Topic 2: Factor analysis concepts and interpretation

Factor analysis

Let $\boldsymbol{Y}_i, i = 1, 2, .., n$ be independent $N_p(\boldsymbol{\mu}, \Sigma)$ variables (think of the $\boldsymbol{Y}_i$s as a results of a battery of p tests applied to the ith individual). Fundamental assumption in factor analysis:

$$\boldsymbol{Y}_i = \Lambda \boldsymbol{f}_i + \boldsymbol{e}_i \tag{4.3}$$

$\Lambda \in M_{p,k}$ factor loading matrix (full rank);

$\boldsymbol{f}_i \in \mathbb{R}^k (k < p)$ factor variable. The components of $\boldsymbol{f}_i$ are thought to be the (latent) factors. Usually $\boldsymbol{f}_i$ are taken to be independent $N(\boldsymbol{\alpha}, I_k)$ (i.e., "orthogonal") but also "oblique" factors are considered sometimes with a covariance matrix $\neq I_k$.

$\boldsymbol{e}_i$ independent $N(\boldsymbol{\theta}, \Sigma_e)$ withΣ_e **diagonal**, i.e., $\Sigma_e = \mathrm{diag}(\sigma_1^2, \sigma_2^2, .., \sigma_p^2)$.

Also, the $\boldsymbol{e}$s are independent from the $\boldsymbol{f}$s.

Then,

$$\boldsymbol{\mu} = \Lambda \boldsymbol{\alpha} + \boldsymbol{\theta}; \Sigma = \Lambda \Lambda^\top + \Sigma_e,$$

or, componentwise:

$$\mathrm{Var}(Y_{ir}) = \sum_{j=1}^{k} \lambda_{rj}^2 + \sigma_r^2 = \text{communality} + \text{uniqueness}.$$

$$\mathrm{Cov}(Y_{ir}, Y_{is}) = \sum_{j=1}^{k} \lambda_{rj} \lambda_{sj}.$$

The fundamental idea of factor analysis is to describe the **covariance relationships** among **many** variables (p "large") in terms of few (k "small") underlying, not observable (latent) random quantities (the **factors**). The model is motivated by the following argument: suppose variables can be grouped by their correlations. That is, all variables in a particular group are highly correlated among themselves but have relatively small correlations with variables in a different group. It is then quite reasonable to assume that each group of variables represents a single underlying construct (**factor**) that is "responsible" for the observed correlations.

Optional viewing: Factor Analysis - an introduction

Ben Lambert. (2014). Factor Analysis - an introduction. Retrieved from: https://youtu.be/WV_jcaDBZ2I

Important notes

- The model (4.3) is similar to a linear regression model but the key differences are that $\boldsymbol{f}_i$ are **random and are not observable**.
- If we knew the Λ (or have found estimates of them), then using properties of orthogonal projections on the linear space spanned by the columns of Λ, we would get:

$$\hat{\boldsymbol{\alpha}} = (\Lambda^\top \Lambda)^{-1} \Lambda^\top \bar{\boldsymbol{Y}}; \hat{\boldsymbol{\theta}} = \bar{\boldsymbol{Y}} - \Lambda \hat{\boldsymbol{\alpha}}.$$

  Because of the above observation, we can consider only $\boldsymbol{\mu}$, Λ, and σ_i^2, $i = 1, 2, \ldots, p$ as unknown parameters when parameterising the factor analysis model. Note also that primary interest in factor analysis is focused on estimating Λ.
- **There is a fundamental indeterminacy** in this model even when we require that $\mathrm{Var}(\boldsymbol{f}) = I_k$ since, if $P \in \mathcal{M}_{k,k}$ **is any** orthogonal matrix then obviously

$$\Lambda\Lambda^\top = \Lambda P (\Lambda P)^\top; \ \Lambda \boldsymbol{f}_i = (\Lambda P)(P^\top \boldsymbol{f}_i).$$

Hence replacing Λ by ΛP and $\boldsymbol{f}_i$ by $P^\top \boldsymbol{f}_i$ leads to the same equations.

Maximum Likelihood Estimation

The likelihood function for the n observations $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_n \in \mathbb{R}^p$ is

$$\begin{aligned} L(\boldsymbol{Y}; \boldsymbol{\mu}, \Lambda, \sigma_1^2, \sigma_2^2, .., \sigma_p^2) &= (2\pi)^{-np/2} |\Sigma|^{-n/2} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{Y}_i - \boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{Y}_i - \boldsymbol{\mu})\right] \\ &= (2\pi)^{-np/2} |\Sigma|^{-n/2} \exp\left[-\frac{n}{2}\left(\operatorname{tr}(\Sigma^{-1}\boldsymbol{S}) + (\bar{\boldsymbol{Y}} - \boldsymbol{\mu})^\top \Sigma^{-1}(\bar{\boldsymbol{Y}} - \boldsymbol{\mu})\right)\right] \end{aligned}$$

with $\boldsymbol{S} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{Y}_i - \bar{\boldsymbol{Y}})(\boldsymbol{Y}_i - \bar{\boldsymbol{Y}})^\top$, and keeping in mind that Σ is a function of Λ and Σ_e (and therefore of $\sigma_1^2, \ldots, \sigma_p^2$). Taking $\log L$, we get:

$$\log L(\boldsymbol{Y}; \boldsymbol{\mu}, \Lambda, \sigma_1^2, \sigma_2^2, .., \sigma_p^2) = -\frac{np}{2}\log(2\pi) - \frac{n}{2}\log(|\Sigma|) - \frac{n}{2}\left[\operatorname{tr}(\Sigma^{-1}\boldsymbol{S}) + (\bar{\boldsymbol{Y}} - \boldsymbol{\mu})^\top \Sigma^{-1}(\bar{\boldsymbol{Y}}\right.$$

After some vector calculus and matrix algebra, we find that,

$$(\Sigma_e^{-1/2}\boldsymbol{S}\Sigma_e^{-1/2})\Sigma_e^{-1/2}\Lambda = \Sigma_e^{-1/2}\Lambda(I + \Lambda^\top\Sigma_e^{-1}\Lambda). \tag{4.4}$$

Recall the note about indeterminacy of Λ. This can be a blessing in disguise, in particular (at least one) solution is one for which $\Lambda^\top\Sigma_e^{-1}\Lambda$ **is diagonal**. Then (4.4) implies that the matrix $\Sigma_e^{-1/2}\Lambda$ has as its columns k eigenvectors that correspond to the k eigenvalues of $\Sigma_e^{-1/2}\boldsymbol{S}\Sigma_e^{-1/2}$. More subtle analysis shows that to obtain the maximum likelihood estimator, these have to be the eigenvectors that correspond to the **largest** eigenvalues of $\Sigma_e^{-1/2}\boldsymbol{S}\Sigma_e^{-1/2}$.

Based on this fact, the following iterative solution (due to Lawley) has been proposed that can be described algorithmically as follows:

1. With an initial guess $\tilde{\Sigma}_e$, calculate $\tilde{\Sigma}_e^{-1/2}\tilde{\Lambda}$ by using the eigenvectors of the k largest eigenvalues of $\tilde{\Sigma}_e^{-1/2}\boldsymbol{S}\tilde{\Sigma}_e^{-1/2}$.
2. Then from $\tilde{\Sigma}_e^{-1/2}\tilde{\Lambda}$, get a (first iteration) value for $\tilde{\Lambda}$.
3. With this value of $\tilde{\Lambda}$ we can calculate the value of $\tilde{Q}(\tilde{\Sigma}_e) = \frac{1}{2}\log|\tilde{\Lambda}\tilde{\Lambda}^\top + \tilde{\Sigma}_e| + \frac{1}{2}\operatorname{tr}(\tilde{\Lambda}\tilde{\Lambda}^\top + \tilde{\Sigma}_e)^{-1}\boldsymbol{S}$ (which is the value of the functional). This functional only depends on the p nonzero values of $\tilde{\Sigma}_e$ and there are several powerful numerical procedures to find its minimum.
4. If it is achieved at Σ_e^*, then update $\tilde{\Sigma}_e$ with the new guess Σ_e^* and repeat from Step 1 to convergence.