Verallgemeinerungen der Rasch-Modells

Vera Völkl

Seminar *Psychometrische Modelle: Theorie und Anwendungen*
Institut für Statistik, LMU

6. Juni 2014

Gliederung

1. Birnbaum-Modelle
   - Modellgleichungen
   - Parameterschätzung
   - Simulation
2. Weitere Verallgemeinerungen
   - Linear-logistisches Testmodell
   - Mehrdimensionale Rasch-Modelle
   - Partial-Credit-Modell
3. Anwendungsbeispiel aus der Genetik
   - Daten
   - Vorgehen

Gliederung

1. Birnbaum-Modelle
   - Modellgleichungen
   - Parameterschätzung
   - Simulation
2. Weitere Verallgemeinerungen
   - Linear-logistisches Testmodell
   - Mehrdimensionale Rasch-Modelle
   - Partial-Credit-Modell
3. Anwendungsbeispiel aus der Genetik
   - Daten
   - Vorgehen

Notation

Bezeichnung	**Bedeutung**
Laufindizes:	
$i = 1, ..., n$	Personen
$j = 1, ..., p$	Aufgaben/Items
Modellparameter:	
$\theta = (\theta_1, ..., \theta_n)^T$	Personenparameter
$\beta = (\beta_1, ..., \beta_p)^T$	Aufgaben-/Itemparameter
$\delta = (\delta_1, ..., \delta_p)^T$	Steigungsparameter
$\gamma = (\gamma_1, ..., \gamma_p)^T$	Rateparameter
u_{ij}	Antwort der Person i auf Item j

Birnbaum-Modelle

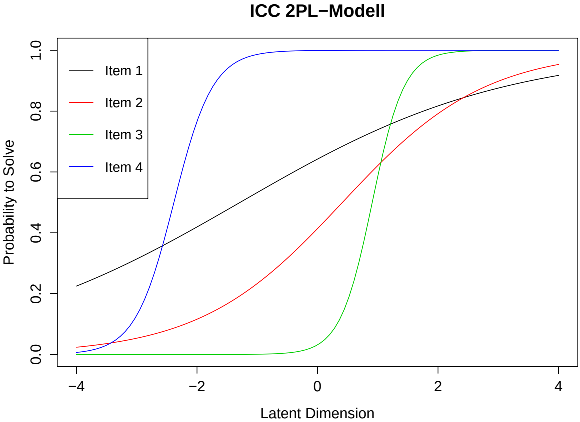
- 1968 von Allan Birnbaum vorgestellt
- Verzicht auf Annahme spezifischer Objektivität
- 2-parametrisches logistisches Modell (2PL-Modell)
- 3-parametrisches logistisches Modell (3PL-Modell)

2PL-Modell

$$P(U_{ij} = 1|\theta_i, \beta_j, \delta_j) = \frac{\exp(\delta_j \cdot (\theta_i - \beta_j))}{1 + \exp(\delta_j \cdot (\theta_i - \beta_j))}$$

- Aufnahme eines itemspezifischen Steigungsparameters δ_j
- Rasch-Modell als Spezialfall mit $\delta_j = 1 \; \forall \; j = 1, ..., p$

2PL-Model



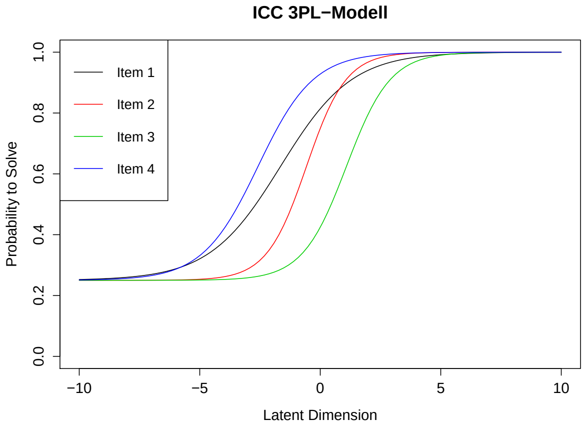
- Unterschiedliche Trennschärfen der Items
- Vergleich zweier Items vom gewählter Person abhängig

3PL-Modell

$$P(U_{ij} = 1|\theta_i, \beta_j, \delta_j, \gamma_j) = \gamma_j + (1 - \gamma_j) \cdot \frac{\exp(\delta_j \cdot (\theta_i - \beta_j))}{1 + \exp(\delta_j \cdot (\theta_i - \beta_j))}$$

- Aufnahme eines itemspezifischen Rateparameters γ_j
- 2PL-Modell als Spezialfall mit $\gamma_j = 0 \ \forall \ j = 1, ..., p$
- Bei der Auswertung von Multiple-Choice-Tests sinnvoll
- γ_j können vorgegeben oder geschätzt werden

3PL-Model



- ICCs entlang der y-Achse um γ_j verschoben
- hier: $\gamma_j = 0,25 \; \forall \; j = 1, ..., 4$

Parameterschätzung

- Rasch-Modell:

  1. Aufstellen der Likelihood bedingt auf Personenscore (suffiziente Statistik für θ)

  $\rightarrow$ unabhängige Schätzung der itemspezifischen Parameter

  2. Schätzung der personenspezifischen Parameter
- Birnbaum-Modelle: Personenscore keine suffiziente Statistik für θ

  $\rightarrow$ bedingte ML-Schätzung nicht wie im Rasch-Modell möglich

Parameterschätzung

- Marginale Schätzung:

  1. Aufstellen der marginalen Likelihood

  $$mL(u, \beta) = \int P(u|\theta, \beta) \cdot f(\theta) \; d\theta$$

  mit geeigneter Randdichte $f(\theta)$

  $\rightarrow$ unabhängige Schätzung der itemspezifischen Parameter

  2. Schätzung der personenspezifischen Parameter

- Problem: Wahl der Randdichte

Simulation

- Wie problematisch ist die Anwendung des Rasch-Modells bei Verletzung spezifischer Objektivität?
- Ist es sinnvoll immer das 2PL-Modell zu verwenden, um eventuelle Verletzung spezifischer Objektivität zu berücksichtigen?
- Ziel der Simulation: Vergleich von Rasch-und 2PL-Modell anhand von Rasch- und 2PL-konformen Daten

Simulation

- Verteilungsannahmen treffen

$$\begin{aligned} \theta_i &\sim \mathcal{N}(0,1) \quad \forall i = 1, ..., n \\ \beta_j &\sim \mathcal{N}(0,1) \quad \forall j = 1, ..., p \\ \delta_j &\sim ln\mathcal{N}(0,\sigma_j^2) \quad \forall j = 1, ..., p \;\; \text{mit } \sigma_j^2 \sim \mathcal{U}(0, 0.5) \end{aligned}$$

- Rasch-konform: $\delta_j = 1 \ \forall j = 1, ..., p$

Simulation

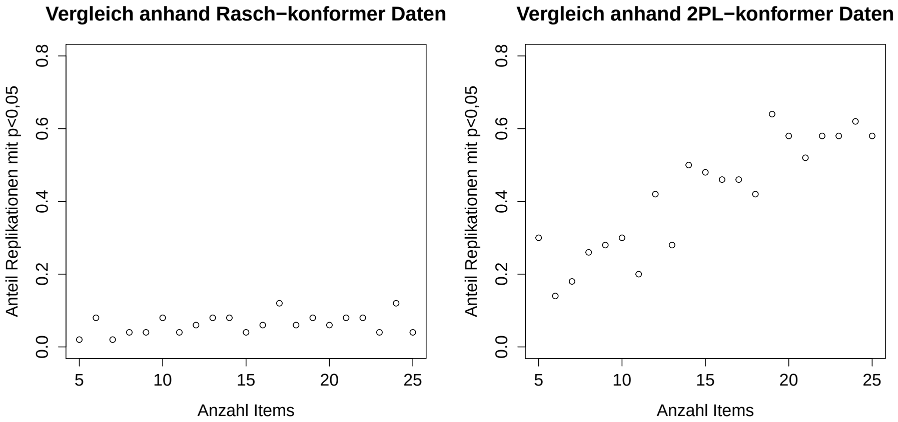
- Parameter aus jeweiligen Verteilungen ziehen
- Lösungswahrscheinlichkeiten berechnen:

$$P(U_{ij} = 1|\theta_i, \beta_j, \delta_j) = \frac{\exp(\delta_j \cdot (\theta_i - \beta_j))}{1 + \exp(\delta_j \cdot (\theta_i - \beta_j))}$$

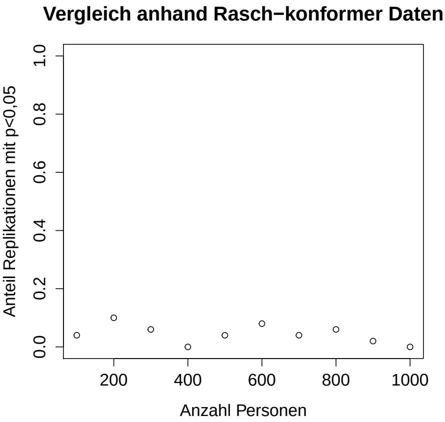
  $\rightarrow$ Ziehung von u_{ij} aus $\{0, 1\}$

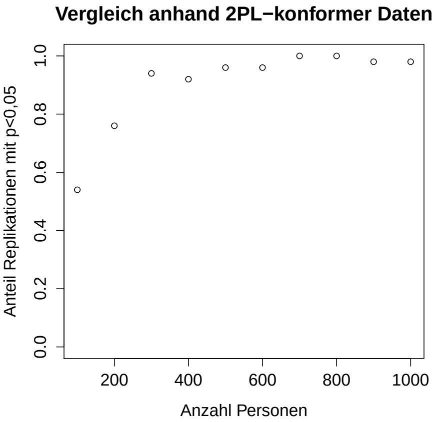
- Modelle anhand der simulierten Daten berechnen (R-Paket ltm)
- Modellvergleich mit LQ-Test
- Untersuchung der Auswirkung unterschiedlicher p/n
  $\rightarrow$ jeweils 50 Replikationen

Ergebnisse der Simulation ($n = 100$)



Ergebnisse der Simulation ($p = 20$)





Fazit

- Rasch-konforme Daten
  - 2PL-Modell nur sehr selten bessere Anpassung
  - Keine Tendenzen für steigende n oder p erkennbar
- 2PL-konforme Daten
  - 2PL-Modell trotz Verletzung der spezifischen Objektivität nicht immer die bessere Anpassung
  - Anpassung des 2PL-Modells bessert sich für größere Datensätze

Gliederung

1 Birnbaum-Modelle
- Modellgleichungen
- Parameterschätzung
- Simulation

2 Weitere Verallgemeinerungen
- Linear-logistisches Testmodell
- Mehrdimensionale Rasch-Modelle
- Partial-Credit-Modell

3 Anwendungsbeispiel aus der Genetik
- Daten
- Vorgehen

Linear-logistisches Testmodell

- 1973 von Fischer vorgestellt
- Zerlegung der Aufgabenparameter in gewichtete Teilkompetenzen:

$$\beta_j = \sum_l w_{jl} \cdot \eta_l$$

wobei η_l dem Beitrag der Teilkompetenz entspricht und

$$w_{jl} = \begin{cases} 1, \text{falls Teilkompetenz zur Lösung der Aufgabe benötigt wird} \\ 0, \text{sonst} \end{cases}$$

- Schätzung der η_l mit bedingter Likelihood (R-Paket eRm)

Mehrdimensionale Rasch-Modelle

- Zerlegung der Personenparameter
  $\rightarrow$ Verzicht auf Forderung nach Eindimensionalität
- Unterscheidung in Modelle mit
  - *between-item multidimensionality*
  - *within-item multidimensionality*
- Parameterschätzung mit EM-Algorithmus (R-Paket MultiLCIRT)

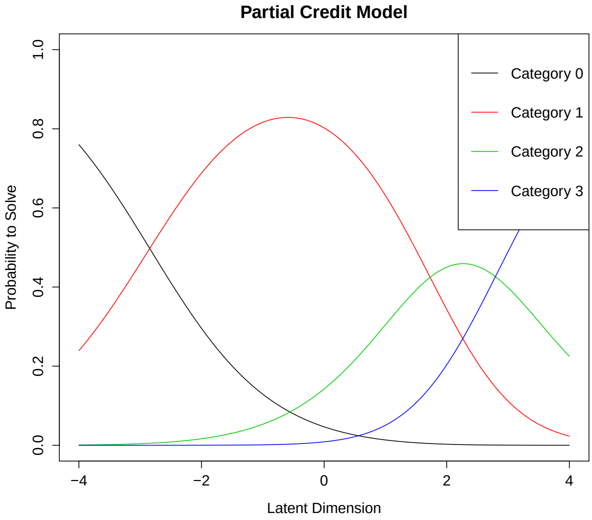
Partial-Credit-Modell

- Anwendung bei Tests mit m_j Antwortkategorien statt binärem Response
- Wahrscheinlichkeit der Person i in Aufgabe j die Kategorie c zu erreichen:

$$P(U_{ij} = c|\theta_i, \beta_j) = \frac{\exp(c \cdot \theta_i - \beta_{jc})}{\sum\limits_{l=0}^{m_j-1} \exp(l \cdot \theta_i - \beta_{jl})}$$

- Rasch-Modell als Spezialfall mit 2 Kategorien
- Parameterschätzung mit bedingter Likelihood (R-Paket eRm)

Partial-Credit-Modell



Gliederung

1 Birnbaum-Modelle
- Modellgleichungen
- Parameterschätzung
- Simulation

2 Weitere Verallgemeinerungen
- Linear-logistisches Testmodell
- Mehrdimensionale Rasch-Modelle
- Partial-Credit-Modell

3 Anwendungsbeispiel aus der Genetik
- Daten
- Vorgehen

Anwendungsbeispiel

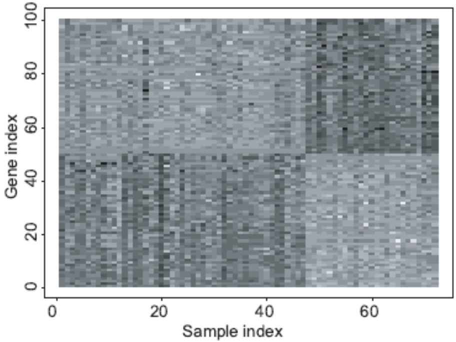
- Li und Hong (2001)
- Ziel: von Genexpressionen auf interessierenden Phänotyp schließen
- Kombination aus
  - Clusteranalyse,
  - Partial-Credit-Modell und
  - Diskriminanzanalyse/Regression

Daten

- 72 Gewebeproben von Patienten mit akuter Leukämie
  - 47 Patienten mit akuter lymphatischer Leukämie (ALL)
  - 25 Patienten mit akuter myeloischer Leukämie (AML)
- Simultane Messung der Genexpressionen mit Microarrays (hier: $\sim$ 4000 Gene)
- Ziel: Klassifizierung zukünftiger Gewebeproben in ALL/AML

Vorbereitung

- Y_i: Phänotyp der Gewebeprobe i
- X_{ij}: Genexpression des j-ten Gens in der i-ten Gewebeprobe
- Auswahl der jeweils 50 überexprimiertesten Gene aus AML-/ALL-Gewebeproben mit Wilcoxon-Rangsummen-Test



Quelle: Li und Hong (2001)

Clustering

- Annahme: Gene mit ähnlicher Expression erfassen denselben latenten Faktor
- Einteilung der 100 Gene basierend auf x_{ij} in k Cluster
- Optimales k wird durch Clusterverfahren festgelegt (hier: $k = 2$)

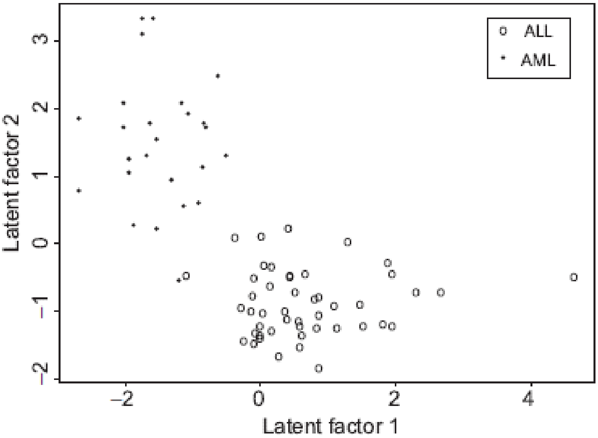
Partial-Credit-Modell

- Aus stetiger ZV X_{ij} diskrete ZV $U_{ij} \in \{0, ..., m-1\}$ bilden (hier: m=4)
- Nach Cluster getrennt 2 Partial-Credit-Modelle rechnen mit
  - U_{ij}: Genexpressionslevel
  - Gewebeprobe $\widehat{=}$ Person

    $\rightarrow \theta_{iK}$: latenter Faktor der i-ten Gewebeprobe im Cluster K
  - Gen $\widehat{=}$ Item

    $\rightarrow \beta_{jK}$: genspezifischer Parameter im Cluster K

Partial-Credit-Modell



Quelle: Li und Hong (2001)

Klassifizierung

- Hier: Diskriminanzanalyse
  - Geschätzter Prädiktionsfehler: 3%
  - Gute Trennung von ALL-/AML-Gewebeproben
- Alternativ: Regressionsmodell
  - Aufnahme der geschätzten latenten Faktoren als Kovariablen
  - Art der Regression abhängig vom untersuchten Phänotyp

Prädiktion

- Gewebeprobe mit Genexpressionen x_{neuj}
- Phänotyp y_{neu} unbekannt
  $\rightarrow$ ML-Schätzer $\hat{\theta}_{neu1}$ und $\hat{\theta}_{neu2}$ bestimmen
  $\rightarrow$ Gewebeprobe mithilfe von Diskriminanzanalyse/Regression klassifizieren

Fazit

Vorteile:

- Vereinfachung komplexer Microarray-Daten
- Messfehler von Microarrays werden durch Verwendung von Quantilen reduziert

Nachteile:

- Informationsverlust durch Diskretisieren
- Clustering basierend auf Expressionsprofilen nicht immer sinnvoll

Literatur

Birnbaum, A. (1968). Some Latent Trait Models and Their Use in Inferring an Examinees Ability, in F. M. Lord and M. Novick (eds), Statistical Theories of Mental Test Scores, 2 edn, Addison-Wesley, Reading, pp. 397-424.

Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research, Acta Psychologica 37(6): 359-374.

Li, H. and Hong, F. (2001). Cluster-rasch models for microarray gene expression data, Genome Biology 2(8).

Molenaar, I. (1995). Estimation of Item Parameters, in G. Fischer and I. Molenaar (eds), Rasch-Models: Foundations, Recent Developements, and Applications, 1 edn, Springer, New York, pp. 39-52.

Literatur

R Development Core Team (2013). R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria.
URL: http://www.R-project.org

Rizopoulos, D. (2006). ltm: An r package for latent variable modelling and item response theory analyses, Journal of Statistical Software 17(5): 1-25.
URL: http://www.jstatsoft.org/v17/i05/

Rost, J. (1996). Lehrbuch Testtheorie, Testkonstruktion, 1 edn, Huber, Bern.

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

Suárez-Falcón, J. C. and Glas, C. (2003). Evaluation of global testing procedures for item t to the rasch model, British Journal of Mathematical and Statistical Psychology 56: 127-143.

Vielen Dank für die Aufmerksamkeit!