

LUDWIG-MAXIMILLIAN-UNIVERSITÄT MÜNCHEN
INSTITUT FÜR STATISTIK

Seminararbeit

Rasch-Modelle und Verallgemeinerung

Seminar: Ausgewählte Aspekte der Wirtschafts- und
Sozialstatistik

Seminarleiter: Prof. Dr. Thomas Augustin

Seminarbetreuer: Eva Endres

Autor: Nataliia Semenenko

28. Juli 2015

Abstract

Im Rahmen des Seminars sind ausgewählte Aspekte der Wirtschafts- und Sozialstatistik zu besprechen. Insbesondere sind die Themengebiete zu bearbeiten, die in unserer Zeit relevant sind, für die Existenz und Entwicklung der Gesellschaft. Mithilfe der statistischen Methoden sind die Probleme z.B im Bildungssystem, Wirtschaft sowie Volkswirtschaft zu untersuchen und wissenschaftlich zu lösen. So sind für dieses Seminar vor allem die folgende Themen zu bearbeiten: Anonymisierungsverfahren, Faktorenanalyse, Ausgewählte Studien (u.a. PISA-Studie), Rasch-Modelle und Verallgemeinerung, Stichprobenplan des Zensus und Gefühlte Inflation.

In dieser Arbeit wird das Rasch-Modell und deren Verallgemeinerungen dargestellt, die für die Analysen in PISA-Studie sowie in der Psychologie verwendet werden.

Erstmal wird es kleine Einleitung zum Thema *Item-Response-Theorie* sowie Wichtigkeit der Testkonstruktion geben. Dann wird die Grundform des Rasch-Modells, Modellannahmen und -Eigenschaften sowie die Schätzung des Parameters dargestellt. Zunächst werden einige Verallgemeinerungen wie z.B. das linear-logistische-Testmodell, Birnbaum-Modelle, das Partial-Credit-Modell kennengelernt. Schließlich anhand der simulierten Daten werden der graphische Modelltest, Likelihood-Quotiententest und Wald-Test durchgeführt und deren Ergebnisse verglichen.

Notation

Grundmodell

n - Anzahl Personen

m - Anzahl Aufgaben

$u \in \{0, 1\}$ - beobachtete Variable Aufgabe gelöst (1 = ja, 0 = nein)

θ - Personen-Parameter, z.B. Fähigkeit einer Person

β - Aufgaben-Parameter, z.B. Schwierigkeit einer Aufgabe

B - Nullpunkt auf gemeinsamen Skala für θ und β (wird festgelegt)

A - die Einheit der gemeinsamen Skala für θ und β (wird festgelegt)

$i = 1, \dots, n$ - Laufindex für Personen, die an einem Test teilnehmen

$j = 1, \dots, m$ - Laufindex für Aufgaben im Test

$u_{i,j} \in \{0, 1\}$ - beobachtete Variable Aufgabe gelöst (1 = ja, 0 = nein),
Eintrag für die i-te Person und j-te Aufgabe

$U_{i,j} \in \{0, 1\}$ - Zufallsvariable Aufgabe gelöst (1 = ja, 0 = nein), Eintrag
für die i-te Person und j-te Aufgabe

$P(U_{i,j} = u_{i,j})$ - Wahrscheinlichkeit, dass die i-te Person bei j-ter Aufgabe
genau Ergebnis $u_{i,j}$ erzielt

$r_i = \sum_{j=1}^m u_{i,j}$ - Zeilenrandsumme für i-te Person

$s_j = \sum_{i=1}^n u_{i,j}$ - Spaltenrandsumme für j-te Aufgabe

$L_{u_i}(\theta_i, \beta)$ - Likelihood für eine Person über alle Aufgaben

$L_u(\theta, \beta)$ - Gemeinsame Likelihood

$h(u|r, \beta)$ - Bedingte Likelihood

$L_u(\beta)$ - Marginale Likelihood

$f(\theta)$ - Randdichte für Personenparameter

Das linear-logistische Testmodell

$l = 1, \dots, L$ - Laufindex für nötige Teilkompetenzen für eine Aufgabe

$\omega_{j,l} \in \{0, 1, 2, \dots\}$ - Gewicht, Teilkompetenz in der Aufgabe enthalten
(0 = nein, 1 = einmal, 2 = zweimal,...)

η_l - Schwierigkeit der Teilkompetenz

Birnbaum-Modelle

δ_j - Diskriminationsparameter, Steigung j-ter Aufgabe

γ_j - Regulierungs-Parameter für Lösungswahrscheinlichkeiten

Das Partial-Credit-Modell

C - Anzahl Antwortkategorien

c - bestimmte c-te Antwortkategorie

$k = 0, \dots, C$ - Laufindex für die Antwortkategorien einer Aufgabe

$\beta_{j,c}$ - Parameter für c-te Antwortkategorie der j-ten Aufgabe

$\tau_{j,k}$ - Schwellenwert

$u \in \{0, \dots, C\}$ - beobachtete Variable Aufgabe gelöst (0 = nein,
1 = 1. Teilschritt, 2 = 1. und 2. Teilschritte,...) im Partial-Credit-
Modell

$u_{i,j} \in \{0, \dots, C_j\}$ - beobachtete Variable Aufgabe gelöst (0 = nein,
1 = 1. Teilschritt, 2 = 1. und 2. Teilschritte,...) Eintrag für die i-te
Person und j-te Aufgabe

$U_{i,j} \in \{0, \dots, C_j\}$ - unbekannte Variable Aufgabe gelöst (0 = nein,
1 = 1. Teilschritt, 2 = 1. und 2. Teilschritte,...) Eintrag für die i-te
Person und j-te Aufgabe

$k = 1 \dots K$ - Laufindex für den Gruppen im LQ-Test

Inhaltsverzeichnis

1	Einleitung	1
2	Grundform des Rasch-Modells	2
2.1	Datendarstellung	2
2.2	Modellgleichung	3
2.3	Aufgaben- und Personencharaktereistische Kurven	4
2.4	Modellannahmen und Modelleigenschaften	7
2.5	Parameterschätzung	10
3	Verallgemeinerungen des Rasch-Modells	13
3.1	Das linear-logistische Testmodell	13
3.2	Birnbaum-Modelle	15
3.3	Das Partial-Credit-Modell	18
3.4	Weitere Verallgemeinerungen	20
4	Simulation des Rasch-Modells	21
5	Modellgeltungstests	23
5.1	Der Graphische Modelltest	23
5.2	Likelihood-Quotienten-Test	25
5.3	Wald-Test	26
6	Zusammenfassung	28
A	Parameterschätzung	29
A.1	Herleitung Formel 2.5	29
A.2	Herleitung Formel 2.6	30
A.3	Herleitung Formel 2.7	30
A.4	Herleitung Formel 2.8	33

Kapitel 1

Einleitung

Das Rasch-Modell ist ein Modell, das von dem dänischen Statistiker Georg Rasch entwickelt wurde (1960). Das ist sogenanntes Grundmodell aller Modelle der *Item-Response-Theorie* (IRT), die sich als eine eigene Teildisziplin der psychologischen Testtheorie entwickelt hat. Das beschreibt eine Reaktion (*Response*) einer oder mehreren Person/-en auf eine Aufgabe oder Frage (*Item*). (Vgl. [Kubinger, Rasch und Yanagida \(2011\)](#), S.555)

Es handelt sich dabei um eine latente Eigenschaft. Um diese Eigenschaft zu erfassen, werden den Personen die Aufgaben oder Fragen gestellt. Anhand gelösten (bzw. nicht) Aufgaben oder beantworteten (bzw. nicht) Fragen wird ein Aufschluss über interessierende Eigenschaft gegeben. Am Ende des Testens bekommt die Person eine Schätzung über die nicht beobachtete Fähigkeit.

In der IRT ist es wichtig zu verstehen, wie ein Test konstruiert werden soll und welche Aufgaben verwendet werden müssen, um objektive Schätzungen zu bekommen. Um die Verzerrungen zu vermeiden, müssen die Testaufgaben gründlich auf die bestimmte Eigenschaften überprüft und bei Bedarf mithilfe des Rasch-Modells aussortiert (bzw. ersetzt) werden. Z.B. im Test für Mathematikkompetenz bei Textaufgaben schneiden sich nicht Muttersprachler schlechter als Kommilitonen, obwohl sie genau so gut oder besser in der Mathematik sind. Es kann auch passieren, dass die bestimmten Gruppen von Probanden einzelne Aufgaben bevorzugen (bzw. benachteiligen), was auch zu den nicht objektiven Schätzungen führt. Nur psychologische Tests, die eine strenge Überprüfung bestanden haben, sind objektiv und erlauben faire Vergleiche.

Das Rasch-Modell wird nicht nur zur Konstruktion von neuen Tests verwendet (der berühmteste Einsatz ist **PISA-Studie** der OECD), sondern auch, um die Tests nachträglich auf die Geltung des Rasch-Modells zu überprüfen. (Vgl. [Strobl \(2010\)](#), S. 1-3)

Kapitel 2

Grundform des Rasch-Modells

2.1 Datendarstellung

Ein Test soll eine nicht direkt beobachtete Eigenschaft (bzw. Fähigkeit) messen. Den in einem Test teilnehmenden Personen werden unterschiedliche Aufgaben gegeben oder verschiedene Fragen gestellt. Für jeder Teilnehmer wird es notiert, ob er jede Aufgabe richtig (bzw. falsch) beantwortet hat oder jeder Frage zugestimmt (bzw. nicht) hat. Für die richtig gelöste Aufgabe (bzw. korrekt zugestimmte Frage) bekommt die Person einen Eintrag von 1 und für falsch gelöste Aufgabe (bzw. unrichtig zugestimmte Frage) erhält die Person eine 0. Aus den Antworten der Personen ergibt sich die Datenmatrix, die aus Nullen und Einser besteht. In die Zeilen sind die Personen und in die Spalten die Aufgaben (bzw. Fragen) eingetragen.

Person	Aufgabe						
	1	2	3	4	5	6	7
1	0	1	0	1	0	0	0
2	0	1	1	1	0	1	0
3	1	1	0	0	0	1	1
4	0	0	0	0	1	1	0
5	0	0	0	0	0	1	0
6	1	1	1	1	1	1	0
7	1	0	0	0	1	0	0

Tabelle 2.1: Rasch-Modell. Datenmatrix Bsp. für 7 Personen und 7 Aufgaben

Die Tabelle 2.1 stellt die Antwortmatrix für 7 Personen und 7 Aufgaben dar. Daraus kann man z. B. ablesen, dass die erste Person die Aufgaben 2 und 4 richtig gelöst hat und die Aufgabe 5 von den vierten, sechsten und

siebten Personen richtig gelöst wurde. Die Allgemeine Darstellung der Daten ist in der Tabelle 2.2 präsentiert: für $i = 1, \dots, n$ Personen und $j = 1, \dots, m$ Aufgaben in der i -ten Zeile und der j -ten Spalte steht ein Eintrag $u_{i,j}$ dafür, ob die i -te Person die j -te Aufgabe tatsächlich richtig (bzw. falsch) gelöst hat.

Person	Aufgabe						
	1	2	3	...	j	...	m
1	$u_{1,1}$	$u_{1,2}$	$u_{1,3}$...	$u_{1,j}$...	$u_{1,m}$
2	$u_{2,1}$	$u_{2,2}$.		.
3	$u_{3,1}$.		.		.
.
i	$u_{i,1}$.	.	.	$u_{i,j}$.
.						.	.
n	$u_{n,1}$	$u_{n,m}$

Tabelle 2.2: Rasch-Modell. Datenmatrix Allgemeine Form

Das Resultat einer Person in einem Test hängt nicht deterministisch von ihrer Fähigkeit ab, sondern kann auch von dem Zufall abhängen. Bei der gleich bleibender Fähigkeit kann eine Person dieselben Aufgaben an unterschiedlichen Tagen anders lösen und damit unterschiedliche Noten (bzw. Punktezahl) bekommen. Es existiert eine latente Zufallsvariable $U_{i,j}$ für das voraussichtliche Resultat vor der Bearbeitung der Aufgaben. Die Wahrscheinlichkeit, dass i -te Person bei der Beantwortung der j -ten Aufgabe einen bestimmten Ergebnis erzielt $P(U_{i,j} = u_{i,j})$, werden wir als Lösungswahrscheinlichkeit bezeichnen. (Vgl. [Strobl \(2010\)](#), S. 5-7)

2.2 Modellgleichung

Die Modellgleichung soll die Wahrscheinlichkeit beschreiben, dass eine Person mit der bestimmten Fähigkeit eine Aufgabe mit dem bestimmten Schwierigkeitsgrad richtig beantwortet.

Natürlich hängt die Lösungswahrscheinlichkeit sowohl von der Fähigkeit der Person als auch von der Schwierigkeit der Aufgabe ab. Mehr begabte Person wird die Fragestellung eher lösen, als eine wenig begabte. Die leichten Aufgaben werden alle Personen mit hoher Wahrscheinlichkeit lösen. Bei der Definition der Modellgleichung müssen wir auch beachten, dass es ein positiver Zusammenhang zwischen der Personenfähigkeit und der Lösungswahrscheinlichkeit gibt. Je fähiger die Person ist, desto eher wird sie die Aufgabe lösen. Da

es sich hier um die Wahrscheinlichkeiten handelt, muss die Funktion zwischen 0 und 1 begrenzt werden. (Vgl. [Strobl \(2010\)](#), S. 7)

Die Modellgleichung

$$P(U_{i,j} = 1|\theta_i, \beta_j) = \frac{e^{\theta_i - \beta_j}}{1 + e^{\theta_i - \beta_j}} \quad (2.1)$$

erfüllt alle Forderungen, die wir bereits besprochen haben. Die Lösungswahrscheinlichkeit ist eine bedingte Wahrscheinlichkeit, die von der Personenfähigkeit θ_i und Aufgabenschwierigkeit β_j abhängt. D.h. die ersten 2 Forderungen sind erfüllt. Der Bruch beschreibt einfache Logistische Funktion

$$P(Y_i = 1) = \pi_i = \frac{e^{\zeta_i}}{1 + e^{\zeta_i}} \quad (2.2)$$

mit ζ_i - der lineare Prädiktor. (Vgl. [Fahrmeir, Alfred und Tutz \(1996\)](#), S. 247-248)

Im Rasch-Modell der lineare Prädiktor ist dargestellt als die Differenz $\theta_i - \beta_j$. An diesem Ausdruck, der sowohl im Zähler als auch im Nenner der Gleichung 2.1 steht, erkennt man positive Differenz, falls $\theta_i > \beta_j$ (die Person ist fähiger als die Aufgabe schwer). Dies bedeutet, dass die Wahrscheinlichkeit, dass die Person diese Aufgabe richtig löst, sehr groß ist. Im Gegensatz dazu wird Differenz negativ, falls $\theta_i < \beta_j$ (die Aufgabe ist schwerer als die Person fähig). Also, die Wahrscheinlichkeit, dass die Person diese Aufgabe richtig löst, ist sehr klein. Somit erfüllt die Modellgleichung die dritte Forderung. Die letzte vierte Forderung ist die Beschränktheit. Der Wertebereich der logistischen Funktion und somit der Rasch-Modell-Funktion liegt zwischen 0 und 1. Das sagt aus, dass die logistische Funktion gut für die Abbildung der Wahrscheinlichkeiten geeignet ist. Für die Beschreibung der Rasch-Modell-Gleichung könnte man auch andere S-förmige Funktionen wie z.B. die kumulierte Dichte der Normal-Verteilung anwenden, was zu den Verletzungen der Rasch-Modell-Annahmen und -Eigenschaften (Siehe Kap. 2.4) führen würde. (Vgl. [Strobl \(2010\)](#), S. 7-9)

2.3 Aufgaben- und Personencharaktereistische Kurven

Die im Rasch-Modell verwendete logistische Funktion beschreibt die Veränderung der Lösungswahrscheinlichkeiten für eine Aufgabe in der Abhängigkeit

von der Fähigkeit der Person. Die Abb. 2.1 stellt uns solche Funktion dar.

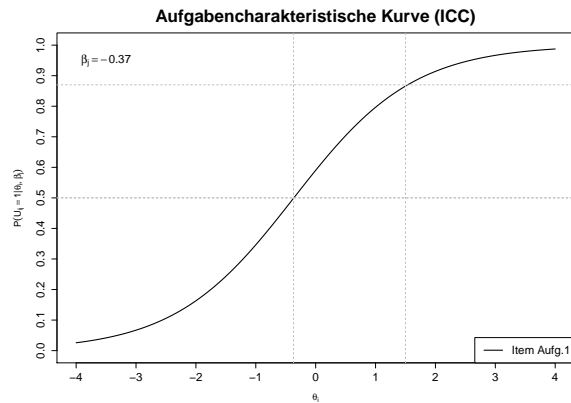


Abbildung 2.1: Aufgabencharakteristische Kurve für eine Aufgabe mit der Schwierigkeitsgrad $\beta_j = -0.37$

Diese nennt man die aufgabencharakteristische Kurve und bezeichnet man mit ICC (*Item Characteristic Curve*). Im allgemeinen gilt: die Personen mit der bestimmten Fähigkeit θ_i gleich der Aufgabenschwierigkeit β_j werden diese Aufgabe mit der Wahrscheinlichkeit von 0.5 lösen. Aus der Abb. 2.1 kann man z.B. ablesen, dass die Personen mit der Fähigkeit $\theta_i = 1.5$, was deutlich größer als $\beta_j = -0.37$ ist, die Aufgabe mit sehr hohen Wahrscheinlichkeit $P = 0.866$ lösen werden. Wohingegen die Personen mit der Fähigkeit von -3 werden diese Aufgabe mit sehr niedriger Wahrscheinlichkeit von 0.067 lösen. (Vgl. Strobl (2010), S. 10)

In der Wirklichkeit besteht ein Test aus mehreren Aufgaben. Die ICCs werden in einer Grafik nebeneinander abgebildet. In der Abb. 2.2 sieht man, dass alle ICCs parallel zueinander verlaufen. In der Abhängigkeit von der Schwierigkeitsgrad der Aufgabe werden die Aufgabencharakteristische Kurven entsprechend nach rechts (Aufgabe 9 mit $\beta_j = 0.57$) oder nach links (Aufgabe 3 mit $\beta_j = -1.65$) verschoben. Das liegt daran, dass es in der Gleichung für Rasch-Modell kein zusätzlichen Parameter für die Steigung der Funktion gibt. D.h. bei der Geltung des Rasch-Modells müssen alle Aufgaben etwas gemeinsames (die Steigung) haben. (Vgl. Strobl (2010), S. 11)

Das ist eine von der wichtigsten Annahmen des Rasch-Modells, die wir in der nächsten Sektion 2.4 betrachten werden.

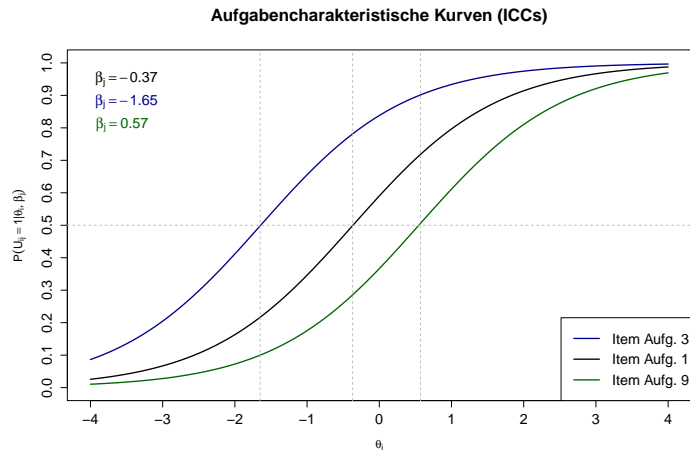


Abbildung 2.2: Aufgabencharakteristische Kurven für drei Aufgaben

Die Steigung der ICC im mittleren Bereich nennt man Trennschärfe. Je höher die Trennschärfe einer Aufgabe ist, desto genauer kann man mit Hilfe dieser Aufgabe zwischen den Personen mit unterschiedlichen Fähigkeiten unterscheiden.

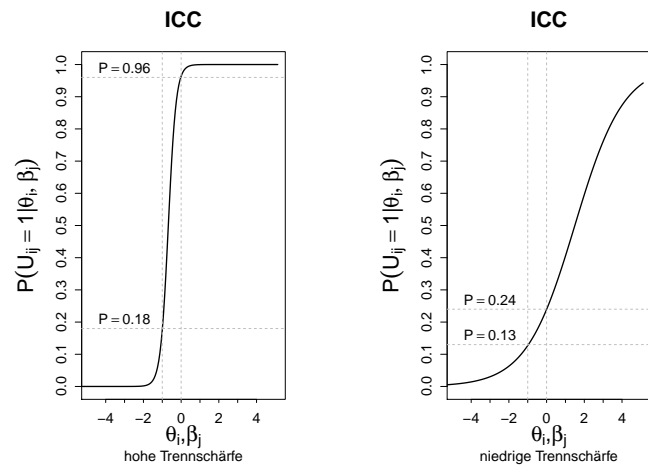


Abbildung 2.3: ICCs mit unterschiedlicher Trennschärfe

Die Abb. 2.3¹ (links) stellt uns die ICC mit der hohen Trennschärfe dar. Die Person mit der Fähigkeit $\theta_i = -1$ löst die Aufgabe mit der Wahrscheinlichkeit von 0.18 und die Person mit der Fähigkeit $\theta_i = 0$ löst diese Aufgabe mit der viel höheren Wahrscheinlichkeit von 0.96. Anhand dieser Aufgabe kann man deutlich zwischen beiden Personen mit nicht so unterschiedlichen Fähigkeiten trennen. Hingegen die Abb. 2.3 (rechts) stellt uns die ICC mit der niedrigen Trennschärfe dar. Die beiden Personen mit der Fähigkeiten -1 und 0 werden die Aufgabe mit der niedrigen Wahrscheinlichkeit lösen. Diese Aufgabe ist weniger geeignet, um zwischen den beiden Personen zu unterscheiden. (Vgl. Strobl (2010), S. 11-12)

2.4 Modellannahmen und Modelleigenschaften

Bis jetzt haben wir einige Eigenschaften und Forderungen des Rasch-Modells besprochen. Es gibt allerdings mehrere Besonderheiten.

„Während es didaktisch einfacher ist, sich diese Besonderheiten als Eigenschaften des Rasch-Modells vorzustellen, die aus der Modellgleichung folgen und die wir uns z.T. auch einfach daran veranschaulichen können, ist es aus messtheoretischer Sicht umgekehrt: Aus bestimmten theoretischen Annahmen kann man die Modellgleichung des Rasch-Modells ableiten.“ (Strobl (2010), S. 14). Die Zentrale Annahmen des Rasch-Modells werden wir zunächst diskutieren.

Suffiziente Statistiken

Eine Statistik heißt suffizient, wenn sie alle Informationen der Stichprobe bezüglich des Parameters enthält (Henning (1974), S. 79) und die Datenreduktion zu keinem Informationsverlust führt (Strobl (2010), S. 15). Man darf nicht die Begriffe erwartungstreu (Vgl. Fahrmeir und Fahrmeir-Künstler-Pigeot-Tutz (2011), Kap. 9.2.1) und suffizient verwechseln.

Wie wir im Beispiel 2.1 sehen, eine Statistik kann erwartungstreu dennoch nicht suffizient sein (x^* enthält keine Informationen über x_2 und x_4).

¹Die Abbildung wurde ähnlich der Abb. 2.4 (Strobl (2010), S. 12) ausgefertigt.

Beispiel 2.1 (Suffizient und Erwartungstreu für n=5:)

- $\bar{x} = \frac{1}{5} \sum_{i=1}^5 x_i = \frac{1}{5}(x_1 + x_2 + x_3 + x_4 + x_5)$

- erwartungstreu ✓

- suffizient ✓

- $x^* = \frac{1}{3}(x_1 + x_3 + x_5)$

- erwartungstreu ✓

- **nicht suffizient**

Vorteil der suffizienten Statistik ist, dass man nicht komplette Daten kennen muss, um alle relevante Informationen über den Parametern zu erhalten. Die ganzen Angaben sind schon in der suffizienten Statistik als Einzelwert zusammengefasst. Im Rasch-Modell sind solche suffizienten Statistiken Spalten- und Zeilenrandsummen. Die Zeilenrandsumme r_i enthält für die i-te Person die gesamte Information über den Personenparameter θ_i und die Spaltenrandsumme s_j enthält für die j-te Aufgabe die gesamte Information über den Aufgabenparameter β_j . (Vgl. [Strobl \(2010\)](#), S. 15-16)

Lokale Stochastische Unabhängigkeit

Die nächste Annahme ist lokale stochastische Unabhängigkeit. Unter stochastischer Unabhängigkeit im Rasch-Modell wird sowohl Personen- als auch Aufgabenunabhängigkeit gemeint. Die Personenunabhängigkeit bedeutet, dass die Lösungswahrscheinlichkeiten von zwei Personen für j-te Aufgabe nicht voneinander abhängen dürfen. In der Praxis bedeutet dies, dass die Personen voneinander nicht abschreiben dürfen. Bei der Aufgabenunabhängigkeit geht es um die ähnliche Bedeutung. Lösungswahrscheinlichkeiten der i-ten Person für zwei Aufgaben dürfen nicht voneinander abhängen. In der Tat dürfen die Lösungen der Aufgaben nicht aufeinander aufgebaut werden. Im Rasch-Modell wird von der lokalen stochastischen Unabhängigkeit gesprochen. Lokal bedeutet hier, dass die Unabhängigkeit der Aufgaben muss gelten, solange man eine Person (bzw. mehrere Personen mit gleichen Fähigkeiten) betrachtet. Natürlich kann es passieren, dass die mehr begabten Personen alle Aufgaben mit höherer Wahrscheinlichkeit lösen als wenig begabten. (Vgl. [Strobl \(2010\)](#), S. 16-20)

Spezifische Objektivität

Die spezifische Objektivität im Rasch-Modell gilt für den Vergleich von Personen ebenso wie für den Vergleich von der Aufgaben. Aussagen über die

Fähigkeiten von 2 Personen hängen nicht davon ab, anhand welcher Aufgabe sie verglichen werden. In der Abb. 2.4² sind 2 Personen mit unterschiedlichen Fähigkeiten ($\theta_a = -1$ und $\theta_b = 0$) und 2 Aufgaben mit verschiedenem Schwierigkeitsniveau ($\beta_{Aufg.3} = -1.7$ und $\beta_{Aufg.1} = -0.4$) gezeichnet. Person a wird beide Aufgaben mit der niedrigeren Wahrscheinlichkeiten als Person b lösen. D.h. der Aufgabenwahl im Rasch-Modell spielt keine Rolle beim Vergleich von Personen. Genauso die Aussagen über die Schwierigkeit von 2 Aufgaben sind unabhängig davon, welche Person gewählt wird.

In der Abb. 2.4 sieht man, dass die leichtere Aufgabe (Item Aufg. 3) von beiden Personen mit höherer Wahrscheinlichkeit als schwierigere Aufgabe (Item Aufg. 1) gelöst wird. Auch der Personenwahl im Rasch-Modell spielt keine Rolle beim Vergleich von Aufgaben. (Vgl. Strobl (2010), S. 20-23)

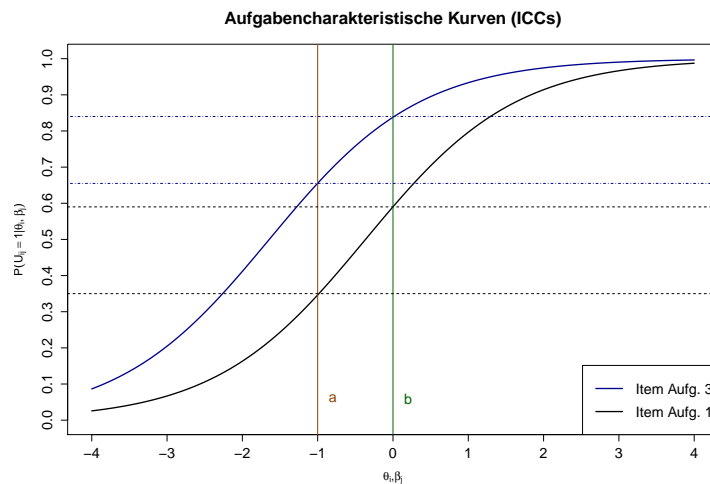


Abbildung 2.4: Spezifische Objektivität

Eindimensionalität

Noch eine zentrale Annahme im Rasch-Modell ist Eindimensionalität. Hier wird es angenommen, dass die Personenparameter θ und Aufgabenparameter β auf **einer** gemeinsamen latenten Dimension liegen. Formal in der Modellgleichung erkennt man das an der Differenz $\theta_i - \beta_j$. Die Aufgabenparameter werden direkt von der Personenparameter abgezogen. Inhaltlich bedeutet es,

²Die Abbildung wurde anhand der simulierten Daten erstellt

dass z.B. in einem Test zur Lesekompetenz nur die Lesekompetenz gemessen wird. (Vgl. [Strobl \(2010\)](#), S. 23)

Messniveau

Die Personen- und Aufgabenparameter haben gemeinsame latente Skala, die keinen absoluten Nullpunkt hat. Wie es auch diskutiert wurde, wird der Nullpunkt durch die Wahl eines Wertes B festgelegt. Allgemeinere Gleichung des Rasch-Modells ist folgendermaßen dargestellt:

$$P(U_{i,j} = 1|\theta_i, \beta_j) = \frac{e^{A(\theta_i - \beta_j) + B}}{1 + e^{A(\theta_i - \beta_j) + B}} \quad (2.3)$$

Hier sieht man, dass der Wert B die gesamte Parameterskala nach links oder rechts verschiebt. Es wird auch die Einheit der Skala durch den Wert von A festgelegt. Der Wert A ausweitet oder zusammendrückt die einzelne Abschnitte der Skala. Man erkennt sofort die lineare Transformation, die von mindestens Intervall-Skalen erlaubt sind. D.h. im Rasch-Modell sind die Messungen von zumindest Intervallskalenniveau zulässig. (Vgl. [Strobl \(2010\)](#), S. 23-25)

2.5 Parameterschätzung

Wie am Anfang schon gesagt wurde, ist das Ziel, die beobachtete Reaktionen durch das unbeobachtete Eigenschaften zu erklären und diese Eigenschaften zu schätzen. D.h. im Rasch-Modell alle Parameter θ_i und β_j . Die grundlegende Methode dafür ist Maximum-Likelihood-Schätzung (ML-Schätzung). (Zur ML-Methode Vgl. [Toutenburg, Heumann, Nittner und Scheid \(2003\)](#), Kap 10.2.4, [Fahrmeir, Alfred und Tutz \(1996\)](#), Kap 2.3 und [Fahrmeir und Fahrmeir-Künstler-Pigeot-Tutz \(2011\)](#), Kap 9.3.1)

Man unterscheidet allerdings im Rasch-Modell die Vorgehensweisen der Schätzung. Je nachdem, ob man die Parameter gleichzeitig oder nacheinander schätzt, werden unterschiedliche Schätzansätze verwendet. Für die gleichzeitige Schätzung der Personen- und Aufgabenparameter verwendet man die Gemeinsame ML-Methode. Und wenn man die Parameter nacheinander schätzt, benutzt man die Bedingte oder Marginale ML-Schätzung.

Der erste Schritt der ML-Schätzung ist die Bildung der Maximum-Likelihood-Funktion. Dafür braucht man die allgemeine Form des Rasch-Modells

$$P(U_{i,j} = u_{i,j} | \theta_i, \beta_j) = \frac{e^{u_{i,j}(\theta_i - \beta_j)}}{1 + e^{\theta_i - \beta_j}} \quad (2.4)$$

Man bildet ein Produkt über alle Aufgabenparameter und erhält Likelihood für eine Person i und alle Aufgaben³

$$L_{u_i}(\theta_i, \beta) = \frac{e^{r_i \theta_i - \sum_{j=1}^m u_{i,j} \beta_j}}{\prod_{j=1}^m (1 + e^{\theta_i - \beta_j})} \quad (2.5)$$

Diese Likelihood-Funktion ist die Grundlegende Funktion für alle Schätzansätze. (Vgl. [Strobl \(2010\)](#), S. 27-28)

Gemeinsame ML-Schätzung

Bei dieser Vorgehensweise bildet man gemeinsame Likelihood⁴

$$L_u(\theta, \beta) = \frac{e^{\sum_{i=1}^n r_i \theta_i - \sum_{j=1}^m s_j \beta_j}}{\prod_{i=1}^n \prod_{j=1}^m (1 + e^{\theta_i - \beta_j})} \quad (2.6)$$

und daraus schätzt man die Personen- und Aufgabenparameter gleichzeitig. Diese Vorgehensweise ist allerdings nicht die beste, weil es hier das große Problem auftritt. Man ist immer früher daraus ausgegangen, dass mit dem wachsenden Stichprobenumfang die Varianz kleiner und die Schätzungen genauer werden. In unserem Fall mit der wachsenden Stichprobengröße steigt die Anzahl zu schätzenden Parameter. D.h. mit jeder zusätzlich in die Stichprobe aufgenommenen Person, muss man um einen Personenparameter mehr schätzen. Man bekommt nie die konstante Parameteranzahl mit dem wachsenden Stichprobenumfang. Die Schätzer sind in diesem Fall nicht konsistent. (Vgl. [Strobl \(2010\)](#), S. 28-29 und zur Konsistenz der Schätzer [Fahrmeir und Fahrmeir-Künstler-Pigeot-Tutz \(2011\)](#), Kap. 9.2.2)

Bedingte ML-Schätzung

Bei dieser Methode werden zuerst die Aufgabenparameter geschätzt. Dafür wird die Logarithmus der Likelihood gebildet, zunächst diese Funktion abgeleitet und gleich Null gesetzt. Als nächstes die geschätzte Aufgabenschwierigkeiten als deterministisch gegeben in die jeweilige Likelihood-Funktion einge-

³Herleitung siehe Anhang A Parameterschätzung, Abschnitt A.1 (Vgl. [Strobl \(2010\)](#))

⁴Herleitung siehe Anhang A Parameterschätzung, Abschnitt A.2 (Vgl. [Strobl \(2010\)](#))

setzt und daraus schon die Personenparameter geschätzt. Diese ML-Methode ist unproblematisch, weil die bedingte Likelihood⁵

$$h(u|r, \beta) = \frac{e^{-\sum_{j=1}^m s_j \beta_j}}{\prod_{i=1}^n \gamma_{r_i}(\beta)} \quad (2.7)$$

schon nicht von der Personenparameter abhängig ist. (Vgl. [Strobl \(2010\)](#), S. 29-33)

Marginale ML-Schätzung

Bei der Marginalen ML-Schätzung ebenso wie bei der Bedingten ML-Methode werden die Personenparameter nach den Aufgabenparametern geschätzt. Der Unterschied besteht daran, dass die Personenparameter aus der Funktion ausintegriert werden. Man multipliziert erst die gemeinsame Likelihood (Formel 2.6) mit der Randdichte der Personenparameter $f(\theta)$ und als nächstes integriert diese über θ . Als Ergebnis bekommt man die Marginale Likelihood⁶

$$L_u(\beta) = \int P(u, \theta|\beta) \partial\theta \quad (2.8)$$

Bei dieser Vorgehensweise kann auch ein Problem auftreten. Man braucht die Annahme über marginale Randverteilung. Wenn diese Vermutung für Randdichte $f(\theta)$ falsch ist, bekommt man eine verzerrte Schätzung. (Vgl. [Strobl \(2010\)](#), S. 33–34)

⁵Herleitung siehe Anhang A Parameterschätzung, Abschnitt A.3 (Vgl. [Strobl \(2010\)](#))

⁶Herleitung siehe Anhang A Parameterschätzung, Abschnitt A.4 (Vgl. [Strobl \(2010\)](#))

Kapitel 3

Verallgemeinerungen des Rasch-Modells

Neben dem Modell, dass wir hier als Grundmodell kennengelernt haben, gibt es viele Erweiterungen mit unterschiedlichen Zielsetzungen. Diese verallgemeinerte Modelle werden um die zusätzlichen Komponente erweitert und deswegen werden die unterschiedliche Annahmen und Forderungen des Grundmodells in einzigen Fällen verletzt.

3.1 Das linear-logistische Testmodell

Das linear-logistische Testmodell (LLTM) von Fischer (1973) gibt ausführlich den Einfluss von Teilkompetenzen auf die Lösungswahrscheinlichkeit an. Die Aufgaben-Parameter sind als Linearkombination der Parameter für die Teilkompetenzen dargestellt

$$\beta_j = \sum_l \omega_{j,l} \cdot \eta_l \quad (3.1)$$

Wobei $\omega_{j,l} \in \{0, 1, 2, \dots\}$ bezeichnet Gewicht für die Teilkompetenz, die in der Aufgabe enthalten ist (0 = nein, 1 = einmal, 2 = zweimal, ...), und η_l steht für die Schwierigkeitsgrad der Teilkompetenz. Die Gewichte werden bei der Konstruktion der Aufgaben festgelegt. Jede Teilkompetenz, die zur Lösung der Aufgabe benötigt wird, erhöht die Schwierigkeit der Aufgabe um ein bestimmten Beitrag η_l , sodass sich die Gesamtschwierigkeit der Aufgabe als die Summe der einzelnen gewichteten Beiträgen ergibt. Diese Vorgehensweise ermöglicht die Schätzung von Schwierigkeitsgrad der einzelnen Teilkomponenten. (Vgl. [Strobl \(2010\)](#), S. 49-50, [Kubinger, Rasch und Yanagida \(2011\)](#), S. 564 und [Moosbrugger, Kelava und Moosbrugger-Kelava \(2012\)](#), S. 267-268)

Die Tabellen 3.1 und 3.2 stellen die Allgemeine Form der Gewichtung und der Teilkompetenzparameter dar. Im Gegensatz zur Datendarstellung (Vgl. Tabelle 2.2) stehen die Aufgaben in beiden Tabellen in den Zeilen. Die Teilkompetenzen sind hier jeweils in die Spalten eingetragen.

Aufgabe	Teilkompetenzen						
	1	2	3	...	l	...	L
1	$\omega_{1,1}$	$\omega_{1,2}$	$\omega_{1,3}$...	$\omega_{1,j}$...	$\omega_{1,L}$
2	$\omega_{2,1}$	$\omega_{2,2}$.		.
3	$\omega_{3,1}$.		.		.
.
j	$\omega_{j,1}$.	.	.	$\omega_{j,l}$.
.						.	.
m	$\omega_{m,1}$	$\omega_{m,L}$

Tabelle 3.1: LLTM. Gewichtung-Darstellung Allgemeine Form

Aufgabe	Teilkompetenzen						
	1	2	3	...	l	...	L
1	$\eta_{1,1}$	$\eta_{1,2}$	$\eta_{1,3}$...	$\eta_{1,j}$...	$\eta_{1,L}$
2	$\eta_{2,1}$	$\eta_{2,2}$.		.
3	$\eta_{3,1}$.		.		.
.
j	$\eta_{j,1}$.	.	.	$\eta_{j,l}$.
.						.	.
m	$\eta_{m,1}$	$\eta_{m,L}$

Tabelle 3.2: LLTM. Teilkompetenzparameter-Darstellung Allgemeine Form

Z.B. in einem Test für Mathematik-Kompetenz, der aus 3 Aufgaben besteht, kann jede Aufgabe 3 Teilkompetenzen zu Kommutativ- Assoziativ- und Distributivgesetz enthalten. In den Tabellen 3.3 und 3.4 sind die Gewichte und die Schwierigkeitsparameter für jede Teilkompetenz eingetragen. Das Beispiel 3.1 demonstriert, wie sich die Gesamtschwierigkeiten der 3 Aufgaben aus einzelnen Teilkompetenzparameter gegeben Gewichtung berechnen lässt.

Beispiel 3.1 (LLTM. β_j für 3 Aufgaben und 3 Teilkompetenzen)

Die Schwierigkeitsparameter β_j ergeben sich aus Tabellen 3.3 und 3.4 folgender weise:

$$\beta_1 = 0 \cdot \eta_{1,1} + 1 \cdot \eta_{1,2} + 2 \cdot \eta_{1,3} = \eta_{1,2} + 2\eta_{1,3}$$

Aufgabe	Teilkompetenzen		
	Kommutativ- Gesetz	Assoziativ- Gesetz	Distributiv Gesetz
1	0	1	2
2	3	1	0
3	0	0	1

Tabelle 3.3: LLTM. Gewichtung-Darstellung Bsp. für 3 Aufgaben und 3 Teilkompetenzen

Aufgabe	Teilkompetenzen		
	Kommutativ- Gesetz	Assoziativ- Gesetz	Distributiv Gesetz
1	$\eta_{1,1}$	$\eta_{1,2}$	$\eta_{1,3}$
2	$\eta_{2,1}$	$\eta_{2,2}$	$\eta_{2,3}$
3	$\eta_{3,1}$	$\eta_{3,2}$	$\eta_{3,3}$

Tabelle 3.4: LLTM. Teilkompetenzparameter-Darstellung Bsp. für 3 Aufgaben und 3 Teilkompetenzen

$$\beta_2 = 3 \cdot \eta_{2,1} + 1 \cdot \eta_{2,2} + 0 \cdot \eta_{2,3} = 3\eta_{2,1} + \eta_{2,2}$$

$$\beta_3 = 0 \cdot \eta_{3,1} + 0 \cdot \eta_{3,2} + 1 \cdot \eta_{3,3} = \eta_{3,3}$$

3.2 Birnbaum-Modelle

Das zwei-parametrische Birnbaum-Modell (2PLM)

In das zwei-parametrische Birnbaum-Modell von Kubinger und Draxler (2006)

$$P(U_{i,j} = 1 | \theta_i, \beta_j, \delta_j) = \frac{e^{\delta_j(\theta_i - \beta_j)}}{1 + e^{\delta_j(\theta_i - \beta_j)}} \quad (3.2)$$

wird der zusätzliche Parameter für die Steigung δ_j aufgenommen. δ_j ist ein Diskriminationsparameter, der "die richtige Antworten bei der Bestimmung des Testwerts unterschiedlich gewichtet" (Kubinger, Rasch und Yanagida (2011), S. 563).

Bisher war dieser Parameter gleich 1 und jetzt $\delta_j \neq 1$, sodass die alle Aufga-

ben nicht mehr die gleiche Steigung aufweisen können (Abb. 3.1¹). Es wird die Forderung von Spezifischen Objektivität (siehe Kap. 2.4) aufgegeben. "Die Diskriminationsparameter geben an, wie stark sich die Lösungswahrscheinlichkeiten in Abhängigkeit von der Personenfähigkeit verändern" (Moosbrugger, Kelava und Moosbrugger-Kelava (2012), S. 251).

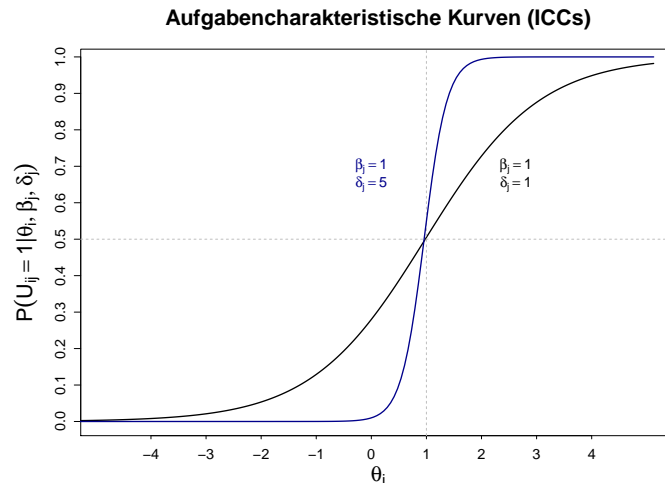


Abbildung 3.1: Zwei ICCs im zwei-parametrischen Birnbaum-Modell (2PLM)

Welches Modell besser passt, hängt von der Zielsetzung ab. Will man den neuen Test konstruieren, muss man die ungeeignete Aufgaben solange aussortieren, bis der Test alle Forderungen des Rasch-Modells erfüllt und somit alle Schätzungen objektiv sind. Wenn man hingegen die Daten aus schon bestehenden Test möglichst gut beschreiben möchte, kann dafür 2PLM oder andere Modelle besser geeignet werden. Allerdings in 2PLM Fall sind die Aufgaben- und Personenparameter nicht mehr trennbar, was zu nicht mehr suffizienten Randsummen führt, sodass es hier keine bedingte sondern marginale ML-Schätzung möglich ist. Um nachzuweisen, ob die zusätzliche Parameter für die unterschiedlicher Trennschärfe nötig sind, kann man die Modelle mithilfe z.B. eines LQ-Tests vergleichen. (Vgl. Strobl (2010), S. 50-52)

¹Die Abbildung wurde ähnlich der Abb. 5.1 (Strobl (2010), S. 52) ausgefertigt.

Das Birnbaum-Modell mit zusätzlichem Rateparameter (3PLM)

Ausgangspunkt für das Birnbaum-Modell mit zusätzlichem Rateparameter ist 2PLM. Hierbei wird noch ein zusätzlicher Parameter γ_j ins Modellgleichung aufgenommen

$$P(U_{i,j} = 1 | \theta_i, \beta_j, \delta_j, \gamma_j) = \gamma_j + (1 - \gamma_j) \cdot \frac{e^{\delta_j(\theta_i - \beta_j)}}{1 + e^{\delta_j(\theta_i - \beta_j)}} \quad (3.3)$$

Bisher war dieser Parameter gleich 0, im 3PLM ist $\gamma_j \neq 0$. Der Rateparameter γ_j regelt, ob es bestimmte untere Schranke > 0 für die Lösungswahrscheinlichkeit gibt.

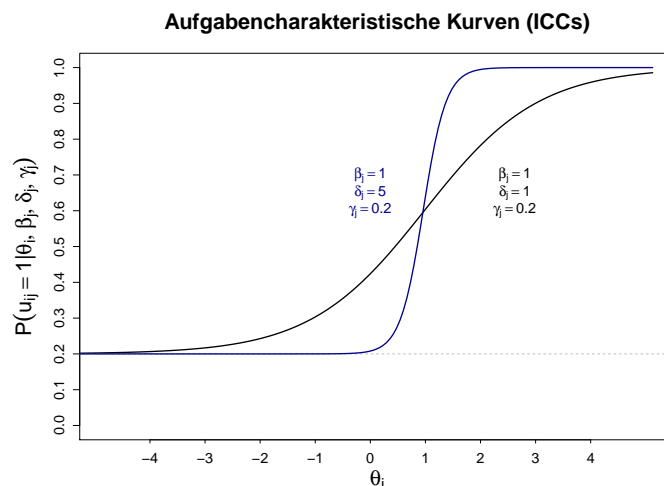


Abbildung 3.2: Zwei ICCs im Birnbaum-Modell mit zusätzlichem Rateparameter (3PLM)

Anders ausgedrückt, ob die Person mit unendlich niedriger Fähigkeit die Aufgabe mit einer gewisser positiver Wahrscheinlichkeit lösen kann. Ein gutes Beispiel dafür ist Multiple-Choice-Test, bei dem Richtige Antwortkombination durch zufälliges Ankreuzen geraten werden kann.

Die Abb. 3.2² veranschaulicht den Verlauf von Aufgabencharakteristischen

²Die Abbildung wurde ähnlich der Abb. 5.2 (Strobl (2010), S. 53) ausgefertigt.

Kurven mit gegebenem Parameter $\gamma_j = 0.2$. Praktisch ist es interessanter den Rateparameter nicht vorzugeben, sondern zu schätzen. (Vgl. Strobl (2010), S. 52-54)

Setzt man $\gamma_j = 0$, bekommt man 2PLM. Wenn man dazu noch $\delta_j = 1$ setzt, bekommt man das Rasch-Modell (so genanntes 1PLM). (Vgl. (Kubinger, Rasch und Yanagida, 2011), S. 563)

3.3 Das Partial-Credit-Modell

Bisher haben wir binäre Zufallsvariable $u_{i,j} \in \{0, 1\}$ betrachtet (Vgl. Kap. 2.1). Das Partial-Credit-Modell von Masters (1982)

$$P(U_{i,j} = c | \theta_i, \beta_j) = \frac{e^{c \cdot \theta_i - \beta_{j,c}}}{\sum_{k=0}^{C_j} e^{k \cdot \theta_i - \beta_{j,k}}} \quad (3.4)$$

mit $\beta_{j,c} = \sum_{l=1}^c \tau_{j,l}$ und $\beta_{j,0} = 0$ stellt uns den Fall mit mehrstufigen Antwortkategorien dar. Jetzt steht $u_{i,j} \in \{0, \dots, C_j\}$ für die Zufallsvariable, ob die i-te Person die j-te Aufgabe ganz oder teilweise gelöst hat (0 = nein, 1 = 1. Teilschritt, 2 = 1. und 2. Teilschritte, ...).

Die Gleichung des Partial-Credit-Modells (Formel 3.4) gibt die Wahrscheinlichkeit dafür an, dass die i-te Person in der j-ten Aufgabe mit der Kategorie c antwortet. Genauer gesagt, dass die i-te Person den bestimmten Anteil der j-ten Aufgabe löst. Der Zähler $e^{c \cdot \theta_i - \beta_{j,c}}$, der den Ausdruck für bestimmte Kategorie c darstellt, steht im Verhältnis zu der Summe über alle Antwortkategorien $\sum_{k=0}^{C_j} e^{k \cdot \theta_i - \beta_{j,k}}$, was im Nenner der Gleichung steht. Man erkennt, dass in dieser Formel für jede Antwortkategorie einen eigenen Aufgabenparameter $\beta_{j,c} = \sum_{l=1}^c \tau_{j,l}$ gibt, der als Summe von der überschrittenen Schwellenwerte dargestellt ist. D.h. die Modellgleichung sieht für jede Antwortkategorie anders aus.

Beispiel 3.2 (Das Partial-Credit-Modell für 4 Kategorien)

Eine Aufgabe bestehe aus 3 Teilschritten. Die Person bekommt 0 Punkte für nicht gelöste Aufgabe. 1 Punkt für einen richtig gelösten Teilschritt, 2 Punkte für zwei fehlerfrei gelösten Teilschritten und 3 Punkte für komplett gelöste Aufgabe.

Die Abb. 3.3³ veranschaulicht ICCs für 4 mögliche Antwortkategorien aus dem Beispiel 3.2. Die blaue Kurve stellt uns die Wahrscheinlichkeit dar, dass die i-te Person diese Aufgabe nicht löst und 0 Punkte erzielt. Die Wahrscheinlichkeit wird hoch für die wenig begabten Personen und umgekehrt

³Die Abbildung wurde ähnlich der Abb. 5.3 (Strobl (2010), S. 56) ausgefertigt.

niedrig für die mehr begabten Personen. Die rosa und die grüne Kurven bezeichnen die Wahrscheinlichkeiten dafür, dass die i -te Person nur teilweise diese Aufgabe löst und 1 bzw. 2 Punkte bekommt. Die Wahrscheinlichkeit wird niedrig für Personen mit hohen und niedrigen Fähigkeiten und hingegen hoch für die Personen mit mittleren Fähigkeiten. Die schwarze Kurve bildet die Wahrscheinlichkeit dafür ab, dass die i -te Person diese Aufgabe komplett löst und 3 Punkte erreicht. Aus der Abb. 3.3 lässt sich ablesen, dass z.B. die Person mit der Fähigkeit $\theta_i = 5$ mit höchster Wahrscheinlichkeit in die dritte Kategorie landet und somit 2 Punkte erzielt, während die Wahrscheinlichkeiten für zweite und vierte Kategorien deutlich kleiner sind und für die erste Kategorie fast bei Null liegt. Die grauen gestrichelten Linien sind die Schwellenwerte, die Schnittpunkte zwischen den Kurven Markieren. Alle Personen mit der Fähigkeiten, die zwischen den 2 Schwellenwerten liegen ($\tau_{j,l} < \theta_i < \tau_{j,l+1}$), haben die größte Wahrscheinlichkeit für diese Kategorie. Die Anzahl der erreichten Punkten entspricht der Zahl der überschrittenen Schwellenwerten. Es ist deutlich, welches Fähigkeitsniveau muss man haben um bestimmte Anzahl der Punkte zu erzielen. Je höher Personenfähigkeit, desto mehr Schwellenwerte werden überschritten und desto höhere Kategorie wird erreicht. (Vgl. Strobl (2010), S. 54-56 und Moosbrugger, Kelava und Moosbrugger-Kelava (2012), S. 265-266)

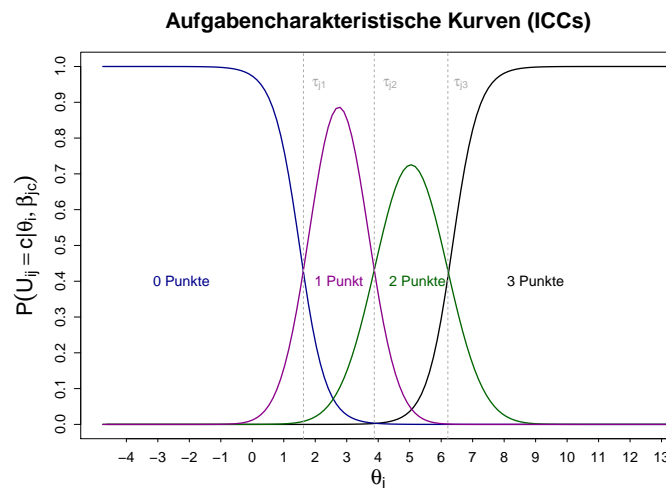


Abbildung 3.3: ICCs für 4 Antwortkategorien im Partial-Credit-Modell

3.4 Weitere Verallgemeinerungen

Es gibt mehrere verallgemeinerten Modellen, die auf Rasch-Modell oder anderen Verallgemeinerungen aufgebaut wurden. So z.B. das Rating-Scale-Modell von Andrich (1978) ist ein Spezialfall des Partial-Credit-Modells, in dem alle Aufgaben die gleiche Anzahl von Antwortkategorien haben (Vgl. Strobl (2010), S. 57). Das dichotome Rasch-Modell mit Rateparametern stellt ein Spezialfall des 3PLMs mit dem Steigungsparameter $\delta_j = 1$ und der unteren Schranke $\gamma_j \neq 0$ dar (Vgl. (Moosbrugger, Kelava und Moosbrugger-Kelava, 2012), S. 252). Das Mischverteilungs-Rasch-Modell von Rost (1990) enthält unbekannte Anzahl von latenten Personengruppen mit unterschiedlichen Aufgabeparametern. Man modelliert unterschiedliche Anzahl von Gruppen und sucht nach beste Modellanpassung mithilfe AIC oder BIC (Zu AIC- und BIC-Prinzip Fahrmeir, Kneib und Lang (2009), S. 161-180). Es gibt noch die Mehrdimensionale Rasch-Modelle. Hierbei wird nicht nur eine Dimension, sondern mehrere Dimensionen betrachtet. In einem Test zu Mathematikkompetenz können sowohl die mathematisch-logischen als auch verbale Aufgaben (Deutsch-Kenntnisse nötig) enthalten werden. Es ist angemessener beide Bereiche als getrennte Dimensionen mit der entsprechender Gewichtung zu betrachten. (Vgl. (Strobl, 2010), S. 61-62)

Auf diese Weise z.B. in PISA 2012 wurden drei latente Personeneigenschaften (in Mathematik-, Lese- und Naturwissenschaft-Kompetenzbereich) untersucht und analysiert. Es wurde die Mehrdimensionale Rasch-Skalierung verwendet, in dem man die gesamte Aufgabenmenge gemeinsam analysiert und durch die mehrdimensionale Rasch-Modellgleichung erweitert wurde. (Vgl. Prenzel, Sälzer, Klieme und Köller (2013), S. 334-335)

Kapitel 4

Simulation des Rasch-Modells

Um das Rasch-Modell sowie die verallgemeinerte Modelle untersuchen zu können und auf dem praktischen Beispiel die Geltung (bzw. nicht) des Rasch-Modells zeigen zu können, braucht man die Daten des tatsächlich durchgeführten Tests oder simulierte Daten.

Mithilfe des Programms R ([R Core Team \(2014\)](#)) unter der Verwendung des Paketes eRm ([Mair und Hatzinger \(2007\)](#)) kann man die nötigen Daten für Rasch-Modell simulieren.

Funktion für Simulation des Rasch-Modells

Funktion *Rasch.Simulation* erzeugt Rasch-Modell-Datensätze mit den Spalten- und Zeilennamen. Die eigentliche Simulation der Daten wird mit Hilfe der Funktion *sim.rasch* (R-Pakete eRm) durchgeführt, die eine dichotome Rasch-Modell-Datenmatrix für die gewünschte Stichprobengröße (Argument *persons*) und die gewünschte Anzahl der Aufgaben (Argument *items*) modelliert. Es wird die zufällige Wahrscheinlichkeitsmatrix in die resultierende 0/1-Matrix transformiert (Argument *cutpoint*). Auch kann man hier die Anfangsbedingung für die Zufallszahlenerzeugung (Argument *seed*) angeben, was dafür sorgt, dass die Werte genau dieser Simulation vorkommen.

Für weitere Modellgeltungstests wurde der Datensatz mit 200 Personen (Zeilennamen: Pers.1,...,Pers.200), 15 Aufgaben (Spaltennamen: Aufg.1,..., Aufg.15) simuliert. Dabei wurde Argument *cutpoint* = "randomized" gewählt. Eine randomisierte Zuweisung bedeutet, dass für jede Zelle eine zusätzliche Zufallszahl gezogen wird. Wenn die Modellwahrscheinlichkeit größer als dieser Wert ist, die Person erhält 1 an diesem Punkt, wenn kleiner, wird der Person einen 0 zugeordnet. Für die Simulation wurde Anfangsbedingung *seed* = 49 ausgewählt.

Simulation und Faktorisierung der Variablen Geschlecht und Schultyp

Desweiteren für 200 Personen wurden 2 Variablen Geschlecht mit 2 Kategorien (männlich, weiblich) und Schultyp mit 4 Kategorien (Gymnasium, Realschule, Hauptschule, Integrierte Gesamtschule) erzeugt. Mithilfe der Funktion *sample* wurden erst die zufällige numerische Variablen simuliert und mithilfe der Funktion *factor* wurden diese numerische Variablen in Nominal-skalierte umgewandelt. Dabei wurde Anfangsbedingung *seed = 49* ausgewählt, weil in diesem Fall wurden auch DIF-Aufgaben mit im Datensatz drin, die wir für weitere Analysen benötigten.

Kapitel 5

Modellgeltungstests

Es gibt unterschiedliche statistische Tests, mit deren Hilfe prüft man, ob das Rasch-Modell gilt. Man kontrolliert, ob es systematische Unterschiede in der geschätzten Aufgaben-Parameter zw. den Gruppen vorliegen. Bei der Geltung des Rasch-Modells gibt es keine systematische Unterschiede und wenn die Forderung von Spezifischen Objektivität (siehe Kap. 2.4) des Rasch-Modells verletzt wurde, wird eine Aufgabe für unterschiedlichen Personen-Gruppen unterschiedlich schwer, d.h. liegt DIF (*Differential Item Functioning*) vor. (Vgl. [Strobl \(2010\)](#), S. 39)

5.1 Der Graphische Modelltest

Der Graphische Modelltest basiert auf dem Vergleich geschätzten Aufgaben-Parameter in 2 Gruppen. Bei der Geltung des Rasch-Modells stimmen die Werte in beiden Gruppen bis auf lineare Transformationen überein. Man trägt die geschätzte Werte in beiden Gruppen in ein Koordinatensystem ein. Idealerweise bei der Übereinstimmung sollen die Werte auf der Winkelhalbierenden liegen. Praktisch weichen diese von der Diagonale ab. (Vgl. [Kubinger, Rasch und Yanagida \(2011\)](#), S. 556 und [Strobl \(2010\)](#), S. 39-41)

Anhand der Simulierten Daten mit 200 Personen und 15 Aufgaben wurde es zum einen hohe Leistung gegen niedrige getestet und zum anderen Männer gegen Frauen. Für die Schätzungen und Vergleiche in R wurden die Funktionen *RM* und *LRtest* aus dem Paket "*eRm*" ([Mair und Hatzinger \(2007\)](#)) verwendet. Zugrunde liegender Test für die grafische Darstellung ist der LQ-Test (Siehe Kap. 5.2). Mithilfe der Funktion *plotGOF* aus dem Paket "*Hmisc*" ([Frank und Harrell \(2015\)](#)) wurden die Grafiken erzeugt. In der Abb. 5.1 sind links die geschätzte Werte für die Leistung und rechts für Geschlecht abgetragen. Um daraus zu schließen, ob die Abweichungen systematisch oder

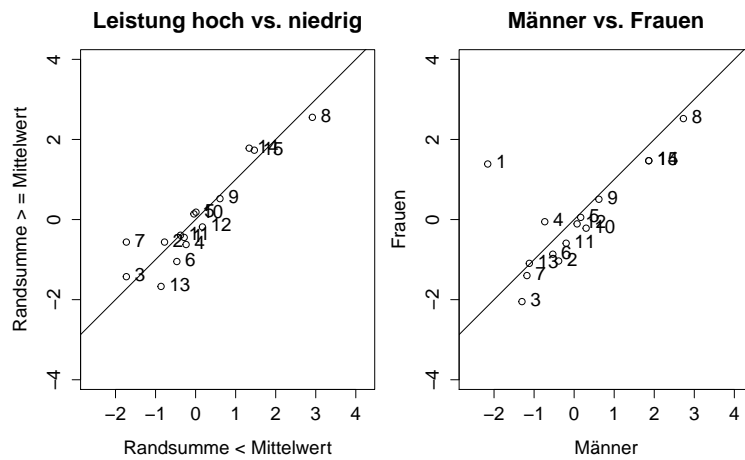


Abbildung 5.1: Der Grafische Modelltest

zufällig sind, trägt man die Konfidenzregionen (zweidimensionale Konfidenzintervalle) meistens zum Signifikanzniveau $\alpha = 0.05$ ab. Schneiden sich die Konfidenzregionen mit der Winkel-halbierenden nicht, liegt es DIF vor. (Vgl. (Strobl, 2010), S. 39-41)

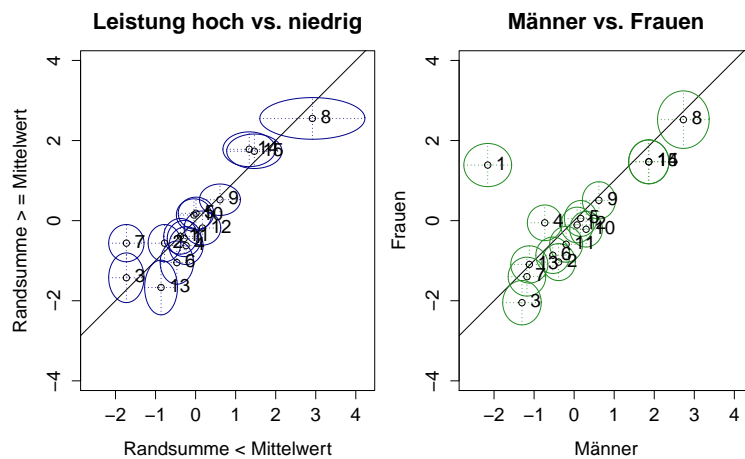


Abbildung 5.2: Der Grafische Modelltest mit Konfidenzregionen

In der Abb. 5.2 links weisen DIF die Aufgaben 7 und eventuell 13 vor, rechts die Aufgaben 1, 4, und eventuell 2 und 3. Es lässt sich auch ablesen, dass z.B. Aufgabe 1 (Abb. 5.2 rechts) oberhalb der Gerade liegt. Diese Aufgabe ist für die Männer leichter zu lösen als für die Frauen. Die unterhalb der

Winkel-halbierenden liegende Aufgabe 3 ist hingegen für die Frauen leichter zu lösen als für die Männer.

5.2 Likelihood-Quotienten-Test

Die Idee von dem Likelihood-Quotienten-Test (LQ-Test) von Andersen (1972) ist genauso wie bei dem Grafischen Modelltest mit dem einzigen Unterschied, dass man beliebig viele Gruppen K wählen kann. Der Likelihood-Quotient

$$LQ = \frac{L_u(r, \hat{\beta})}{\prod_{k=1}^K L_{u_k}(r_k, \hat{\beta}_k)} \quad (5.1)$$

vergleicht die gemeinsame Schätzung der Parameter (gemeinsame Likelihood im Zähler des Bruches) mit den Schätzungen getrennt nach Gruppen (Produkt der einzelnen Likelihood-Funktionen für jeder Gruppe im Nenner des Bruches). Bei der Geltung des Rasch-Modells passen die gemeinsam geschätzte Parameter zu den Daten genauso gut wie die geschätzte Parameter für K -Gruppen getrennt.

Somit lassen sich die Hypothesen

$$H_0 : LQ = 1 \text{ vs. } H_1 : LQ \neq 1 \quad (5.2)$$

für LQ-Test leicht formulieren. Die Nullhypothese ist dabei, dass in jeder Teilgruppe dieselben Schwierigkeitsparameter gelten. In diesem Fall werden der Zähler und der Nenner gleich groß und somit $LQ = 1$. Wenn es in jeder Teilgruppe unterschiedliche Schwierigkeitsparameter gelten, werden die Bruchteile nicht gleich und folglich $LQ \neq 1$ ist die Gegenhypothese.

Allerdings wird der Likelihood-Quotient nicht direkt zum Testen verwendet, sondern die aus dem LQ berechnete Teststatistik

$$T = -2 \cdot \ln(LQ) \sim \chi^2(K - 1) \cdot (m - 1) \quad (5.3)$$

Die Nullhypothese wird abgelehnt, wenn Teststatistik größer, als $z_{1-\alpha}$ -Quantil der χ^2 -Verteilung. (Vgl. Strobl (2010), S. 41-43)

In R erfolgt der LQ-Test mithilfe der Funktion *LRtest* aus dem Paket "eRm" (Mair und Hatzinger (2007)). Es wurde noch Mal die Leistung (hohe gegen niedrige) sowie Geschlecht (Männer gegen Frauen) getestet. Außerdem wurden die Personen nach Schultyp (Gymnasium, Realschule, Hauptschule und Integrierte Gesamtschule) aufgeteilt und LQ-Test für 4 Gruppen durchgeführt.

Leistung		Geschlecht		Schultyp	
Andersen LR-test:		Andersen LR-test:		Andersen LR-test:	
LR-value:	24.807	LR-value:	111.721	LR-value:	49.155
Chi-square df:	14	Chi-square df:	14	Chi-square df:	42
p-value:	0.037	p-value:	0	p-value:	0.208

Tabelle 5.1: Likelihood-Quotienten-Test

In der Tabelle 5.1 sind die Ergebnisse der 3 Tests zum Signifikanzniveau $\alpha = 0.05$ dargestellt. Es wurden signifikante Unterschiede in den Gruppen für Leistung und Geschlecht (p-value: 0.037 bzw. 0 kleiner als 0.05) erkannt. D.h. unsere Nullhypothese kann verworfen werden. Die Aufgaben weisen DIF auf. Hingegen hat der LQ-Test keine signifikante Unterschiede (p-value: 0.208 > 0.05) für die Aufteilung nach Schultyp erkannt. Die Nullhypothese, dass das Rasch-Modell gilt, kann nicht abgelehnt werden.

5.3 Wald-Test

Der Wald-Test mit den Hypothesen

$$H_0 : \hat{\beta}_{j,k} = \hat{\beta}_{j,l} \text{ vs. } H_1 : \hat{\beta}_{j,k} \neq \hat{\beta}_{j,l} \text{ für } k \neq l \quad (5.4)$$

erlaubt direkte Vergleiche der Schätzungen in 2 Gruppen. Die Nullhypothese ist wiederum die Übereinstimmung der geschätzten Parameter in beiden Gruppen. Die H_0 ist abzulehnen, falls die Teststatistik im Test für eine Aufgabe

$$T = \text{sign}(\hat{\beta}_{j,1} - \hat{\beta}_{j,2}) \cdot \sqrt{W_j} \sim N(0, 1) \quad (5.5)$$

mit

$$W_j = \frac{(\hat{\beta}_{j,1} - \hat{\beta}_{j,2})^2}{\hat{\sigma}_{j,1}^2 + \hat{\sigma}_{j,2}^2}$$

bzw. Teststatistik im Test für alle Aufgaben gleichzeitig

$$T = (\hat{\beta}_1 - \hat{\beta}_2)'(\hat{\Sigma}_1 - \hat{\Sigma}_2)^{-1}(\hat{\beta}_1 - \hat{\beta}_2) \sim \chi^2(m - 1) \quad (5.6)$$

größer als $z_{1-\alpha}$ -Quantil der Normal- bzw. χ^2 -Verteilung ist. (Vgl. Strobl (2010), S. 44-45)

Praktisch in R wird der Test mithilfe der Funktion `Waldtest` R-Paket "eRm" (Mair und Hatzinger (2007)) durchgeführt. Der Wald-Test wurde wiederum

für die Leistung und für das Geschlecht durchgeführt. Der Test liefert gleiche Ergebnisse (Tabelle 5.2) wie der Grafische- und der LQ-Test. Die Aufgaben 7 und 13 für die Leistung und die Aufgaben 1, 2, 3 und 4 für Geschlecht weisen DIF auf. Diese müssen aus dem Test ausgeschlossen bzw. ersetzt werden.

Leistung hoch vs. niedrig			Männer vs. Frauen		
Wald test on item level (z-values):			Wald test on item level (z-values):		
	z-statistic	p-value		z-statistic	p-value
beta Aufg. 1	-0.512	0.609	beta Aufg. 1	8.643	0.000
beta Aufg. 2	0.677	0.499	beta Aufg. 2	-2.039	0.041
beta Aufg. 3	0.780	0.435	beta Aufg. 3	-1.991	0.046
beta Aufg. 4	-1.186	0.236	beta Aufg. 4	2.139	0.032
beta Aufg. 5	0.573	0.567	beta Aufg. 5	-0.354	0.723
beta Aufg. 6	-1.667	0.095	beta Aufg. 6	-1.049	0.294
beta Aufg. 7	3.570	0.000	beta Aufg. 7	-0.638	0.523
beta Aufg. 8	-0.504	0.614	beta Aufg. 8	-0.420	0.675
beta Aufg. 9	-0.262	0.794	beta Aufg. 9	-0.351	0.725
beta Aufg. 10	0.599	0.549	beta Aufg. 10	-1.672	0.095
beta Aufg. 11	-0.044	0.965	beta Aufg. 11	-1.260	0.208
beta Aufg. 12	-1.114	0.265	beta Aufg. 12	-0.590	0.555
beta Aufg. 13	-1.986	0.047	beta Aufg. 13	0.076	0.939
beta Aufg. 14	1.107	0.268	beta Aufg. 14	-1.041	0.298
beta Aufg. 15	0.640	0.522	beta Aufg. 15	-1.041	0.298

Tabelle 5.2: Wald-Test

Kapitel 6

Zusammenfassung

Das Rasch-Modell und Verallgemeinerungen spielen grundlegende Rolle in der *Item-Response-Theorie*. In der Testentwicklung mithilfe des Rasch-Modells kann man die für ein Test nicht geeignete Aufgaben bestimmen. Unter nicht geeigneten Aufgaben versteht man auch diejenige Aufgaben, die DIF aufweisen. Die bekanntesten Tests dafür sind der grafische Modell-Test, LQ-Test, Wald-Test sowie χ^2 -Anpassungstest, der als weitere Fragestellung für DIF-Analysen betrachtet werden kann. Das Grundmodell ist allerdings nicht immer für die Analysen sowie für die Datenbeschreibung gut geeignet. In dem man das Rasch-Modell mit den neuen Annahmen erweitert oder auf schon existierte Forderungen, Eigenschaften und Annahmen verzichtet, ergeben sich die verallgemeinerten Modelle. Diese Verallgemeinerungen sind manchmal besser für die Analysen von durchgeführten Tests. Für weitere Analysen wäre es wichtig weitere Verallgemeinerungen des Rasch-Modells detailliert betrachten zu können sowie anhand Simulationen von verallgemeinerten Modellen weitere Analysen durchzuführen. Von besonderer Interesse stehen die unterschiedliche Methoden, wie z.B. Maentel-Haenszel-Verfahren, mit deren Hilfe DIF-Aufgaben bestimmt werden können.

Anhang A

Parameterschätzung

A.1 Herleitung Formel 2.5

Likelihood-Funktion für **alle** Aufgaben **einer** Person:

$$L_{u_i}(\theta_i, \beta) =$$

$$P(U_{i,1} = u_{i,1}, \dots, U_{i,m} = u_{i,m} | \theta_i, \beta_1, \dots, \beta_m) =$$

$$P(U_i = u_i | \theta_i, \beta) =$$

$$\prod_{j=1}^m P(U_{i,j} = u_{i,j} | \theta_i, \beta_j) =$$

$$\prod_{j=1}^m \frac{e^{u_{i,j}(\theta_i - \beta_j)}}{1 + e^{\theta_i - \beta_j}} =$$

$$\frac{e^{r_i \theta_i - \sum_{j=1}^m u_{i,j} \beta_j}}{\prod_{j=1}^m (1 + e^{\theta_i - \beta_j})}$$

A.2 Herleitung Formel 2.6

Likelihood-Funktion für **alle** Aufgaben und **alle** Personen:

$$L_u(\theta, \beta) =$$

$$P(u|\theta, \beta) =$$

$$P(U_1 = u_1, \dots, U_n = u_n | \theta_1, \dots, \theta_n, \beta) =$$

$$P(U_i = u_i | \theta_i, \beta) =$$

$$\prod_{i=1}^n \prod_{j=1}^m P(U_{i,j} = u_{i,j} | \theta_i, \beta_j) =$$

$$\prod_{i=1}^n \frac{e^{r_i \theta_i - \sum_{j=1}^m u_{i,j} \beta_j}}{\prod_{j=1}^m (1 + e^{\theta_i - \beta_j})} =$$

$$\frac{e^{\sum_{i=1}^n r_i \theta_i - \sum_{j=1}^m s_j \beta_j}}{\prod_{i=1}^n \prod_{j=1}^m (1 + e^{\theta_i - \beta_j})}$$

A.3 Herleitung Formel 2.7

Die Likelihood für eine Person und **alle** Aufgaben:

$$L_{u_i}(\theta_i, \beta) = P(u_i | \theta_i, \beta) = \frac{e^{r_i \theta_i - \sum_{j=1}^m u_{i,j} \beta_j}}{\prod_{j=1}^m (1 + e^{\theta_i - \beta_j})}$$

Die Likelihood für eine Person und **alle** Aufgaben wegen Unabhängigkeit kann man als Produkt von 2 Funktionen darstellen:

$$L_{u_i}(\theta_i, \beta) = g(r_i|\theta_i, \beta) \cdot h(u_i|r_i, \theta_i, \beta)$$

\Leftrightarrow

$$h(u_i|r_i, \theta_i, \beta) = \frac{L_{u_i}(\theta_i, \beta)}{g(r_i|\theta_i, \beta)}$$

wobei

$h(u_i|r_i, \theta_i, \beta)$ - bedingte auf r_i Likelihood für i-te Person

$g(r_i|\theta_i, \beta)$ - Wahrscheinlichkeit, eine bestimmte r_i zu beobachten

Wahrscheinlichkeit für die bestimmte Zeilenrandsumme:

$$g(r_i|\theta_i, \beta) =$$

$$\sum_{\sum_j u_{i,j}=r_i} P(u_i|\theta_i, \beta) =$$

$$\sum_{\sum_j u_{i,j}=r_i} \frac{e^{r_i\theta_i - \sum_{j=1}^m u_{i,j}\beta_j}}{\prod_{j=1}^m (1 + e^{\theta_i - \beta_j})} =$$

$$\frac{e^{r_i\theta_i}}{\prod_{j=1}^m (1 + e^{\theta_i - \beta_j})} \cdot \sum_{\sum_j u_{i,j}=r_i} e^{-\sum_{j=1}^m u_{i,j}\beta_j} =$$

$$\frac{e^{r_i\theta_i}}{\prod_{j=1}^m (1 + e^{\theta_i - \beta_j})} \cdot \gamma_{r_i}(\beta)$$

Bedingte Likelihood für i-te Person:

$$h(u_i|r_i, \theta_i, \beta) =$$

$$\frac{L_{u_i}(\theta_i, \beta)}{g(r_i|\theta_i, \beta)} =$$

$$\frac{e^{r_i\theta_i - \sum_{j=1}^m u_{i,j}\beta_j}}{\prod_{j=1}^m (1 + e^{\theta_i - \beta_j})} \cdot \frac{e^{r_i\theta_i} \gamma_{r_i}(\beta)}{\prod_{j=1}^m (1 + e^{\theta_i - \beta_j})} =$$

$$\frac{e^{r_i\theta_i - \sum_{j=1}^m u_{i,j}\beta_j}}{\prod_{j=1}^m (1 + e^{\theta_i - \beta_j})} \cdot \frac{\prod_{j=1}^m (1 + e^{\theta_i - \beta_j})}{e^{r_i\theta_i} \gamma_{r_i}(\beta)} =$$

$$\frac{e^{-\sum_{j=1}^m u_{i,j}\beta_j}}{\gamma_{r_i}(\beta)} = h(u_i|r_i, \beta)$$

Bedingte Likelihood für gesamte Daten:

$$h(u|r, \beta) =$$

$$\prod_{i=1}^n h(u_i|r_i, \beta) =$$

$$\prod_{i=1}^n \frac{e^{-\sum_{j=1}^m u_{i,j}\beta_j}}{\gamma_{r_i}(\beta)} =$$

$$\frac{e^{-\sum_{i=1}^n \sum_{j=1}^m u_{i,j}\beta_j}}{\prod_{i=1}^n \gamma_{r_i}(\beta)} =$$

$$\frac{e^{-\sum_{j=1}^m s_j\beta_j}}{\prod_{i=1}^n \gamma_{r_i}(\beta)}$$

A.4 Herleitung Formel 2.8

Multiplikation der Likelihood mit der Randdichte:

$$L_u(\theta, \beta) \cdot f(\theta) = P(u|\theta, \beta) \cdot f(\theta) = P(u, \theta|\beta)$$

Marginale Likelihood für die Aufgabenparameter:

$$L_u(\beta) = \int P(u, \theta|\beta) \partial\theta$$

Literaturverzeichnis

- Fahrmeir, L., Alfred, H. und Tutz, G. (1996). *Multivariate statistische Verfahren*, 2. überarb. edn, de Gruyter, Berlin.
- Fahrmeir, L. und Fahrmeir-Künstler-Pigeot-Tutz (2011). *Statistik: Der Weg zur Datenanalyse*, korrigierter nachdruck 7. edn, Springer, Berlin.
- Fahrmeir, L., Kneib, T. und Lang, S. (2009). *Regression: Modelle, Methoden und Anwendungen*, 2. edn, Springer-Verlag Berlin Heidelberg, Berlin, Heidelberg.
- Frank, E. und Harrell, J. (2015). *Hmisc: Harrell Miscellaneous*.
URL: <http://CRAN.R-project.org/package=Hmisc>
- Henning, H. J. (1974). *Skalenanalyse und RASCH-Modell: Univ., Diss.–Bonn, 1975*.
- Kubinger, K. D., Rasch, D. und Yanagida, T. (2011). *Statistik in der Psychologie: Vom Einführungskurs bis zur Dissertation ; [Lehrbuch]*, Hogrefe, Göttingen.
- Mair, P. und Hatzinger, R. (2007). *Extended Rasch modeling: The eRm package for the application of IRT models*.
URL: <http://www.jstatsoft.org/v20/i09>
- Moosbrugger, H., Kelava, A. und Moosbrugger-Kelava (2012). *Testtheorie und Fragebogenkonstruktion: Mit 66 Abbildungen und 41 Tabellen*, 2. aktualisierte und überarb. edn, Springer, Berlin.
- Prenzel, M., Sälzer, C., Klieme, E. und Köller, O. (2013). *PISA 2012: Fortschritte und Herausforderungen in Deutschland*, Waxmann, Münster.
- R Core Team (2014). *R: A Language and Environment for Statistical Computing*.
URL: <http://www.R-project.org/>

Strobl, C. (2010). *Das Rasch-Modell: Eine verständliche Einführung für Studium und Praxis*, 1. edn, Hampp, München.

Toutenburg, H., Heumann, C., Nittner, T. und Scheid, S. (2003). *Lineare Modelle: Theorie und Anwendungen*, 2. neu bearb. und erw. edn, Physica-Verlag, Heidelberg.