Partial Credit Model und Tutz Model

Martina Maida, Maria Weiler

November 22, 2011

Background

Psychologische Tests verfolgen das Ziel, mit wissenschaftlichen Methoden quantitative Aussagen über den relativen Grad der individuellen Ausprägung eines Merkmals zu treffen. Um eine solche Aussage fundiert zu ermöglichen, basieren psychologische Tests auf einer Testtheorie. Sie beschreibt den Zusammenhang zwischen dem zu erfassenden Merkmal und dem Testverhalten (Rost, 1996).

Klassische Testtheorie

- Zentrales Konzept ist Reliabilität
- Messfehlertheorie $X = T + E$
- Reliabilität kann nicht direkt berechnet werden
- Reliabilität ist streng genommen das einzige Gütekriterium
- Messungen basieren auf Korrelationen
- -> Stichprobenabhängigkeit

Item Response Theory

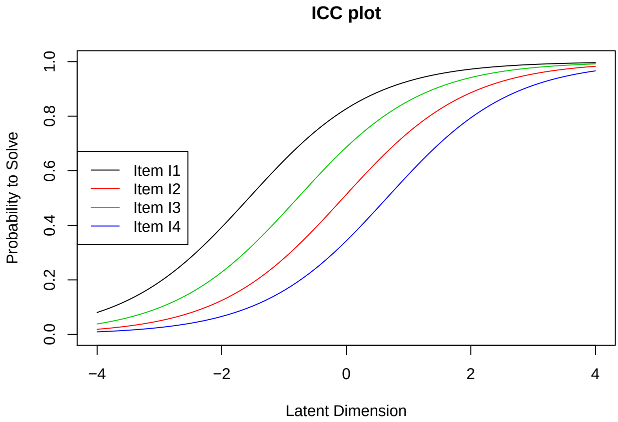
- Probabilistische Theorie oder latent trait theory
- misst latente Variablen
- Messung der Wahrscheinlichkeit von korrekten Antworten
- Mathematische Funktion von Personen und Items
- nach van der Linden Hambleton (1997) über 100 IRT Modelle

1 PLM Parameter Logistic Model

- 1 PLM
- $P(X_{\upsilon i} = 1|\theta_\upsilon, \beta_i) = \dfrac{\exp(\theta_\upsilon - \beta_i)}{1 + \exp(\theta_\upsilon - \beta_i)}$

$X_{\upsilon i}$ ...person υ agrees to statement i, i = 1 ... k, $\upsilon = 1$ ... n
θ_υ ... location of person υ on latent trait (amount of trait)
β_i ... location of item i on latent trait (stimulative nature)
in case of performance tests:
$X_{\upsilon i}$ ...person υ gives correct answer to item i
θ_υ ... 'ability' of person υ
β_i ... 'difficulty' of item i

IRT-Einteilung



4 Items 1PLM (zB Rasch Modell)

Rasch - Assumptions /Properties

- Lokale stochastische Unabhängigkeit
- Suffizienz der Randsummen
- Monotonie
- Eindimensionalität

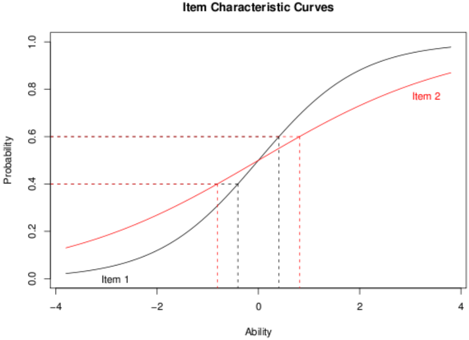
2 PLM Parameter Logistic Model

- 2 PLM
- $P(X_{vi} = 1|\theta_v, \beta_i, \alpha_i) = \dfrac{\exp((\theta_v - \beta_i)\alpha_i)}{1 + \exp((\theta_v - \beta_i)\alpha_i)}$

α ... item discriminating power,
zeigt den Anstieg der ICC für item i bei $p = 0.5$

2 PLM ICC - Grafik: Psychometric Methods 2010/11

'discrimination' for Item 1: $\alpha_1 = 1$, for Item 2: $\alpha_2 = 0.5$



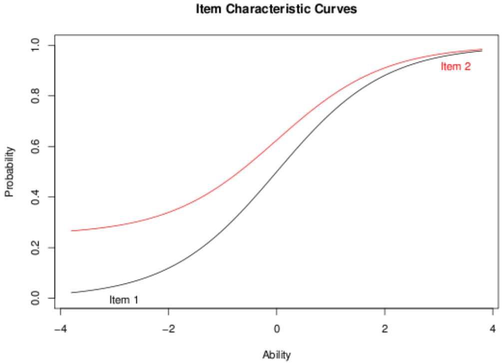
3 PLM Parameter Logistic Model

- 3 PLM
- $P(X_{vi} = 1 | \theta_v, \beta_i, \alpha_i) = c_i + (1 - c_i) \dfrac{\exp((\theta_v - \beta_i)\alpha_i)}{1 + \exp((\theta_v - \beta_i)\alpha_i)}$

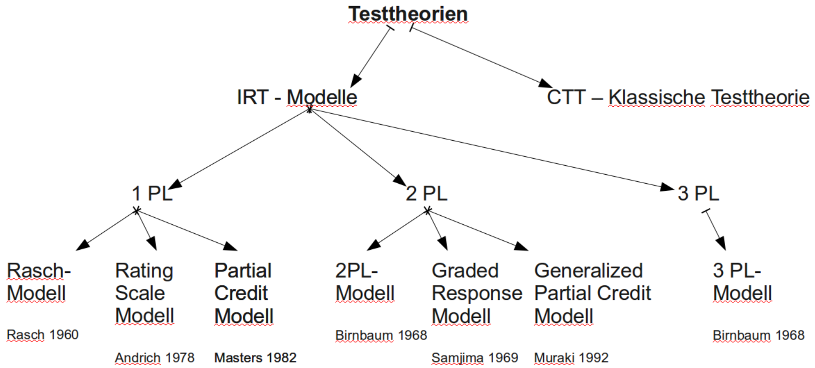
c_i ... Ratewahrscheinlichkeit

3 PLM ICC - Grafik: Psychometric Methods 2010/11

'guessing' probability is 0.25



Überblick


Testtheorien
IRT - Modelle
CTT – Klassische Testtheorie
1 PL
2 PL
3 PL
Rasch-Modell
Rasch 1960
Rating Scale Modell
Andrich 1978
Partial Credit Modell
Masters 1982
2PL-Modell
Birnbaum 1968
Graded Response Modell
Samjima 1969
Generalized Partial Credit Modell
Muraki 1992
3 PL-Modell
Birnbaum 1968

Modellübersicht

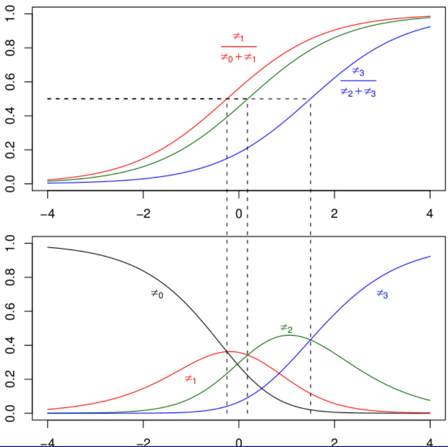
- RSM (Rating Scale Model) Polytome Erweiterung des Rasch-Modells, Einführung von Item-Kategorien
- PCM (Partial Credit Model) Erweiterung des Rating Scale Modells, Verschiedene Anzahl von Kategorien pro Item
- GRM (Graded Response Model) Erweiterung des 2PL-Modells auf politome Daten
- GPCM (Generalized PCM) 2PL-Modell, inkludiert den Diskriminationsparameter α

Partial Credit Model

Masters (1982) erweitert das Rasch Modell um polytome Items mit geordneten Kategorien.

$$P(X_{\upsilon ih} = 1) = \frac{\exp[h\theta_\upsilon - \beta_{ih}]}{\sum_{l=0}^{m_i} \exp(l\theta_\upsilon - \beta_{il})}$$

Category Probability Curves



Datenstruktur

Startpunkt im PCM (wie auch bei Rasch-Modellen) ist die Personen x Item Matrix.
Beispiel:
7 Aufsätze wurden von 3 Lehrpersonen mit Noten bewertet...

PCM - Herleitung

	[,1]	[,2]	[,3]
[1,]	1	1	3
[2,]	1	2	5
[3,]	1	2	4
[4,]	2	5	5
[5,]	2	4	4
[6,]	1	1	2
[7,]	1	3	4

PCM - Herleitung

Rasch Modell für dichotome Daten

$$\frac{P_{iv1}}{P_{iv0} + P_{iv1}} = \frac{\exp(\theta_v - \beta_i)}{1 + \exp(\theta_v - \beta_i)}$$

PCM - Herleitung

Erweiterung auf mehr als zwei Antwortkategorien:

$$\frac{P_{i\upsilon x}}{P_{i\upsilon x}-1+P_{i\upsilon x}} = \frac{\exp(\theta_\upsilon - \beta_{ix})}{1+\exp(\theta_\upsilon - \beta_{ix})}, x = 1, 2, ..., m_i$$

PCM - Herleitung

Die Konvention gibt vor, dass man dies als unbedingte Wahrscheinlichkeit für jedes mögliche Ergebnis umformuliert:

$$\sum_{h=0}^{m_i} P_{i\upsilon h} = 1$$

Die Wahrscheinlichkeit der Person υ, dass sie bei Item i x erzielt, wird wie folgt formuliert:

$$P_{i\upsilon x} = \frac{\exp \sum_{k=0}^{x}(\theta_\upsilon - \beta_{ik})}{\sum_{h=0}^{m_\upsilon} \exp \sum_{k=0}^{h}(\theta_\upsilon - \beta_{ik})}$$

Parameterschätzung

- Mit der Parameterschätzung werden die Item- und Personenparameter geschätzt.
- Es gibt verschiedene Möglichkeiten der Schätzung:
  - Conditional Maximum Likelihood
  - Joint Maximum Likelihood (relativ einfach, aber produziert Bias)
  - Marginal Maximum Likelihood (benötigt Annahmen über die Verteilung der Personenparameter)

Conditional Maximum Likelihood

$$L_C = \exp(-\sum_i \beta_i s_i) / \prod_r \sum_{x|r} \exp(-\sum_i x_i \beta_i)^{n_r}$$

- Keine Annahmen über Personenparameter nötig, weil sie in der Conditional Likelihood nicht vorkommen
- Konsistente Schätzer der Itemparameter
- Stichprobenunabhängigkeit
- ...in general the procedure of choice

Goodness of Fit

- Prinzipiell gibt es Tests of Itemfit, Personfit und Globalfit
- item fit identifiziert problematische Items
- person fit identifiziert Personen mit Antworten, die nicht dem allgemeinen Antwortmuster folgen
- global fit zeigt den allgemeinen Fit des Datensatzes auf das PCM
- Für die Berechnung mit R werden die auch χ^2 basierenden Funktionen itemfit und personfit verwendet
- Zusätzlich zu diesen statistischen Tests ist eine grafische Darstellung empfehlenswert

Analysis using eRm

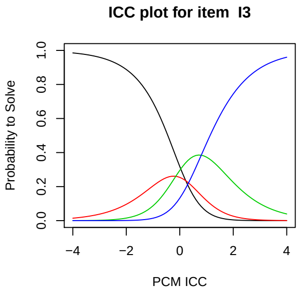
```
> library("eRm")
> data("pcmdat")
> res.pcm<-PCM(pcmdat)
```

Beispiel

```
> res.pcm

Results of PCM estimation:

Call:  PCM(X = pcmdat)

Conditional log-likelihood: -121.0979
Number of iterations: 16
Number of parameters: 19

Item (Category) Difficulty Parameters (eta):
              I1.c2     I1.c3      I2.c1      I2.c2    I2.c3     I3.c1      I3.c2     I3.c3      I4.c1
Estimate -0.9368481 0.3303217 -0.2508642 -0.0650128 1.431592 0.2239408 0.07447576 0.8782792 -0.4229829 0.2
Std.Err   0.5652102 0.8094052  0.5254578  0.5797920 1.023935 0.5609934 0.56794601 0.7839333  0.5467276 0.6
              I4.c3     I5.c1       I5.c2      I5.c3     I6.c1      I6.c2      I6.c3      I7.c1     I7.c2
Estimate -0.1395695 0.9803568 -0.09620567 -0.4030339 0.5861060 -0.5041350 -1.0995166 -0.8493303 0.2980869
Std.Err   0.6497043 0.7880101  0.58461877  0.5788289 0.8264099  0.6344952  0.5869572  0.5044312 0.7005385
```

Beispiel

```
> plotICC(res.pcm, item.subset = 3, xlab = "PCM ICC", mplot = TRUE, legpos = FALSE)
```

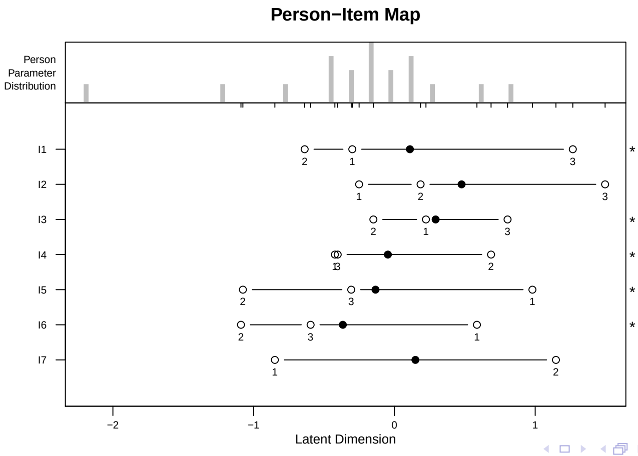


Beispiel

```
> thresholds(res.pcm)

Design Matrix Block 1:
   Location Threshold 1 Threshold 2 Threshold 3
I1  0.11011    -0.29916    -0.63769     1.26717
I2  0.47720    -0.25086     0.18585     1.49661
I3  0.29276     0.22394    -0.14947     0.80380
I4 -0.04652    -0.42298     0.68648    -0.40307
I5 -0.13434     0.98036    -1.07656    -0.30683
I6 -0.36651     0.58611    -1.09024    -0.59538
I7  0.14904    -0.84933     1.14742          NA
```

Beispiel

```
> plotPImap(res.pcm)
```



Beispiel

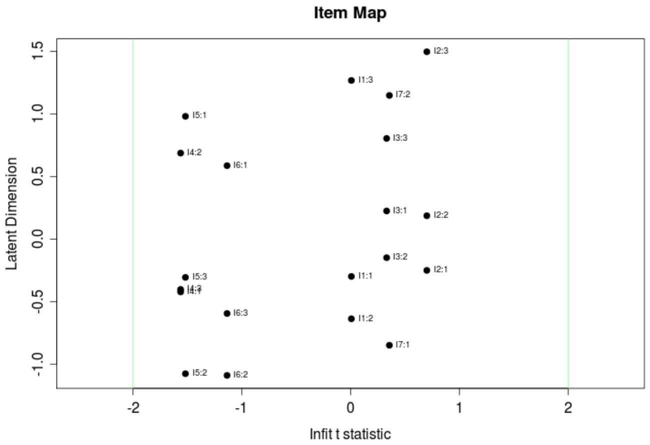
```
> LRtest(res.pcm)

I2 I4 I1 I5
Full and subgroup models are estimated without these items!

Andersen LR-test:
LR-value: 2.567
Chi-square df: 7
p-value:  0.922
```

Beispiel

```
> pp<-person.parameter(res.pcm)
> personfit(pp)

Personfit Statistics:
    Chisq df p-value Outfit MSQ Infit MSQ Outfit t Infit t
P1  9.105  6   0.168      1.301     1.113     1.19    0.57
P2  5.801  6   0.446      0.829     0.721    -0.61   -1.16
P3  5.048  6   0.538      0.721     0.751    -0.72   -0.63
P4  9.581  6   0.143      1.369     1.167     0.94    0.57
P5  7.903  6   0.245      1.129     1.233     0.60    1.01
P6  6.131  6   0.409      0.876     0.924    -0.35   -0.19
P7  3.948  6   0.684      0.564     0.660    -0.06    0.00
P8  4.215  6   0.648      0.602     0.705    -1.68   -1.25
P9  4.085  6   0.665      0.584     0.452    -1.75   -2.50
P10 9.718  6   0.137      1.388     1.259     1.58    1.16
P11 7.399  6   0.286      1.057     1.050     0.35    0.33
P12 8.544  6   0.201      1.221     1.098     0.84    0.47
P13 5.388  6   0.495      0.770     0.805    -0.42   -0.36
P14 4.677  6   0.586      0.668     0.728    -1.21   -0.94
P15 7.301  6   0.294      1.043     1.207     0.29    0.91
P16 7.983  6   0.239      1.140     1.037     0.64    0.27
P17 4.671  6   0.587      0.667     0.707    -0.57   -0.54
P18 4.426  6   0.619      0.632     0.720    -1.37   -0.97
P19 6.068  6   0.416      0.867     0.881    -0.44   -0.40
P20 4.534  6   0.605      0.648     0.672    -1.44   -1.43
```

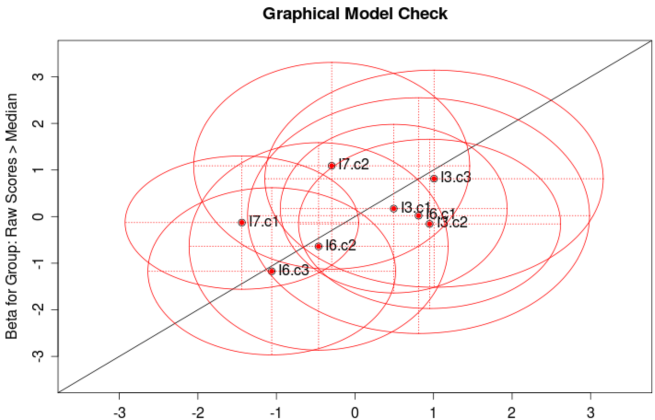
```
> itemfit(pp)

Itemfit Statistics:
    Chisq df p-value Outfit MSQ Infit MSQ Outfit t Infit t
I1 18.689 19   0.477      0.934     0.987    -0.26    0.01
I2 24.268 19   0.186      1.213     1.110     1.12    0.70
I3 19.352 19   0.434      0.968     1.041    -0.04    0.33
I4 13.553 19   0.809      0.678     0.754    -1.71   -1.56
I5 14.376 19   0.761      0.719     0.786    -1.06   -1.52
I6 14.303 19   0.766      0.715     0.813    -1.11   -1.14
I7 21.986 19   0.285      1.099     1.048     0.63    0.36
```

Beispiel


Item Map
Latent Dimension
Infit t statistic
I2:3
I1:3
I7:2
I5:1
I3:3
I4:2
I6:1
I3:1
I2:2
I3:2
I2:1
I5:3
I1:1
I4:3
I4:1
I6:3
I1:2
I7:1
I5:2
I6:2
1.5
1.0
0.5
0.0
-0.5
-1.0
-2
-1
0
1
2

Beispiel

```
> lr<-LRtest(res.pcm, se = T)
I2 I4 I1 I5
Full and subgroup models are estimated without these items!
> plotGOF(lr, conf = list(), xlim = c(-3.5, 3.5), ylim = c(-3.5, 3.5))
```



Sequential Models for Ordered Responses

Sequential Models for Ordered Responses

- Auch hier sind die Antworten ordinale Kategorien
- Im SM geht es um das sequentielle Lösen von Teilproblemen
- Die Lösung des Problems erfolgt in sukzessiven Schritten
  - Der jeweilige Folgeschritt kann erst nach erfolgreicher Absolvierung der vorhergehenden Stufe erreicht werden
  - Der Übergang von einem Level zum nächsten wird modelliert
  - Die Itemparameter werden direkt, entsprechend der Schwierigkeit dieses Übergangs, interpretiert
- Das PCM verfolgt genauso das Konzept des stufenweisen Vorgehens, diese Stufen sind aber lokal
- Sie sind bedingt durch die Antwort in der Kategorie h-1 oder h

Datensatz

- Bereits in der Aufgabenstellung, die zum Datensatz führt, ist sequenzielles Vorgehen notwendig.
- Beispiel: Es gibt 6 Fragen mit 3 Stufen, die 99 Studierenden präsentiert werden.
- Die Aufgabenstellung aus einem Test aus der Lehrveranstaltung Statistik ist
  - 1. Finde die passende Methode zu vorgegebenen Variablen/ oder Problemen
  - 2. Berechne mit Hilfe der Methode die Werte
  - 3. Interpretiere die Werte

Datenstruktur

Startpunkt im Tutz-Modell (wie auch bei Rasch-Modellen) ist die Personen x Item Matrix. Die Items werden hier mit 0 nicht bestanden bzw. 1 bestanden angegeben.

Andrich "Rating Scale" model
Masters "Partial Credit" model

Rating level	Step 1	Step 2	Step 3	Steps Tried	Successes Scored
3-Hard	1	1	1	3	3
2	1	1	0	3	2
1	1	0	0	3	1
0-Easy	0	0	0	3	0

Key: 1=Passed 0=Failed M=Not administered

Herleitung Sequential Model

Um die beobachteten Punkte und Fähigkeiten zu erhalten, müssen die Schritte modelliert werden. Ein erfolgreicher Übergang wird dargestellt als

$U*_{i\ \upsilon h} \equiv \theta_\upsilon + \varepsilon_{i\upsilon h}$
$\theta =$ Fähigkeit der Person υ, $\varepsilon =$ noise Variable.

Ein weiterer erfolgreicher Übergang wird modelliert durch:

$U_{i\upsilon,h+1} = 1$ für $U*_{i\ \upsilon,h+1} \geq b_{i,h+1}$
$b_{i,h+1}$ ist die Schwierigkeit des $h+1$sten Schrittes.

Herleitung Sequential Model

Der $h + 1$ste Schritt ist durch ein dichotomes Response Model gegeben:

$$P(U_{iv,h+1} = 1) = F(\theta_v - b_{i,h+1})$$

F ist eine Funktion, die den Noise Variablen ε zu Grunde liegt.

Herleitung Sequential Model

Wenn diese eine Logistische Funktion ist
$F(x) = \exp(x)/(1 + \exp(x))$,
kann jeder Schritt mit einem dichotomen Rasch Modell modelliert werden.

Das Sequential Model hat dann die Form:
$P(U_{i\upsilon} > h | U_{i\upsilon} \geq h) = F(\theta_\upsilon - b_{i,h+1}), h = 0, ..., m_i - 1$

Parameterschätzung

- Mit der Parameterschätzung werden die Itemparameter geschätzt
- Diese sind durch Personenparameter bedingt (Es gibt keine Personenparameterschätzung)
- Es gibt verschiedene Möglichkeiten der Schätzung:
  - Joint Maximum Likelihood (relativ einfach, aber produziert Bias)
  - Marginal Maximum Likelihood (benötigt Annahmen über die Verteilung der Personenparameter)
- Da man eine Verteilung für die Personen annimmt, kommt letztlich die Marginal Maximum Likelihood in Frage

GoF

Goodness of Fit Checks gründen auf der dichotomen Natur der Übergänge in höhere Level. Für dichotome Modelle kommen viele Methoden in Frage (siehe Andersen (1973), Kelderman (1984), Hembleton und Swaminathan (1985)).

GoF

Wählt man zB. eine log-likelihood-ratio Statistik, kann man die GoF bei den Übergängen messen.
Ein Vergleich mit der asymptotischen χ^2 Verteilung zeigt, ob das Modell zu den Steps passt.
Durch die daraus resultierende Vielfalt der Tests wird empfohlen, ein Signifikanzlevel hinzu zu fügen.
GoF muss auf die genaue Erhebung abgestimmt werden.

Übungsbeispiele

Viel Erfolg!