Some Factor Analysis

Estimation for the Simplest Factor-Analytic Model

The common factor model:

$$\boldsymbol{y}_i = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{f}_i + \boldsymbol{\epsilon}_i$$

for person *i*, in which *f* is a latent variable and the matrix of "factor loadings" (regression parameters for the *y*s on the *f*s) and the (independent) residual variances are to be estimated.

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda & \lambda & \dots & \lambda \\ \lambda & \lambda & \dots & \lambda \\ \lambda & \lambda & \dots & \lambda \\ \vdots & \vdots & & \vdots \\ \lambda & \lambda & \dots & \lambda \end{bmatrix} \quad \boldsymbol{\Delta} = \begin{bmatrix} \sigma^2_{\epsilon_i} & 0 & \dots & 0 \\ 0 & \sigma^2_{\epsilon_i} & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \sigma^2_{\epsilon_i} \end{bmatrix}$$

This implies that the observed variables as distributed in multivariate normal form with mean $\boldsymbol{\mu}$ and covariance matrix:

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Delta}$$

Among the readings, Bock & Bargmann (1966), Jennrich & Robinson (1969), Jöreskog(1969, 1971), and Rubin & Thayer (1982) variously develop what is now commonly called the "Wishart" likelihood that may be maximized to estimate the parameters.

The multivariate normal likelihood is:

$$L = \prod_{i=1}^{N} \frac{|\boldsymbol{\Sigma}|^{-1/2}}{(2\pi)^{p/2}} \exp[-\frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu})]$$

The maximum likelihood estimate of $\boldsymbol{\mu}$ is (either) obviously (or see Anderson, 1958, p. 47) the mean vector $\boldsymbol{y}.$ If we let the sum of products of the N observations corrected to the sample mean be

$$N \cdot \boldsymbol{S} = \sum_{i=1}^{N} \boldsymbol{y}\boldsymbol{y}' - N\boldsymbol{y}.\boldsymbol{y}.',$$

and the loglikelihood to be maximized to estimate the parameters that yield $\boldsymbol{\Sigma}$ is:

$$\log L = -\frac{Np}{2}\log 2\pi - \frac{N}{2}\log|\boldsymbol{\Sigma}| - \frac{N}{2}\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{S}.$$

For maximization, we compute only the part that involves the data:

$$\ell = -\frac{N}{2}\log|\boldsymbol{\Sigma}| - \frac{N}{2}\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{S}$$

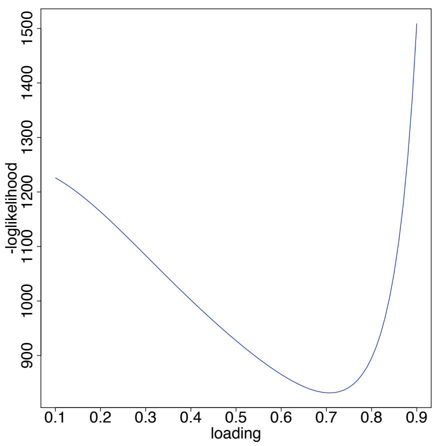
Sometimes the criterion function is modified to include all parts of the (asymptotically chi-square distributed) likelihood-ratio goodness of fit statistic:

$$F = (N-1)(\log|\boldsymbol{\Sigma}| + \mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{S} - \log|\boldsymbol{S}| - p)$$

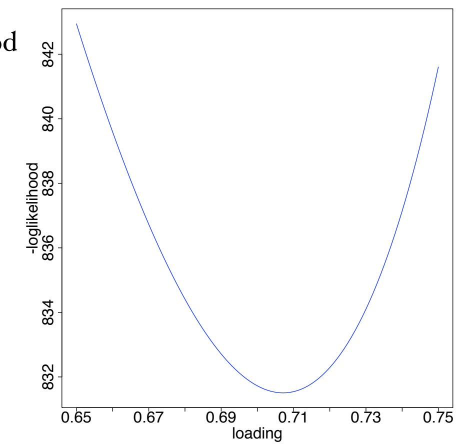
Some examples

- Naive estimation, one factor equal loadings
- Jennrich & Robinson's suggestion
- Jöreskog's first "congeneric tests" example (for this we do three models, and then the unrestricted model using the the EM algorithm of Rubin & Thayer)
- And finally back to Bock & Bargmann

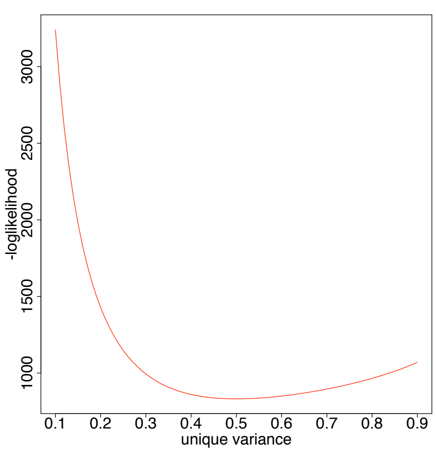
The loglikelihood for the (single, equal) factor loading; all correlations equal 0.5



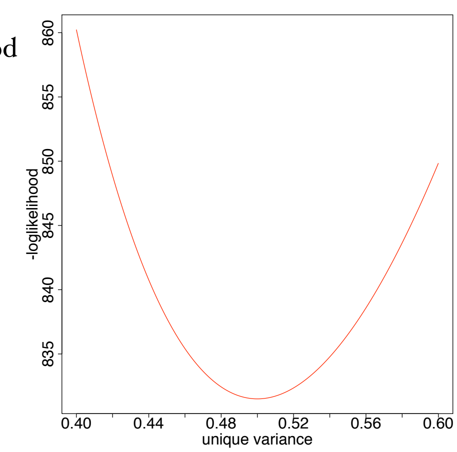
The loglikelihood for the (single, equal) factor loading; all correlations equal 0.5



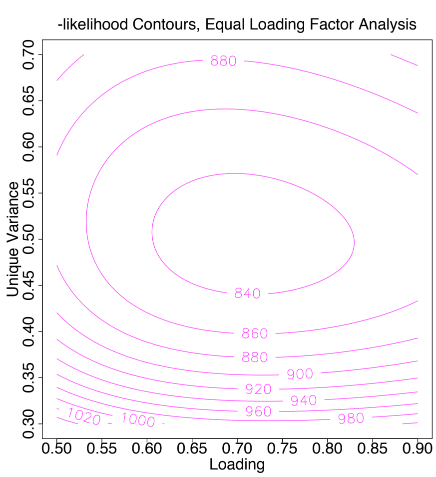
The loglikelihood for the (single, equal) unique variance; all correlations equal 0.5



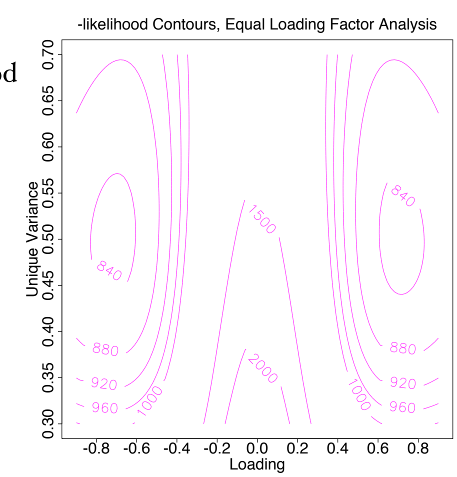
The loglikelihood for the (single, equal) unique variance; all correlations equal 0.5



The loglikelihood for both the loading and the unique variance; all correlations equal 0.5



The loglikelihood for both the loading and the unique variance; all correlations equal 0.5



Some Factor Analysis

EM Estimation for the Factor-Analytic Model

A purely statistical (not psychological) description of factor analysis in Rubin & Thayer's notation:

Υ is the (centered) ***n*** x ***p*** observed data matrix, and *Z* an ***n*** x ***q*** unobserved matrix of factor scores.
Each column of *Z* is *N*(0,1) with correlation matrix *R* among columns.
The conditional distribution (given *Z*) of the *i*th row of *Υ* is normal with mean

$$\alpha + Z_i\beta$$

and residual covariance

$$\tau^2 = \text{diag}(\tau_1^2, \ldots, \tau_p^2)$$

In (our) traditional language, the βs are factor loadings (regression coefficients of the *Υ*s on the *Z*s) and the τ^2s are the unique variances.

We will consider only their "case 1" with *R* = *I* (orthogonal factors) and unrestricted

For the EM algorithm, consider the *Z*s "missing data" and then estimate the regression parameters anyway. For that we'll need covariance matrices *C*:

$$C_{yy} = \sum_1^n \frac{Y_i'Y_i}{n} \qquad C_{yz} = \sum_1^n \frac{Y_i'Z_i}{n} \qquad C_{zz} = \sum_1^n \frac{Z_i'Z_i}{n}$$

E-Step:

$$E(C_{yy}|Y, \tau^2, \beta, R) = C_{yy}$$
$$E(C_{yz}|Y, \tau^2, \beta, R) = C_{yy}\delta$$
$$E(C_{zz}|Y, \tau^2, \beta, R) = \delta' C_{yy}\delta + \Delta$$

For the simplest special case, *R* = *I*,

$$\delta = (\tau^2 + \beta'\beta)^{-1}\beta' \qquad \Delta = I - \beta(\tau^2 + \beta'\beta)^{-1}\beta'$$

Why? The δs are the regression coefficients of the *Z*s on the *Υ*s.

$$\delta = (\Sigma_{YY})^{-1}\Sigma_{YZ} \qquad \Delta = I - \Sigma_{ZY}(\Sigma_{YY})^{-1}\Sigma_{YZ}$$

$$Z_i \sim N(\delta Y_i, \Delta)$$

M-Step:

$$\beta* = [\delta' C_{yy}\delta + \Delta]^{-1}(C_{yy}\delta)'$$

$$\tau*^2 = \text{diag}(C_{yy} - C_{yy}\delta[\delta' C_{yy}\delta + \Delta]^{-1}(C_{yy}\delta)'$$

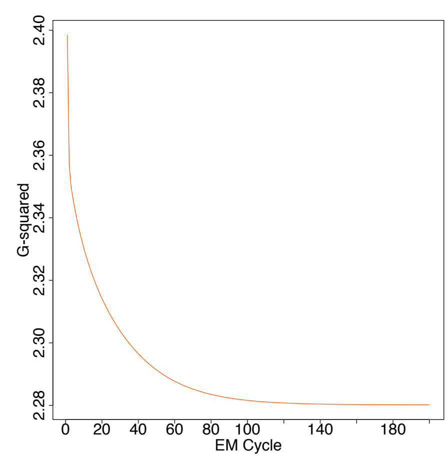
Why?

$$\beta* = [E(C_{zz}|Y, \tau^2, \beta, R)]^{-1}[E(C_{yz}|Y, \tau^2, \beta, R)]'$$

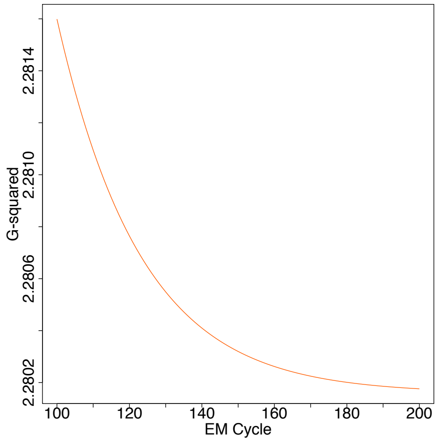
$$\tau*^2 = \text{diag}(C_{yy} - [E(C_{yz}|Y, \tau^2, \beta, R)][E(C_{zz}|Y, \tau^2, \beta, R)]^{-1}[E(C_{yz}|Y, \tau^2, \beta, R)]')$$

Regression.

EM computations ar quick per-cycle, but it takes man cycles



EM computations ar quick per-cycle, but it takes man cycles



The EM algorithm "climbs a local hill" from the starting values.

Rubin & Thayer rather go on about how revealing this is about the (often) multi-modal nature of factor analysis likelihoods.

Bock & Bargmann

Case I: The Quasi-Simplex

For repeated measurements (learning trials—the example variables involve scores at stages of learning on a two-hand coordination task):

"According to the simplex model, each of these variables incorporates a new component of skill at that stage of practice. These components are assumed to combine additively to determine the score of each subject at the respective stage of practice."

Bock & Bargmann, p. 523

Algebraically, that gives a model for the test scores

$$\boldsymbol{y}_i = \boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{\xi}_i + \boldsymbol{\epsilon}_i$$

for person i, in which $\boldsymbol{\xi}_i$ is a latent variable and $\boldsymbol{A}$ is fixed and known:

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 1 & 1 & \dots & 1 \end{bmatrix}$$

$$\boldsymbol{y}_i = \boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{\xi}_i + \boldsymbol{\epsilon}_i$$

implies that the observed variables as distributed in multivariate normal form with mean $\boldsymbol{\mu}$ and covariance matrix:

$$\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{\Phi}\boldsymbol{A}' + \boldsymbol{\Gamma}$$

Case I, the only one we'll discuss, restricts the latent variables to be uncorrelated,

$$\boldsymbol{\Phi} = \mathrm{diag}[\phi_1, \phi_2, \dots, \phi_m]$$

and the error variances to be homoscedastic:

$$\boldsymbol{\Gamma} = \gamma \boldsymbol{I}$$

The multivariate normal likelihood is:

$$L = \prod_{i=1}^{N} \frac{|\boldsymbol{\Sigma}|^{-1/2}}{(2\pi)^{p/2}} \exp[-\frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu})]$$

The maximum likelihood estimate of $\boldsymbol{\mu}$ is (either) obviously (or see Anderson, 1958, p. 47) the mean vector $\boldsymbol{y}.$ If we let the sum of products of the N observations corrected to the sample mean be

$$N \cdot \boldsymbol{S} = \sum_{i=1}^{N} \boldsymbol{y}\boldsymbol{y}' - N\boldsymbol{y}.\boldsymbol{y}.',$$

and the loglikelihood to be maximized to estimate the parameters that yield $\boldsymbol{\Sigma}$ is:

$$\log L = -\frac{Np}{2}\log 2\pi - \frac{N}{2}\log|\boldsymbol{\Sigma}| - \frac{N}{2}\mathrm{tr}\boldsymbol{\Sigma}^{-1}\boldsymbol{S}.$$

Bock and Bargmann show how to figure the derivatives of that loglikelihood, and use a Newton-Raphson algorithm to find parameter estimates that maximize it.

Among residuals from the first class, we have:

- Derivative-free R
- R with derivatives
- C++ (with derivatives; no choice here)