Factor indeterminacy and its implications for Item Response Theory

Alwin Stegeman

a.w.stegeman@rug.nl
www.gmw.rug.nl/∼stegeman

New Version - 2 March 2015

university of groningen

faculty of behavioural and social sciences

Overview

Factor indeterminacy and its implications for IRT

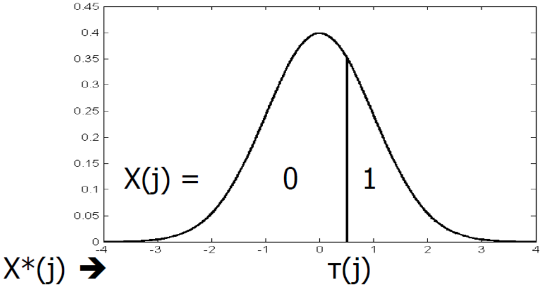
1. The link between Common Factor Analysis and IRT
2. Factor indeterminacy in Common Factor Analysis
3. The implications of factor indeterminacy for IRT
4. Taking into account factor indeterminacy in IRT
5. Exploratory IRT via Common Factor Analysis

Common Factor Analysis and Item Response Theory

We assume dichotomous items and one latent trait.

Let $X_j \in \{0, 1\}$ be the score variable for item j.

Let X_j^* be a $N(0, 1)$ variable such that: $\quad X_j^* > \tau_j \iff X_j = 1$



Common Factor Analysis and Item Response Theory

Let X_j^* satisfy the 1-factor model: $$X_j^* = \lambda_j F + u_j E_j$$

For F and E_j independent and $E_j \sim N(0,1)$, we obtain:

$$P(X_j = 1 \mid F = f) = P(X_j^* > \tau_j \mid F = f) = \Phi(\, a_j\,(f - b_j)\,)$$

$$\text{with } a_j = \lambda_j / u_j \text{ and } b_j = \tau_j / \lambda_j$$

This is the **Normal Ogive** item response function, which is closely approximated by the **2-PL**.

Note: $\mathrm{Corr}(E_j, E_k) = 0$ for $j \neq k \iff$ local independence

Lord & Novick (1968), Takane & De Leeuw (1987)

Item Factor Analysis

Heuristic method to estimate Normal Ogive or 2-PL item parameters.

1. Estimate thresholds $\tau_j = \Phi^{-1}(\text{proportion } X_j = 0)$.
2. Estimate tetrachoric correlations $\mathrm{Corr}(X_j^*, X_k^*)$ for $j \neq k$.
3. Fit the 1-factor model to the tetrachoric correlation matrix.

(or more complicated algorithms...)

overview: Wirth & Edwards (2007)

Note: latent trait scores (IRT) $\iff$ factor scores (FA)

Factor indeterminacy in the Common Factor model

For n observations and p items, the 1-factor model is $\mathbf{X}^* = \mathbf{F}\,\boldsymbol{\lambda}^T + \mathbf{E}\,\mathbf{U}$, with $\mathbf{X}^*$ $(n \times p)$, $\mathbf{F}$ $(n \times 1)$, $\boldsymbol{\lambda}$ $(p \times 1)$, $\mathbf{E}$ $(n \times p)$, and $\mathbf{U}$ diagonal $(p \times p)$.

Assumptions: $n^{-1}\mathbf{E}^T\mathbf{F} = \mathbf{O}$, $n^{-1}\mathbf{E}^T\mathbf{E} = \mathbf{I}_p$.

Factor scores $\mathbf{F}$ and unique part $\mathbf{E}$ are **not uniquely determined**!

$$\begin{aligned}\mathbf{F}, \mathbf{E} &= \text{determinate part} + \text{indeterminate part} \\ &= \text{regression on } \mathbf{X}^* + \text{residual}\end{aligned}$$

Indeterminate parts of $\mathbf{F}$ and $\mathbf{E}$ are linked such that the assumptions hold.

Wilson (1928), Guttman (1955)

Factor indeterminacy in the Common Factor model

For two alternative factors $\mathbf{F}_1$ and $\mathbf{F}_2$, we have a minimal correlation

$$\mathrm{Corr}(\mathbf{F}_1, \mathbf{F}_2) \geq 2\, R^2 - 1\,,$$

with $R^2 = \boldsymbol{\lambda}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\lambda}$ of the regression of $\mathbf{F}$ on $\mathbf{X}^*$, and $\boldsymbol{\Sigma} = \mathrm{Corr}(\mathbf{X}^*)$.

Factors are interpreted via the loadings $\boldsymbol{\lambda}$, but what if the minimal correlation is very small or negative?

Minimal correlations should be reported when using Common Factor Analysis.

Guttman (1955), Grice (2001)

Remarks on factor indeterminacy

- Factor indeterminacy is caused by having more factors (including the unique part) than observed variables.
- The Common Factor model is usually estimated in covariance form and factor indeterminacy is ignored.
- Factor indeterminacy is a controversial issue ("stop using the factor model" versus "assume infinitely many items measuring F, then the factor scores are unique in the limit")

Steiger (1979), Maraun (1996)

The implications of factor indeterminacy for IRT

When using the Normal Ogive or 2-PL models:

- There is no unique true latent trait score to estimate.
- Maximum Likelihood ‘estimation’ of latent trait scores ?
- Can standard errors of latent trait score estimates be trusted ?
- Is the interpretation of the latent trait valid ?

Can we take into account factor indeterminacy in IRT estimation?

Taking into account factor indeterminacy in IRT

Via Item Factor Analysis? Some observations:

- To obtain an 'estimate' of $\mathbf{F}$ (including indeterminate part), we must fit the 1-factor model to $\mathbf{X}^*$ and not to Corr($\mathbf{X}^*$).
- We do not know $\mathbf{X}^*$, but we can generate many $\widetilde{\mathbf{X}}^*$ that correspond to the dichotomous data $\mathbf{X}$.
- By fitting the 1-factor model to each $\widetilde{\mathbf{X}}^*$, we obtain many 'estimates' of $\mathbf{F}$ and of the item parameters and the minimal correlation. We can compute the mean and std for each parameter.

Direct-Fitting Item Factor Analysis (DIFIFAC)

Stegeman (2015):

1. Estimate thresholds $\tau_j = \Phi^{-1}(\text{proportion } X_j = 0)$.
2. Estimate tetrachoric correlation matrix $\mathbf{\Sigma} = \text{Corr}(\mathbf{X}^*)$.
3. Generate many $\widetilde{\mathbf{X}}^*$ such that $\widetilde{X}^*_{ij} > \tau_j \Leftrightarrow X_{ij} = 1$, and $\text{Corr}(\widetilde{\mathbf{X}}^*) = \mathbf{\Sigma}$.
4. Fit the 1-factor model to each $\widetilde{\mathbf{X}}^*$ (Stegeman & Kiers, 2014).
5. Compute mean and std of 'estimates' of $\mathbf{F}$, and of estimates of item parameters and minimal correlation.

Next step: simulation study to evaluate performance.

Note: Estimated loadings $\boldsymbol{\lambda}$, unique stds $\mathbf{U}$, and minimal correlations are the same for all $\widetilde{\mathbf{X}}^*$.

DIFIFAC Simulation Study - Setup

We consider 10 dichotomous items, with true loadings

$$\boldsymbol{\lambda}^T = (0.6 \;\; 0.7 \;\; 0.8 \;\; 0.6 \;\; 0.7 \;\; 0.8 \;\; 0.6 \;\; 0.7 \;\; 0.8 \;\; 0.6),$$

true unique variances $u_j^2 = 1 - \lambda_j^2$, and true thresholds

$$\tau_j = 0,\ 0.5,\ 1,\ -0.5,\ 0,\ -1,\ 0,\ 0.5,\ 0,\ 1.$$

True $\mathbf{F}$ and $\mathbf{E}$ are random $N(0,1)$ such that model assumptions hold.

True data is $\mathbf{X}^* = \mathbf{F}\,\boldsymbol{\lambda}^T + \mathbf{E}\,\mathbf{U}$, with minimal correlation 0.83.

Generate 100 datasets $\mathbf{X}^* + \sigma\,\mathbf{N}$, with $\mathbf{N}$ random $N(0,1)$ and $\sigma = 0.4$.

For each $\mathbf{X}^* + \sigma\,\mathbf{N}$, make a dichotomous $\mathbf{X}$ by using thresholds τ_j.

For each dichotomous $\mathbf{X}$, apply DIFIFAC.

DIFIFAC Simulation Study - Results

	$n = 300$ (97 cases)	$n = 1000$ (100 cases)
$\boldsymbol{\lambda}$	0.07 (0.01)	0.05 (0.01)
$\mathbf{U}$	0.06 (0.01)	0.04 (0.01)
$\mathbf{F}$ (det)	0.33 (0.01)	0.32 (0.01)
min corr	0.04 (0.03)	0.03 (0.01)

Mean and std of 100 mean absolute deviation (MAD) values.

DIFIFAC used 50 generated $\widetilde{\mathbf{X}}^*$ per run.

DIFIFAC Application - NPV-J IN subscale

Dutch Personality Questionnaire Junior (NPV-J)
$n = 866$ adolescents
105 dichotomous items in 5 subscales
We only use the Inadequacy (IN) subscale, $p = 28$ items

DIFIFAC used 50 generated $\widetilde{\mathbf{X}}^*$

Total fit of the 1-factor model is 56.0%

Total explained common variance is 48.0% (unidimensionality)

Explained common variance per item is 15.6% - 76.1% (item fit)

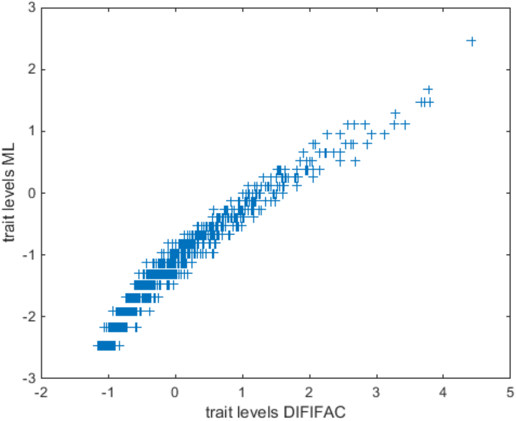
Minimal correlation is 0.98 (factor indeterminacy)

DIFIFAC Application - NPV-J IN subscale

	mean	std
loadings $\boldsymbol{\lambda}$	0.35 - 0.86	-
unique std $\mathbf{U}$	0.00 - 0.67	-
factor scores $\mathbf{F}$	-1.15 - 4.44	0.17 - 0.61

Mean and std of estimates for 50 generated $\widetilde{\mathbf{X}}^*$

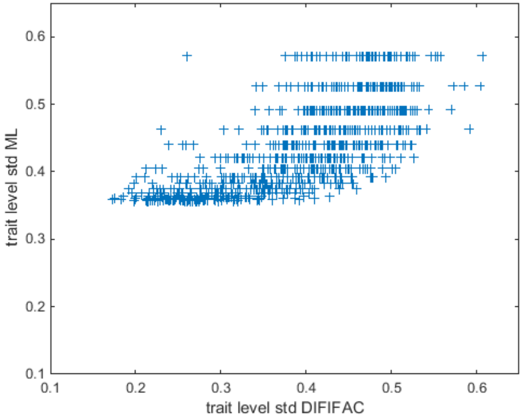
DIFIFAC Application - NPV-J IN subscale



Factor score estimates for DIFIFAC (x-axis) and 2-PL (y-axis)

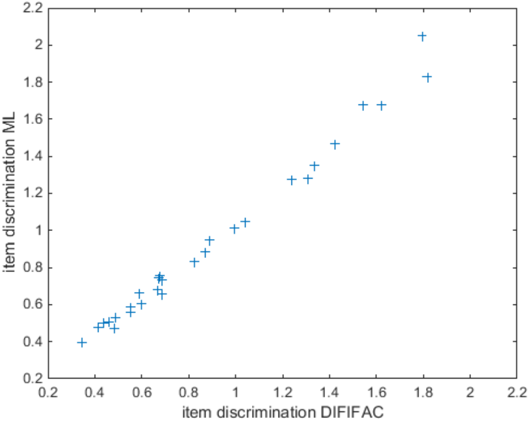
2-PL fit: Meijer & Tendeiro (2012)

DIFIFAC Application - NPV-J IN subscale



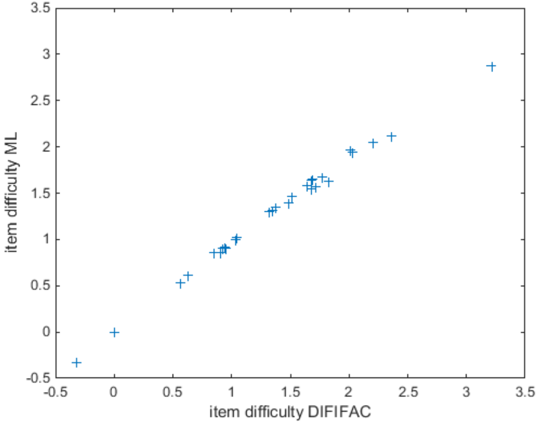
Factor score std estimates for DIFIFAC (x-axis) and 2-PL (y-axis)

DIFIFAC Application - NPV-J IN subscale



Item discrimination estimates for DIFIFAC (x-axis) and 2-PL (y-axis)

DIFIFAC Application - NPV-J IN subscale



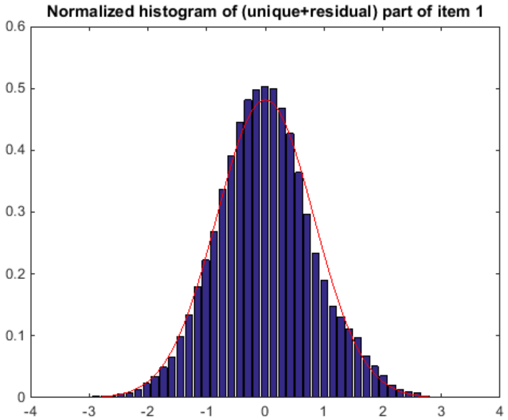
Item difficulty estimates for DIFIFAC (x-axis) and 2-PL (y-axis)

DIFIFAC as Exploratory IRT tool

By considering $\widetilde{\mathbf{X}}^* - \mathbf{F}\,\boldsymbol{\lambda}^T$, we can check:

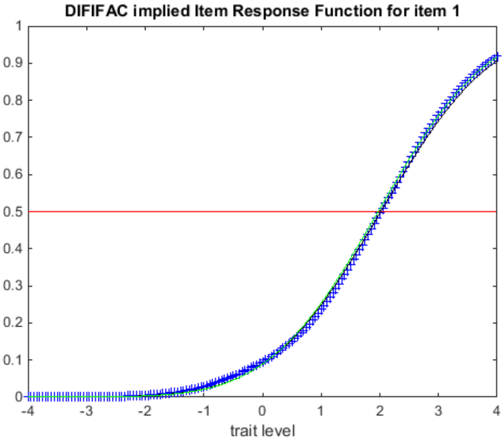
- Normality of the (unique+residual) part per item.
- Nonparametric item response function per item.
- Correlations of the (unique+residual) parts of different items (local independence check).
- Person fit indices by adding the squared standardized columns (chi-square(p) under normality and local independence).

DIFIFAC Application - NPV-J IN subscale



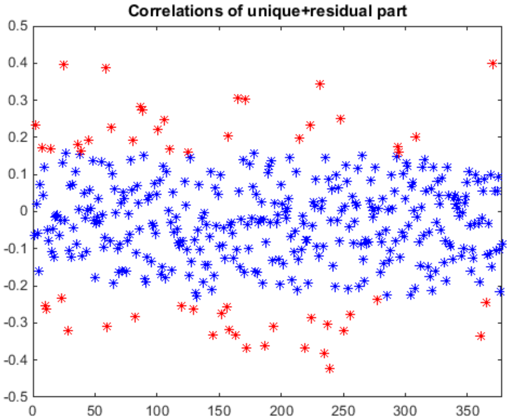
Normalized histogram of the (unique+residual) part of item 1 compared to normal density

DIFIFAC Application - NPV-J IN subscale



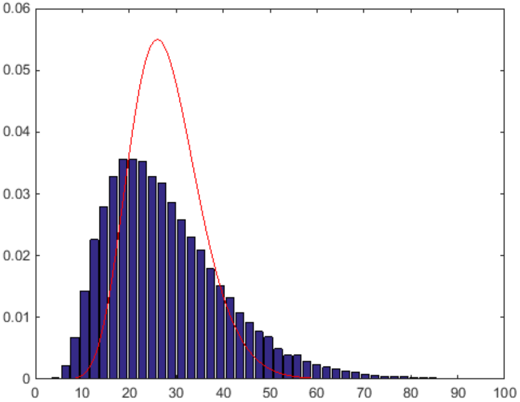
Nonparametric item response function for item 1 (blue), compared to closest 2-PL (black), and estimated 2-PL (green)

DIFIFAC Application - NPV-J IN subscale



Correlations of the (unique+residual) parts per item pair
(red = significant at 0.05 level, based on simulations: 14 percent)

DIFIFAC Application - NPV-J IN subscale



Normalized histogram of person fit statistics, compared to chi-square(28) density (red)

Concluding Remarks

- Factor indeterminacy is less (larger minimal correlation) when more similar items are used, but this may violate local independence.
- Factor score standard errors are different when factor indeterminacy is taken into account, but are not much larger for the NPV-J IN subscale.
- The DIFIFAC procedure may be used as an exploratory IRT tool.
- DIFIFAC also works for polytomous items, and multiple latent traits.

Thank You !

References

- Grice, J.W. (2001). Computing and evaluating factor scores. *Psychological Methods*, *6*, 430-450.
- Guttman, L. (1955). The determinacy of factor score matrices with implications for five other basic problems of common factor theory. *The British Journal of Statistical Psychology*, *8*, 65-81.
- Lord, F.M., & Novick, M.R. (1968). *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.
- Maraun, M.D. (1996). Metaphor taken as math: indeterminacy in the factor analysis model. With comments and reply. *Multivariate Behavioral Research*, *31*, 517-689.
- Meijer, R.R., & Tendeiro, J.N. (2012). The use of the l_z and l_{z*} person-fit statistics and problems derived from model misspecification. *Journal of Educational and Behavioral Statistics*, *37*, 758-766.

References

- Steiger, J.H. (1979). Factor indeterminacy in the 1930s and the 1970s, some interesting parallels. *Psychometrika*, *44*, 157-167.
- Stegeman, A., & Kiers, H.A.L. (2014). Direct-fitting common factor analysis. In preparation.
- Stegeman, A. (2015). Exploratory item response theory via direct-fitting item factor analysis. In preparation.
- Takane, Y., & De Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, *52*, 393-408.
- Wilson, E. (1928). On hierarchical correlation systems. *Proceedings of the National Academy of Sciences*, *24*, 283-291.
- Wirth, R.J., & Edwards, M.C. (2007). Item factor analysis: current approaches and future directions. *Psychological Methods*, 12, 58-79.