \frac{\partial}{\partial b_{k-1}} \sum_{i=1}^m \left( -\frac{x_{ik} P_{ik}^* Q_{ik}^*}{p_{ik}} + \frac{x_{ik+1} P_{ik}^* Q_{ik}^*}{p_{ik+1}} \right) \\
&= \frac{\partial}{\partial b_{k-1}} \sum_{i=1}^m \left( -\frac{x_{ik} P_{ik}^* Q_{ik}^*}{p_{ik}} \right) \\
&= \sum_{i=1}^m \left( -x_{ik} \frac{0 \cdot p_{ik} - P_{ik}^* Q_{ik}^* P_{ik-1}^* Q_{ik-1}^*}{p_{ik}^2} \right) \\
&= \sum_{i=1}^m \left( x_{ik} \frac{P_{ik}^* Q_{ik}^* P_{ik-1}^* Q_{ik-1}^*}{p_{ik}^2} \right)
\end{aligned}$$



5)

$$\begin{aligned}
& \frac{\partial^2}{\partial b_k \partial b_{k+1}} \ln f \\
&= \frac{\partial}{\partial b_{k+1}} \sum_{i=1}^m \left( -\frac{x_{ik} P_{ik}^* Q_{ik}^*}{p_{ik}} + \frac{x_{ik+1} P_{ik}^* Q_{ik}^*}{p_{ik+1}} \right) \\
&= \frac{\partial}{\partial b_{k+1}} \sum_{i=1}^m \left( \frac{x_{ik+1} P_{ik}^* Q_{ik}^*}{p_{ik+1}} \right) \\
&= \sum_{i=1}^m \left( x_{ik+1} \frac{0 \cdot p_{ik+1} - P_{ik}^* Q_{ik}^* (-1) P_{ik+1}^* Q_{ik+1}^*}{p_{ik+1}^2} \right) \\
&= \sum_{i=1}^m \left( x_{ik+1} \frac{P_{ik}^* Q_{ik}^* P_{ik+1}^* Q_{ik+1}^*}{p_{ik}^2} \right)
\end{aligned}$$

6) Mit

$$\frac{\partial}{\partial a} \ln f = \sum_{i=1}^m \sum_{k=1}^K x_{ik} \frac{\frac{\partial}{\partial a} p_{ik}}{p_{ik}}$$

(s. unter Punkt 2) gilt

$$\begin{aligned}
\frac{\partial^2}{\partial a \partial a} \ln f &= \frac{\partial}{\partial a} \left( \sum_{i=1}^m \sum_{k=1}^K x_{ik} \frac{\frac{\partial}{\partial a} p_{ik}}{p_{ik}} \right) \\
&= \sum_{i=1}^m \sum_{k=1}^K x_{ik} \frac{\partial}{\partial a} \frac{\frac{\partial}{\partial a} p_{ik}}{p_{ik}} \\
&= \sum_{i=1}^m \sum_{k=1}^K x_{ik} \frac{p_{ik} \frac{\partial^2}{\partial a \partial a} p_{ik} - \frac{\partial}{\partial a} p_{ik} \frac{\partial}{\partial a} p_{ik}}{p_{ik}^2} \\
&= \sum_{i=1}^m \sum_{k=1}^K x_{ik} \left( \frac{\frac{\partial^2}{\partial a \partial a} p_{ik}}{p_{ik}} - \frac{\left( \frac{\partial}{\partial a} p_{ik} \right)^2}{p_{ik}^2} \right).
\end{aligned}$$

Da

$$\frac{\partial}{\partial a} p_{ik} = \theta_i (P_{i,k-1}^* Q_{ik-1}^* - P_{ik}^* Q_{ik}^*)$$

und

$$\begin{aligned}
\frac{\partial^2}{\partial a \partial a} p_{ik} &= \frac{\partial}{\partial a} \theta_i (P_{i,k-1}^* Q_{ik-1}^* - P_{ik}^* Q_{ik}^*) \\
&= \theta_i^2 [P_{i,k-1}^* Q_{ik-1}^* (1 - 2P_{i,k-1}^*) - P_{ik}^* Q_{ik}^* (1 - 2P_{ik}^*)]
\end{aligned}$$

gilt

$$\begin{aligned}
\frac{\partial^2}{\partial a \partial a} \ln f &= \sum_{i=1}^m \sum_{k=1}^K x_{ik} \left\{ \frac{\theta_i^2 [P_{i,k-1}^* Q_{ik-1}^* (1 - 2P_{i,k-1}^*) - P_{ik}^* Q_{ik}^* (1 - 2P_{ik}^*)]}{p_{ik}} - \right. \\
&\quad \left. \frac{\theta_i^2}{p_{ik}^2} (P_{i,k-1}^* Q_{ik-1}^* - P_{ik}^* Q_{ik}^*)^2 \right\}
\end{aligned}$$

7)

$$\begin{aligned}\frac{\partial^2 \ln f}{\partial b_k \partial a} &= \frac{\partial}{\partial a} \sum_{i=1}^m \left( -\frac{x_{ik} P_{ik}^* Q_{ik}^*}{p_{ik}} + \frac{x_{ik+1} P_{ik}^* Q_{ik}^*}{p_{ik+1}} \right) \\ &= \sum_{i=1}^m \left( -x_{ik} \frac{p_{ik} \frac{\partial}{\partial a} (P_{ik}^* Q_{ik}^*) - P_{ik}^* Q_{ik}^* \frac{\partial}{\partial a} p_{ik}}{p_{ik}^2} + x_{ik+1} \frac{p_{ik+1} \frac{\partial}{\partial a} (P_{ik}^* Q_{ik}^*) - P_{ik}^* Q_{ik}^* \frac{\partial}{\partial a} p_{ik+1}}{p_{ik+1}^2} \right)\end{aligned}$$

mit

$$\begin{aligned}\frac{\partial}{\partial a} (P_{ik}^* Q_{ik}^*) &= \theta_i P_{ik}^* Q_{ik}^* (1 - 2P_{ik}^*) \\ \frac{\partial}{\partial a} p_{ik} &= \frac{\partial}{\partial a} (P_{ik-1}^* - P_{ik}^*) = \theta_i (P_{ik-1}^* Q_{ik-1}^* - P_{ik}^* Q_{ik}^*).\end{aligned}$$

## A.7 Berechnung der Erwartungswerte der zweiten Ableitungen der logistischen CIRCCs (S. 23)

1) Es gilt

$$\begin{aligned}\frac{\partial^2 \ln f}{\partial b_k \partial b_k} &= \sum_{i=1}^m \left( \left( -x_{ik} \frac{P_{ik}^* Q_{ik}^* (1 - 2P_{ik}^*) p_{ik} - P_{ik}^* Q_{ik}^* P_{ik}^* Q_{ik}^*}{p_{ik}^2} \right) + \right. \\ &\quad \left. + \left( x_{i,k+1} \frac{P_{ik}^* Q_{ik}^* (1 - 2P_{ik}^*) p_{ik+1} - P_{ik}^* Q_{ik}^* P_{ik}^* Q_{ik}^*}{p_{ik+1}^2} \right) \right) \\ &= \sum_{i=1}^m \left( \left( -x_{ik} \frac{P_{ik}^* Q_{ik}^* (1 - 2P_{ik}^*)}{p_{ik}} - x_{ik} \frac{(P_{ik}^* Q_{ik}^*)^2}{p_{ik}^2} \right) + \right. \\ &\quad \left. + \left( x_{i,k+1} \frac{P_{ik}^* Q_{ik}^* (1 - 2P_{ik}^*)}{p_{ik+1}} - x_{i,k+1} \frac{(P_{ik}^* Q_{ik}^*)^2}{p_{ik+1}^2} \right) \right) \\ &= \sum_{i=1}^m \left( P_{ik}^* Q_{ik}^* (1 - 2P_{ik}^*) \left( \frac{-x_{ik}}{p_{ik}} + \frac{x_{i,k+1}}{p_{ik+1}} \right) + \right. \\ &\quad \left. + (P_{ik}^* Q_{ik}^*)^2 \left( \frac{-x_{ik}}{p_{ik}^2} - \frac{x_{i,k+1}}{p_{ik+1}^2} \right) \right)\end{aligned}$$

Mit

$$E \left( \frac{-X_{ik}}{p_{ik}} + \frac{X_{i,k+1}}{p_{ik+1}} \right) = -\frac{p_{ik} n_i}{p_{ik}} + \frac{p_{ik+1} n_i}{p_{ik+1}} = 0$$

und

$$\begin{aligned}E \left( (P_{ik}^* Q_{ik}^*)^2 \left( \frac{-X_{ik}}{p_{ik}^2} - \frac{X_{i,k+1}}{p_{ik+1}^2} \right) \right) &= (P_{ik}^* Q_{ik}^*)^2 \left( -\frac{n_i p_{ik}}{p_{ik}^2} - \frac{n_i p_{ik+1}}{p_{ik+1}^2} \right) \\ &= -n_i (P_{ik}^* Q_{ik}^*)^2 \left( \frac{1}{p_{ik}} + \frac{1}{p_{ik+1}} \right)\end{aligned}$$

ergibt sich

$$E \left( \frac{\partial^2 \ln f}{\partial b_k \partial b_k} \right) = -\sum_{i=1}^m n_i (P_{ik}^* Q_{ik}^*)^2 \left( \frac{1}{p_{ik}} + \frac{1}{p_{ik+1}} \right).$$

$$2) E \left( \frac{\partial^2}{\partial b_k \partial b_{k-1}} \ln f \right) =$$

$$\begin{aligned} & E \left( \sum_{i=1}^m \left( X_{ik} \frac{P_{ik}^* Q_{ik}^* P_{ik-1}^* Q_{ik-1}^*}{p_{ik}^2} \right) \right) \\ &= \sum_{i=1}^m p_{ik} n_i \frac{P_{ik}^* Q_{ik}^* P_{ik-1}^* Q_{ik-1}^*}{p_{ik}^2} \\ &= \sum_{i=1}^m n_i \frac{P_{ik}^* Q_{ik}^* P_{ik-1}^* Q_{ik-1}^*}{p_{ik}} \end{aligned}$$

$$3) E \left( \frac{\partial^2}{\partial b_k \partial b_{k+1}} \ln f \right) \text{ wird analog zu 2 berechnet.}$$

$$4) \text{ Es gilt } E \left( \frac{\partial^2}{\partial a \partial a} \ln f \right)$$

$$\begin{aligned} &= E \left( \sum_{i=1}^m \sum_{k=1}^K x_{ik} \left\{ \frac{\theta_i^2 \left[ P_{i,k-1}^* Q_{ik-1}^* (1 - 2P_{i,k-1}^*) - P_{ik}^* Q_{ik}^* (1 - 2P_{ik}^*) \right]}{p_{ik}} - \right. \right. \\ &\quad \left. \left. \frac{\theta_i^2}{p_{ik}^2} (P_{i,k-1}^* Q_{ik-1}^* - P_{ik}^* Q_{ik}^*)^2 \right\} \right) \\ &= \sum_{i=1}^m \sum_{k=1}^K \theta_i^2 p_{ik} n_i \left\{ \frac{\left[ P_{i,k-1}^* Q_{ik-1}^* (1 - 2P_{i,k-1}^*) - P_{ik}^* Q_{ik}^* (1 - 2P_{ik}^*) \right]}{p_{ik}} - \right. \\ &\quad \left. \frac{1}{p_{ik}^2} (P_{i,k-1}^* Q_{ik-1}^* - P_{ik}^* Q_{ik}^*)^2 \right\} \\ &= \sum_{i=1}^m \theta_i^2 n_i \sum_{k=1}^K \left\{ \left[ P_{i,k-1}^* Q_{ik-1}^* (1 - 2P_{i,k-1}^*) - P_{ik}^* Q_{ik}^* (1 - 2P_{ik}^*) \right] - \right. \\ &\quad \left. \frac{1}{p_{ik}} (P_{i,k-1}^* Q_{ik-1}^* - P_{ik}^* Q_{ik}^*)^2 \right\} \end{aligned}$$

und da

$$\sum_{k=1}^K \frac{\partial^2}{\partial a \partial a} p_{ik} = \sum_{k=1}^K \left[ P_{i,k-1}^* Q_{ik-1}^* (1 - 2P_{i,k-1}^*) - P_{ik}^* Q_{ik}^* (1 - 2P_{ik}^*) \right] = 0,$$

denn  $P_{i0}^* Q_{i0}^* = 0 = P_{iK}^* Q_{iK}^*$  und z.B.  $0 - c + c - d + d - f + f - 0 = 0$ , gilt

$$E \left( \frac{\partial^2 \ln f}{\partial b_k \partial a} \right) = \sum_{i=1}^m \theta_i^2 n_i \sum_{k=1}^K \frac{1}{p_{ik}} (P_{i,k-1}^* Q_{ik-1}^* - P_{ik}^* Q_{ik}^*)^2$$

5)

$$\begin{aligned}
& \sum_{i=1}^m \left( -x_{ik} \frac{p_{ik} \frac{\partial}{\partial a} (P_{ik}^* Q_{ik}^*) - P_{ik}^* Q_{ik}^* \frac{\partial}{\partial a} p_{ik}}{p_{ik}^2} + x_{ik+1} \frac{p_{ik+1} \frac{\partial}{\partial a} (P_{ik}^* Q_{ik}^*) - P_{ik}^* Q_{ik}^* \frac{\partial}{\partial a} p_{ik+1}}{p_{ik+1}^2} \right) \\
&= \sum_{i=1}^m \left( -x_{ik} \frac{p_{ik} \theta_i P_{ik}^* Q_{ik}^* (1 - 2P_{ik}^*) - \theta_i P_{ik}^* Q_{ik}^* (P_{ik-1}^* Q_{ik-1}^* - P_{ik}^* Q_{ik}^*)}{p_{ik}^2} + \right. \\
&\quad \left. x_{ik+1} \frac{p_{ik+1} \theta_i P_{ik}^* Q_{ik}^* (1 - 2P_{ik}^*) + \theta_i P_{ik}^* Q_{ik}^* (P_{ik}^* Q_{ik}^* - P_{ik+1}^* Q_{ik+1}^*)}{p_{ik+1}^2} \right) \\
&= \sum_{i=1}^m \left( -x_{ik} \frac{\theta_i P_{ik}^* Q_{ik}^* (1 - 2P_{ik}^*)}{p_{ik}} + x_{ik+1} \frac{\theta_i P_{ik}^* Q_{ik}^* (1 - 2P_{ik}^*)}{p_{ik+1}} + \right. \\
&\quad \left. x_{ik} \frac{\theta_i P_{ik}^* Q_{ik}^* (P_{ik-1}^* Q_{ik-1}^* - P_{ik}^* Q_{ik}^*)}{p_{ik}^2} + x_{ik+1} \frac{\theta_i P_{ik}^* Q_{ik}^* (P_{ik}^* Q_{ik}^* - P_{ik+1}^* Q_{ik+1}^*)}{p_{ik+1}^2} \right)
\end{aligned}$$

und da

$$E \left( \sum_{i=1}^m \left( -x_{ik} \frac{\theta_i P_{ik}^* Q_{ik}^* (1 - 2P_{ik}^*)}{p_{ik}} + x_{ik+1} \frac{\theta_i P_{ik}^* Q_{ik}^* (1 - 2P_{ik}^*)}{p_{ik+1}} \right) \right) = 0$$

(s. Argumentation unter 4)) gilt

$$\begin{aligned}
& E \left( \sum_{i=1}^m \left( x_{ik} \frac{\theta_i P_{ik}^* Q_{ik}^* (P_{ik-1}^* Q_{ik-1}^* - P_{ik}^* Q_{ik}^*)}{p_{ik}^2} + x_{ik+1} \frac{\theta_i P_{ik}^* Q_{ik}^* (P_{ik}^* Q_{ik}^* - P_{ik+1}^* Q_{ik+1}^*)}{p_{ik+1}^2} \right) \right) \\
&= \sum_{i=1}^m P_{ik}^* Q_{ik}^* n_i \left( p_{ik} \frac{\theta_i (P_{ik-1}^* Q_{ik-1}^* - P_{ik}^* Q_{ik}^*)}{p_{ik}^2} + p_{ik+1} \frac{\theta_i (P_{ik}^* Q_{ik}^* - P_{ik+1}^* Q_{ik+1}^*)}{p_{ik+1}^2} \right) \\
&= \sum_{i=1}^m P_{ik}^* Q_{ik}^* n_i \theta_i \left( \frac{(P_{ik-1}^* Q_{ik-1}^* - P_{ik}^* Q_{ik}^*)}{p_{ik}} + \frac{(P_{ik}^* Q_{ik}^* - P_{ik+1}^* Q_{ik+1}^*)}{p_{ik+1}} \right) = E \left( \frac{\partial^2 \ln f}{\partial b_k \partial a} \right)
\end{aligned}$$

## A.8 Ableitungen der logistischen CIRCCs nach den Fähigkeiten - Berechnung des Erwartungswertes der zweiten Ableitung (Seite 24)

Mit A.5, S. 44, und

$$\frac{\partial p_{ik}}{\partial \theta_i} = \frac{\partial}{\partial \theta_i} (P_{jk-1}^* - P_{jk}^*) = a_j P_{jk-1}^* Q_{jk-1}^* - a_j P_{jk}^* Q_{jk}^*$$

gilt

$$\begin{aligned}
\frac{\partial \ln f}{\partial \theta_i} &= \sum_{j=1}^n \sum_{k=1}^{K_j} \frac{x_{jk}}{p_{jk}} \frac{\partial}{\partial \theta_i} p_{ik} \\
&= \sum_{j=1}^n \sum_{k=1}^{K_j} x_{jk} \frac{a_j P_{jk-1}^* Q_{jk-1}^* - a_j P_{jk}^* Q_{jk}^*}{p_{jk}} \\
&= \sum_{j=1}^n \sum_{k=1}^{K_j} a_j x_{jk} \frac{P_{jk-1}^* Q_{jk-1}^* - P_{jk}^* Q_{jk}^*}{p_{jk}}.
\end{aligned}$$

Die zweite Ableitung ergibt analog zu 6) auf Seite 47

$$\begin{aligned}
\frac{\partial^2 \ln f}{\partial \theta_i \partial \theta_i} &= \frac{\partial}{\partial \theta_i} \sum_{j=1}^n \sum_{k=1}^{K_j} x_{jk} a_j^2 \frac{P_{jk-1}^* Q_{jk-1}^* - P_{jk}^* Q_{jk}^*}{p_{ik}} \\
&= \sum_{j=1}^n \sum_{k=1}^{K_j} x_{jk} \left\{ \frac{a_j^2 \left[ P_{i,k-1}^* Q_{ik-1}^* (1 - 2P_{i,k-1}^*) - P_{ik}^* Q_{ik}^* (1 - 2P_{ik}^*) \right]}{p_{ik}} \right. \\
&\quad \left. - \frac{a_j^2}{p_{jk}} (P_{jk-1}^* Q_{jk-1}^* - P_{jk}^* Q_{jk}^*)^2 \right\}.
\end{aligned}$$

Für den Erwartungswert der zweiten Ableitung erhält man analog zu 4) auf Seite 49:

$$\begin{aligned}
E \left( \frac{\partial^2 \ln f}{\partial \theta_i \partial \theta_i} \right) &= - \sum_{j=1}^n \sum_{k=1}^{K_j} \frac{n_{jk} p_{jk} a_j^2}{p_{jk}^2} (P_{jk-1}^* Q_{jk-1}^* - P_{jk}^* Q_{jk}^*)^2 \\
&= - \sum_{j=1}^n \sum_{k=1}^{K_j} \frac{n_{jk} a_j^2}{p_{jk}} (P_{jk-1}^* Q_{jk-1}^* - P_{jk}^* Q_{jk}^*)^2.
\end{aligned}$$



## Index

- Äquivalenzklasse, 19, 37
- Abstandsverhältnis, 37, 40
- affine Transformation, 37
- Akaike's Informationskriterium, 31
- Anpassung
  - Güte der, 16, 26
- Anspruchsniveau, 19
- asymmetrisch, 19, 37
  
- Bayessches Informationskriterium, 31
- bedingte Dichtefunktion, 14
- bernoulli-verteilt, 10, 12
- Bindungen, 16
- Bootstrap, 17
  
- Cauchy
  - Funktionalgleichung, 40
- Chance, 4
- Charakteristische Item-Kurve, 2
- Charakteristischen Item-Kategorie-Kurven, 26
- Chi-Quadrat-Koeffizient, 16
- CIRCC, 19
- Cronbachs Alpha, 26, 29
  
- Devianz-Residuen, 16
  
- Einstellung, 1
- Einstellungsstärke, 2, 19, 20, 24, 25
- empirische Logits, 6
- empirische Verknüpfung, 38
- Erwartungswert
  - Komponenten der Hesse-Matrix, 9, 25
- exakter Test von Fisher, 16
  
- Fähigkeit, 2
- Fähigkeits-Parameter, 13
- Faktor
  - im R-Paket, 28
- Fisher-Scoring, 6, 10, 22, 25
- Funktionalgleichung von Cauchy, 40
  
- Güte der Anpassung, 16, 26
- Gamma, 16, 26, 30
- GIR-Modelle, 19
- Graded Item Response-Modell, 19
  
- Hauptdiagonale, 30
- Hesse-Matrix, 7, 9, 22
  
- ICC, 3
  - logistische, 4
- Information, 17, 27, 33
- Informationsfunktion, 17, 27, 33
- Informationskriterium
  - Akaike's, 31
  - Bayessche, 31
- Informationsmatrix, 9
- Intervallskala, 4, 37
  - IRT, 39
- IRCCC, 26
- irreflexiv, 19, 37
- IRT, 1
  - Intervallskala, 39
  - Normalverteilung, 4
- Item, 2
- Item-Parameter, 10, 13, 15, 17, 20, 24
- Iteminformationsfunktion, 18, 27, 33
- Itemkategorie-Informationsfunktion, 27
- Itemkategorien, 19
- Iteration, 6, 7, 24, 25
  
- JMLE, 12
- Joint Maximum Likelihood-Schätzmethode, 12
  
- Kendalls Tau, 16, 26, 30
- Konsistenz
  - bei JMLE, 13
  - bei MMLE, 14
- Korrelation, 26, 29
- Kovarianzmatrix, 9
- Kumulative Charakteristische Item-Kategorie-Kurven, 19
  
- Likelihood-Ratio-Test, 16, 32
- lineare Regression, 6
- LL-Funktion, 6, 10, 12, 15, 22, 24
- Log-Likelihood-Funktion, 6
- Log-Odds, 4, 16
- logistische
  - ICC, 4
  - Regression, 4, 5, 16, 17

- Regressionsfunktion, 16
- Verteilungsfunktion, 3
- logistisches Modell, 20
- Logit, 4, 15
  - empirisches, 6, 16
- Marginal Maximum Likelihood-Schätzmethode, 14
- Maximum-Likelihood-Methode, 5
- Messskala, 37
- ML-Methoden, 5, 20
- MMLE-Schätzmethoden, 28
- Multinomialverteilung, 21
- Newton-Raphson-Verfahren, 6, 22
  - Fähigkeitsparameter, 10
  - JMLE, 12
- Nominalskala, 37
- Normalverteilung, 4
- Odds, 4
- Odds-Ratio, 16
- Ordinalskala, 19, 37
- Orthogonalität, 13
- parametrischer Bootstrap-Pearson-Test, 17
- Pearson-Korrelation, 16
- Pearson-Residuen, 16
- PISA-Studie, 2, 28
- polytome Variable, 19
- R-Paket ltm, 28
- Randverteilung, 14
- Regression
  - lineare, 6
  - logistische, 4
- Regressionsfunktion
  - logistische, 16
- Relation, 19, 37
- residuale Devianz, 16
- Scaled Deviance, 16
- Schwierigkeit, 2, 3
- Schwierigkeitsparameter
  - Information, 18
  - logistische ICC, 4
- Skala, 37
  - Intervall-, 37
- Nominal-, 37
- Ordinal-, 37
- Standardisierung, 13
- Start-Fähigkeitsvektor, 13, 25
- Taylor-Polynom 2. Grades, 6, 11
- Test
  - exakter von Fisher, 16
  - Odds-Ratio, 16
  - parametrischer Bootstrap-Pearson, 17
  - residuale Devianz, 16
- Testinformationsfunktion, 17, 27, 33
- Transformation
  - affine, 37
  - injektive, 37
  - strikt monoton steigende, 37
- transitiv, 19, 37
- Trennschärfe, 2, 3
  - bei polytomen Variablen, 19
  - Definition, 3
  - logistische ICC, 4
- Trennschärfeparameter, 1, 9, 20, 30
  - Information, 18
- Unabhängigkeit
  - Antwortverhalten, 10
  - lokale, 14
- Varianz, 1, 26
  - Fähigkeitsparameter, 11, 25, 27
  - Information, 17
  - Itemparameter, 9
  - Stichprobengrösse, 29
- Verkettung
  - von Funktionen, 37
- Verteilung
  - gemeinsame, 10, 12
- Verteilungsfunktion, 3
- Zerlegung, 37



## Glossar

$:=$	rechter Term definiert linken Term	3
$=:$	linker Term definiert rechten Term	22
$A \times B$	kartesisches Produkt von $A$ und $B$	5
$A \setminus B$	Differenzmenge von $A$ und $B$	38
$A^n$	$n$ -faches kartesisches Produkt der Menge $A$	6
$C_k$	Kumulative charakteristische Item-Kategorie-Kurve	20
$E(X)$	Erwartungswert der Zufallsvariable $X$	9
$I(A)$	Information der Statistik $A$	18
$ICC_j$	Charakteristische Kurve des Items $j$	2
$P(X = x)$	Wahrscheinlichkeit, dass die Zufallsvariable $X$ den Wert $x$ annimmt	10
$P_{ik}^*$	logistische Itemkategoriefunktion	22
$Q_{ik}^*$	$1 - P_{ik}^*$	22
$T_I$	Testinformationsfunktion	18
$V(A)$	Varianz der Statistik $A$	9
$\iff$	genau dann, wenn	39
$\Phi$	Verteilungsfunktion der Standardnormalverteilung	4
$\alpha$	Trennschärfeparameter der logistischen ICC	4
$\beta$	Schwierigkeitsparameter eines Items der logistischen ICC	4
$\circ$	Verkettung von Funktionen	37
$\exp(x)$	$e^x$	5
$\frac{\partial f(x)}{\partial x}$	partielle Ableitung der Funktion $f$ nach $x$	7
$\hat{a}$	geschätzter Wert des Parameters $a$	6
$\ln(x)$	Logarithmus zur Basis $e$ von $x$	4
$\mathbb{N}$	Menge der natürlichen Zahlen mit der Null	5
$\mathbb{N}^*$	Menge der natürlichen Zahlen ohne die Null	5
$\mathbb{N}_k$	$\{m \mid m \in \mathbb{N} \text{ und } m \leq k\}$	5
$\mathbb{N}_k^*$	$\{m \mid m \in \mathbb{N}^* \text{ und } m \leq k\}$	5
$\mathbb{R}$	Menge der reellen Zahlen	2
$\mathbb{Q}$	Menge der rationalen Zahlen	38
$\mathbb{R}^+$	Menge der positiven reellen Zahlen mit der Null	38
$\mathbf{A}^{-1}$	Matrix $\mathbf{A}$ invertiert	7
$\mathbf{A}^t$	Matrix $\mathbf{A}$ transponiert	6
$\mathbf{H}$	Hesse-Matrix	7
$\mathbf{a}^t$	Vektor $\mathbf{a}$ transponiert	6
$\mathbf{q}$	fetter Buchstabe für Vektor	7

$\prod_{i=1}^k A_i$	$(\dots((A_1 \times A_2) \times A_3) \times \dots) \times A_k$	5
$\pi$	logistische Regressionsfunktion	5
$\prod_{i=1}^m x_i$	Produkt der $x_i$ von 1 bis $m$	6
$\sum_{i=1}^m x_i$	Summe der $x_i$ von 1 bis $m$	6
$\theta$	Variable über Fähigkeiten	2
$]a, b[$	offenes Intervall	3
$]a, b]$	linksoffenes Intervall; rechtsabgeschlossenes Intervall	6
$[a, b]$	abgeschlossenes Intervall	2
$e$	Eulersche Zahl	5
$f : A \rightarrow B$	Abbildung von $A$ nach $B$	2
$g'(x)$	Ableitung von $g$ nach $x$	7
$p$	logistische ICC	4
<b>IRT</b>	Item Response Theory	1
<b>CIRCC</b>	Kumulative charakteristische Itemkurven	20
<b>GIRT</b>	Graded IRT-Modell	19
<b>IRCCC</b>	Charakteristische Item-Response-Kategoriekurve	21
<b>JMLE</b>	Joint Maximum Likelihood-Schätzmethode	12
<b>LL</b>	Log-Likelihood	6
<b>ML</b>	Maximum-Likelihood	5
<b>MMLE</b>	Marginal Maximum Likelihood-Schätzmethode	4

## Literatur

- Agresti, A. (1990). *Categorical Data Analysis*. New York: John Wiley.
- Agresti, A. (2002). *Categorical Data Analysis* (2. Aufl.). New York: John Wiley.
- Baker, F. B. (2001). *The Basics of Item Response Theory*. University of Maryland: ERIC Clearinghouse on Assessment and Evaluation.
- Baker, F. B. & Kim, S.-H. (2004). *Item Response Theory: Parameter Estimation Techniques*. Basel: Dekker.
- Birnbaum, A. (1968). Some Latent Trait Models and their Use in Inferring an Examinee's Ability. In F. Lord & M. Novick (Hrsg.), *Statistical Theories of Mental Test Scores* (S. 397-479). Reading: Addison-Wesley.
- Bock, R. & Aitkin, M. (1981). Marginal Maximum Likelihood Estimation of Item Parameters: Application of an EM Algorithm. *Psychometrika*, 46, 443-459.
- Bock, R. & Lieberman, M. (1970). Fitting a Response Model for  $n$  Dichotomously Scored Items. *Psychometrika*, 35, 179-197.
- Cronbach, L. J. (1951). Coefficient Alpha and the Internal Structure of Tests. *Psychometrika*, 16 (3), 297-334.
- Fischer, G. H. & Molenaar, I. W. (Hrsg.). (1995). *Rasch Models: Foundations, Recent Developments and Applications*. Berlin: Springer.
- Hosmer, D. W. & Lemeshow, S. (2000). *Applied Logistic Regression*. New York: Wiley.
- Kale, B. (1962). On the Solution of Likelihood Equations by Iteration Process: The Multiparametric Case. *Biometrika*, 49, 479-486.
- Maxwell, A. (1959). Maximum Likelihood Estimates of Item Parameters using the Logistic Function. *Psychometrika*, 24, 221-227.
- Müller, P. H. (1991). *Lexikon der Stochastik*. Berlin: Akademie Verlag.
- Neyman, J. & Scott, E. (1948). Consistent Estimates Based on Partially Consistent Observations. *Econometrica*, 16, 1-32.
- OECD. (2007). *PISA 2006 - Schulleistungen im internationalen Vergleich Naturwissenschaftliche Kompetenzen für die Welt von morgen*. Paris: OECD Publishing.
- OECD. (2009). *PISA 2006 - Technical Report*. Paris: OECD Publishing.
- Perline, R., Wright, B. D. & Wainer, H. (1979). The Rasch Model as Additive Conjoint Measurement. *Applied Psychological Measurement*, 3 (2), 237-255.
- R Development Core Team. (2006). *R: A Language and Environment for Statistical Computing*. Wien: R Foundation for Statistical Computing; URL <http://www.R-projekt.org/>.
- Rasch, G. (1980). *Probabilistic Models for some Intelligence and Attainment Tests*. Chicago: The University of Chicago Press. (Erweiterte Auflage der Ausgabe des Danish Institute for Educational Research (1960) in Kopenhagen)
- Rizopoulos, D. (2006, November). ltm: An R Package for Latent Variable Modeling and Item Response Theory Analysis. *Journal of Statistical Software*, 17 (5).
- Samejima, F. (1969). Estimation of Latent Ability using a Response Pattern of Graded Scores. *Psychometrika Monograph Supplement*, 17.
- Stegmüller, W. (1970). *Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie: Erfahrung, Festsetzung, Hypothese und Einfachheit in der wissenschaftlichen Begriffs- und Theorienbildung* (Bd. 2 - Theorie und Erfahrung). Berlin: Springer. (Studienausgabe Teil A)
- van der Linden, W. J. & Hambleton, R. K. (Hrsg.). (1997). *Handbook of Modern Item Response Theory*. New York: Springer.