UNIVERSITY OF TWENTE.

Statistical Analyses Using CAT Data

Cees Glas

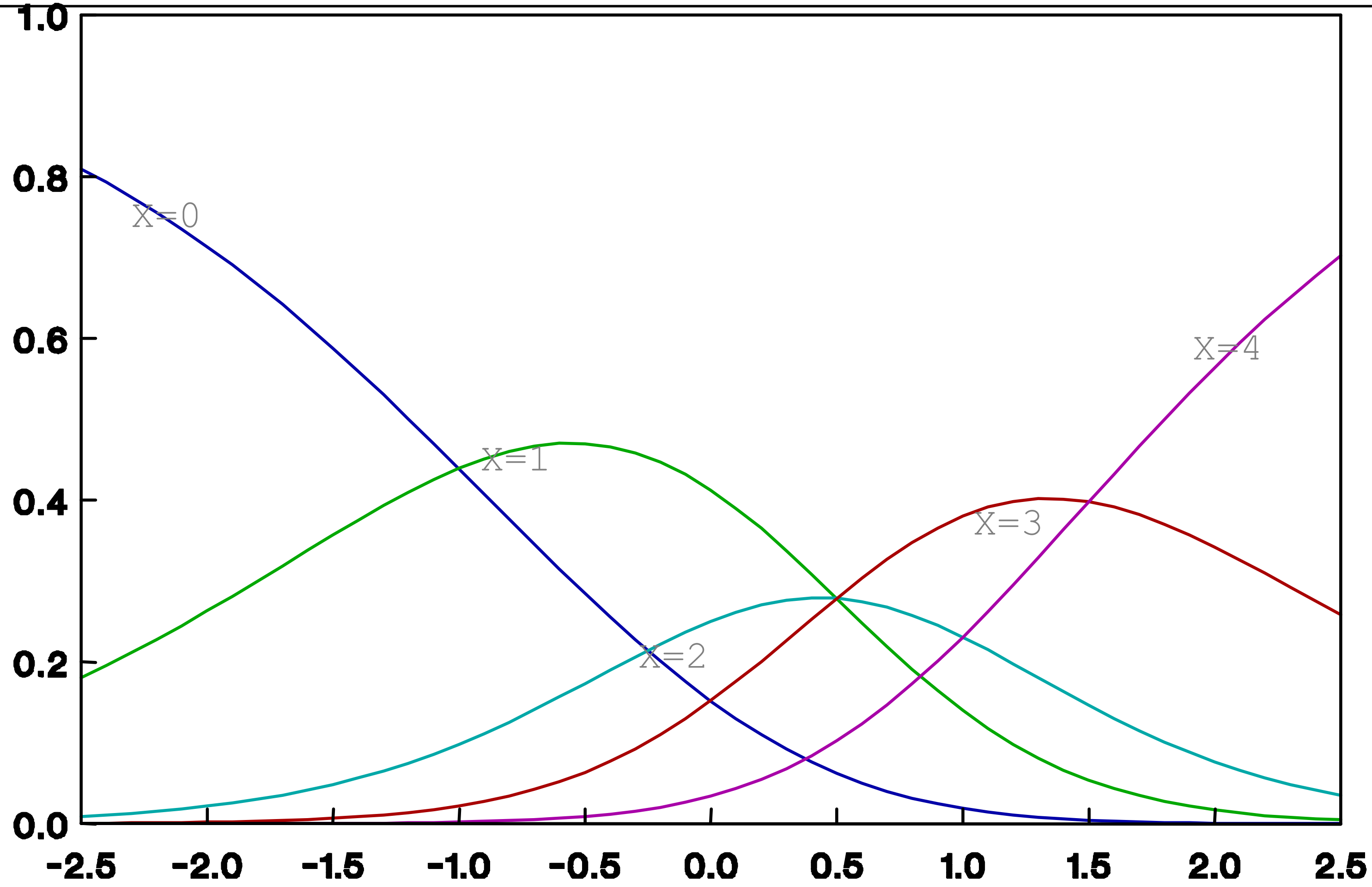
Overview

- Model Fit
  - How
  - ….and then?
    - Group specific item parameters
    - Multidimensional CAT
- Using theta-estimates in statistical analyses

IRT Model for Item with 5 response categories

Probability

Response Category



Latent Ability Scale

Item Calibration Design

	Cat	Total	Score	Design
Item1	0	234	78	1100
	1		79	
	2		62	
	3		15	
Item2	0	234	46	1100
	1		152	
	2		36	
Item3	0	472	16	1111
	1		173	
	2		268	
	3		15	
Item4	0	114	12	0010
	1		11	
	2		13	
	3		38	
	4		40	

Parameter Estimation

	Par	Cat	Estimate	Se
Item1	A		0.171	0.097
	B	1	-0.028	0.097
		2	0.234	0.138
		3	1.417	0.145
Item2	A		3.398	0.304
	B	1	-2.183	0.261
		2	2.121	0.304
Item3	A		0.171	0.128
	B	1	-2.406	0.150
		2	-0.455	0.206
		3	2.873	0.128
Item4	A		2.584	0.235
	B	1	-2.021	0.094
		2	-1.654	0.092
		3	-1.992	0.260
		4	-0.414	0.235

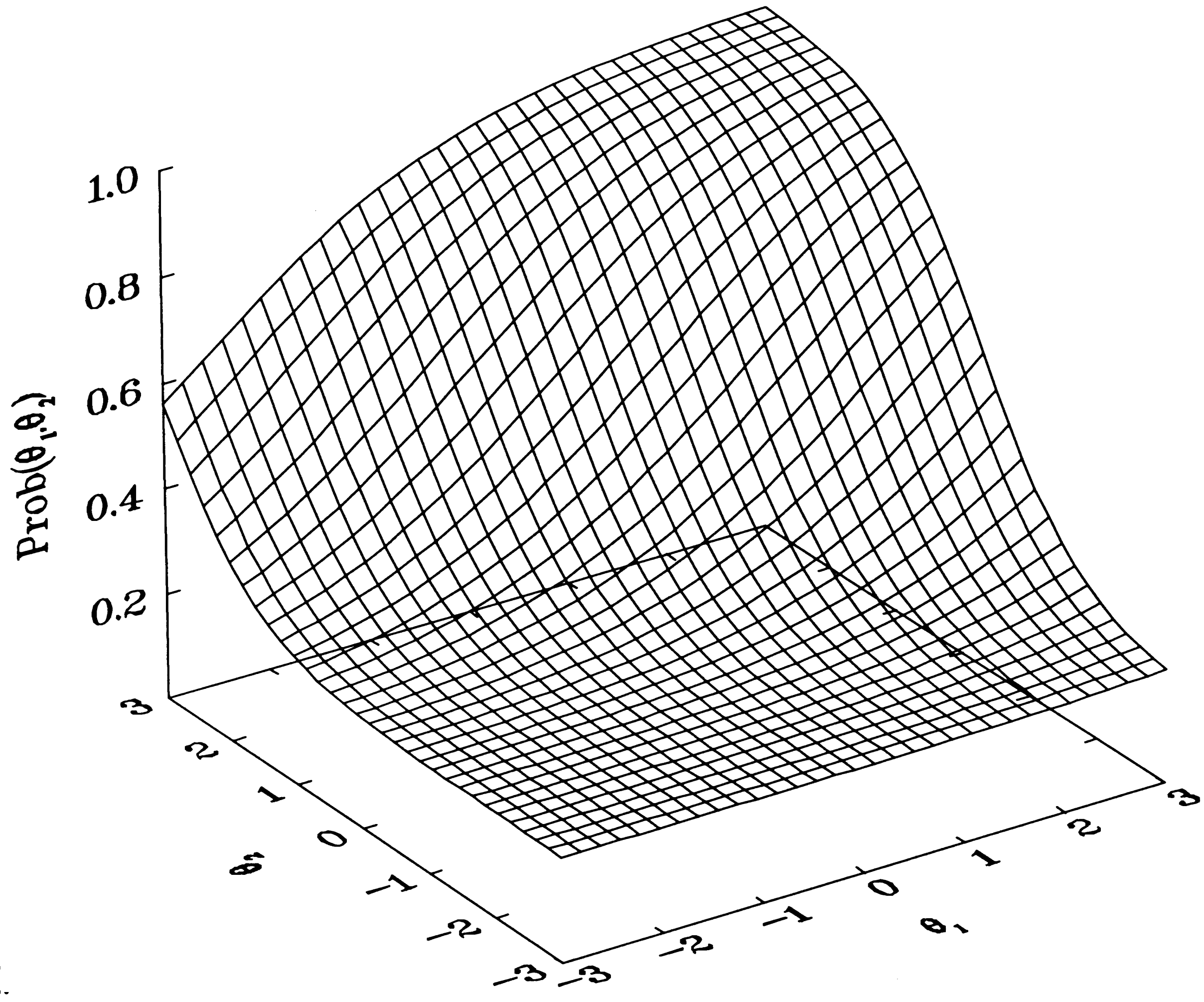
Model Fit

	LM	df	Prob	Groups 1 Obs.	1 Exp.	2 Obs.	2 Exp.	3 Obs.	3 Exp.	Abs. Dif.
Item1	6.04	2	0.03	1.12	0.95	1.40	1.03	0.82	1.18	0.37
Item2	2.24	2	0.33	0.75	0.76	1.00	0.91	1.14	1.20	0.09
Item3	1.93	2	0.38	1.68	1.59	1.73	1.59	1.73	1.71	0.14
Item4	1.14	1	0.29	2.60	2.61	2.95	2.86			0.09

If the model does not fit

- Go to a more complicated model:
  - Multidimensional Model
    - Distinguish Subscales: Between items multidimensionality
    - Allow items to load on more dimensions: within items multidimensionallity
- Group-specific item parameters

Multidimensional IRT model



Solving item bias using group-specific item parameters

The problem: item bias (differential item functioning)

An item has different response probabilities across groups conditional on a fixed theta-level

Example: items pertaining to football in a language comprehension test disadvantage girls

Solution: group-specific item parameters

Application: disease-specific item parameters

Solving item bias using group-specific item parameters

Assumption: a number of items in the scale is free of cultural bias.

Some items measure the same construct over countries but have a different metric.

Example: the number of cars owned by a family is a proxy for wealth, but the scales are different in Enschede and Amsterdam.

Solution: different weights, i.e. item parameters.

Solving item bias using group-specific item parameters

Important advantage: different groups can still be compared on the same theta-scale for the same construct

Weisscher, Glas, Vermeulen, & de Haan (2010). The use of an item response theory-based disability item bank across diseases: accounting for differential item functioning. Journal of Clinical Epidemiology, 63, 543-549.

van Groen, ten Klooster, Taal, van de Laar, & Glas (2010). Application of the Health Assessment Questionnaire Disability Index to various rheumatic diseases. Quality of Life Research, 19, 1255-1263.

IRT structural modeling

$$X_n = \begin{cases} 1 \text{ if person } n \text{ is male} \\ 0 \text{ if person } n \text{ is female} \end{cases}$$

$$\theta_n = \mu + x_n \beta + \varepsilon_n \qquad \varepsilon_n \sim N(0, \sigma)$$

Testing of hypothesis: $\beta = 0$

IRT structural modeling

$$X_{1n} = \begin{cases} 1 \text{ if person } n \text{ is male} \\ 0 \text{ if person } n \text{ is female} \end{cases}$$

$$X_{n} = \begin{cases} 1 \text{ if person } n \text{ is overweigh} \\ 0 \text{ if person } n \text{ is not overweight} \end{cases}$$

$$\theta_n = \mu + x_{1n}\beta_1 + x_{2n}\beta_2 + x_{1n}x_{2n}\beta_{12} + \varepsilon_n$$

$$\varepsilon_n \sim N(0, \sigma)$$

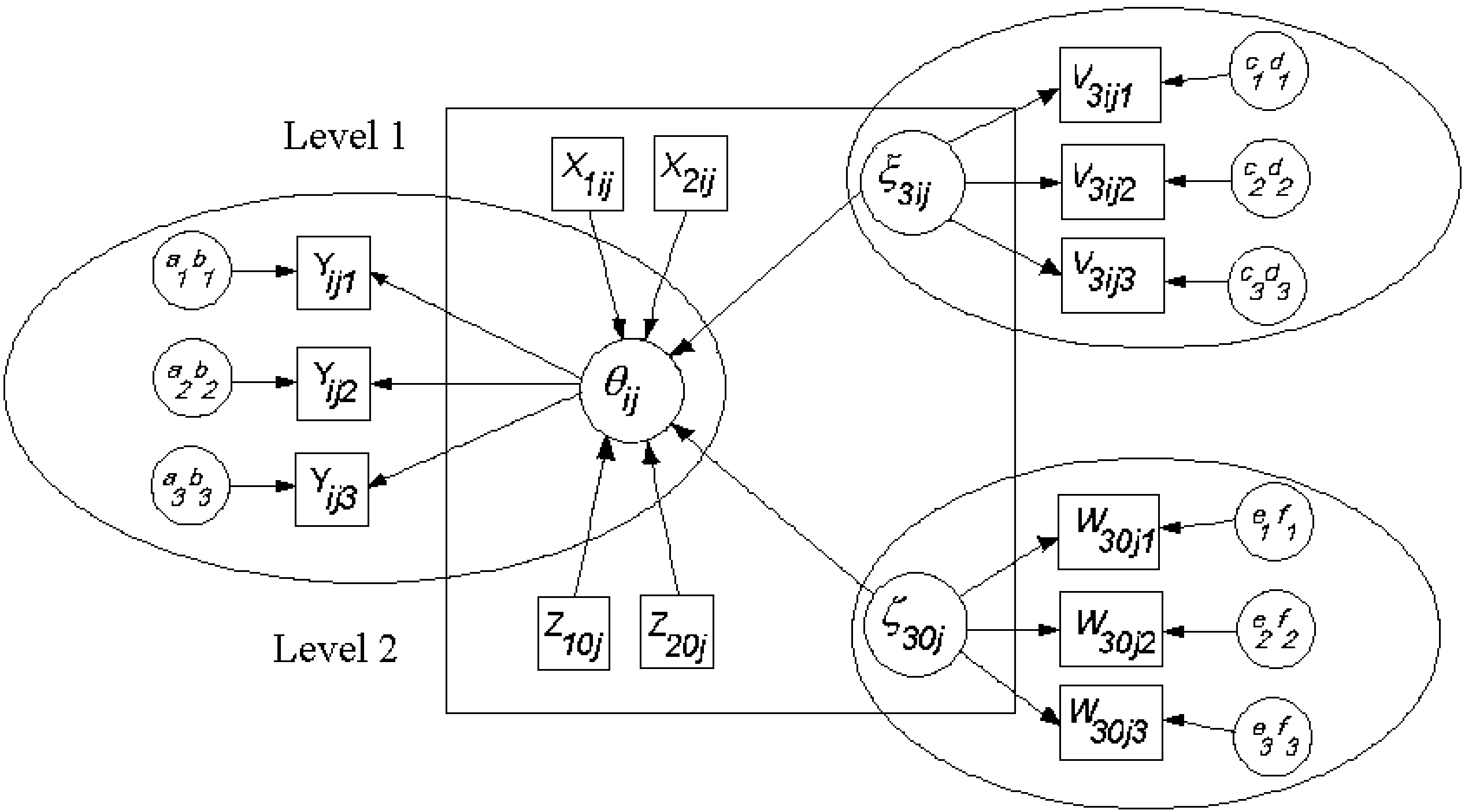


Figure 1
Path diagram of a multilevel IRT model

Using theta-estimates in statistical analyses

- The problem:
  - Theta’s are no observations but estimated parameters
  - The standard error (i.e. the reliability of the estimate) depends on:
    - Test length
    - Distance between population and item parameters

Using theta-estimates in statistical analyses

- Conclusion: do not analyze thetas with standard statistical software (SPPS, SAS, ect).
- The solution:
  - Advanced software which can estimate an IRT model and a structural model on Theta’s simultaneously (Mplus, Gllamm, Mx)
  - Plausible value imputation

Using theta-estimates in statistical analyses: Plausible value imputation

- For every theta: draw 5-10 plausible values from the posterior distribution of the estimates
- This models the reliability of the estimates
- Replicate a standard analysis (SPPS, SAS) for every replication and pool the estimates

Conclusion

- IRT and CAT support flexible measurement using modern technology
- But it also demands more explicit validation of the measures
- And subsequent statistical analyses need careful attention

Statistical Background of CAT

- Practitioners and test takers generally accept the outcomes of CAT, though the statistical foundation of CAT is not simple
- Due to the statistical foundation of CAT, on-line calibration of items must be considered carefully
- The basis of CAT is the ignorability principle

	Child						
Family	1	2	3	k	.	.	K
1	0	1	-	-	-	-	-
2	1	-	-	-	-	-	-
3	0	0	0	1	-	-	-
i	0	0	0	0	1	-	-
.	.	.	.	.	.		.
.	0	0	0	0	0	0	1
N	0	0	1	-	-	-	-

Suppose parents keep getting children until the first son is born

- Is the probability of a boy still 50%?

Statistical Background of CAT

Data matrix X

Missing data indicator D

For persons i and items k, D has elements

$$d_{ik} = \begin{cases} 0 & \text{if } x_{ik} \text{ was not observed} \\ 1 & \text{if } x_{ik} \text{ was observed} \end{cases}$$

Missing at Random (MAR, Rubin)

$$p(D \mid x_{mis}, x_{obs}, \varphi, y) = p(D \mid x_{obs}, \varphi, y)$$

The missing data indicator does not depend on the (unobserved) missing data

In CAT this is trivially true:

$$p(D \mid x_{mis}, x_{obs}, \beta) = p(D \mid x_{obs}, \beta) = 1$$

Theorem (Donald Rubin, 1976)

- If MAR and distinctness (θ and ϕ independent) hold, then
- We can make inferences based on a model which ignores the process giving rise to the missing data
- Applied to the family-example: the design is completely determined by the observations so the proportion of boys is still 50%

Item review: no violation of ignorability in the operational phase making inferences on individuals

$$P(x_{obs}, x_{mis}, d; \theta, \beta) =$$

$$P(x_{obs}; \theta, \beta) P(x_{mis}; \theta, \beta) P(d \mid x_{obs}, x_{mis}, \theta, \beta) =$$

$$P(x_{obs}; \theta, \beta)$$

No bias in operational phase

Item Selection Mode	θ	Bias	Standard Error
CAT with Item Review	-2.0	0.06	0.57
	-1.0	0.07	0.48
	0.0	0.10	0.47
	1.0	0.10	0.46
	2.0	0.11	0.46
CAT with Item Review	-2.0	0.15	0.72
If θ > 0.0	-1.0	0.01	0.49
	0.0	0.00	0.46
	1.0	0.14	0.47
	2.0	0.16	0.73

But large bias in calibration phase

Item Selection Mode	β	Bias	Standard Error	Mean
CAT with Item Review	-2.0	0.64	0.22	-1.33
	-1.0	0.34	0.29	-0.65
	0.0	0.01	0.07	-0.01
	1.0	0.28	0.22	0.71
	2.0	0.60	0.18	1.39
CAT with Item Review	-2.0	0.07	0.21	-1.90
If θ > 0.0	-1.0	0.15	0.29	-0.84
	0.0	0.01	0.07	0.01
	1.0	0.20	0.19	0.79
	2.0	0.52	0.22	1.47

Violation ignorability in item calibration

$$
\begin{aligned}
p(x, d, \beta) \quad &= \quad p(x_{obs} \mid \theta, \beta)\, p(x_{mis} \mid \theta, \beta)\, g(\theta) \\
&\propto \quad p(x_{obs} \mid \theta, \beta)\, p(\theta \mid x_{mis}, \beta) \\
&\neq \quad p(x_{obs} \mid \theta, \beta)\, g(\theta)
\end{aligned}
$$

Conclusion

- In the operational CAT phase violations of Ignorability (item review, ignored covariates) do not interfere with the estimates of ability
- If these data are used for calibration or estimation of population effects, the ignored data and covariates should be taken into account