



Generalized Partial Credit Model (GPCM) Graded Response Model (GRM)

Alexandra Grand, Klaus Sommerauer

► Überblick IRT-Modelle

IRT - Modelle

1PL			2PL		
RM (Rasch,1960)	→ RSM (Andrich,1978)	PCM (Masters,1982)	2PL-Modell (Birnbaum,1968)	→ GPCM (Muraki,1992)	GRM (Samejima,1969)

Antwortformat

dichotom	polytom	polytom	dichotom	polytom	polytom
----------	---------	---------	----------	---------	---------

(Rasch)-Annahmen

<ul style="list-style-type: none"> ✓ Unidimensionalität ✓ Monotonie ✓ suffiziente Statistik ✓ lokale stochastische Unabhängigkeit ✓ spezifische Objektivität ✓ „stichprobenunabhängig“ 	<ul style="list-style-type: none"> ✓ Unidimensionalität ✓ Monotonie x keine suffiziente Statistik ✓ lokale stochastische Unabhängigkeit x keine spezifische Objektivität x keine „Stichprobenunabhängigkeit“
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------



adjacent categories logistic model

cumulative logistic model

▶ 2PL-Modelle (für polytome Daten)

2 PL (Parameter Logistic) Modelle

beschreiben probabilistische Beziehung zwischen der beobachtbaren Antwort auf ein Item (manifeste Variable) und der latenten Variable auf Basis der jeweiligen **Item-Characteristic Funktion**

$$P(X_{vj} = h \mid \beta_{jh}, \alpha_j, \theta_v) = f_{jh}(\theta_v)$$

in Abhängigkeit von θ , mit **zwei** Itemparameter:

- Itemschwierigkeitsparameter bzw. Schwellenparameter β_{jh}
- Itemdiskriminationsparameter α_j



Generalized Partial Credit Models (GPCM)

▶ Generalized Partial Credit Model (GPCM)

Muraki (1992)

- GPCM = Erweiterung des PCM um Itemdiskriminationsparameter α_j ; α_j 's können zwischen Items variieren
- Messung von latenten Fähigkeiten, Eigenschaften anhand polytomer Items
- Items mit geordneten Antwortkategorien $h = 0, 1, \dots, H$; gemeinsame Analyse unterschiedlicher Antwortformate möglich
- adjacent categories logistic model
- Schwellenparameter müssen nicht zwingend geordnet sein
- Kategoriewahrscheinlichkeiten sind direkt berechenbar

► Generalized Partial Credit Model (GPCM)

Herleitung:

(Schwellen)Wahrscheinlichkeit, dass bei Item j Kategorie h

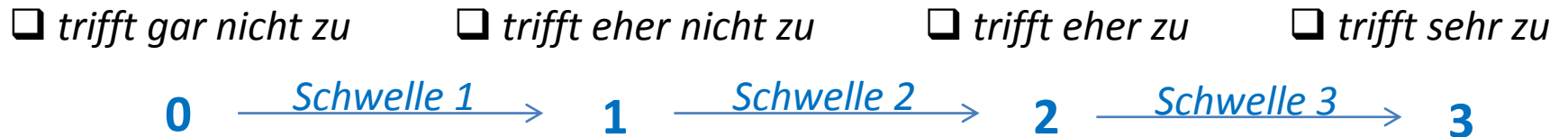
(gegenüber Kategorie $h-1$) gewählt wird → *2 PL-Birnbaum-Modell*

$$P_{jh|h-1,h}(\theta) = \frac{P_{jh}(\theta)}{P_{jh-1}(\theta) + P_{jh}(\theta)}$$

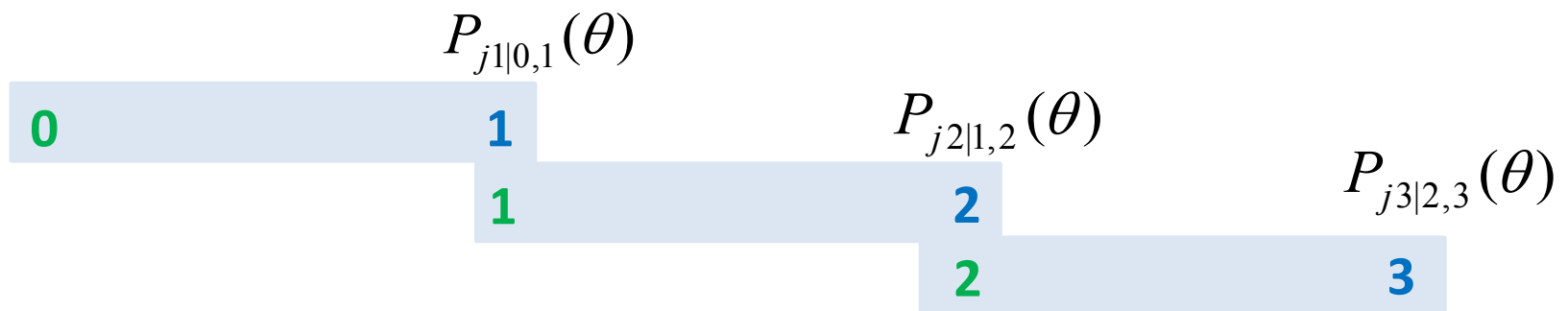
▶ GPCM – Beispiel/Schwellenwahrscheinlichkeit

Heute war ich wütend. (adaptiert von Baker, Rounds, Zevon, 2000)

Antwortkategorien:



Dichotome Vergleiche der Antwortkategorien mittels 2PL – Birnbaum-Modell



► GPCM – Schwellenwahrscheinlichkeit

$$P(X_{vj} = h \mid \theta_v, \beta_{jh}, \alpha_j) = \frac{\exp((\theta_v - \beta_{jh})\alpha_j)}{1 + \exp((\theta_v - \beta_{jh})\alpha_j)}$$

$$h = 0, 1, \dots, H; \beta_{j0} \equiv 0$$

$$P(X_{vj} = h \mid \theta_v, \beta_j, d_{jh}, \alpha_j) = \frac{\exp((\theta_v - \beta_j + d_{jh})\alpha_j)}{1 + \exp((\theta_v - \beta_j + d_{jh})\alpha_j)}$$

α_j Steigungsparameter (*slope* parameter) des Items j ; $-\infty$ bis ∞
(sollten zwischen 0.8 und 2.5 liegen) Ayala, 2009

β_{jh} Schwellenparameter (*threshold* parameter) der Schwelle $h-1/h$ des Items j

β_j Lokationsparameter (*location* parameter) des Items j

d_{jh} Category Parameter des Items j

θ_v Fähigkeitsparameter (latent trait parameter) der Person v ; $-\infty$ bis $+\infty$

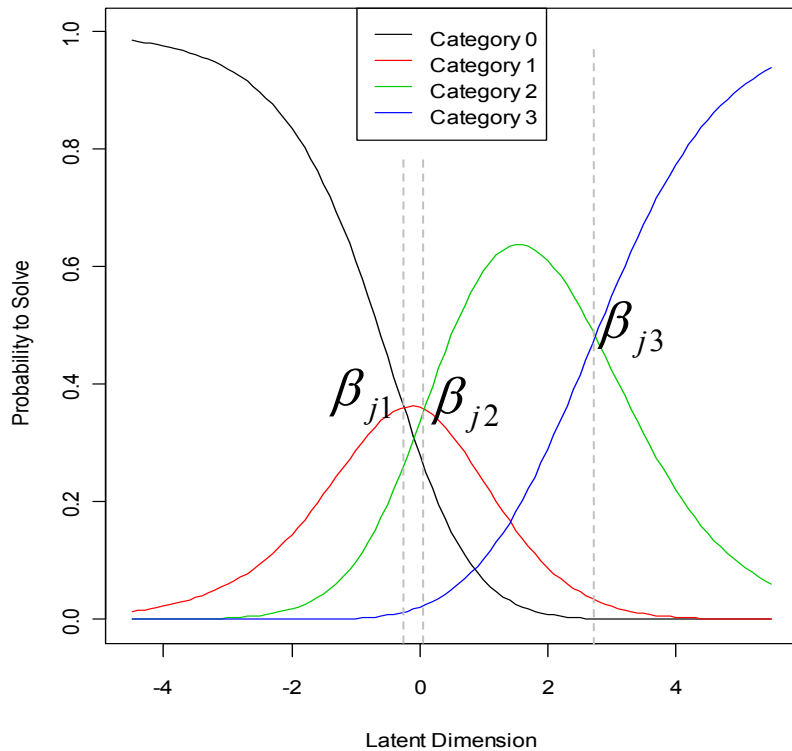
▶ GPCM – Kategoriewahrscheinlichkeit

ICRF – Item-Category Response Function

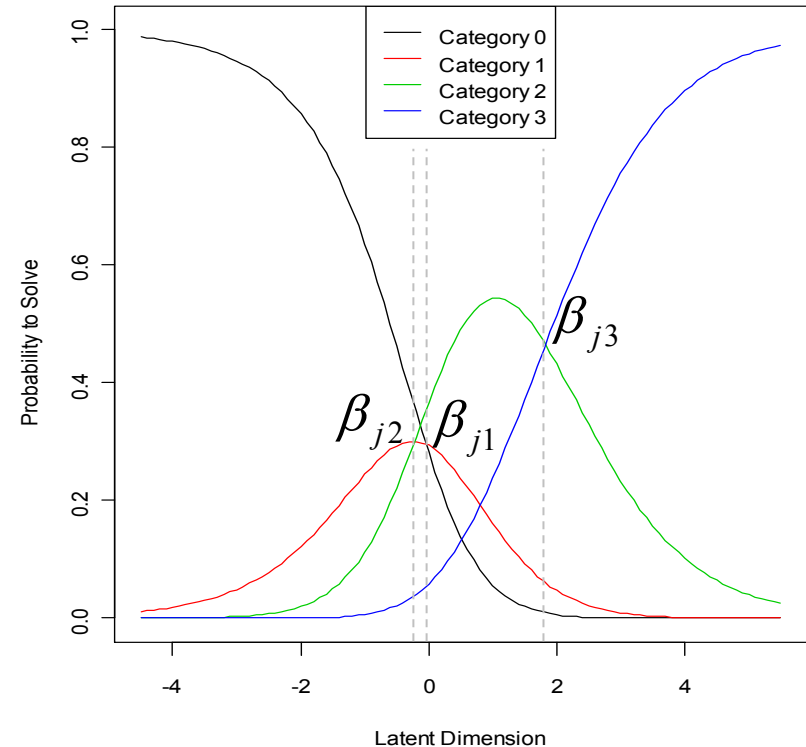
$$P_{jh}(\theta_v) = \frac{\exp\left[\sum_{x=0}^h (\theta_v - \beta_{jx})\alpha_j\right]}{\sum_{c=0}^H \exp\left[\sum_{x=0}^c (\theta_v - \beta_{jx})\alpha_j\right]}$$

▶ GPCM – ICRF (Item-Category Response Function)

- lokale Schätzung der Schwellenparameter (wie beim PCM)
- Schwellenparameter: Schnittpunkte benachb. Kategoriewahrsch.kurven
- wenn zB $\theta = \beta_{j2}$ dann beträgt Schwellenwahrscheinlichkeit **0.5**
→ Kategoriewahrscheinlichkeit $P_{j2} = P_{j1}$



Schwellenparameter geordnet

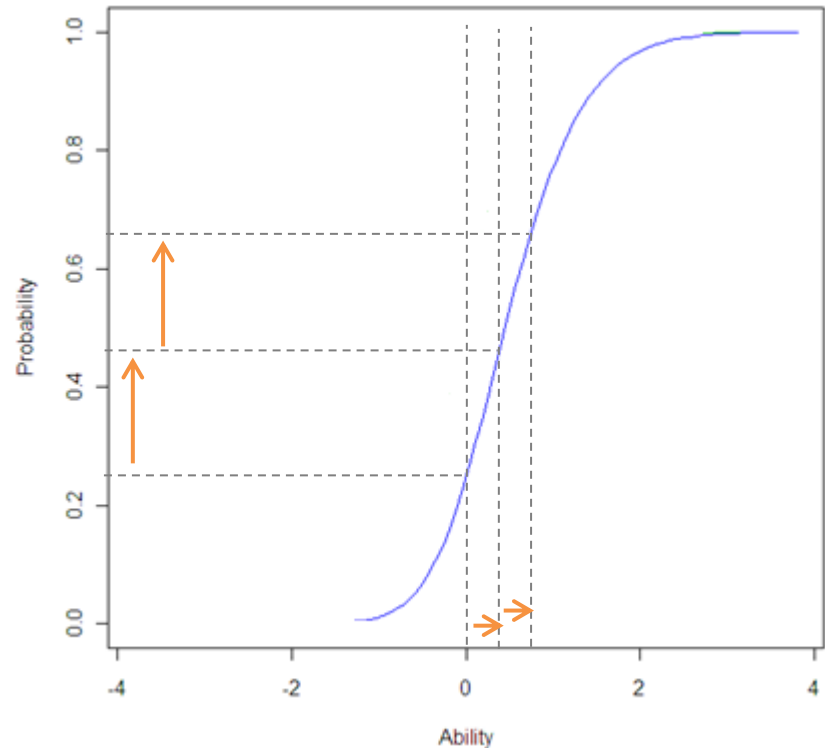
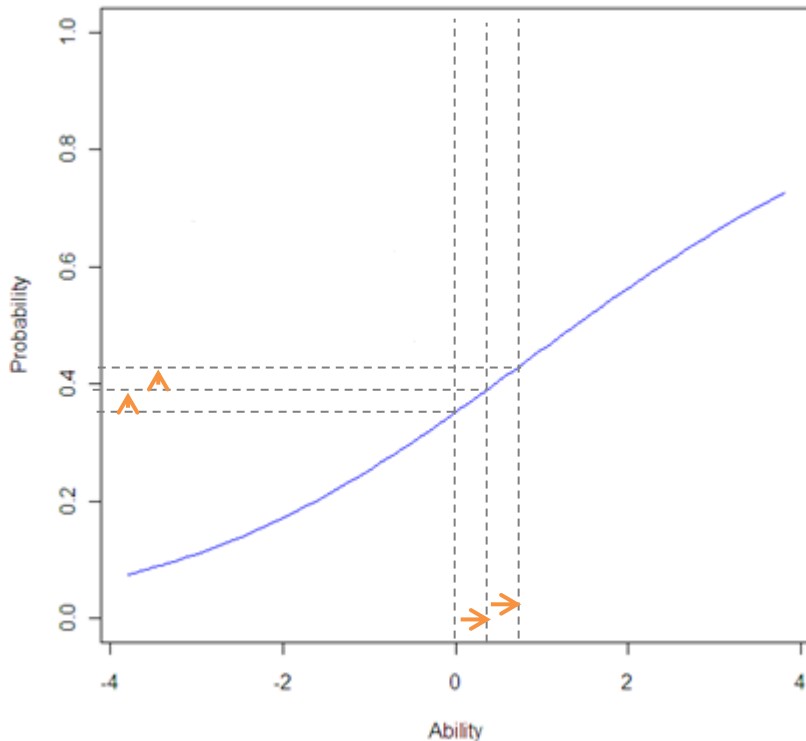


Schwellenparameter ungeordnet

▶ GPCM – Diskriminationsparameter

Je größer α_j desto:

- steiler der Anstieg der (eingipfeligen) Itemfunktion
- enger liegen die Schwellenparameter beieinander (bei polytomen Items)
- besser die Unterscheidung dicht beieinander liegender Fähigkeits- bzw. Eigenschaftsausprägungen im Bereich des steilsten Anstieges
- höher die Iteminformation und somit niedriger der Standardfehler



▶ GPCM – Informationsfunktion

In welchem Bereich der latenten Variable bietet ein Test oder Items den höchsten Informationswert bzw. eine präzise Messung?

- *zB Auswahlverfahren für „Schumpeter-HAK“ – kognitiver Test:
→ Test bzw. Items mit hohem Informationswert im höheren Fähigkeitsbereich
wichtig*

▶ GPCM – TIF, IIF, ICIF

TIF – Test Information Function $I(\theta)$



IIF – Item Information Function $I_j(\theta)$



ICIF – Item-Category Information Function $I_{jh}(\theta)$

▶ TIF – Test Information Function

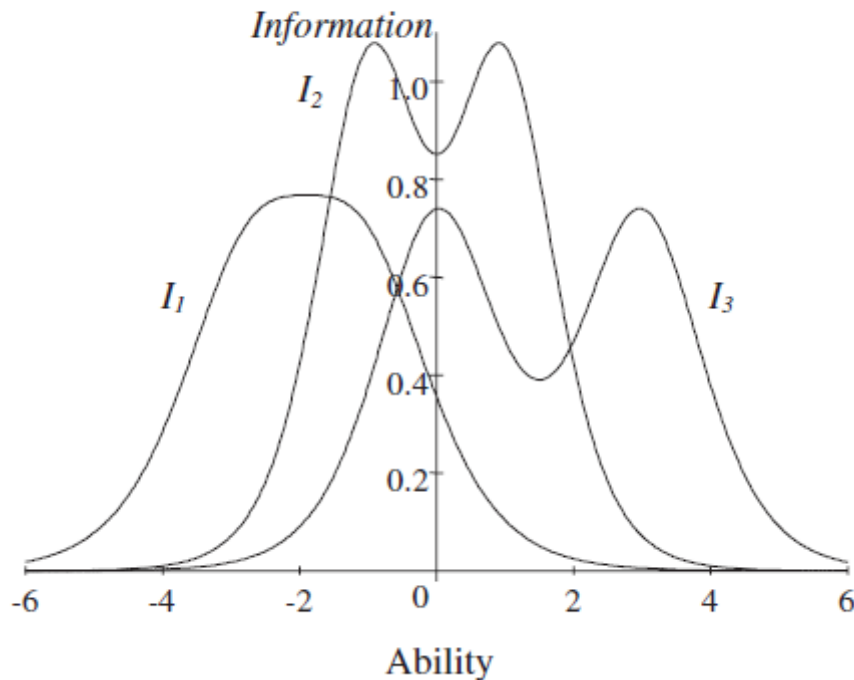
für n polytome Items: $I(\theta) = \sum_{j=1}^n I_j(\theta)$
(Samejima, 1969)

▶ GPCM – IIF

▶ IIF – Item Information Function $I_j(\theta) = a^2 \text{var}(X_j)$ (Donoghue, 1994)

- Summe der IIF = TIF
- kann so viele Gipfel wie Schwellenparameter haben
- > 1 Gipfel nur dann wenn Schwellenparameter geordnet sind

Information Functions of Three Trinary Generalized Partial Credit Model (GPCM) Items



$$\begin{aligned}\alpha_1 &= 1.5, \beta_{11} = -2.8, \beta_{12} = -1.0 \\ \alpha_2 &= 2.0, \beta_{21} = -1.0, \beta_{22} = 1.0 \\ \alpha_3 &= 1.7, \beta_{31} = 0, \beta_{32} = 3.0\end{aligned}$$

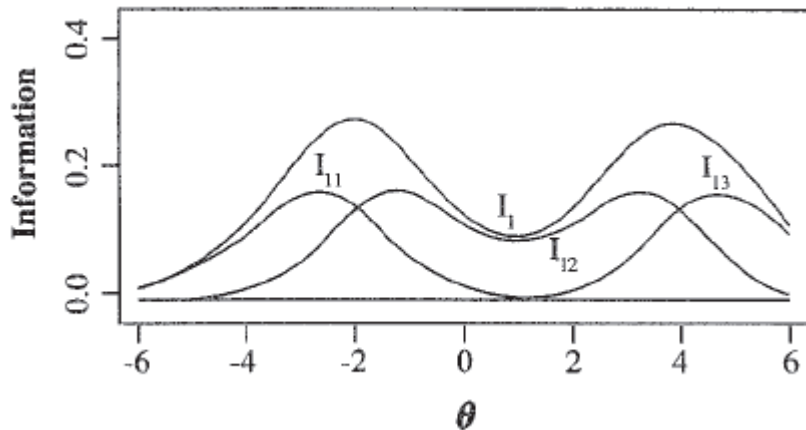
▶ GPCM – ICIF

▶ **ICIF – Item-Category Information Function** $I_{jh}(\theta) = P_{jh}(\theta)I_j(\theta)$ (Muraki, 1993)

- Summe der ICIF = IIF $I_j(\theta) = \sum_{h=0}^H I_{jh}(\theta)$

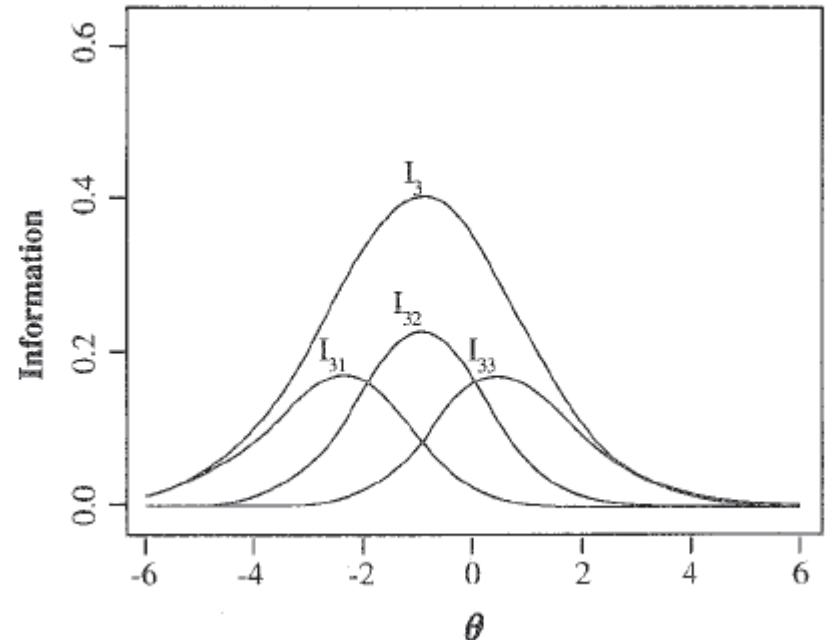
- Schnittpunkte der ICIF's sind die Schwellenparameter β_{jh}

$\alpha_1 = 1, \beta_{12} = -2.0, \beta_{13} = 4.0$



(Muraki, 1993)

$\alpha_3 = 1, \beta_{32} = -2.0, \beta_{33} = 0$



(Muraki, 1993)

▶ GPCM – Itemparameterschätzung I

MML (Marginal Maximum Likelihood)

marginal probability des Antwortpatterns i_v :
$$P(i_v) = \int_{-\infty}^{\infty} P(i_v | \theta, \beta, \alpha) dG(\theta)$$

$$L_m = \frac{N!}{\prod_{i=1}^{h^n} r_i!} \prod_{i=1}^{h^n} \{P(i_v)\}^{r_i}$$



$$\ln L_m = \ln N! - \sum_{i=1}^{h^n} \ln r_i! + \sum_{i=1}^{h^n} r_i \ln P(i_v)$$

- beobachtetes response pattern i_v ; $i = 1, \dots, h^n$ Antwortpattern
 - Verteilung d. latenten Variable: Annahme $\theta \sim N(0,1)$
 - r_i Anzahl der Personen mit pattern i
 - N Gesamtanzahl der befragten Personen
- r_i multinomial-verteilt mit den Parametern N und $P(i_v)$

▶ GPCM – Itemparameterschätzung II

MML (Marginal Maximum Likelihood)-EM Algorithmus

→ zu schätzende Parameter: α_j, β_{jh}

1. partielle Ableitung von $\ln L$ nach den gesuchten Parametern und „Null setzen“;
 θ -Parameter mit Verteilungsfunktion (Annahme: Normalverteilung)

$$\frac{\partial \ln L}{\partial \alpha_j} = 0 \quad \frac{\partial \ln L}{\partial \beta_{jh}} = 0$$

! Problem: Integral

✓ Lösung: Approximation mittels **Gauss-Hermite-Quadrature** unter Anwendung des **EM-Algorithmus**

„E“ – Schritt 1: Berechnung: vorläufig erwartete Häufigkeit d. Kat. h bei Item j
und vorläufig erwartete Stichprobengröße

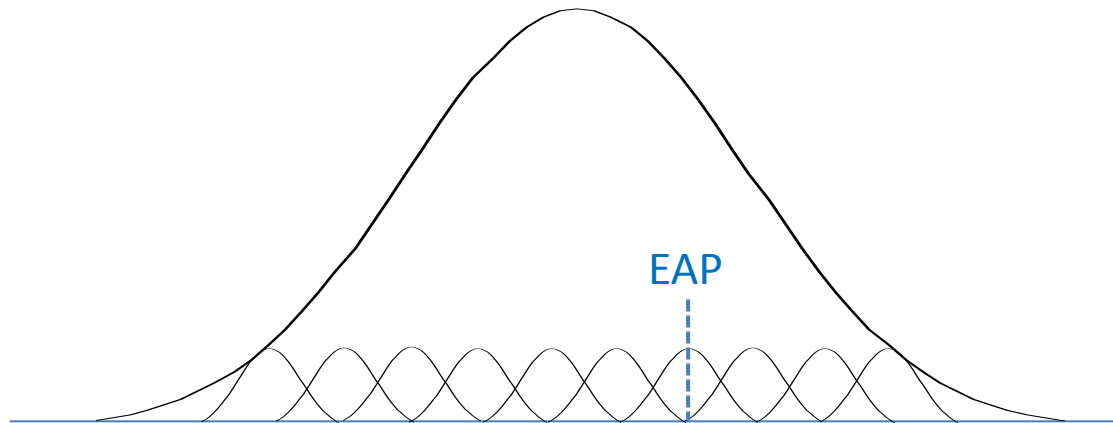
„M“ – Schritt 2: MML-Schätzungen

„E“ und „M“ werden wiederholt (EM-cycle) bis alle Schätzungen stabil sind

▶ GPCM – Personenparameterschätzung I

▶ **EAP** (*Expected A Posteriori*) – Schätzer $EAP = E(\theta | x_v, \beta, \alpha, f(\theta))$

- Schätzwerte der θ -Verteilungsparameter (aus MML) = Basis für Berechnung der Erwartungswerte einzelner Personen- bzw. Fähigkeitsparameter (=expected a posteriori Schätzer)
- EAP-Werte = Mittelwerte der *a-posteriori-Verteilung* von θ_v



vgl. Rost, 2004

- Bayesianisches Verfahren (Parameterschätzung für *0/max score* möglich)

▶ GPCM – Personenparameterschätzung II

- ▶ **MAP (Maximum A Posteriori) Schätzer** bzw. Bayes Modal Estimate
 - Bayesianisches Verfahren
 - ähnlich wie EAP
 - Berechnung des MODUS der *a-posteriori-Verteilung*
- ▶ **ML (Maximum Likelihood) Schätzer**
 - Fähigkeitsparameter zB mittels Newton-Raphson Algorithmus angepasst
 - Maximierung der log Likelihood der Antwortpatterns (Daten) unter Bedingung der Modellparameter
- ▶ **WML (Weighted Maximum Likelihood) Schätzer**
 - Warm (1989)
 - Bayesianisches Verfahren
 - Maximierung der log Likelihood der Personenparameter unter Bedingung der Daten und Itemparameter
- ▶ **JML (Joint Maximum Likelihood) Schätzer**
 - gemeinsame Schätzung von Item- und Personenparameter

▶ GPCM vs. Rasch – Annahmen

- Unidimensionalität
- Lokale Unabhängigkeit
- Monotonie
- keine Suffizienz – erschöpfende Statistik
nicht Randsummen, sondern (mit Diskriminationsparametern) gewichteter Summenscore → „Innere“ der Matrix wird benötigt
- keine spezifische Objektivität
Item- und Personenparameter nicht separierbar; Vergleich zweier Personen v und v' hängt von dem Item (bzw. dessen Diskriminationsparameter α_j) ab; Rangfolge der Itemschwierigkeit ist abhängig von Personenstichprobe
- keine „Stichprobenunabhängigkeit“
- keine Item-Stichprobenunabhängigkeit

▶ GPCM – Goodness of Fit I

▶ Likelihood-Ratio Chi-Squared Statistic

$$G^2 = -2 \ln \Lambda$$
$$= 2 \sum \text{beobachtet} \ln \frac{\text{beobachtet}}{\text{angepasst}}$$

- GoF auf **Itemlevel**; GoF gesamt = \sum LR-Statistiken d. einzelnen Items
- Klassifizierung der Personen in homogene Gruppen W (hinsichtlich d. geschätzten θ -Parameter)

$$G_j^2 = 2 \sum_{w=1}^{W_j} \sum_{h=1}^{H_j} r_{wjh} \ln \frac{r_{wjh}}{N_{wj} P_{jh}(\bar{\theta}_w)}$$

Muraki In Linden, 1997

W Intervalle auf dem latenten Kontinuum θ

r_{wjh} beobachtete Häufigkeit d. Antw. in Kat. h bei Item j im w -ten Intervall

N_{wj} Anzahl der Personen im w -ten Intervall

$\bar{\theta}_w$ Intervall-Mittelwert

$P_{jh}(\bar{\theta}_w)$ Werte der angepassten ICRF-Funktion

▶ GPCM – Goodness of Fit II

▶ „nonparametric“ Fit Statistic for GPCM (Liang, Wells, 2009)

- Erweiterung von Douglas and Cohen (2001)
 - Vergleich von nichtparametrischen IRF mit parametrischen IRF
 - Ist Distanz zwischen den zwei IRF's signifikant?
 - Root Integrated Squared Error (RISE); Signifikanzniveau mittels bootstrapping
-

▶ Modellwahl

PCM ist in GPCM genestet

→ LR-Test (anova)

▶ GPCM - Anwendungsbereiche

- Education: *zB*
 - [ISAT](#)-Illinois Standards Achievement Test (2011)
 - Information-seeking behaviour (Timmers, Glas, 2010)
 - ...
- Psychologie: *zB*
 - CAT (Van Rijn, Eggen, Hemker, Sanders, 2002)
 - ADHD (Gomez, 2008)
 - ...
- Marketing/Tourismus: *zB*
 - siehe [Anwendungsbeispiel in R](#)
 - ...
- ...

Anwendungsmöglichkeiten

1. Messung von latenten Fähigkeiten, kognitiven Prozessen, Eigenschaften usw.
2. Testkonstruktion –analyse, Entwicklung von Skalen, Itemselektion, ...

Mathematics Short-Response Scoring Rubric

List *one* set of 5 whole numbers for which the following is true:

- The mode is 8 and the mean is 6.

Show your work.

SCORE LEVEL	DESCRIPTION
2	Completely correct response, including correct work shown and/or correct labels/units if called for in the item
1	Partially correct response
0	No response, or the response is incorrect

Quelle: http://www.isbe.state.il.us/assessment/htmls/sample_books.htm

→ Long-Response Scoring von Score Level 0 - 4

▶ GPCM – Anwendung in R

- R-Package *itm*

Version 0.9-7

Rizopoulos (2011)

- Datensatz

Erhebung zum Thema „Urlaub in Australien“; Informationssuche – travel blog

(VK5-Gruppe, Tourismus u. Freizeitmarketing, 2009)

Teildatensatz „Exploratory Behaviour“

n = 579

4 Items (6-stufige Likert-items)

▶ GPCM – Datensatz

> `descript(exbehav)`

Descriptive statistics for the 'exbehav' data-set

Sample:

4 items and 579 sample units; 0 missing values

Proportions for each level of response:

	1	2	3	4	5	6
website1_r10	0.1071	0.2038	0.1658	0.1675	0.2211	0.1347
website1_r11	0.0397	0.1520	0.1503	0.2453	0.2608	0.1520
website1_r12	0.1278	0.2003	0.1572	0.1744	0.2142	0.1261
website1_r13	0.2453	0.2003	0.1831	0.1675	0.1209	0.0829

Frequencies of total scores:

	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Freq	3	10	10	23	25	26	30	37	41	60	50	57	33	33	44	29	22	15	10	10	11

Cronbach's alpha:

	value
All Items	0.6791
Excluding website1_r10	0.7650
Excluding website1_r11	0.5644
Excluding website1_r12	0.5265
Excluding website1_r13	0.5558

...

► GPCM – Modellanpassung

Itemparameterschätzung mittels MML (Marginal Maximum Likelihood)

```
> res_gpcm<- gpcm(exbehav,constraint=c("gpcm"))
```

```
> summary(res_gpcm)
```

```
> coef(res_gpcm)
```

```
Call:  
gpcm(data = exbehav, constraint = c("gpcm"))
```

Model Summary:

log.Lik	AIC	BIC
-3786.303	7620.606	7725.278

	Catgr.1	Catgr.2	Catgr.3	Catgr.4	Catgr.5	Dscrmn
ws1_r10	-4.163	1.060	-0.077	-1.513	3.254	0.168
ws1_r11	-2.592	-0.680	-0.750	0.145	1.294	0.961
ws1_r12	-1.288	-0.248	-0.102	0.240	1.345	1.286
ws1_r13	-0.429	-0.083	0.317	0.968	1.508	0.998

Coefficients:

\$website1_r10

	value	std.err	z.value
Catgr.1	-4.163	1.182	-3.522
Catgr.2	1.060	0.868	1.222
Catgr.3	-0.077	0.861	-0.090
Catgr.4	-1.513	0.881	-1.718
Catgr.5	3.254	1.018	3.197
Dscrmn	0.168	0.034	4.955

...

▶ GPCM – Personenparameterschätzung/EB

```
> theta<-factor.scores(res_gpcm, resp.patterns = exbehav, method=c("EB"))
```



methods for class *gpcm*:

- EB „Empirical Bayes“
- EAP „Expected a Posteriori“
- MI „Multiple Imputation“

(siehe auch Rizopoulos, 2011)

```
> theta
```

Call:

```
gpcm(data = exbehav, constraint = c("gpcm"))
```

Scoring Method: Empirical Bayes

Factor-Scores for specified response patterns:

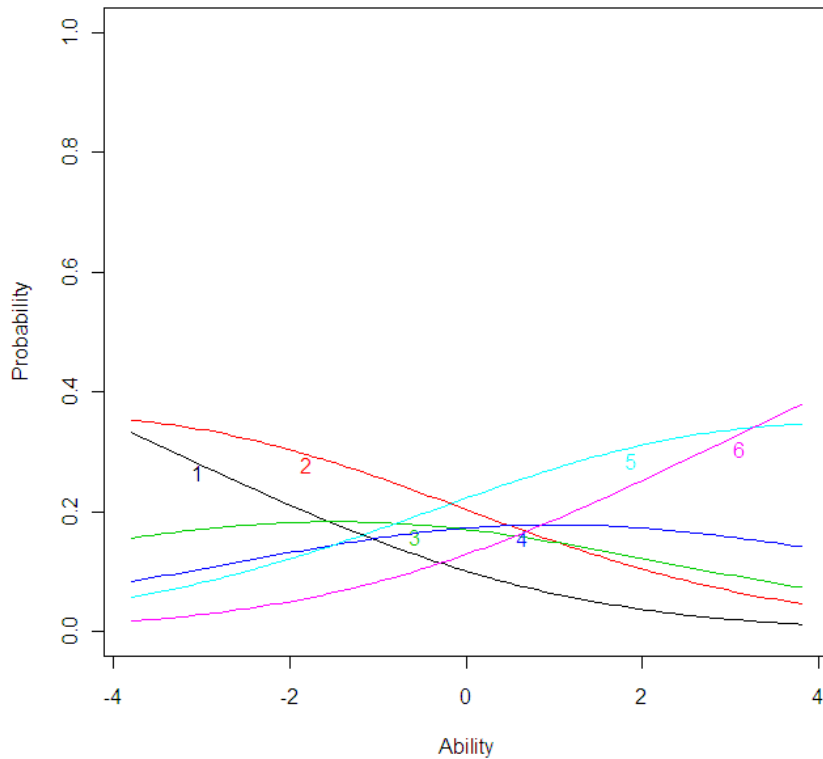
	website1_r10	website1_r11	website1_r12	website1_r13	Obs	Exp	z1	se.z1
1	1	6	6	6	4	0.459	1.626	0.568
2	5	4	3	4	1	0.964	0.094	0.404
3	2	6	4	4	1	0.643	0.562	0.434
4	3	5	3	1	1	0.674	-0.292	0.409
5	4	5	4	3	3	1.259	0.274	0.412

...

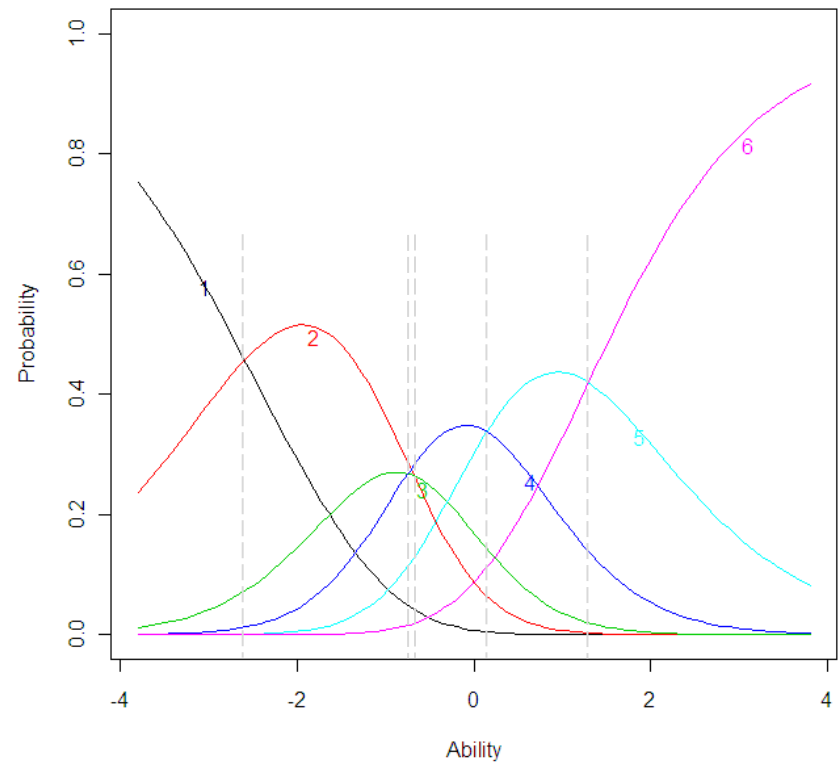
► GPCM – Item Response Category Characteristic Curves I

```
> plot(res_gpcm, type=c("ICC"))
```

Item Response Category Characteristic Curves - Item: website1_r10



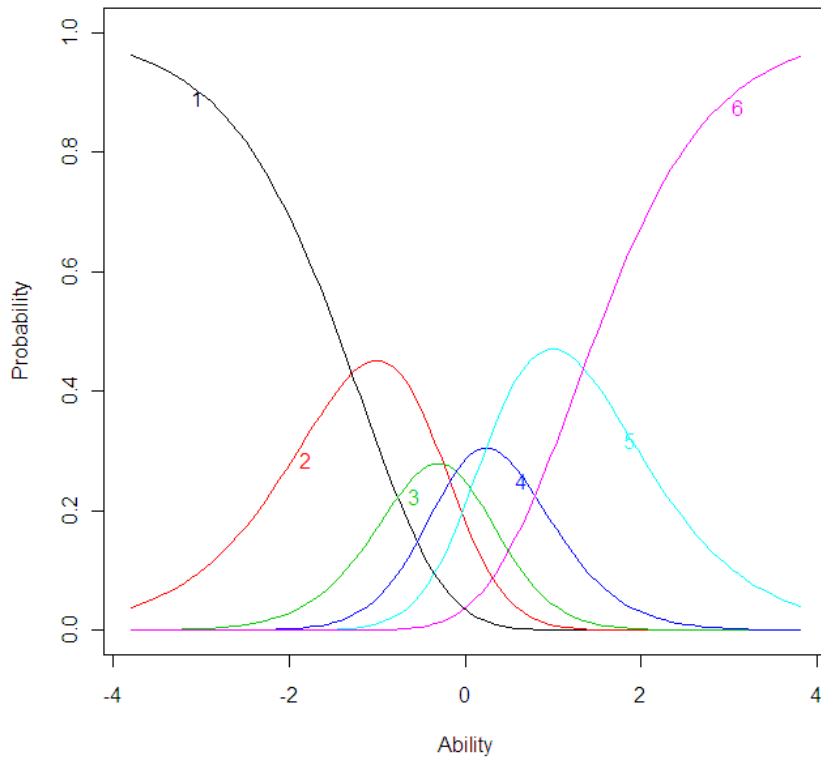
Item Response Category Characteristic Curves - Item: website1_r11



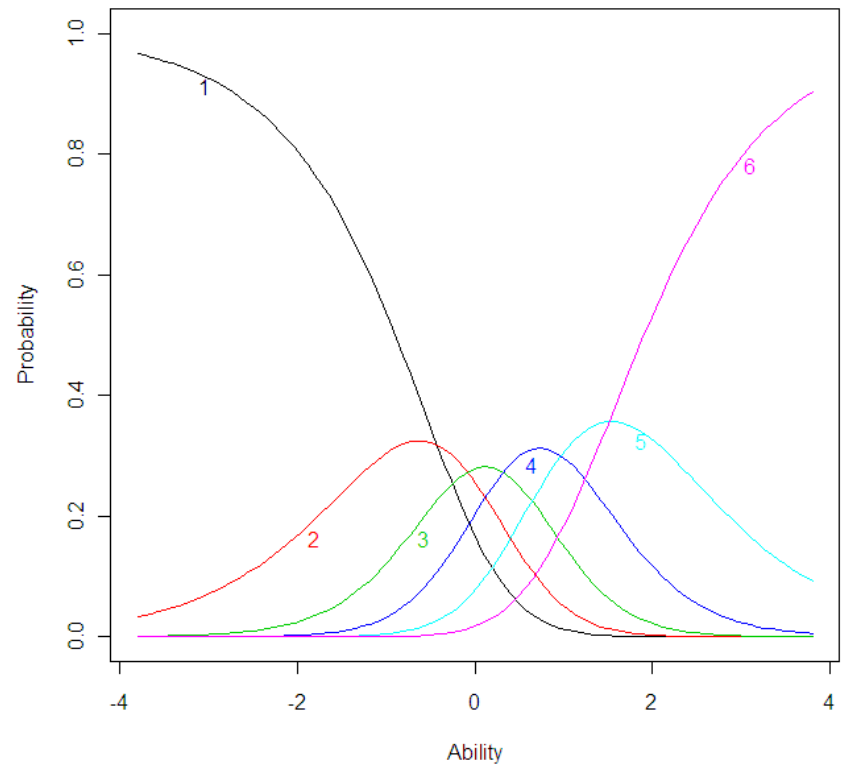
▶ GPCM – Item Response Category Characteristic Curves II

```
> plot(res_gpcm, type=c("ICC"))
```

Item Response Category Characteristic Curves - Item: website1_r12

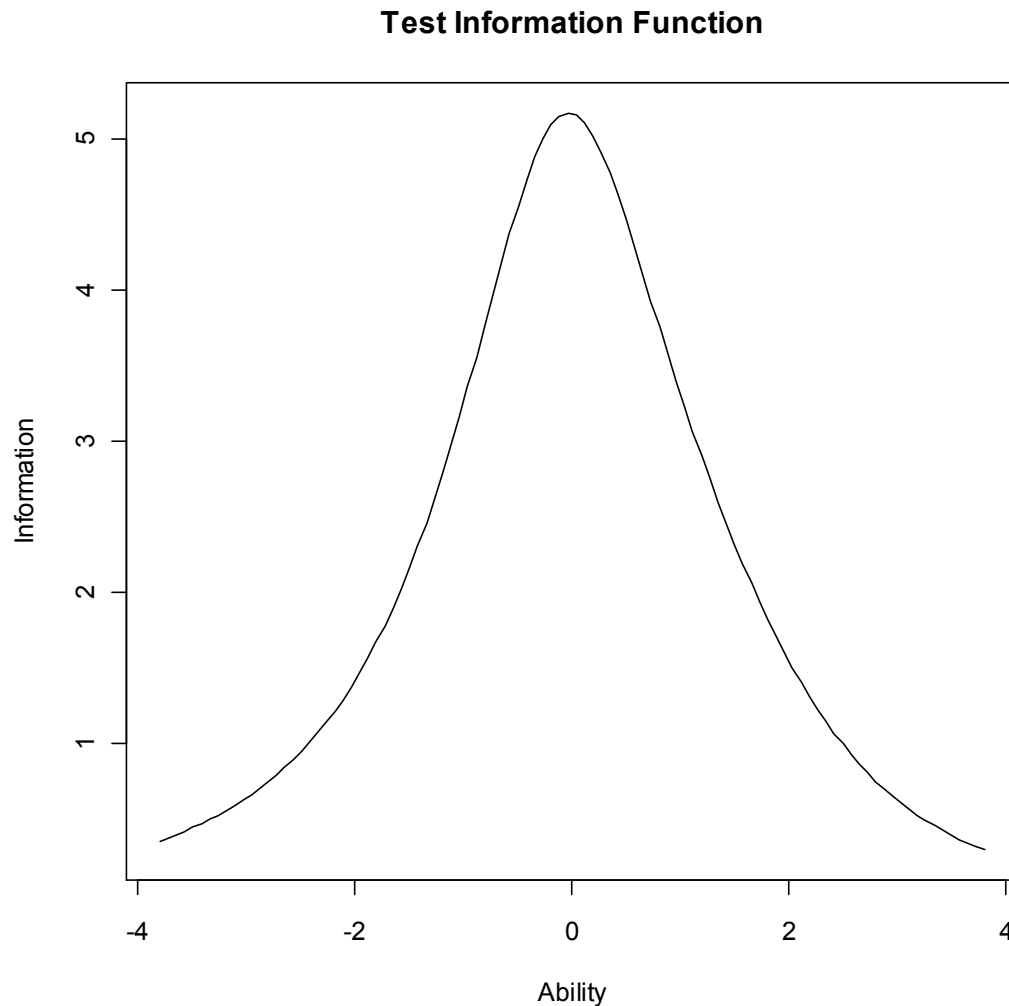


Item Response Category Characteristic Curves - Item: website1_r13



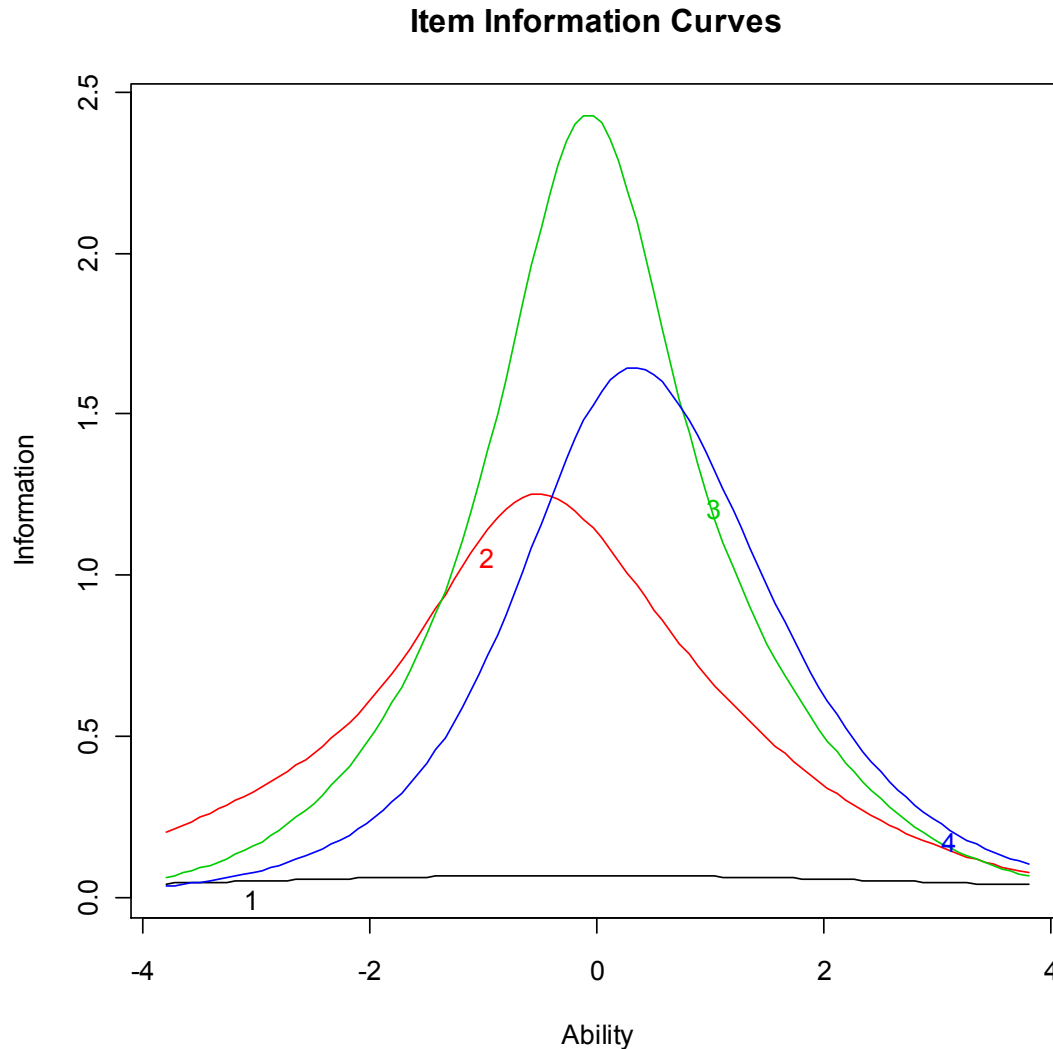
▶ GPCM – TIF (Test Information Function)

```
> plot(res_gpcm, type=c("IIC"), items=0)
```



► GPCM – IIF (Item Information Function)

```
> plot(res_gpcm, type=c("IIC"))
```



► GPCM – GoF (Pearson chi-squared statistic)

$$\sum_{r=1}^{2^P} \frac{\{O(r) - E(r)\}^2}{E(r)}$$

$O(r)$ beobachtete Häufigkeit des Antwortpatterns r
 $E(r)$ erwartete Häufigkeit des Antwortpatterns r
 P Anzahl der Items

Rizopoulos, 2011

```
> GoF.gpcm(res_gpcm, simulate.p.value = FALSE)
```

Pearson chi-squared Goodness-of-Fit Measure

Call:

```
gpcm(data = exbehav, constraint = c("gpcm"))
```

Tobs: 2416.84

df: 1271

p-value: <0.001

▶ GPCM – GoF (Parametric Bootstrap)

> GoF.gpcm(res_gpcm)

Parametric Bootstrap Approximation to Pearson chi-squared Goodness-of-Fit Measure

Call:

```
gpcm(data = exbehav, constraint = c("gpcm"))
```

Tobs: 2416.84

data-sets: 100

p-value: 0.03

Achtung! Bootstrap-Verfahren:
nach einigen Wiederholungen:
p-value variiert zwischen 0.01,
0.02 und 0.03, ...

▶ GPCM – Exploratory Behaviour - Items

GoF-Ergebnis → GPCM-Modell passt nicht gut, daher ev.:

- nachträgliche Zusammenlegung von benachbarten Antwortkategorien
- Itemselektion (siehe zB Rijn et al., 2002)
- Anpassung und Prüfung eines PCM, GRM (Graded Response Model)
Modellvergleich PCM/GPCM zB mittels LR-test
Modellvergleich PCM/GPCM/GRM zB mittels Informationskriterien
- nachträgliche Dichotomisierung der 6-stufigen Antwortkategorien

► Modellvergleich PCM / GPCM

```
> res_pcm<- gpcm(exbehav,constraint=c("rasch"))  
> anova(res_pcm, res_gpcm)
```

Likelihood Ratio Table

	AIC	BIC	log.Lik	LRT	df	p.value
res_pcm	7838.08	7925.31	-3899.04		20	
res_gpcm	7620.61	7725.28	-3786.30	225.48	24	<0.001

```
> coef(res_pcm)
```

	Catgr.1	Catgr.2	Catgr.3	Catgr.4	Catgr.5	Dscrmn
website1_r10	-1.388	-0.155	-0.049	0.008	1.182	1
website1_r11	-2.323	-0.539	-0.676	0.106	1.132	1
website1_r12	-1.157	-0.090	-0.120	0.102	1.240	1
website1_r13	-0.301	-0.051	0.263	0.834	1.303	1



Graded Response Model (GRM)

► Graded Response Model (GRM)

Samejima (1969)

- Messung von latenten Fähigkeiten, Eigenschaften anhand polytomer Items
- Items mit geordneten Antwortkategorien $h = 0, 1, \dots, H$
- zB
 - Likert-Items
 - Schulnoten
 - partial credits/Teilpunkte (bei Leistungstests, ...)
- die Anzahl der Kategorien kann für die einzelnen Items variieren
- „abgegrenzte“ Regionen auf dem latenten Kontinuum („category boundaries“)
- α_j - Diskriminationsparameter können (wie beim GPCM) bei allen n Items variieren
- kumulatives Modell
- Schwellenparameter sind zwingend geordnet (beim PCM bzw. GPCM nicht)
- Kategoriewahrscheinlichkeiten nur indirekt berechenbar; sogenanntes „Differenzmodell“

▶ Graded Response Model (GRM)

- Homogene Modelle
 - normal ogive GRM (Lord and Novick, 1968)
 - logistic GRM (Birnbaum, 1968) ✓
- Heterogene Modelle

Samejima's (logistic) Graded Response Model

Herleitung:

(kumulative Schwellen)Wahrscheinlichkeit, dass bei Item j Kategorie h *oder* $> h$ (gegenüber Kategorie $< h$) gewählt wird → 2 PL-Birnbaum-Modell

$$P(X_{vj, \geq h}) = \frac{P(X_{vj, \geq h})}{P(X_{vj, < h}) + P(X_{vj, \geq h})}$$

► GRM – Beispiel

Kauf von 25 Stück Statistikbücher: Nettopreis € 27,50 je Stk., Bezugskosten € 30,00 (netto), Ust 10%. Wie hoch ist der Bruttoeinkaufspreis?

Lösungsschritte

keine Antwort,
falscher Ansatz

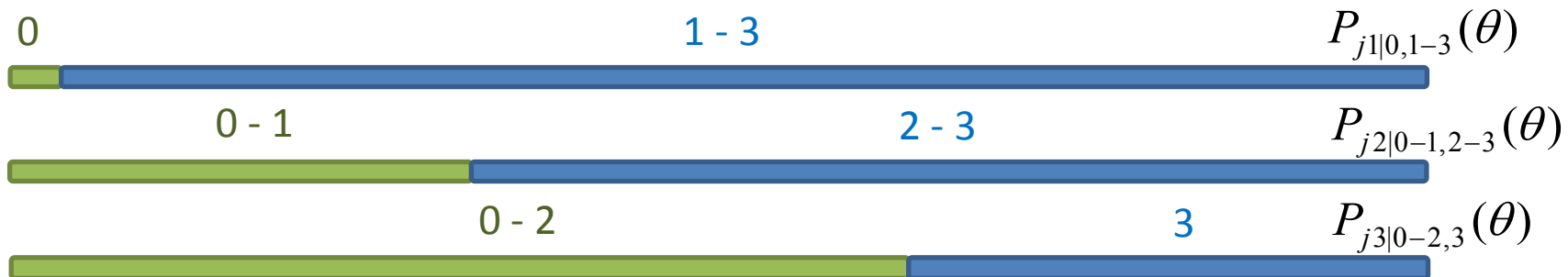
$$\begin{array}{l} 27,50 * 25 = \\ 687,50 \end{array}$$

$$\begin{array}{l} 687,50 + 30 = \\ 717,50 \end{array}$$

$$\begin{array}{l} 717,50 * 1,1 = \\ 789,25 \end{array}$$



kumulativ-dichotome Vergleiche mittels 2PL – Birnbaum-Modell



▶ GRM – Schwellenparameter (*thresholds*)

- Schwellenparameter GPCM \neq GRM
an der Stelle des latenten Kontinuums, wo die **Wahrscheinlichkeit**,
mindestens in **Kategorie h** zu antworten, **0.5** beträgt
- Anzahl d. Schwellenparameter = $h - 1$ Antwortkategorien (wie beim GPCM)
- geordnet: $\beta_{j4} > \beta_{j3} > \beta_{j2} > \beta_{j1}$
- Schwellenparameter sind nicht die Schnittpunkte zweier benachbarter
Kategoriewahrscheinlichkeitskurven
- zwischen den Schwellenparametern liegt in der Mitte das Maximum der
jeweils dazwischenliegenden Kategoriewahrscheinlichkeit

► GRM – Cumulative Category Response Function

Kumulative Wahrscheinlichkeit für Person v bei Item j in Kategorie h oder höher zu antworten:

$$\pi_{vj, \geq h}^* = \frac{\exp((\theta_v - \beta_{jh})\alpha_j)}{1 + \exp((\theta_v - \beta_{jh})\alpha_j)}$$

aus zB vier Antwortkategorien resultieren drei kum. Wahrscheinlichkeitskurven:

- $\pi_{vj \geq 1}^*$, P für Antwort in Kategorie 1 oder höher
- $\pi_{vj \geq 2}^*$, P für Antwort in Kategorie 2 oder höher
- $\pi_{vj, 3}^*$, P für Antwort in Kategorie 3

Wie kann die Wahrscheinlichkeit für eine bestimmte Kategorie h berechnet werden?

→ Differenz zweier „benachbarter“ kumulierter Wahrscheinlichkeitskurven

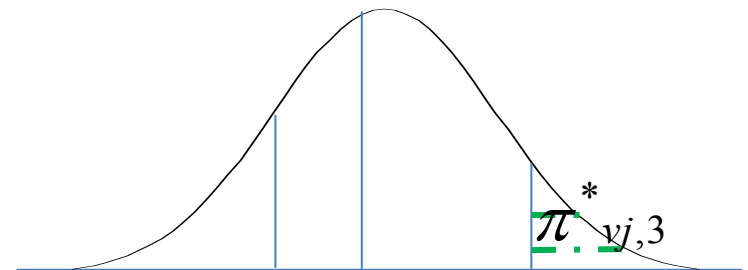
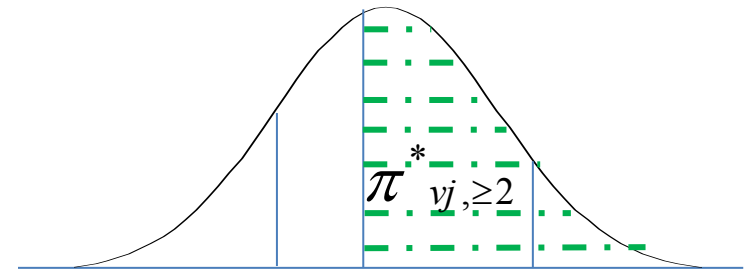
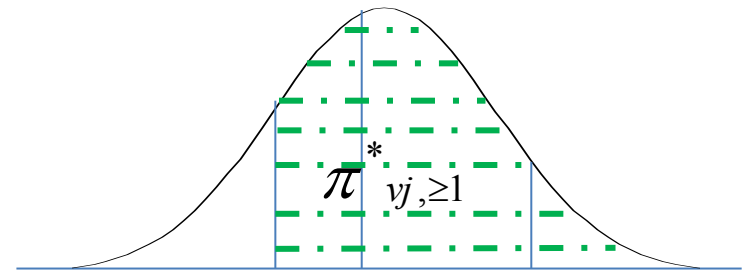
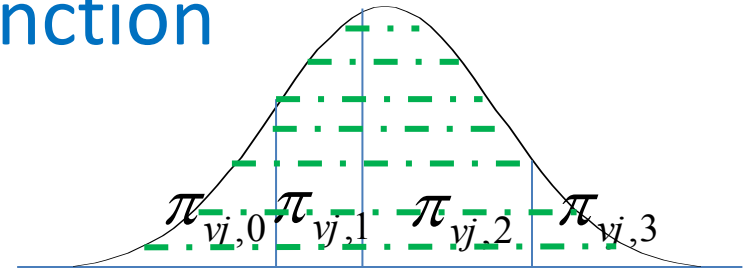
► GRM – Category Response Function

$$\pi_{vj,0} = 1 - \pi^*_{vj,\geq 1}$$

$$\pi_{vj,1} = \pi^*_{vj,\geq 1} - \pi^*_{vj,\geq 2}$$

$$\pi_{vj,2} = \pi^*_{vj,\geq 2} - \pi^*_{vj,\geq 3}$$

$$\pi_{vj,3} = \pi^*_{vj,3}$$



▶ GRM – Kategoriewahrscheinlichkeiten

„*Processing Function*“ (Samejima, 1999)

Wahrscheinlichkeit, dass eine Person ν bei Item j den Schritt h erfolgreich absolviert hat unter der gemeinsamen Bedingung (*joint condition*), dass:

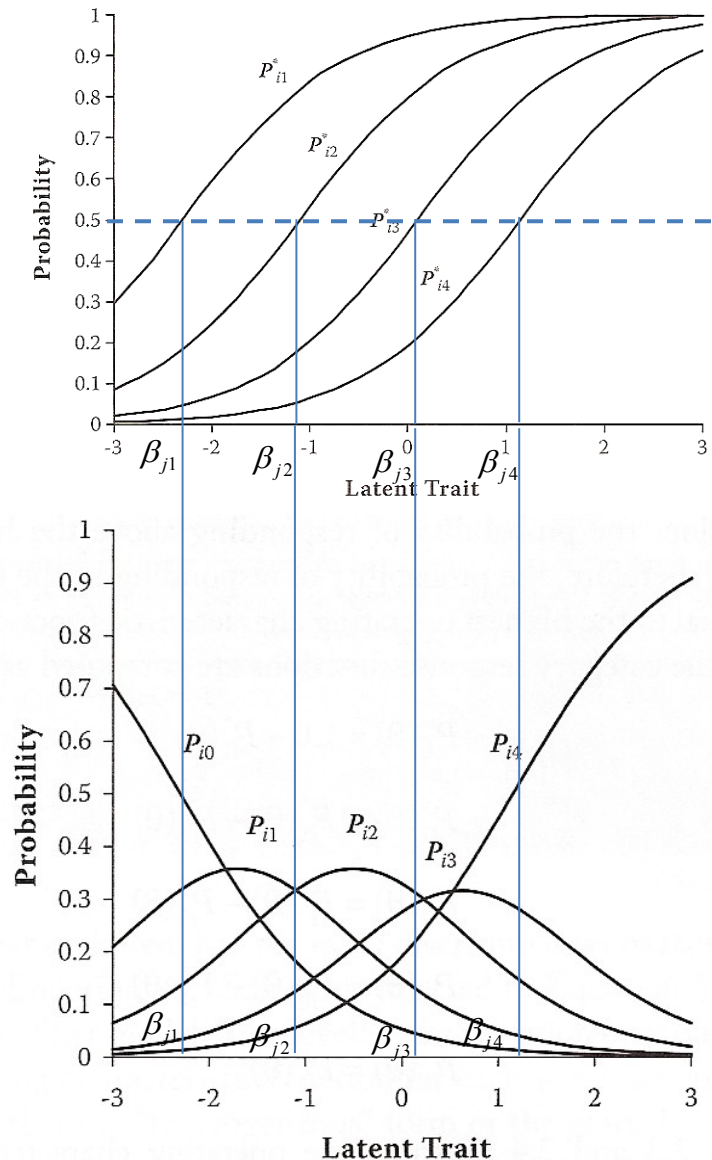
- der Fähigkeitsparameter θ gegeben ist
- die Schritte bis zum Schritt h ($h-1$) bereits erfolgreich absolviert wurden

► GRM – Thresholds/Kategoriewahrscheinlichkeiten

Schwellenwahrscheinlichkeiten
(Operating Characteristic Curves)

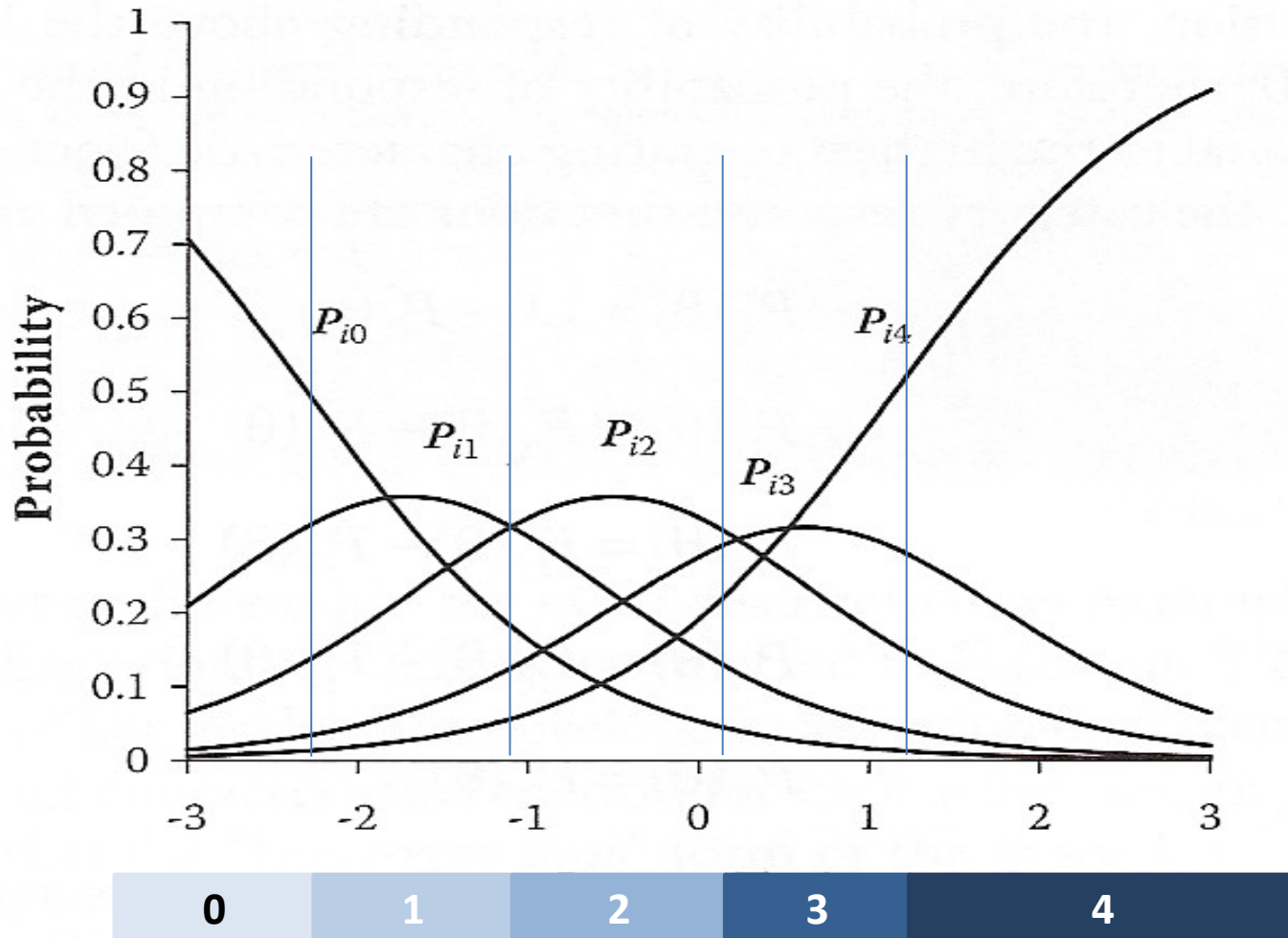
$$\alpha = 1.25, \beta_{j1} = -2.3, \beta_{j2} = -1.1, \beta_{j3} = 0.1, \beta_{j4} = 1.15$$

Kategoriewahrscheinlichkeiten
(Category Response Curves)

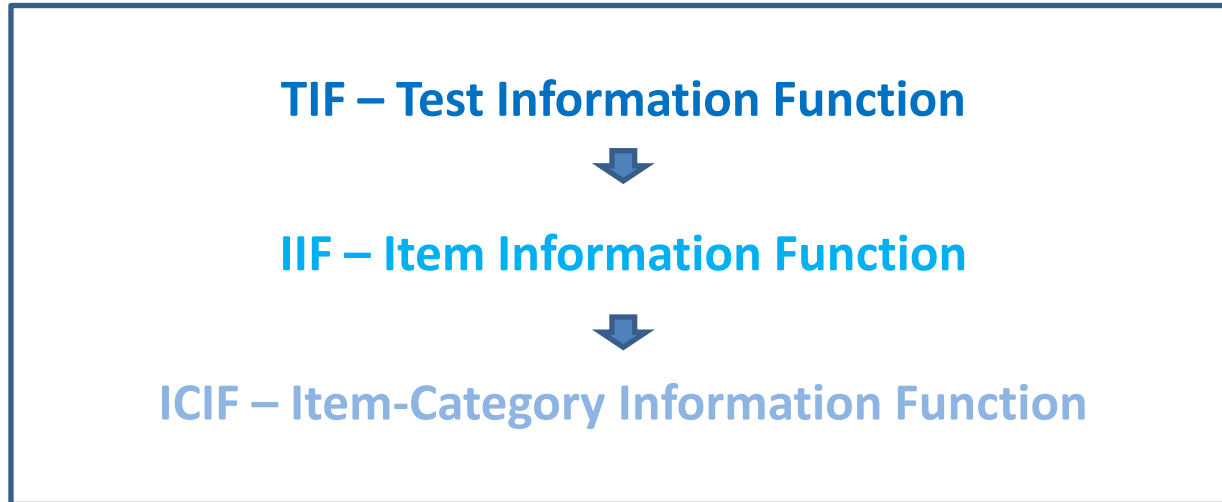


► GRM – „category boundaries“

$$\alpha = 1.25, \beta_{j_1} = -2.3, \beta_{j_2} = -1.1, \beta_{j_3} = 0.1, \beta_{j_4} = 1.15$$



▶ GRM – TIF, IIF, ICIF



siehe zB Samejima, 1969; Samejima, 1977; Nering, Ostini, 2010

▶ GRM – Itemparameterschätzung

MML (Marginal Maximum Likelihood) – **EM** (siehe GPCM)

$$L_m = \prod_v \int_{-\infty}^{\infty} P(i_v | \theta_v, \beta, \alpha) dG(\theta) = \prod_{v=1}^N P_{i_v} = \prod_I P_i^{r_i}$$

N Anzahl der Befragten

$G(\theta)$ Verteilungsfunktion der Fähigkeitsparameter (Annahme: Normalverteilung)

i_v Antwortpattern der Person v

r_i Häufigkeit mit der ein bestimmtes Antwortpattern i auftritt

▶ GRM – Personenparameterschätzung / GOF

Personenparameter

- MAP (Maximum A Posteriori) Schätzer
- ML (Maximum Likelihood) Schätzer
- WML (Weighted Maximum Likelihood) Schätzer

...

→ *siehe GPCM*

Goodness of Fit

- Likelihood-Ratio Chi-Square Statistic = Σ Chi-Square Statistics der einzelnen Items

→ *siehe GPCM*

Modellvergleich

- LR-Test: Vergleich GPCM und GRM nicht möglich (da nicht genestet)

▶ Graded Response Model (GRM)

Modellannahmen

- unidimensional
- lokale Unabhängigkeit

im Vergleich zu „Rasch“-Annahmen:

- keine suffiziente Statistik
- keine spezifische Objektivität
- keine „Stichprobenunabhängigkeit“

Anwendungsmöglichkeiten

1. Messung von:

- latenten Fähigkeiten, Eigenschaften von Personen
- kognitiven Prozessen (zB Problemlösen in Form von Schritten/Stufen)
 - homogener Fall
 - heterogener Fall (zB „[acceleration model](#)“) siehe zB Samejima In Linden (1997); Samejima (1999)

2. Entwicklung von Skalen, Itemselektion, ...

► GRM – Beispiel „acceleration model“

Nach einer Preissenkung von 15% und der Inanspruchnahme von 2% Skonto wurden, innerhalb der Kassafrist, für ein Statistikbuch € 27,50 bezahlt. Wie hoch war der ursprüngliche Nettopreis?

Lösungsschritte → jeder Schritt enthält > 1 Subschritt (Subprozess)

keine Antwort,
falscher Ansatz

$$\frac{27,50}{0,98} = 28,06$$

$$\frac{28,06}{1,1} = 25,51$$

$$\frac{25,51}{0,85} = 30,01$$



Subprozesse:

- 1)Stufenweise, retrograde Berechnung
- 2)Reihenfolge (1. Skontoabzug)
- 3)Welcher Wert gegeben? → 98%
- 4)Was ist gefragt?
- 5)Prozentrechnung i.H.

Subprozesse:

- 1)Reihenfolge (2. Ust)
- 2)Umsatzsteuer bei Büchern = 10%!
- 3)Welcher Wert gegeben? → 110%
- 4)Was ist gefragt?
- 5)Prozentrechnung a.H.

Subprozesse:

- 1)Abzug des Rabatts; Welcher Wert gegeben? → 85%
- 2)Was ist gefragt?
- 3)Prozentrechnung i.H.

▶ GRM – Anwendungsbereiche

- Psychologie: *zB*
 - NEO PI-R (Reise, Henson, 2000)
 - Subjective Well-Being – „*Mood-Items*“ (Baker, Rounds, Zevon, 2000)
 - CAT (De Ayala, Dodd, Koch, 1990)
 - ...
- Education: *zB*
 - TOEFL (Tang, 1996)
 - ISAT-Illinois Standards Achievement Test (2011)
 - ...
- Marketing/Tourismus: *zB*
 - siehe [Anwendungsbeispiel in R](#)
 - ...
- ...

Write a Sentence



airport terminal / so

SCORE	RESPONSE DESCRIPTION
3	<p>The response consists of ONE sentence that:</p> <ul style="list-style-type: none"> ■ has no grammatical errors, ■ contains forms of both key words used appropriately, AND ■ is consistent with the picture.
2	<p>The response consists of one or more sentences that:</p> <ul style="list-style-type: none"> ■ have one or more grammatical errors that do not obscure the meaning, ■ contain BOTH key words, (but they may not be in the same sentence and the form of the word(s) may not be accurate), AND ■ are consistent with the picture.
1	<p>The response:</p> <ul style="list-style-type: none"> ■ has errors that interfere with meaning, ■ omits one or both key words, OR ■ is not consistent with the picture.
0	<p>The response is blank, written in a foreign language, or consists of keystroke characters.</p>

► GRM – Modellanpassung

Itemparameterschätzung mittels MML

```
> res_grm<- grm(exbehav)
```

```
> summary(res_grm)
```

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

Model Summary:

log.Lik	AIC	BIC
-3766.157	7580.314	7684.985

Coefficients:

```
$website1_r10  
value  
Extrmt1 -4.321  
Extrmt2 -1.634  
Extrmt3 -0.207  
Extrmt4 1.194  
Extrmt5 3.792  
Dscrmn 0.512  
...
```

Vergleich: GRM/GPCM



```
> coef(res_grm)
```

	Extrmt1	Extrmt2	Extrmt3	Extrmt4	Extrmt5	Dscrmn
ws1_r10	-4.321	-1.634	-0.207	1.194	3.792	0.512
ws1_r11	-2.416	-1.145	-0.518	0.295	1.388	1.893
ws1_r12	-1.375	-0.534	-0.044	0.488	1.384	2.512
ws1_r13	-0.876	-0.203	0.362	1.000	1.777	2.186

```
> coef(res_gpcm)
```

	Catgr.1	Catgr.2	Catgr.3	Catgr.4	Catgr.5	Dscrmn
ws1_r10	-4.163	1.060	-0.077	-1.513	3.254	0.168
ws1_r11	-2.592	-0.680	-0.750	0.145	1.294	0.961
ws1_r12	-1.288	-0.248	-0.102	0.240	1.345	1.286
ws1_r13	-0.429	-0.083	0.317	0.968	1.508	0.998

→ GRM: „zwingend“ geordnete
Schwellenparameter

► GRM – Überblick/R-Befehle

Modellanpassung

```
> res_grm<- grm(exbehav)
```

Personenparameterschätzung

```
> theta<-factor.scores(res_grm, resp.patterns = exbehav, method=c("EB"))
```

Item Operation Characteristic Curves

```
> plot(res_grm, type=c("OCCu"))
```

Item Response Category Characteristic Curves

```
> plot(res_grm, type=c("ICC"))
```

Test Information Function

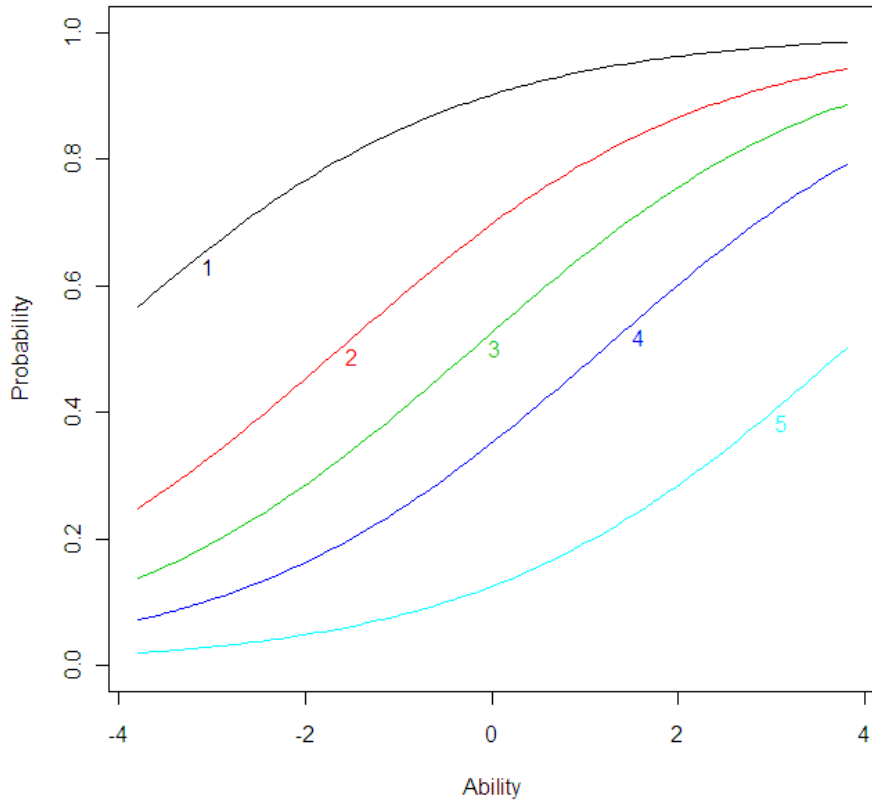
```
> plot(res_grm, type=c("IIC"), items=0)
```

Item Information Function

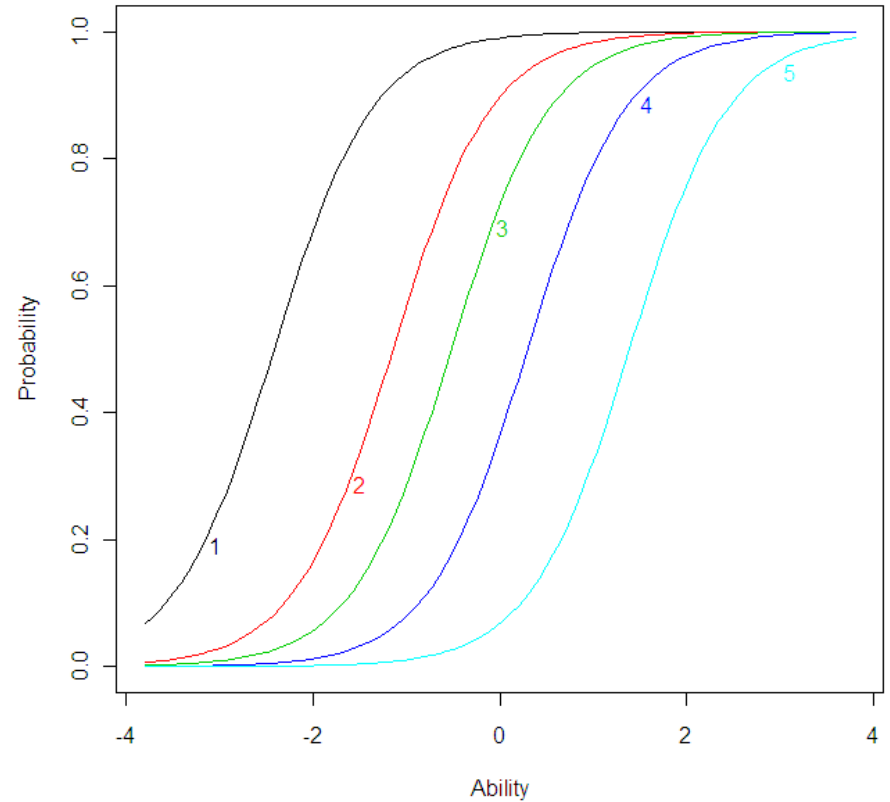
```
> plot(res_grm, type=c("IIC"))
```

▶ IOCC – Item Operation Characteristic Curves I

Item Operation Characteristic Curves - Item: website1_r10

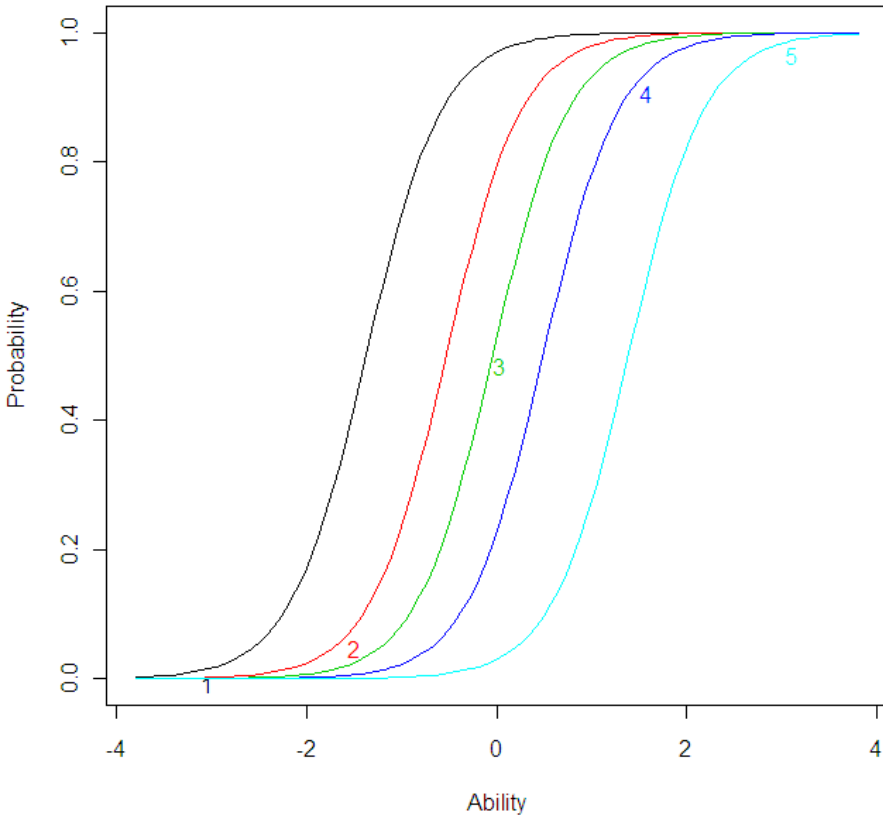


Item Operation Characteristic Curves - Item: website1_r11

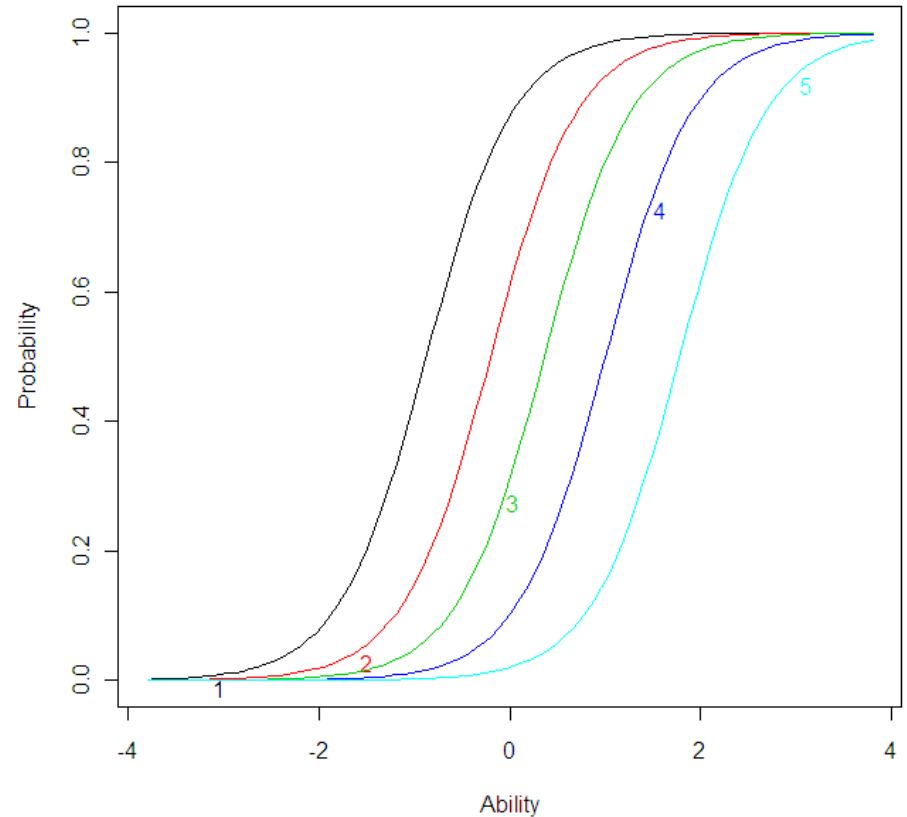


▶ IOCC – Item Operation Characteristic Curves II

Item Operation Characteristic Curves - Item: website1_r12



Item Operation Characteristic Curves - Item: website1_r13

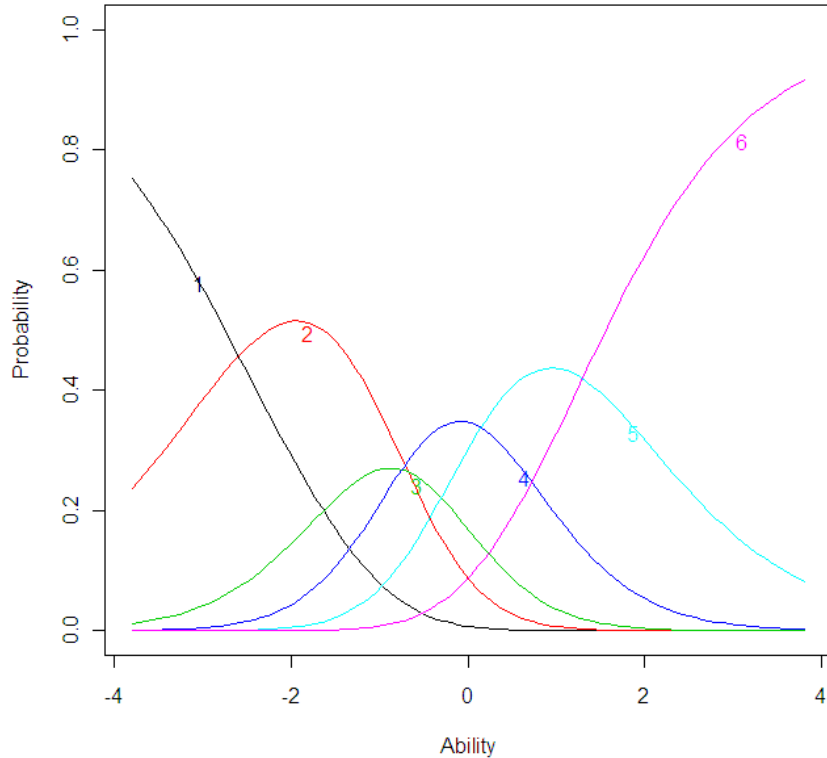


► ICC – Vergleich GPCM/GRM

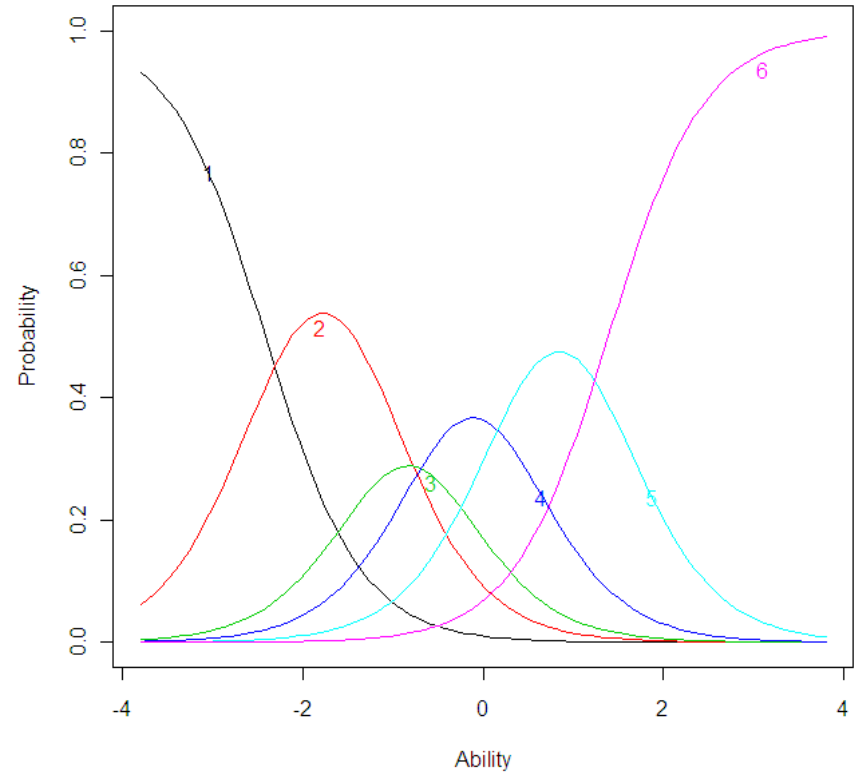
GPCM

GRM

Item Response Category Characteristic Curves - Item: website1_r11



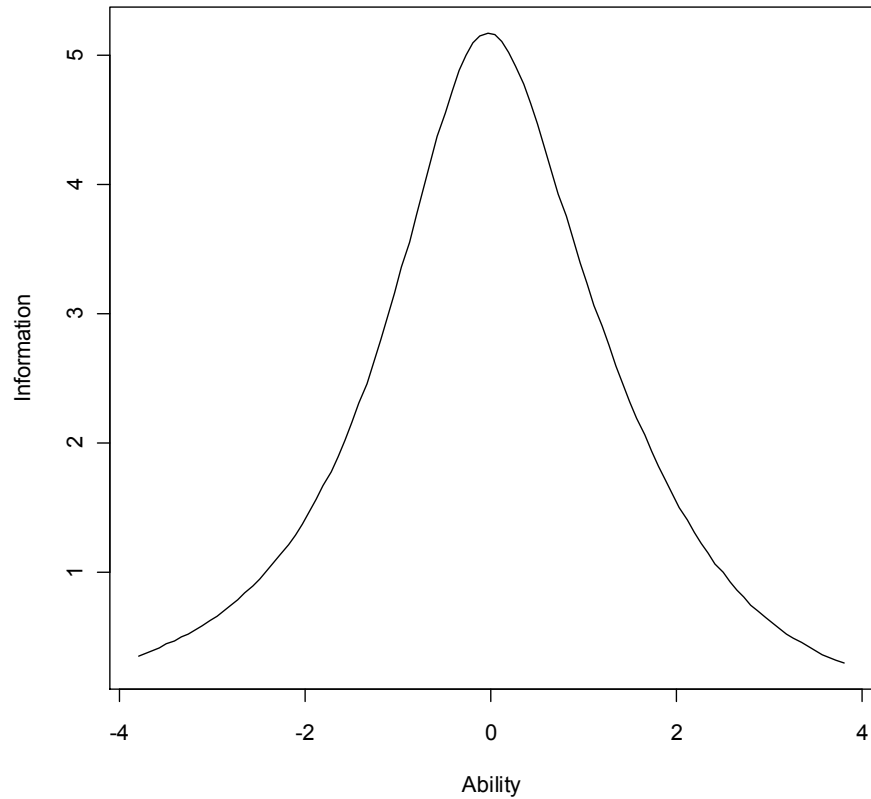
Item Response Category Characteristic Curves - Item: website1_r11



► TIF – Vergleich GPCM/GRM

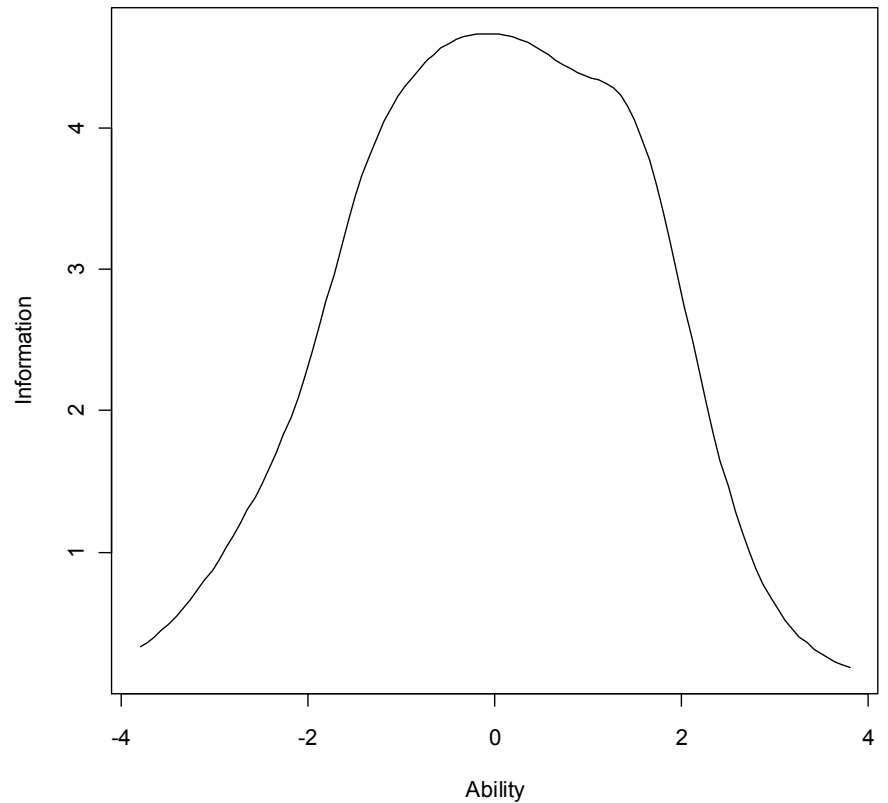
GPCM

Test Information Function



GRM

Test Information Function



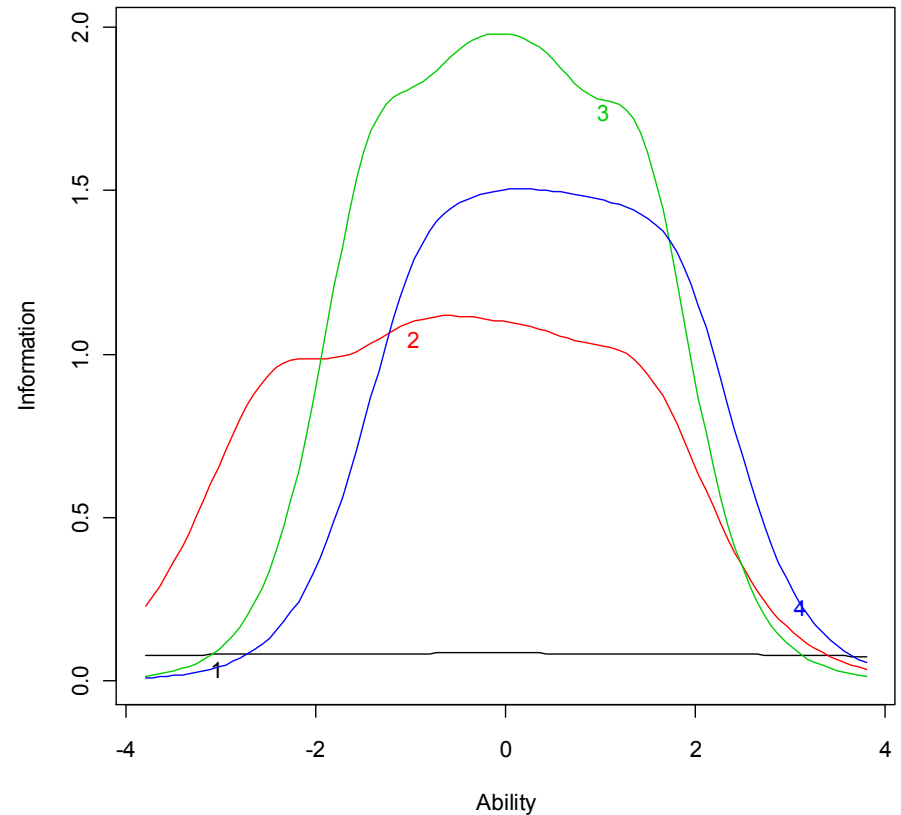
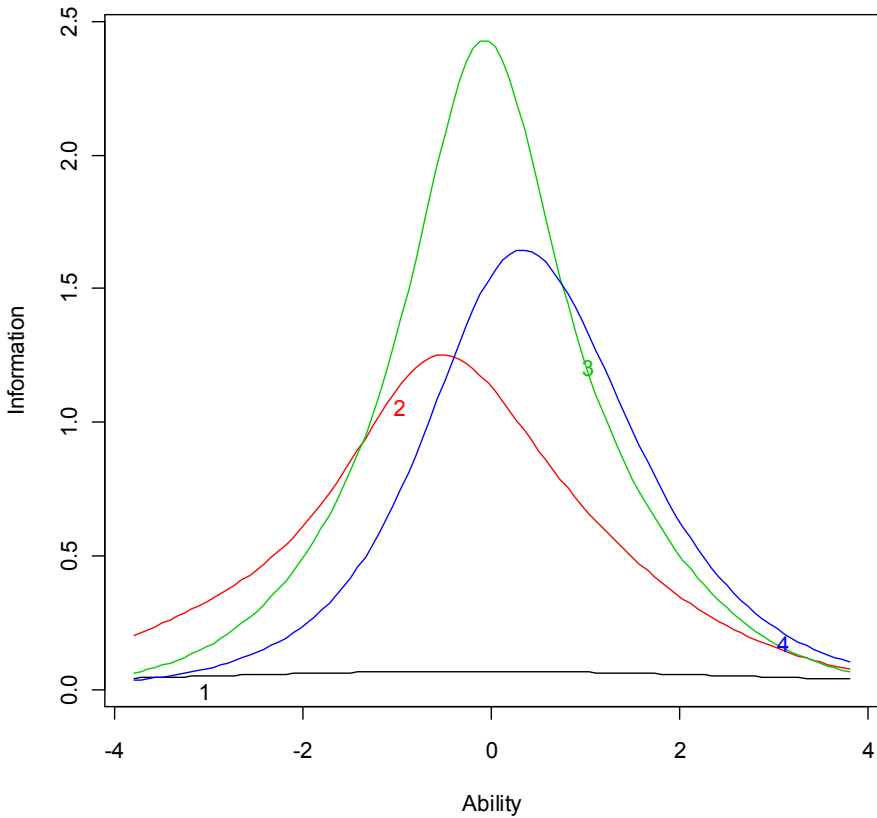
IIF – Vergleich GPCM/GRM

GPCM

GRM

Item Information Curves

Item Information Curves



▶ Modellvergleich PCM / GPCM / GRM

```
> summary(res_pcm)
```

```
Call:  
gpcm(data = exbehav, constraint = c("rasch"))
```

Model Summary:

log.Lik	AIC	BIC
-3899.041	7838.083	7925.309

```
> summary(res_grm)
```

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

Model Summary:

log.Lik	AIC	BIC
-3766.157	7580.314	7684.985

```
> summary(res_gpcm)
```

```
Call:  
gpcm(data = exbehav, constraint = c("gpcm"))
```

Model Summary:

log.Lik	AIC	BIC
-3786.303	7620.606	7725.278

► free Software für GPCM und GRM



- **ltm** (Rizopoulos, 2011)
- **plink** (Weeks, 2011)
- **pp** Personenparameterschätzung (MLE, WLE) für GPCM (Reif, 2011)

WinGen

Windows software that Generates IRT parameters and item responses

Kyung (Chris) T. Han, 2010. URL: <http://www.hantest.net/wingen>



Kyung (Chris) T. Han, 2011. URL: <http://www.hantest.net/irteq>

...



Generalized Partial Credit Model
Graded Response Model

- Baker, J., Rounds, J., and Zevon, M. (2000). A Comparison of Graded Response and Rasch Partial Credit Models with Subjective Well-Being. *Journal of Educational and Behavioral Statistics*, 25 (3), 253-270.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In Lord, F. and Novick, M. (Eds.): *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- De Ayala, R., Dodd, B. and Koch, W. (1990). A computerized simulation of a flexilevel test and its comparison with a Bayesian computerized adaptive test. *Journal of Educational Measurement*, 27, 227-239.
- De Ayala, R. J. (2009). *The Theory and Practice of Item Response Theory*. Guilford Press, New York, London.
- Donoghue (1994). In Rijn, P. W., Eggen, T. J. H. M., Hemker, B. T., Sanders, P. F. (2002). Evaluation of Selection Procedures for Computerized Adaptive Testing With Polytomous Items. *Applied Psychological Measurement*, 26, 393-411.
- Douglas, J. and Cohen, A. (2001). Nonparametric item response function estimation for assessing parametric model fit. *Applied Psychological Measurement*, 25, 234-243.
- Gomez, R. (2008). Parent ratings of the ADHD items of the disruptive behavior rating scale: Analyses of their IRT properties based on the generalized partial credit model. *Personality and Individual Differences*, 45, 181-186.
- Liang, T., Wells, C. (2009). A Model Fit Statistic for Generalized Partial Credit Model. *Educational and Psychological Measurement*, 69, 913-928.
- Lord, F. and Novick, M. (1986). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Masters, G. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.
- Van der Linden, W. and Hambleton, R. (Eds) (1997). *Handbook of modern item response theory*. Springer-Verlag, New York.
- Muraki, E. (1992). A generalized partial credit model: application of an EM algorithm. *Applied Psychological Measurement*, 16, 159-176.
- Muraki, E. (1993). Information Functions of the Generalized Partial Credit Model. *Applied Psychological Measurement*, 17, 351-363.

- Nering, M, Ostini, R. (2010). *Handbook of polytomous item response theory models*. Routledge, New York.
- Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: The University of Chicago Press.
- Reise, S. and Henson, J. (2000). Computerization and adaptive administration of the NEO-PI-R. *Assessment*, 7, 347–364.
- Rijn, P. W., Eggen, T. J. H. M., Hemker, B. T., Sanders, P. F. (2002). Evaluation of Selection Procedures for Computerized Adaptive Testing With Polytomous Items. *Applied Psychological Measurement*, 26. 393-411.
- Rizopoulos, D. (2011). Package ltm. URL: <http://cran.r-project.org/web/packages/ltm/ltm.pdf>.
- 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 (2004). *Testtheorie-Testkonstruktion*. Hans Huber Verlag, Bern.
- Samejima, F. (1969). Estimation of Latent Ability Using a Response Pattern of Graded Scores (Psychometric Monograph No. 17). Richmond, VA: Psychometric Society.
- Samejima, F. (1997). Graded response model. In: van der Linden, W. and Hambleton, R. (Eds): *Handbook of modern item response theory*. New York: Springer-Verlag.
- Samejima, F. (1999). General Graded Response Model. *Annual NCME Meeting*, Montreal, Canada.
- Tang, K. (1996). Polytomous Item Response Theory Models and Their Applications in Large-Scale Testing Programs: Review of Literature. TOEFL Report MS-02.
- Timmers, C., Glas, C. (2010). Developing scales for information-seeking behaviour. *Journal of Documentation*, 66, 46-69.