Eine Einführung in die nichtparametrische Item-Response-Theorie, mit Fokus auf stochastischen Ordnungseigenschaften

Ali Ünlü

Statistische Methoden in den Sozial- und Bildungswissenschaften
Fakultät Statistik
Technische Universität Dortmund

Workshop über Psychometric Computing
Institut für Statistik
Ludwig-Maximilians-Universität München

25. Februar 2010

Contents

Überblick

Nichtparametrische Item-Response-Theorie

Wissensraumtheorie

Zusammenfassung

Contents

Überblick

Nichtparametrische Item-Response-Theorie

Wissensraumtheorie

Zusammenfassung

Die Hauptbeiträge im Überblick

Dieser Vortrag ist über nichtparametrische Item-Response-Theorie (NIRT; erster Teil) und Wissensraumtheorie (WRT; zweiter Teil) in den quantitativen Verhaltenswissenschaften.

Es werden die wichtigsten Grundkonzepte dieser Theorien unter Berücksichtigung der Historie eingeführt.

Ein kurzer Ausblick auf Verallgemeinerungen der WRT in nichtparametrischer Weise rundet den Vortrag ab.

Das Hauptanliegen ist, einen Beitrag zur Vereinheitlichung der bis dato getrennt vorliegenden Testtheorien der Wissensräume und Item-Response-Theorie zu leisten.

Contents

Überblick

Nichtparametrische Item-Response-Theorie

Wissensraumtheorie

Zusammenfassung

technische universität dortmund

Kurt Lewins (1890–1947) Definition von Psychologie

Als Ausgangspunkt für die nachfolgenden Betrachtungen, Lewins (1951) Definition von Psychologie:

Psychologie ist die Wissenschaft, welche Beziehungen (f) zwischen Verhalten (B) einerseits und Person- (P) und Situationscharakteristika (S) andererseits studiert:

$$B = f(P, S).$$

Lewin, K., 1951, Field theory in social science; selected theoretical papers, D. Cartwright (Ed.), New York: Harper & Row.

Latente Konstrukte und qualitative Diskretheit — zwei zentrale Unterschiede der Psychologie zur Physik

Person- und Situationscharacteristika sind in der Regel nicht beobachtbar, d.h. latent. Psychologische/mentale Konstrukte, wie z.B. Intelligenz, induktives Denken, Fähigkeit im Umgang mit arithmetischen Rechenoperationen und Wissenszustand, können nicht direkt gemessen werden. Sie sind Abstraktionen, hypothetisch, und nur indirekt erfassbar, über Instrumente wie psychometrische Tests.

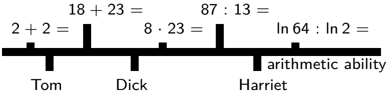
Psychologische Skalen sind sehr oft von qualitativer, diskreter Natur, im Vergleich zu numerischen, stetigen Standardskalen wie Länge, Zeit, Masse und Energie in der Physik.

Die klassische Physik des 19. Jahrhunderts hatte einen prägenden und wegen der genannten Unterschiede oftmals negativen Einfluss auf die Entwicklung der Psychologie/Psychometrie.

Beispiel: Louis Guttmans (1916–1987) Skala

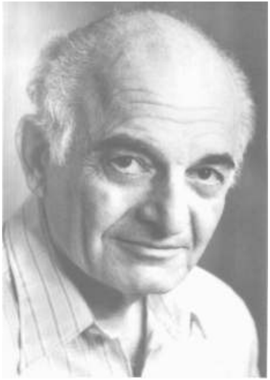
Historischer Vorgänger der NIRT und WRT

Jeder Proband und jedes Item (das psychologische Konstrukt messend) haben bestimmte Positionen auf dem latenten Kontinuum der möglichen Personfähigkeits- bzw. Itemschwierigkeitswerte.



Lineare Anordnung der Personen nach Fähigkeit, der Items nach Schwierigkeit:

$$\theta_T < \theta_D < \theta_H,$$
$$\delta_1 < \delta_2 < \delta_3 < \delta_4 < \delta_5.$$



Guttman, L., 1944, A basis for scaling qualitative data, American Sociological Review, 9, 139–150.

Beispiel: Guttman-Skala — Messeigenschaften, I

Deterministische Guttman-Regel:
Eine richtige Antwort tritt genau dann ein wenn die Personfähigkeit (θ-Wert) mindestens so groß wie die Itemschwierigkeit (δ-Wert) ist.

Zentrale Messeigenschaften der Guttman-Skala:

- Jede Person, die ein schweres Item löst, löst auch ein leichteres Item.
- Eine fähige Person löst alle Items, die eine weniger fähige Person löst.
- Aus dem Totalscore einer Person können eindeutig ihre individuellen Antworten zu allen Items des Tests rekonstruiert werden; die sogenannte Eigenschaft der *Reproduzierbarkeit*.

Beispiel: Guttman-Skala — Messeigenschaften, II

- Für Personen mit verschiedenen Totalscore-Werten gilt, dass deren latente Anordnung entlang des Fähigkeitskontinuums die gleiche ist wie ihre manifeste Anordnung bezüglich des Totalscores. Z.B.:

$$\theta_T < \theta_D < \theta_H,$$

$$Total_T = 1 < Total_D = 2 < Total_H = 4.$$

- Für Items mit verschiedenen Anteilen "Richtig in der Stichprobe" gilt, dass deren latente Anordnung entlang des Schwierigkeitskontinuums umgekehrt gleich ihrer manifesten Anordnung bezüglich des Anteils "Richtig in der Stichprobe" ist. Z.B.:

$$\delta_1 < \delta_2 < \delta_3, \delta_4 < \delta_5,$$

$$p_1 = 1 > p_2 = 2/3 > p_3, p_4 = 1/3 > p_5 = 0.$$

Beispiel: Guttman-Skala — Item-Response-Funktion, I

Generelle Frage der Item-Response-Modellierung:
Wie variiert das Antwortverhalten zu einem Item mit variierenden Fähigkeitswerten der Personen; *Item-Response-Funktion*?

Eingekleidet in Lewins Definition, $B = f(P, S)$: Welche Beziehung f besteht zwischen Verhalten einerseits

$$B \quad : \quad \text{Antwort (0 oder 1) zu Item '}2 + 2 =\text{'}$$

und Personcharakteristika

$$P \quad : \quad \text{arithmetische Fähigkeit } \theta$$

und Situationscharakteristika

$$S \quad : \quad \text{Itemschwierigkeit } \delta_1$$

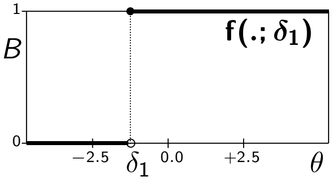
andererseits, unter dem Guttman-Modell?

Beispiel: Guttman-Skala — Item-Response-Funktion, II

Die Guttman-Regel ergibt:

$$\begin{aligned} B &= f(P, S) \\ &= f(\theta; \delta_1) \\ &= \begin{cases} 0 & : \text{ falls } \theta < \delta_1 \\ 1 & : \text{ falls } \theta \geq \delta_1 \end{cases} \end{aligned}$$

$f(.; \delta_1)$ als Funktion der Personfähigkeitswerte θ mit Antwortwerten $B \in \{0, 1\}$ heisst *Item-Response-Funktion* (IRF) von Item '$2 + 2 =$':



Beispiel: Guttman-Skala — Item-Response-Funktion, III

Falls $\mathbb{P}(I_1 = 1|\theta)$ die bedingte Wahrscheinlichkeit notiert, dass eine Person mit Fähigkeitswert θ das Item I_1, '$2 + 2 =$', löst, so gilt

$$\mathbb{P}(I_1 = 1|\theta) = \begin{cases} 0 & : \quad \text{falls } \theta < \delta_1 \\ 1 & : \quad \text{falls } \theta \geq \delta_1 \end{cases}$$

Die IRF $f(.;\delta_1)$ ist somit die bedingte positive Antwortwahrscheinlichkeit $\mathbb{P}(I_1 = 1|\theta)$, interpretiert als eine Funktion der Fähigkeitswerte θ.

Die IRFs der Guttman-Skala sind Treppenfunktionen. Ein Item mit solch einer IRF heisst *Guttman-Item* oder *G-Item*, und es wird auch als *perfektes Item* bezeichnet.

Robert J. Mokken (geb. 1929) — NIRT

Eindimensionalität beibehaltende, probabilistische Erweiterung der Guttman-Skala

NIRT geht davon aus, dass der Zusammenhang zwischen $\mathbb{P}(I_j = 1|\theta)$ und θ, also die IRF, durch Ordnungsrestriktionen beschrieben ist. Im Gegensatz zur PIRT werden in der NIRT keine expliziten parametrischen funktionalen Formen für die IRFs (z.B. logistisch oder normal) vorausgesetzt.

Restriktionen nach Mokken: Monotonie und invariante Item-ordnung.

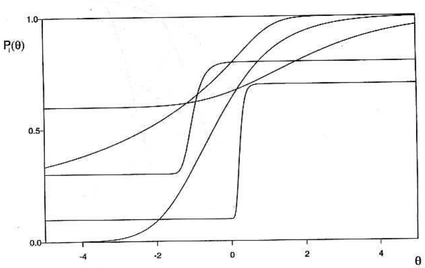
Mokken, R.J., 1971, A theory and procedure of scale analysis, The Hague, Mouton/Berlin: De Gruyter.

NIRT — Monotonie und Monotone-Homogeneity-Modell

Die Annahme der *Monotonie* besagt, dass jede IRF eines Tests $\{I_j : 1 \leq j \leq k\}$ eine monotone Funktion von $\theta \in \Theta$ ist. Für alle $\theta_1, \theta_2 \in \Theta$, $\theta_1 < \theta_2$ gilt

$$\mathbb{P}(I_j = 1|\theta_1) \leq \mathbb{P}(I_j = 1|\theta_2).$$

Erfüllen die IRFs zusätzlich noch die klassischen Annahmen der Eindimensionalität, $\theta \in \Theta \subset \mathbb{R}$, und lokalen Unabhängigkeit, $\mathbb{P}(\prod I_j = i_j|\theta) = \prod \mathbb{P}(I_j = i_j|\theta)$, so spricht man vom *Monotone-Homogeneity-Modell* (MHM).



Beispiel-IRFs die die Annahme der Monotonie erfüllen.

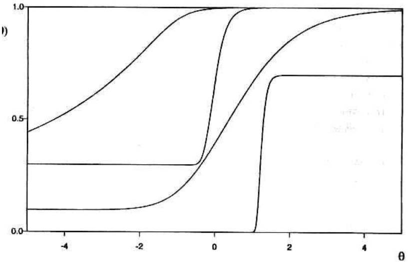
Das MHM ist ein Messmodell für Personen. Es beschreibt dichotome Antwortdaten die von *homogenen* (eindimensionalen) Items herrühren, deren IRFs in *monotoner* Beziehung zum latenten Kontinuum der Fähigkeitswerte stehen.

NIRT — Invariante Itemordnung und Double-Monotonicity-Modell

Die Annahme der *invarianten Itemordnung* besagt, dass die IRFs eines Tests $\{I_j : 1 \leq j \leq k\}$ global über das Kontinuum hinweg sich nicht durchkreuzen, d.h. in Schwierigkeit einheitlich angeordnet werden können. Für alle $\theta \in \Theta$ gilt

$$\mathbb{P}(I_1 = 1|\theta) \leq \cdots \leq \mathbb{P}(I_k = 1|\theta).$$

Erfüllen die IRFs zudem die Annahmen der Eindimensionalität, lokalen Unabhängigkeit und Monotonie, so spricht man vom *Double-Monotonicity-Modell* (DMM).

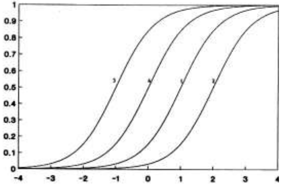


Beispiel-IRFs die die Annahme der invarianten Itemordnung erfüllen.

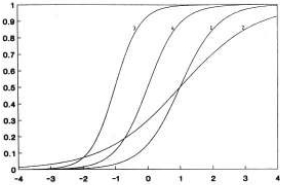
Das DMM ist ein Messmodell für Personen und Items. Neben der *Monotonie der IRFs* wird im DMM eine weitere Monotonie-Annahme, eine *Monotonie unter den Items*, getroffen.

Parametrische Item-Response-Theorie

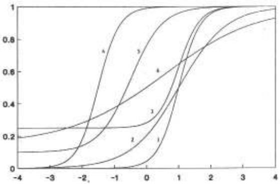
Beispiel: 1-, 2- und 3-Parameter logistische Modelle (1PLM, 2PLM und 3PLM)



Beispiel-IRFs 1PLM (Rasch)



Beispiel-IRFs 2PLM (Birnbaum)



Beispiel-IRFs 3PLM (Birnbaum)

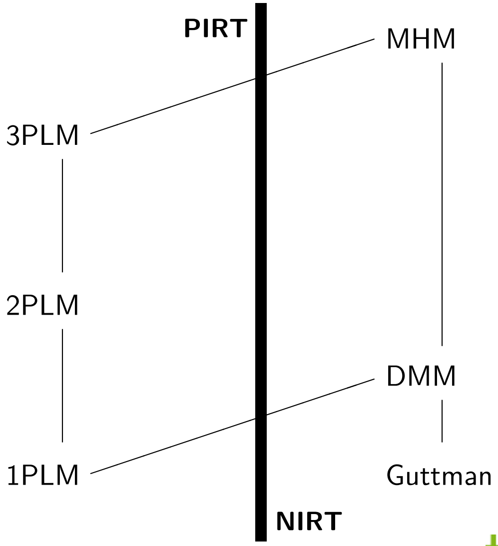
$$\mathbb{P}_{I_j=1|\theta} = \frac{e^{(\theta-\delta_j)}}{1+e^{(\theta-\delta_j)}}$$

$$\mathbb{P}_{I_j=1|\theta} = \frac{e^{\alpha_j(\theta-\delta_j)}}{1+e^{\alpha_j(\theta-\delta_j)}}$$

$$\mathbb{P}_{I_j=1|\theta} = \gamma_j+(1-\gamma_j)\cdot\frac{e^{\alpha_j(\theta-\delta_j)}}{1+e^{\alpha_j(\theta-\delta_j)}}$$

Im Gegensatz zu solchen parametrischen Modellen werden in der NIRT keine *expliziten* parametrischen funktionalen Formen für die IRFs, und somit auch keine *expliziten* Itemparameter wie z.B. Schwierigkeit, Trennschärfe oder Ratewahrscheinlichkeit, vorausgesetzt. NIRT-Modelle liefern in der Regel keine numerischen Punktschätzungen expliziter Modellgrößen. Man erhält vielmehr Messungen auf Ordinalskalenniveau.

Modellhierarchie populärer IRT-Modelle



MHM — Messeigenschaften

Personenskalierung unter dem MHM

Das folgende Theorem ist *das* Schlüsselresultat, welches die Anwendung des MHM als Messmodell für Personen begründet. Die Totalscore-Variable kann zur ordinalen Messung, d.h. linearen Ordnung, der Personen entlang des latenten Kontinuums herangezogen werden.

Theorem. Angenommen das MHM gelte. Sei $c \in \Theta$ ein fixer Latent-Trait-Punkt. Dann gilt für beliebige zwei Werte $s, t \in \mathbb{N}_0$ mit $0 \leq s < t \leq k$ der Totalscore-Variablen $X_+ := \sum_{j=1}^{k} I_j$

$$\mathbb{P}(\theta > c | X_+ = s) \leq \mathbb{P}(\theta > c | X_+ = t).$$

Insbesondere gilt dann für beliebige $s, t \in \mathbb{N}_0$ mit $0 \leq s < t \leq k$

$$\mathbb{E}(\theta | X_+ = s) \leq \mathbb{E}(\theta | X_+ = t).$$

DMM — Messeigenschaften

Itemskalierung unter dem DMM

Das folgende Theorem ist *das* Schlüsselresultat, welches die Anwendung des DMM als Messmodell für Items begründet. Die Populationspopularität (Anteil "Richtig in der Population"), bzw. die Stichprobenpopularität (Anteil "Richtig in der Stichprobe") die diese schätzt, kann zur ordinalen Messung, d.h. linearen Ordnung, der Items bzgl. ihrer invarianten Itemordnung entlang des latenten Kontinuums herangezogen werden.

Theorem. Angenommen das DMM gelte. Für beliebige zwei Items I_{j_1} und I_{j_2} $(1 \leq j_1, j_2 \leq k)$ sind die folgenden Aussagen äquivalent:

1. $\mathbb{P}(I_{j_1} = 1) \leq \mathbb{P}(I_{j_2} = 1)$;
2. $\forall \theta \in \Theta : \mathbb{E}(I_{j_1}|\theta) \leq \mathbb{E}(I_{j_2}|\theta)$;
3. $\forall \theta \in \Theta : \mathbb{P}(I_{j_1} = 1|\theta) \leq \mathbb{P}(I_{j_2} = 1|\theta)$.

Contents

Überblick

Nichtparametrische Item-Response-Theorie

Wissensraumtheorie

Zusammenfassung

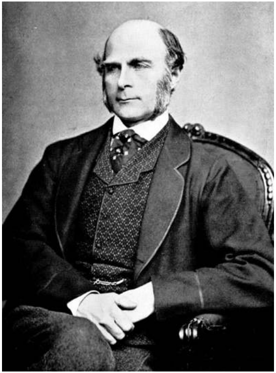
Einfluss der Physik auf die Entwicklung der Psychometrie

Philosophie und Ursprung der modernen Psychometrie, der IRT, gehen zurück auf die klassische Physik des 19. Jahrhunderts, deren Methoden als musterhaft erachtet wurden. In Anlehnung an die Physik mit ihren fundamentalen numerischen Standardskalen wie Länge, Zeit oder Masse werden in der Psychometrie Personen und Items durch reelle Zahlen bzw. Vektoren von reellen Zahlen, die z.B. für Personfähigkeit und Itemschwierigkeit stehen, dargestellt.

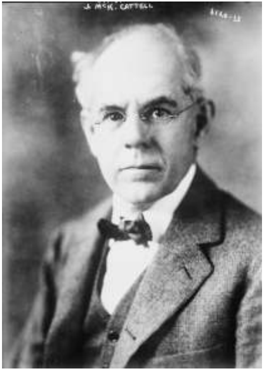
Die Wissensraumtheorie, als eine Disziplin der mathematischen Psychologie, bricht mit dieser historisch erwachsenen Tradition.

Messung im Sinne numerischer Skalen ähnlich den in der Physik als *sine qua non* für Präzision und 'Wissenschaft-sein' war/ist für eine differenzierte Behandlung von Wissenstestung und Wissensvermittlung eher hinderlich als fördernd. Präzision in der Wissensraumtheorie versteht sich als Feingliedrigkeit der repräsentierenden kombinatorischen Strukturen.

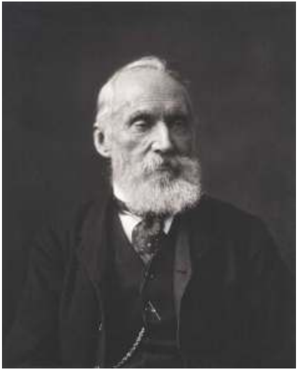
Prominente Advokaten dieser Sichtweise



Sir Francis Galton (1822–1911)
Pearson, K., 1924, The life, letters and labours of Francis Galton, London: Cambridge University Press.



James McKeen Cattell (1860–1944)
Cattell, J.M., 1886, Psychometrische Untersuchungen, Philosophische Studien, 3, 305–335, 452–492.



William Thomson Kelvin (1824–1907)
Kelvin, W.T., 1889, Popular lectures and addresses, Vol. 1: Constitution of matter, Chapter: Electrical units of measurement, London: MacMillan.

Motivation: Grenzen einer rein numerischen Zugangsweise

Angenommen es wird ein 'Athlet-Quotient (AQ)' Test entworfen, der athletisches Können erfassen soll. Drei ausgezeichnete Athleten Michael Jordan, Tiger Woods und Pete Sampras machen den Test und erzielen überragende AQ-Werte. Diese hohen Werte, die die Athleten in dieser Situation gleichstellen, geben ein verzerrtes Bild von der wesentlichen Unterschiedlichkeit der Athleten.

Auch ein multidimensionaler, rein numerischer Ansatz scheitert. Hinzunahme weiterer numerischer Dimensionen zur Trennung der drei Athleten ist nur ein erster Schritt in einer Folge weiterer. Miteinbeziehen von Athleten wie Muhammad Ali und Pele würde weitere Dimensionen erforderlich machen. Zudem, aus methodischer und interpretativer Sicht, sollte die Anzahl der numerischen Dimensionen eher klein sein, so dass auch numerische Vektoren wohl eher schwach aussagekräftig wären.

Jean-Claude Falmagne — WRT

Mehrdimensionalität erlaubende, ordnungstheoretische Erweiterung der Guttman-Skala

Natürlicher Ausgangspunkt für eine Theorie Wissenstestung und -erwerb: Wisseneinheiten implizieren andere Wisseneinheiten. Z.B. kann das Beherrschen einer Algebra-Aufgabe das Beherrschen anderer Algebra-Aufgaben nach sich ziehen.

Solche Abhängigkeiten zwischen Wisseneinheiten können eingesetzt werden, um effiziente adaptive, computergestützte Prozeduren zur Wissensdiagnose und -vermittlung zu entwickeln. Beispielsystem '*A*ssessment and *LE*arning in *K*nowledge *S*paces' (ALEKS), ein vollautomatisierter Mathematik-Tutor im Internet.

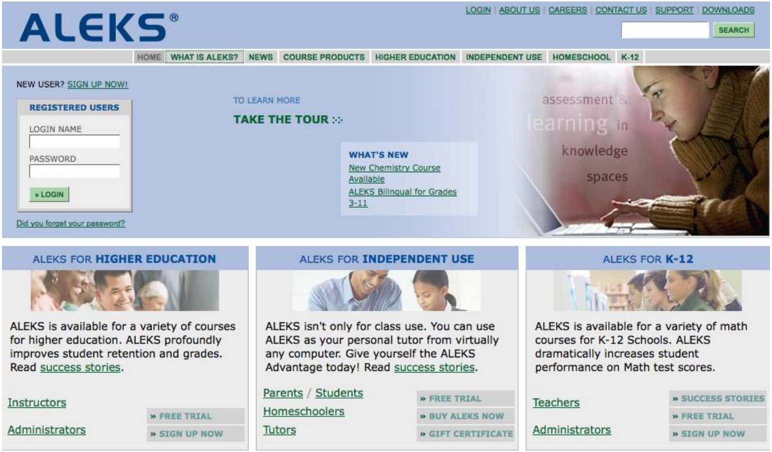
Doignon, J.-P., Falmagne, J.-Cl., 1999, Knowledge spaces, Springer.

Falmagne, J.-Cl., Doignon, J.-P., 2010, Learning spaces, Springer.

Albert, D., Doble, C., Eppstein, D., Falmagne, J.-Cl., Hu, X., 2010, Knowledge spaces: Applications to education, Springer.

ALEKS-System

www.aleks.com



Beispiel-Wissensbereich: Elementary Algebra

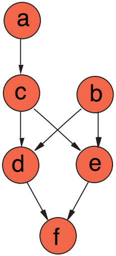
Eine Menge von sechs Algebra-Aufgaben:

a. A car travels on the freeway at an average speed of 52 miles per hour. How many miles does it travel in 5 hours and 30 minutes?
b. Using the pencil, mark the point at the coordinates $(1, 3)$.
c. Perform the multiplication $4x^4y^4 \cdot 2x \cdot 5y^2$ and simplify your answer as much as possible.
d. Find the greatest common factor of the expressions $14t^6y$ and $4tu^5y^8$. Simplify your answer as much as possible.
e. Graph the line with slope -7 passing through the point $(-3, -2)$.
f. Write an equation for the line that passes through point $(-5, 3)$ and is perpendicular to line $8x + 5y = 11$.

Voraussetzungsrelation zwischen den Problemen, I

Elementary Algebra (6 Knoten/Probleme)

Ein plausibles Voraussetzungsdiagramm der Beherrschbarkeitsabhängigkeiten für diese Algebra-Aufgaben:

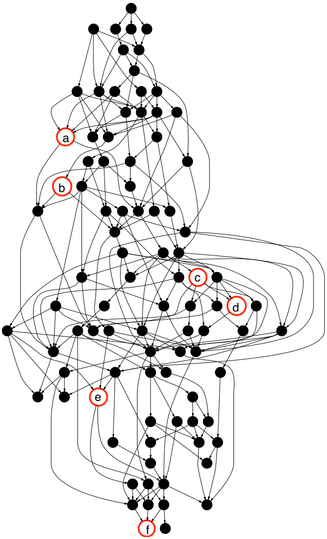


Z.B.: Das Beherrschen von Problem b ist eine Voraussetzung für das Beherrschen von Problem e. Mit anderen Worten: Beherrscht man Problem e, so beherrscht man auch notwendigerweise Problem b.

Voraussetzungsrelation zwischen den Problemen, II

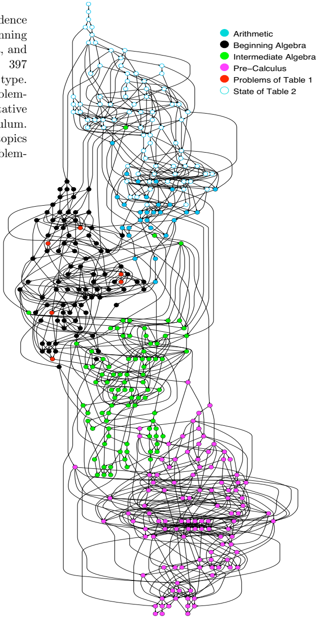
Beginning Algebra (88 Knoten/Probleme)

Vorheriges Diagramm ist Teil eines größeren:



Voraussetzungsrelation zwischen den Problemen, III

Figure 3. Combined precedence diagram for Arithmetic, Beginning Algebra, Intermediate Algebra, and Pre-Calculus. Each of the 397 points represents a problem type. We recall that these 397 problem-types capture only a representative part of the standard curriculum. The full curriculum in these topics would contain around 650 problem-types.



Wissenszustand und Wissensstruktur

Das Voraussetzungsdiagramm für die sechs Algebra-Aufgaben spezifiziert vollständig die zulässigen/möglichen *Wissenszustände*. Falls ein Prüfling z.B. Problem e beherrscht, so muss sie/er auch die Probleme a, b und c beherrschen. Daher besitzt der Prüfling den Wissenszustand $\{a, b, c, e\}$.

Auf diese Weise gibt es insgesamt zehn mögliche Wissenszustände, die mit dem Voraussetzungsdiagramm verträglich sind:

$$\begin{aligned}\mathcal{K} \quad = \quad & \{\emptyset, \{a\}, \{b\}, \{a, b\}, \{a, c\}, \{a, b, c\}, \\ & \{a, b, c, d\}, \{a, b, c, e\}, \{a, b, c, d, e\}, \{a, b, c, d, e, f\}\}.\end{aligned}$$

Diese Menge $\mathcal{K}$ aller möglichen Wissenszustände heisst *Wissensstruktur*.

technische universität dortmund

Grundbegriffe — mathematische Definitionen

Definition. Eine *Wissensstruktur* ist ein Paar $(Q, \mathcal{K})$ bestehend aus einer Itemmenge Q und einer Familie $\mathcal{K}$ von Teilmengen von Q mit $\emptyset, Q \in \mathcal{K}$. Die Elemente $K \in \mathcal{K}$ heißen *Wissenszustände*.

Definition. Jede Quasiordnung, d.h. reflexive und transitive binäre Relation, auf Q heißt *Vermutungsrelation*.

Eine Wissensstruktur ist ein deterministisches Modell auf der Personenebene. Eine Vermutungsrelation ist ein deterministisches Modell auf der Itemebene. Diese zwei Formulierungen sind mehr oder weniger mathematisch äquivalent (Birkhoff, 1937).

Birkhoff, G., 1937, Rings of sets, Duke Mathematical Journal, 3, 443–454.

WRT und NIRT — Annahmen, I

In eigenen Arbeiten wurden die Annahmen und stochastischen Ordnungseigenschaften der eindimensionalen Mokken-Modelle auf allgemeine quasi-geordnete Personen- und Itemräume erweitert und Anwendungen dieses verallgemeinerten Mokkenschen Ansatzes in der WRT beschrieben.

NIRT	WRT
Eindimensionalität $\Theta \subseteq \mathbb{R}$	Diskr. Dimensionalität $\mathcal{K} \subseteq 2^Q$
IRF $\mathbb{P}(I_l = 1 \mid \theta) =: \mathbb{P}_{I_l}(\theta)$	IRF $\mathbb{P}(I_l = 1 \mid K) =: \mathbb{P}_{I_l}(K)$
Lokal Unabhängigkeit in θ	Lokal Unabhängigkeit in K
Monotonie: IRFs sind monotone Funktionen in θ.	Isotonie: IRFs sind isotone Funktionen in K.

Ünlü, A., 2007, Nonparametric item response theory axioms and properties under nonlinearity and their exemplification with knowledge space theory, Journal of Mathematical Psychology, 51, 383–400.

Ünlü, A., 2008, A note on monotone likelihood ratio of the total score variable in unidimensional item response theory, British Journal of Mathematical and Statistical Psychology, 61, 179–187.

WRT und NIRT — Annahmen, II

NIRT

Die Itemmenge Q besitzt eine invariante Itemordnung, wenn gilt

$$\forall \theta \in \Theta : \mathbb{P}_{I_1}(\theta) \leq \cdots \leq \mathbb{P}_{I_k}(\theta).$$

WRT

Sei $\mathcal{S}$ eine Vermutungsrelation auf Q, und sei $\mathcal{K}_{\mathcal{S}}$ die zugehörige Wissensstruktur. Die Itemmenge Q erlaubt eine *invariante Itemordnung*, wenn für alle $(I_1, I_2) \in \mathcal{S}$ gilt

$$\forall K \in \mathcal{K}_{\mathcal{S}} : \mathbb{P}_{I_2}(K) \leq \mathbb{P}_{I_1}(K).$$

Es ergeben sich MHM und DMM entsprechende Mengen von allgemeinen nichtparametrischen Axiomen in der WRT. In weiterer Forschung ist diese Mokken-artige Formulierung der WRT zu untersuchen.

Stochastische Ordnungseigenschaften, I

Monotone likelihood ratio (MLR)

NIRT

MLR von X_+ in $\theta \in \Theta$ gilt, falls für alle $0 \leq x_{+,1} \leq x_{+,2} \leq k$

$$\frac{\mathbb{P}(X_+ = x_{+,2}|\theta)}{\mathbb{P}(X_+ = x_{+,1}|\theta)}$$

monoton in $\theta \in \Theta$ ist.

WRT

MLR von X_+ in $K \in \mathcal{K}$ gilt, falls für alle $0 \leq x_{+,1} \leq x_{+,2} \leq m$

$$\frac{\mathbb{P}(X_+ = x_{+,2}|K)}{\mathbb{P}(X_+ = x_{+,1}|K)}$$

isoton in $K \in \mathcal{K}$ ist.

Analog kann man MLR von $\theta \in \Theta$ oder $K \in \mathcal{K}$ in X_+ definieren.

Die MLR-Eigenschaft ist eher technisch. Sie ist deswegen in der NIRT wichtig, da sie bestimmte *interpretierbare* stochastische Ordnungseigenschaften impliziert.

Stochastische Ordnungseigenschaften, II

Stochastic ordering of the latent trait (SOL)

NIRT	**WRT**
SOL von $\theta \in \Theta$ durch X_+ bedeutet, dass für alle $\theta_0 \in \Theta$ $$\mathbb{P}(\theta \geq \theta_0 \mid X_+ = x_+)$$ monoton in x_+ ist.	SOL von $K \in \mathcal{K}$ durch X_+ gilt, falls für alle $K_0 \in \mathcal{K}$ $$\mathbb{P}(K \supseteq K_0 \mid X_+ = x_+)$$ monoton in x_+ ist.

Analog kann man 'stochastic ordering' der manifesten Variablen (SOM) X_+ durch $\theta \in \Theta$ oder $K \in \mathcal{K}$ definieren.

Die SOL-Eigenschaft ist sehr wichtig in der NIRT. Sie begründet den Einsatz der Totalscore-Variablen zum Schätzen der Ordnung der Personen entlang des latenten Kontinuums.

Fundamentales Theorem der NIRT

Theorem. Unter den Annahmen der Eindimensionalität, lokalen Unabhängigkeit und Monotonie, und der Forderung dass jede IRF nur Werte echt zwischen Null und Eins annimmt, folgt MLR von X_+ in (eindimensionalem) $\theta \in \Theta$.

Ghurye, S.G., Wallace, D.L., 1959, A convolutive class of monotone likelihood ratio families, The Annals of Mathematical Statistics, 30, 1158–1164.

Grayson, D.A., 1988, Two-group classification in latent trait theory: Scores with monotone likelihood ratio, Psychometrika, 53, 383–392.

Huynh, H., 1994, A new proof for monotone likelihood ratio for the sum of independent Bernoulli random variables, Psychometrika, 59, 77–79.

Ünlü, A., 2008, A note on monotone likelihood ratio of the total score variable in unidimensional item response theory, British Journal of Mathematical and Statistical Psychology, 61, 179–187.

Verallgemeinertes fundamentales NIRT-Theorem

Ein *nichtlinearer Latent-Trait-Raum* ist ein Paar $(\Theta, \preceq)$ bestehend aus einer nichtleeren Menge Θ von *Latent-Trait-Punkten* und einer Quasiordnung $\preceq$ auf Θ. Dies nennen wir die Annahme der *nichtlinearen Dimensionalität*.

Theorem. Unter den Annahmen der nichtlinearen Dimensionalität, lokalen Unabhängigkeit und Isotonie, und der Forderung dass jede IRF nur Werte echt zwischen Null und Eins annimmt, folgt MLR von X_+ in (nichtlinear-dimensionalem) $\theta \in \Theta$.

Das WRT-Äquivalent dieses verallgemeinerten Theorems ergibt sich als Spezialfall, wenn man $\Theta := \mathcal{K}$ und $\preceq := \subseteq$ wählt.

Ünlü, A., 2008, A note on monotone likelihood ratio of the total score variable in unidimensional item response theory, British Journal of Mathematical and Statistical Psychology, 61, 179–187.

Ist das verallgemeinerte Theorem für die Praxis brauchbar?

NIRT

Es gilt (Junker, 1993):

1. MLR von (eindimensionalem) θ in X_+ impliziert SOM.
2. MLR von X_+ in (eindimensionalem) θ impliziert SOL.

WRT

Gelten diese Implikationen auch für die verallgemeinerten Äquivalente? Falls nein, versuche Gründe hierfür zu quantifizieren, und spezifiziere Bedingungen unter denen die Implikationen gelten.

Junker, B.W., 1993, Conditional association, essential independence and monotone unidimensional item response models, The Annals of Statistics, 21, 1359–1378.

Ja, teilweise!

Theorem. Es gilt:

1. MLR von (nichtlinear-dimensionalem) θ in X_+ impliziert SOM.
2. MLR von X_+ in (nichtlinear-dimensionalem) θ impliziert nicht notwendigerweise SOL im Allgemeinen.

In Ünlü (2007) auch diskutiert:

- Der Grund für diese Tatsache ist die ordnungstheoretische Eigenschaft der *Vollständigkeit*. Bedingungen, unter denen die Implikation gilt, werden spezifiziert.
- Die Mokken-artige nichtparametrische Formulierung der WRT wird mit der bis dato parametrischen Formulierung der WRT (basierend auf Latente-Klassen-Analyse) verglichen.

Ünlü, A., 2007, Nonparametric item response theory axioms and properties under nonlinearity and their exemplification with knowledge space theory, Journal of Mathematical Psychology, 51, 383–400.

Contents

Überblick

Nichtparametrische Item-Response-Theorie

Wissensraumtheorie

Zusammenfassung

Die Hauptbeiträge nochmals zusammengefasst

Es wurden die wichtigsten Grundkonzepte der NIRT und WRT unter Berücksichtigung der Historie eingeführt.

Ein kurzer Ausblick auf Verallgemeinerungen der WRT in nichtparametrischer Richtung im Sinne der Mokkenschen NIRT wurde gegeben.

Das Hauptanliegen ist, einen Beitrag zur Vereinheitlichung der bis dato getrennt vorliegenden Testtheorien der Wissensräume und Item-Response-Theorie zu leisten.