Visualization
CH09: Factor Analysis

Jong-June Jeon

University of Seoul, Department of Statistics

Contents

- Introduction

- Factor Models

- Estimation of Factor Models: Factor Loadings

- Estimation of Factor Models: Factor Scores

Introduction

Motivation: Understanding Latent Structure

- Many psychological, social, or marketing measurements are inherently latent.
- Factor analysis identifies underlying variables (factors) that explain the pattern of correlations within a set of observed variables.
- We consider a consumer survey assessing satisfaction with a product.

Example: Product Satisfaction Survey

Observed variables X_1 to X_8:

1. X_1: The price of the product is reasonable.
2. X_2: The quality of the product is excellent.
3. X_3: The product meets my expectations.
4. X_4: Customer service is satisfactory.
5. X_5: Delivery is prompt.
6. X_6: The return process is easy.
7. X_7: The brand image is positive.
8. X_8: The product is easy to use.

Goal of Factor Analysis

- Identify latent factors (e.g., "Product Quality", "Service Experience", "Brand Perception") that account for the correlations among these items.
- Each observed variable is modeled as a linear combination of common factors plus a unique factor.
- Useful for dimension reduction and interpretation of survey constructs.
  - X_2, X_3, X_7: Related to **Product Quality Factors**
  - X_4, X_5, X_6: Related to **Service Factors**
  - X_1, X_7: Related to **Price and Brand**

Inter-item Score Correlations

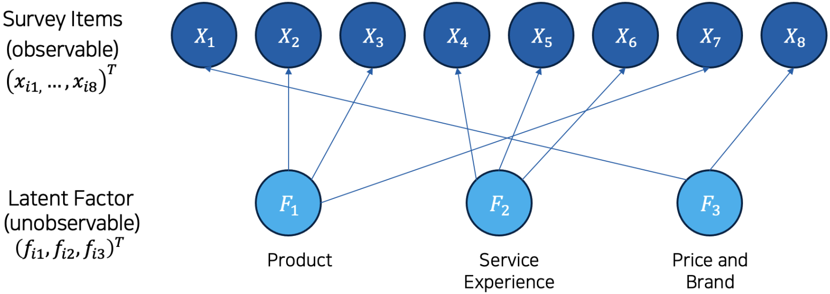

Figure: Visualization of the Factor model

Factor Models

Inter-item Score Correlations

- **Score Standardization: $\mathbf{x}_i - \bar{\mathbf{x}}$**
  - For convenience, we denote $\mathbf{x}_i - \bar{\mathbf{x}}$ as $\mathbf{x}_i$, assuming the mean of the scores is 0.
- Let $\mathbf{x}_i$ be an 8-dimensional multivariate variable, with components x_{ik} for $k = 1, \ldots, 8$ and $i = 1, \cdot, n$.
- $\mathbf{X} \in \mathbb{R}^{n \times p}$: (observed) data matrix where the row i of $\mathbf{X}$ is $\mathbf{x}_i^\top$.
- $\mathbf{F} \in \mathbb{R}^{n \times k}$: (latent) factor matrix where the row i of $\mathbf{X}$ is $\mathbf{f}_i^\top = (f_{i1}, f_{i2}, f_{i3})$
- $\mathbf{\Lambda} \in \mathbb{R}^{p \times k}$: loading matrix where $(\mathbf{\Lambda})_{ij}$ denotes a weight from the factor j to the observed variable i. In Figure the weight denotes the direct link from the factor toe the variable (common across the individuals)

Factor Model Representation

$$\mathbf{X} = \mathbf{F}\boldsymbol{\Lambda}^\top + \boldsymbol{\epsilon}$$

- $\mathbf{X} \in \mathbb{R}^{n \times p}$: p-dimensional vector of observed variables ($p = 8$ here)
- $\mathbf{F} \in \mathbb{R}^{n \times k}$: k-dimensional vector of latent factors ($k < p$)
- $\boldsymbol{\Lambda} \in \mathbb{R}^{p \times k}$: $p \times k$ factor loading matrix
- $\boldsymbol{\epsilon} \in \mathbb{R}^{n \times p}$: error matrix ($\epsilon_i^\top$ denotes the row i of $\boldsymbol{\epsilon}$.)

Factor Model: Covariance-Based Derivation

Factor model:

$$\mathbf{x}_i = \mathbf{\Lambda}\mathbf{f}_i + \epsilon_i$$

- $\mathbf{x}_i \in \mathbb{R}^p$: observed variables
- $\mathbf{f}_i \in \mathbb{R}^k$: common factors
- $\epsilon_i \in \mathbb{R}^p$: unique factors

Factor Model: Covariance-Based Derivation

Assumptions:

- $\mathbb{E}[\mathbf{f}_i] = \mathbf{0}, \quad \mathbb{E}[\boldsymbol{\epsilon}_i] = \mathbf{0}$
- $\mathbb{E}[\mathbf{f}_i \mathbf{f}_i^\top] = \mathbf{I}_k$ (factors are standardized)
- $\mathbb{E}[\boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i^\top] = \boldsymbol{\Psi}$ (diagonal matrix)
- $\mathbb{E}[\mathbf{f}_i \boldsymbol{\epsilon}_i^\top] = \mathbf{0}$ (independence)

Covariance structure:

$$\boldsymbol{\Sigma} = \mathbb{E}[\mathbf{x}_i \mathbf{x}_i^\top] = \boldsymbol{\Lambda} \mathbb{E}[\mathbf{f}_i \mathbf{f}_i^\top] \boldsymbol{\Lambda}^\top + \mathbb{E}[\boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i^\top] = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top + \boldsymbol{\Psi}$$

- $\boldsymbol{\Sigma} \in \mathbb{R}^{p \times p}$: covariance matrix of observed variables

Estimation of Factor Models: Factor Loadings

Estimation of Factor Loadings $\boldsymbol{\Lambda}$

Goal: Given the sample covariance matrix $\mathbf{S} \approx \boldsymbol{\Sigma}$, estimate

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top + \boldsymbol{\Psi}$$

Approaches:

- **Principal Factor Method (PFA):**
  - Replace $\boldsymbol{\Psi}$ with initial estimates (e.g., communality), then perform eigen decomposition on $\mathbf{S} - \boldsymbol{\Psi}$
  - Retain top k eigenvalues/vectors to estimate $\boldsymbol{\Lambda}$
- **Maximum Likelihood Estimation (MLE):**
  - Assume multivariate normality: $\mathbf{x}_i \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$
  - Maximize log-likelihood:

$$\ell(\boldsymbol{\Lambda}, \boldsymbol{\Psi}) = -\frac{n}{2}\left(\log|\boldsymbol{\Sigma}| + \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S})\right)$$

  - Numerical optimization is required

Principal Factor Analysis (PFA)

Step-by-step estimation:

1. Compute the sample covariance matrix $\mathbf{S}$ (estimate of $\boldsymbol{\Sigma}$)
2. Estimate initial uniqueness: $\boldsymbol{\Psi}_0 = \text{diag}(\psi_1, \dots, \psi_p)$ (e.g., $\psi_j = s_{jj} - h_j^2$)
3. Define common variance matrix: $\mathbf{S}_c = \mathbf{S} - \boldsymbol{\Psi}_0$ (estimate of $\boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top$)
4. Eigen-decomposition: $\mathbf{S}_c = \mathbf{V}\mathbf{D}\mathbf{V}^\top$
5. Retain top k components:

$$\boldsymbol{\Lambda} = \mathbf{V}_k \mathbf{D}_k^{1/2}$$

Note: Iterate if desired to update $\boldsymbol{\Psi}$ using residuals.

MLE Optimization for Factor Analysis (1/2)

Challenge: No closed-form solution for

$$\ell(\boldsymbol{\Lambda}, \boldsymbol{\Psi}) = -\frac{n}{2}\left(\log|\boldsymbol{\Sigma}| + \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S})\right)$$

Approaches:

- **EM Algorithm**
  - Treat latent factors $\mathbf{f}_i$ as missing data
  - Iteratively update expected sufficient statistics (E-step) and parameter estimates (M-step)
  - Guarantees non-decreasing likelihood
- **Direct Numerical Optimization**
  - Maximize the likelihood directly over $\boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}$
  - Use algorithms such as Newton-Raphson, Fisher scoring, or BFGS

Gradient of the Log-Likelihood (w.r.t. $\mathbf{\Lambda}$)

Log-likelihood function:

$$\ell(\mathbf{\Lambda}, \mathbf{\Psi}) = -\frac{n}{2}\left(\log|\mathbf{\Sigma}| + \mathrm{tr}\left(\mathbf{\Sigma}^{-1}\mathbf{S}\right)\right) \quad \text{with } \mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Lambda}^\top + \mathbf{\Psi}$$

Gradient with respect to $\mathbf{\Lambda}$:

$$\frac{\partial \ell}{\partial \mathbf{\Lambda}} = -\frac{n}{2}\left(\frac{\partial}{\partial \mathbf{\Lambda}}\log|\mathbf{\Sigma}| + \frac{\partial}{\partial \mathbf{\Lambda}}\mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{S})\right)$$

Using matrix calculus:

$$\frac{\partial \ell}{\partial \mathbf{\Lambda}} = -n\left(\mathbf{\Sigma}^{-1}\mathbf{S}\mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1}\right)\mathbf{\Lambda}$$

Factor Rotation: Motivation

- Factor loadings $\boldsymbol{\Lambda}$ are not unique.
- Any rotation $\mathbf{T}$ of $\boldsymbol{\Lambda}$ preserves the model:

$$\boldsymbol{\Lambda}^* = \boldsymbol{\Lambda}\mathbf{T}, \quad \text{with } \mathbf{T}^\top\mathbf{T} = \mathbf{I}$$

- Goal: Simplify interpretation by achieving a structure where each variable loads highly on only one factor.

Example:

- Without rotation: mixed loadings across all factors
- With rotation: clearer factor-variable associations

Types of Rotation

Orthogonal Rotation (e.g., Varimax):

- $\boldsymbol{\Lambda}^* = \boldsymbol{\Lambda}\mathbf{T}$
- $\mathbf{T}^\top\mathbf{T} = \mathbf{I}$
- Factors remain uncorrelated

Oblique Rotation (e.g., Promax, Oblimin):

- $\mathbf{T}^\top\mathbf{T} \neq \mathbf{I}$
- Factor correlation matrix: $\boldsymbol{\Phi} = \mathbf{T}^\top\mathbf{T}$
- Allows for correlated latent factors

Identifiability under $\mathbf{\Lambda}^\top \mathbf{\Lambda} = \mathbf{D}$

Factor model:

$$\mathbf{X} = \mathbf{\Lambda} \mathbf{f} + \mathbf{e}$$

Assumptions:

- $\mathrm{Cov}(\mathbf{f}) = \mathbf{I}_k$, $\mathrm{Cov}(\mathbf{e}) = \sigma^2 \mathbf{I}_p$
- $\mathrm{Cov}(\mathbf{f}, \mathbf{e}) = \mathbf{0}$
- $\mathbf{\Lambda}^\top \mathbf{\Lambda} = \mathbf{D}$, diagonal

Implication:

- This constraint eliminates orthogonal indeterminacy
- $\mathbf{\Lambda}$ is **identifiable up to sign changes**

Estimation of Factor Models: Factor Scores

Definition of Factor Scores

Factor score: An estimate of the latent factor $\mathbf{f}_i$ for each observation $\mathbf{x}_i$, based on the model:

$$\mathbf{x}_i = \mathbf{\Lambda}\mathbf{f}_i + \boldsymbol{\epsilon}_i$$

- $\mathbf{\Lambda} \in \mathbb{R}^{p \times k}$: factor loading matrix
- $\mathbf{f}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_k)$: latent factors
- $\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{\Psi})$: unique factors (diagonal covariance)

Goal: Estimate $\hat{\mathbf{f}}_i \approx \mathbb{E}[\mathbf{f}_i \mid \mathbf{x}_i]$

Derivation: Conditional Expectation

Joint distribution of $\mathbf{x}_i$ and $\mathbf{f}_i$:

$$\begin{bmatrix} \mathbf{x}_i \\ \mathbf{f}_i \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{\Lambda} \\ \boldsymbol{\Lambda}^\top & \mathbf{I}_k \end{bmatrix} \right), \quad \text{where } \boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top + \boldsymbol{\Psi}$$

From properties of multivariate normal distributions:

$$\mathbb{E}[\mathbf{f}_i \mid \mathbf{x}_i] = \boldsymbol{\Lambda}^\top \boldsymbol{\Sigma}^{-1} \mathbf{x}_i$$

Computation Methods for Factor Scores

1. **Regression method (Thomson, 1939):**

$$\hat{\mathbf{f}}_i = \boldsymbol{\Lambda}^\top \boldsymbol{\Sigma}^{-1} \mathbf{x}_i$$

2. **Bartlett's method (Bartlett, 1937):**

$$\hat{\mathbf{f}}_i = \left(\boldsymbol{\Lambda}^\top \boldsymbol{\Psi}^{-1} \boldsymbol{\Lambda} \right)^{-1} \boldsymbol{\Lambda}^\top \boldsymbol{\Psi}^{-1} \mathbf{x}_i$$

- Bartlett's estimator minimizes the residual unique variance.
- Regression scores may be correlated across factors, while Bartlett scores are uncorrelated but scale-dependent.

Visualizing Factor Scores

Explore the distribution of observations in the latent factor space.

- Each observation i has an estimated factor score vector:

$$\hat{\mathbf{f}}_i = (\hat{f}_{i1}, \hat{f}_{i2}, \ldots, \hat{f}_{ik})$$

- Plot $\hat{f}_{i1}$ vs. $\hat{f}_{i2}$ to visualize patterns:
  - Group separation
  - Outlier detection
  - Interpretation of latent dimensions

Common plots:

- 2D scatter plot of $\hat{f}_{i1}$ and $\hat{f}_{i2}$
- Color by categorical groups or cluster labels
- Add text labels or convex hulls for clusters