ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Exploratory Factor Analysis

Applied Multivariate Statistics – Spring 2012

Latent-variable models

- Large number of observed (manifest) variables should be explained by a few un-observed (latent) underlying variables
- E.g.: Scores on several tests are influenced by “general academic ability”
- Assumes local independence: Manifest variables are independent given latent variables

	Latent variables	
Manifest Variables	Continuous	Categorical
Continuous	Factor Analysis	Latent Profile Analysis
Categorical	Item Response Theory	Latent Class Analysis

Overview

- Introductory example
- The general factor model for x and Σ
- Estimation
- Scale and rotation invariance
- Factor rotation: Varimax
- Factor scores
- Comparing PCA and FA

Introductory example: Intelligence tests

- Six intelligence tests (general, picture, blocks, maze, reading, vocab) on 112 persons
- Sample correlation matrix

	general	picture	blocks	maze	reading	vocab
general	1.0000000	0.4662649	0.5516632	0.3403250	0.5764799	0.5144058
picture	0.4662649	1.0000000	0.5724364	0.1930992	0.2629229	0.2392766
blocks	0.5516632	0.5724364	1.0000000	0.4450901	0.3540252	0.3564715
maze	0.3403250	0.1930992	0.4450901	1.0000000	0.1839645	0.2188370
reading	0.5764799	0.2629229	0.3540252	0.1839645	1.0000000	0.7913779
vocab	0.5144058	0.2392766	0.3564715	0.2188370	0.7913779	1.0000000

- Can performance in and correlation between the six tests be explained by one or two variables describing some general concept of intelligence?

Introductory example: Intelligence tests

Model:

f: Common factor ("ability")

$$x_{1i} = \lambda_1 f_i + u_{1i}$$

$$x_{2i} = \lambda_2 f_i + u_{2i}$$

u: Random disturbance specific to each exam

$$\dots$$

$$x_{6i} = \lambda_6 f_i + u_{6i}$$

λ: Factor loadings - Importance of f on x_j

Key assumption:

u_1, u_2, u_3 are uncorrelated

Thus x_1, x_2, x_3 are conditionally uncorrelated given f

General Factor Model

- General model for one individual:

$$\begin{aligned} x_1 &= \mu_1 + \lambda_{11} f_1 + ... + \lambda_{1q} f_q + u_1 \\ ... \\ x_p &= \mu_p + \lambda_{p1} f_p + ... + \lambda_{pq} f_q + u_p \end{aligned}$$

- In matrix notation for one individual:

$$x = \mu + \Lambda f + u$$

- In matrix notation for n individuals:

$$x_i = \mu + \Lambda f_i + u_i \quad (i = 1, ..., n)$$

- **Assumptions:**
  - Cov(u_j, f_s) = 0 for all j, s
  - E[u] = 0, Cov(u) = Ψ is a diagonal matrix (diagonal elements = «uniquenesses»)
- **Convention:**
  - E[f] = 0, Cov(f) = identity matrix (i.e. factors are scaled)
  Otherwise, Λ and μ are not well determined

To be determined from x:

- Number q of common factors
- Factor loadings Λ
- Specific variances Ψ
- Factor scores f

Representation in terms of covariance matrix

- Using formulas and assumptions from previous slide:

$$x = \mu + \Lambda f + u \quad \Leftrightarrow \quad \Sigma = \Lambda\Lambda^T + \Psi$$

- Factor model = particular *structure imposed* on covariance matrix
- Variances can be split up:

$$var(x_j) = \sigma_j^2 = \sum_{k=1}^{q} \lambda_{jk}^2 + \psi_j$$

"communality": variance due to common factors

"specific variance", "uniqueness"

- "Heywood case" (= kind of estimation error):

$$\psi_j < 0$$

Estimation: MLE

- Assume x_i follows multivariate normal distribution
- Choose Λ, Ψ to maximize the log-likelihood:

$$l = \log(L) = -\frac{n}{2}\log(|\Sigma|) - \frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^T \Sigma^{-1}(x_i - \mu)$$

- Iterative solution, difficult in practice (local maxima)

Number of factors

- MLE approach for estimation provides test:

$$H_q: q - factor\ model\ holds$$
$$vs$$
$$H_u: \Sigma\ is\ unconstrained$$

- Modelling strategy:
  Start with small value of q and increase successively until some H_q is not rejected.
- (Multiple testing problem: Significance levels are not correct)

- Example revisited

Intelligence tests revisited: Number of factors

Part of output of R function “factanal”:

```
Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 6.11 on 4 degrees of freedom.
The p-value is 0.191
```

Hypothesis can not be rejected;
for simplicity, we thus use two factors

Scale invariance of factor analysis

- Suppose $y_j = c_j x_j$ or in matrix notation $y = Cx$ (C is a diagonal matrix); e.g. change of measurement units

$$
\begin{aligned}
Cov(y) &= C\Sigma C^T = \\
&= C(\Lambda\Lambda^T + \Psi)C^T = \\
&= (C\Lambda)(C\Lambda)^T + C\Psi C^T = \\
&= \hat{\Lambda}\hat{\Lambda}^T + \hat{\Psi}
\end{aligned}
$$

  I.e., loadings and uniquenesses are the same if expressed in new units
- Thus, using cov or cor gives basically the same result
- Common practice:
  - use correlation matrix or
  - scale input data
  (This is done in “factanal”)

Rotational invariance of factor analysis

- Rotating the factors yields exactly the same model
- Assume MM^T and transform $f^* = M^T f, \Lambda^* = \Lambda M$
- This yields the same model:
  $x^* = \Lambda^* f^* + u = (\Lambda M)(M^T f) + u = \Lambda f + u = x$
  $\Sigma^* = \Lambda^* \Lambda^{*T} + \Psi = (\Lambda M)(\Lambda M)^T + \Psi = \Lambda\Lambda^T + \Psi = \Sigma$
- Thus, the rotated model is equivalent for explaining the covariance matrix
- Consequence: Use rotation that makes interpretation of loadings easy
- Most popular rotation: **Varimax rotation**
  Each factor should have few large and many small loadings

Intelligence tests revisited: Interpreting factors

Part of output of R function “factanal”:

```
Loadings:
        Factor1 Factor2
general 0.499   0.543
picture 0.156   0.622
blocks  0.206   0.860
maze    0.109   0.468
reading 0.956   0.182
vocab   0.785   0.225
```

Spatial reasoning

Verbal intelligence

Interpretation of factors is generally debatable

Estimating factor scores

- Scores are assumed to be random variables: Predict values for each person
- Two methods:
  - Bartlett (option “Bartlett” in R):
  Treat f as fix (ML estimate)
  - Thompson (option “regression” in R):
  Treat f as random (Bayesian estimate)
- No big difference in practice

Case study: Drug use

Loadings:

	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
cigarettes	0.494				0.407	0.110
beer	0.776				0.112	
wine	0.786					
liquor	0.720	0.121	0.103	0.115	0.160	
cocaine		0.519		0.132		0.158
tranquillizers	0.130	0.564	0.321	0.105	0.143	
drug store medication		0.255				0.372
heroin		0.532	0.101			0.190
marijuana	0.429	0.158	0.152	0.259	0.609	0.110
hashish	0.244	0.276	0.186	0.881	0.194	0.100
inhalants	0.166	0.308	0.150		0.140	0.537
hallucinogenics		0.387	0.335	0.186		0.288
amphetamine	0.151	0.336	0.886	0.145	0.137	0.187

Social drugs Amphetamine Smoking

Hard drugs Hashish Inhalants ?

```
Test of the hypothesis that 6 factors are sufficient.
The chi square statistic is 22.41 on 15 degrees of freedom.
The p-value is 0.0975
```

Significance vs. Relevance:

Might keep less than six factors if fit of correlation matrix is good enough

Comparison: PC vs. FA

- PCA aims at explaining ***variances***, FA aims at explaining ***correlations***
- PCA is exploratory and without assumptions
  FA is based on statistical model with assumptions
- First few PCs will be same regardless of q
  First few factors of FA depend on q
- FA: Orthogonal rotation of factor loadings are equivalent
  This does not hold in PCA
- More mathematically:
  Assume we only keep the PCs in Γ_1
  PCA: $x = \mu + \Gamma_1 z_1 + \Gamma_2 z_2 = \mu + \Gamma_1 z_1 + e$
  FA: $x = \mu + \Lambda f + u$
  Cov(u) is diagonal by assumption, Cov(e) is not
- ! Both PCA and FA only useful if input data is correlated !

Concepts to know

- Form of the general factor model
- Representation in terms of covariance matrix
- Scale and Rotation invariance, varimax
- Interpretation of loadings

R functions to know

- Function “factanal”