

Item response theorie

L. Andries van der Ark

Departement MTO, UvT

a.vdark@uvt.nl

Presentatie gebaseerd op

Sijtsma, K. & Emons, W.H.M. (2008). Statistical models for the development of psychological and educational tests. In T. Rudas (Ed.), *Handbook of probability* (pp. 257-275). Thousand Oaks, CA: Sage.

IRT: Onderzoeksvraag

- **Interesse:** θ (latente trek)

Organizing principle behind a cohesive set of observable behaviors

[1] Kwaliteit van leven na een chirurgische ingreep, [2] Mening over regeringsbeleid, [3] Taalvaardigheid van studenten, [4] SES van buurten.

- θ
 - Niet observeerbaar
 - Meestal verondersteld unidimensioneel
 - Meestal gemeten met groot aantal items ($X_1, \dots, X_j, \dots, X_J$)
vanwege [1] interesse in individuele verschillen, [2] meetprecisie, [3] constructvaliditeit, [4] outlierdetectie, [5] mogelijkheid om eenvoudige vragen te stellen ipv complexe.

IRT: Dataset - Loopproblemen

Lft		Diagnose		itemscores				
40	M	Amputatie O	0	0	0	0	0	...
63	V	Amputatie B	0	0	0	0	0	...
57	M	Amputatie B	0	0	0	0	1	...
67	M	CVA	0	0	0	0	0	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Checklist met 35 items. Meting $X_+ = \sum_j X_j$.

Item 1	In huis lopen: niet andere kamers
Item 2	In huis lopen: niet alle kamers
Item 3	In huis lopen: klein afstanden
⋮	⋮
Item 35	Buiten lopen: snel (bijna) altijd hulpmiddel

IRT Modellen:

Statistische modellen beschrijven relatie tussen $P(X = x|\theta)$ en θ .

IRT-analyse geeft informatie over

- Dimensionaliteit van de data
- Kwaliteit van individuele items
- Kwaliteit van de gehele test
- θ -schattingen (plus s.e.) voor elke observatie.

Nonparametrische IRT: Model van monotone homogeniteit

Assumpties

1. *Unidimensionaliteit:*

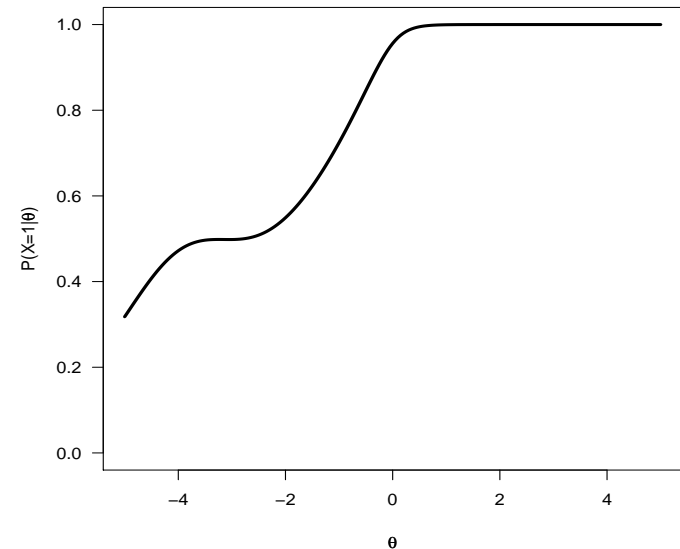
θ eendimensioneel.

2. *Lokale onafhankelijkheid:*

$$P(\mathbf{X} = \mathbf{x}|\theta) = \prod_j P(X_j = x_j|\theta)$$

3. *Monotonie:*

$P(X = 1|\theta)$ niet-dalend in θ



Als MHM geldt dan $E(\theta|X_+ = K) \geq E(\theta|X_+ = C)$ ($K > C$).

Itemscores kunnen zinvol opgeteld worden.

Nonparametrische IRT: Model fit

Assumpties impliceren observeerbare eigenschappen. Deze worden onderzocht.

- *Automatisch Item Selectie Algoritme*: Partitionering items in *Mokken schalen* m.b.v. schalingscoëfficiënten.

Vuistregels

$.0 \leq H < .3$	Geen schaal
$.3 \leq H < .4$	Zwakke schaal
$.4 \leq H < .5$	Medium schaal
$.5 \leq H < 1$	Sterke schaal

- Onderzoek of MHM-model past bij de schalen.

Nonparametrische IRT: Model fit

Alle 35 items: Medium/sterke schaal ($H = .50$), sommige schaalbaarheidscoëfficiënten wat laag.

Items 1, 30 en 31 aparte subschaal (m.b.t. hulpmiddelen); item 2: minder goed schaalbaar (alle kamers)

31 items, sterke schaal $H = .54$.

```
library(mokken)
scale.3 <- aisp(item.scores, verbose=F, lowerbound = .3)
scale.4 <- aisp(item.scores, verbose=F, lowerbound = .4)
coefH(item.scores[,scale.3==1])$H
coefH(item.scores[,scale.4==1])$H
```

Parametrische IRT: Twee-parameter logistisch model

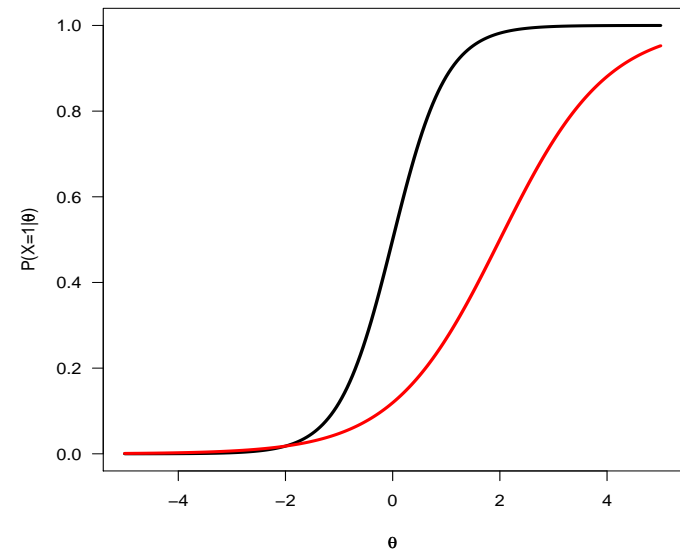
Meten op intervalniveau.

Assumpties

1-3 *Unidimensionaliteit, Lokale onafhankelijkheid, Monotonie.*

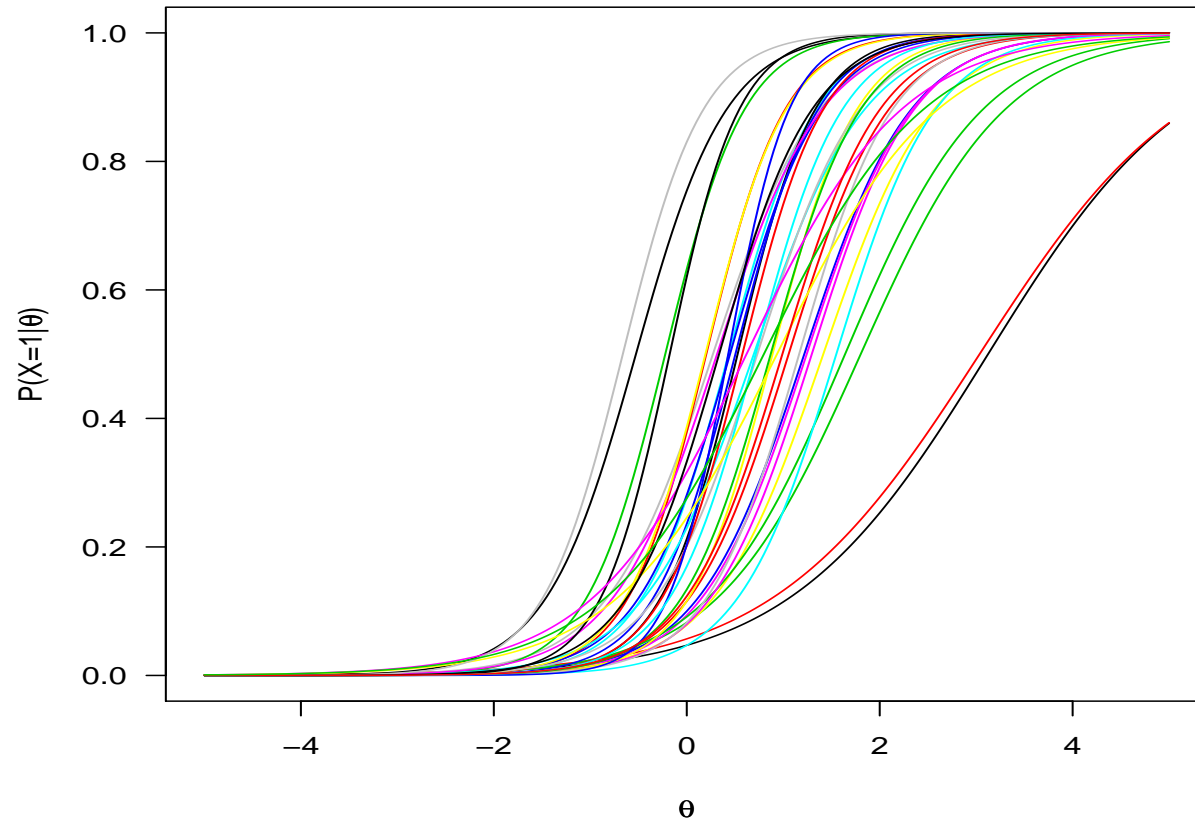
4

$$P(X_j = 1|\theta) = \frac{\exp \alpha_j(\theta - \beta_j)}{1 + \exp \alpha_j(\theta - \beta_j)}$$

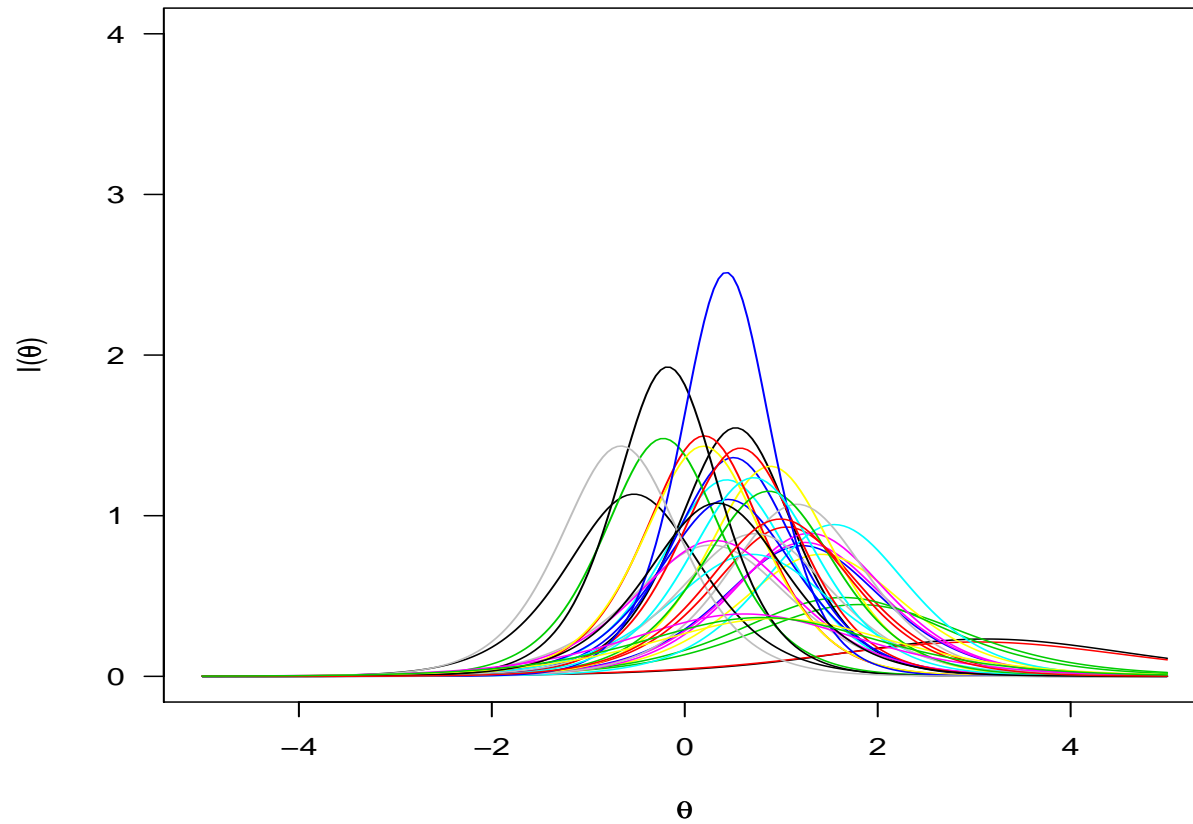


```
library(ltm)
model <- ltm(item.scores ~ z1)
```

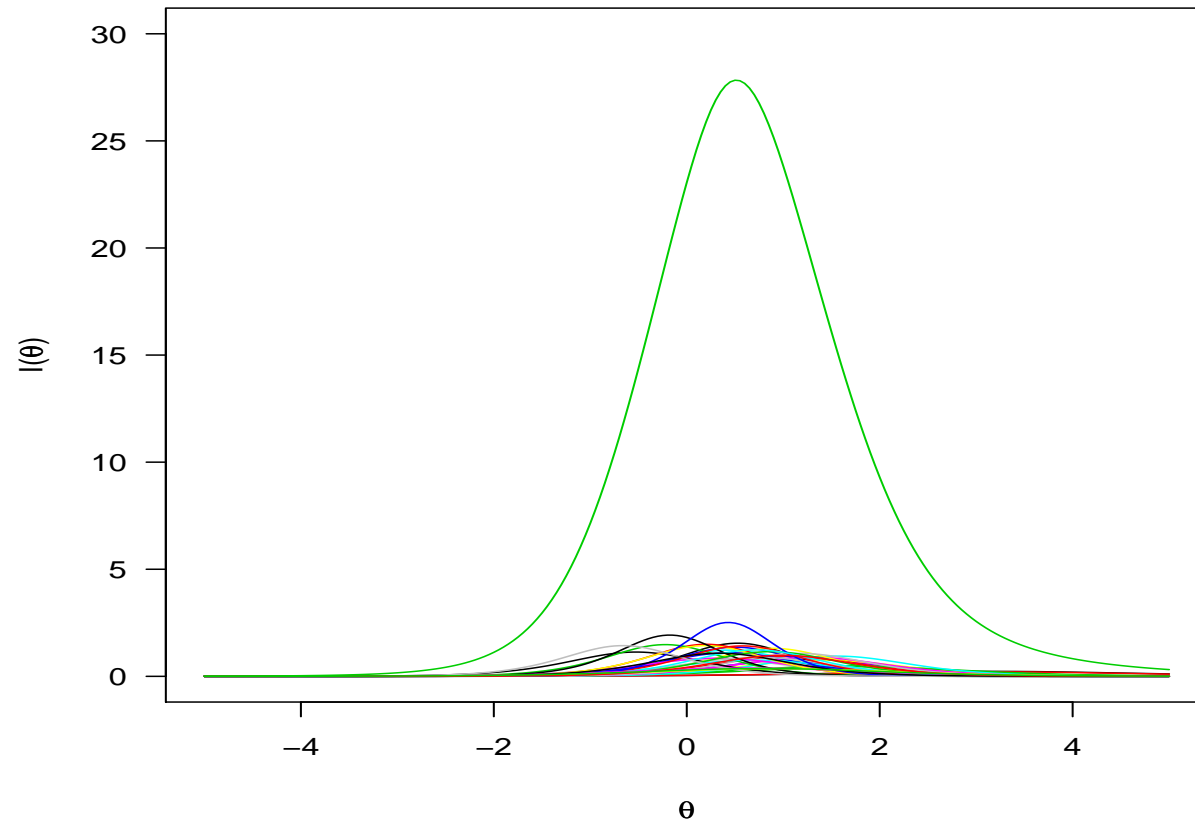
Item-respons-functies (Relatie θ en $P(X = 1|\theta)$)



Item-informatiefuncties (Meetnauwkeurigheid)



Test-informatiefuncties (Meetnauwkeurigheid)



Geschatte θ als meting

Observatie	$\hat{\theta}$	s.e.
1	-1.511	(.506)
2	-1.511	(.506)
3	-1.511	(.506)
4	-0.259	(.226)
5	-0.536	(.260)
6	0.370	(.188)
7	-1.073	(.369)
8	-0.768	(.299)
9	-0.867	(.320)
10	-0.954	(.339)
⋮	⋮	⋮

```
fs <- factor.scores(model,item.scores)
theta <- round(fs$score.dat[, "z1"],3)
se <- round(fs$score.dat[, "se.z1"],3)
```

Werken met geschatte θ 's

Vergelijkbaar met andere type scores.

Groep	N	Gem	se
Amputatie bovenbeen	105	0.285	0.088
Amputatie onderbeen	114	0.105	0.094
Amputatie, overige	20	-0.110	0.253
Dystrofie (CRPS1)	65	0.730	0.086
Beroerte (CVA)	67	0.203	0.099
Artrose heup	201	0.445	0.052
Artrose knie	80	0.474	0.095
Polio	89	0.061	0.088
Diabetische voet	57	-0.839	0.123
Overige voet	183	-0.824	0.056

```
X <- cbind(X, theta)
diagnosis <- as.factor(X[,"diagnosis.1"])
tapply(theta,diagnosis, mean)
```

Books

- Baker, F. B. & Kim, S.-H. (2004). *Item response theory: Parameter estimation techniques* (2nd ed.). New York: Dekker.
- Embretson, S. E. & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.
- Sijtsma, K. & Emons, W.H.M. (2008). Statistical models for the development of psychological and educational tests. In T. Rudas (Ed.), *Handbook of probability* (pp. 257-275). Thousand Oaks, CA: Sage.
- Sijtsma, K. & Molenaar, I. W. (2003). *Introduction to nonparametric item response theory*. Thousand Oaks, CA: Sage.

Software

R packages: ltm (PIRT), mokken (NIRT)
SAS routines (PIRT)
Multilog (PIRT), MSP (NIRT)

Ceterum censeo Carthaginem esse delendam

**CRONBACH'S ALFA IS GEEN MAAT VOOR INTERNE
CONSISTENTIE**

Het een (onder)schatter van de betrouwbaarheid

Twee covariantiematrices

$$\begin{pmatrix} 1.0 & 0.3 & 0.3 & 0.3 \\ 0.3 & 1.0 & 0.3 & 0.3 \\ 0.3 & 0.3 & 1.0 & 0.3 \\ 0.3 & 0.3 & 0.3 & 1.0 \end{pmatrix}$$

1-factor structuur

$$\alpha = .6315$$

$$\begin{pmatrix} 1.0 & 0.9 & 0.0 & 0.0 \\ 0.9 & 1.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.0 & 0.9 \\ 0.0 & 0.0 & 0.9 & 1.0 \end{pmatrix}$$

2-factor structuur

$$\alpha = .6315$$