SEM Algorithm

Munch, J., Schreck, F., Höft, N.

Statistics M. Sc. - Joint masters program
Free University Berlin, Berlin, Germany
Humboldt University Berlin, Berlin, Germany

1. Motivation

- Expectation-Maximization (EM) is one of the most popular tools to obtain maximum likelihood estimates in incomplete data problems
- Problem: EM algorithm does not provide an asymptotic variance-covariance matrix for the parameters
- Such matrix is needed for asymptotically valid inference
- Existing approaches are computationally problematic and not generically applicable
- SEM algorithm provides numerically stable asymptotic variance-covariance matrix for estimates obtained by EM (Meng & Rubin, 1991)

2. EM algorithm

- In latent variable problems, especially in incomplete data problems, maximum likelihood estimation is analytically not possible
- EM algorithm enables convergence to optimum, which could be a local optimum depending on the starting values (Dempster et.al, 1977)
- EM algorithm consists of two iteratively applied steps:

UNTIL $\Delta\theta < \varepsilon$ REPEAT:

E-step:

Find expected complete-data log-likelihood and treat $\theta^{(t)}$ as the true parameter θ

$$Q\left(\theta|\theta^{(t)}\right) = E_{Y_{mis}|Y_{obs},\theta^{(t)}}[logL(\theta; Y_{obs}, Y_{mis})]$$

M-step:

Update $\theta^{(t)}$ by the value that maximizes the expected log-likelihood from the E-step

$$\theta^{(t+1)} = \underset{\theta}{\operatorname{argmax}}\, Q\left(\theta|\theta^{(t)}\right)$$

Calculate change in estimated parameters

$$\Delta\theta = |\theta^{(t+1)} - \theta^{(t)}|$$

3. Supplemented EM algorithm

- Calculation of asymptotic variance-covariance matrix V is done in three steps:

Step 1:

Evaluate adjusted information matrix I_{oc}

$$I_{oc} = I_o(\theta^*|S^*(Y_{obs})), where\ S^* = E[S(Y)|Y_{obs}, \theta]$$

Step 2:

Calculate the elements of rate-of-convergence matrix DM

$$DM = \begin{pmatrix} r_{11} & \cdots & r_{1j} \\ \vdots & \ddots & \vdots \\ r_{i1} & \cdots & r_{ij} \end{pmatrix}$$

FOR $i = 1, \ldots, d$

FOR $j = 1, \ldots, d$

UNTIL $\Delta r_{ij} < \sqrt{\varepsilon}$ REPEAT:

Calculate element in rate-of-convergence matrix DM

$$r_{ij}{}^{(t)} = \frac{[\tilde{\theta}_j^{(t+1)}(i) - \theta_j{}^*]}{[\theta_i{}^{(t)} - \theta_i{}^*]},\ \text{for}\ t = 1, \ldots, T$$

Calculate change in $r_{ij}{}^{(t)}$

$$\Delta r_{ij} = \left|r_{ij}{}^{(t)} - r_{ij}{}^{(t-1)}\right|,\ \text{for}\ t = 1, \ldots, T$$

Step 3:

Calculate asymptotic variance-covariance matrix

$$V = I_{oc}{}^{-1} + \Delta V, where\ \Delta V = = I_{oc}{}^{-1}\mathrm{DM}(\mathrm{I} - \mathrm{DM})^{-1}$$

- Advantages of SEM:
  - + Matrix V not necessarily symmetric, which can be used for diagnostics of EM algorithm
  - + Computationally more stable than numerical differentiation
  - + Applicable in non-iid cases and under complicated missing data patters
  - + Parameters allowed to convergence with different number of steps
  - + Parallel computing possible

4. Simulation study

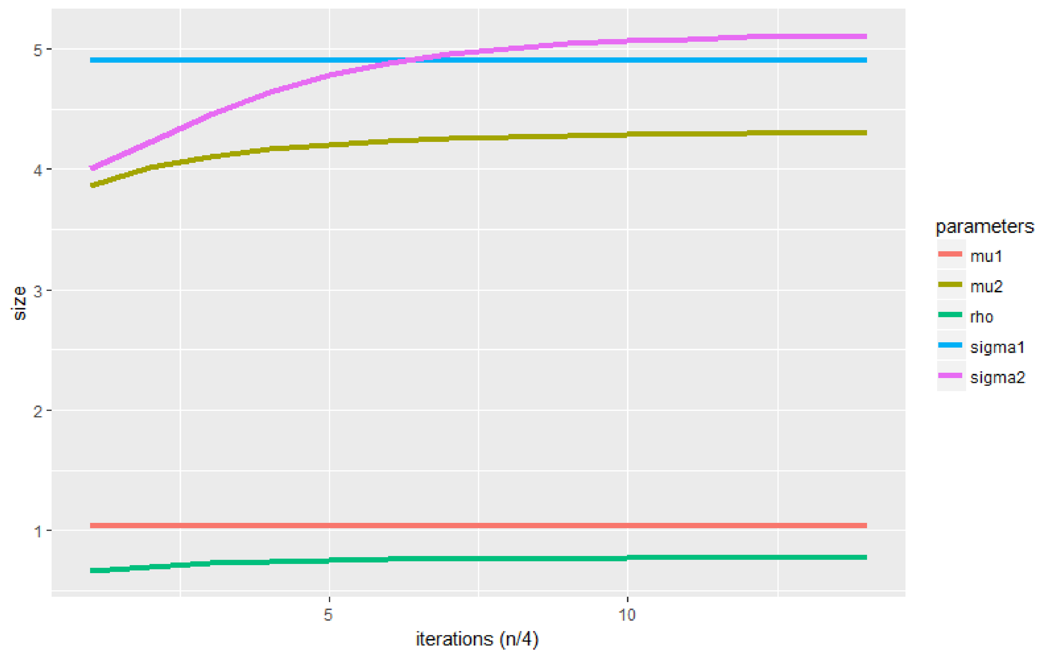
- Data simulation by random draw from bivariate normal distribution:

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 1 \\ 4 \end{pmatrix}, \begin{pmatrix} 4 & 3 \\ 3 & 5 \end{pmatrix}\right)$$

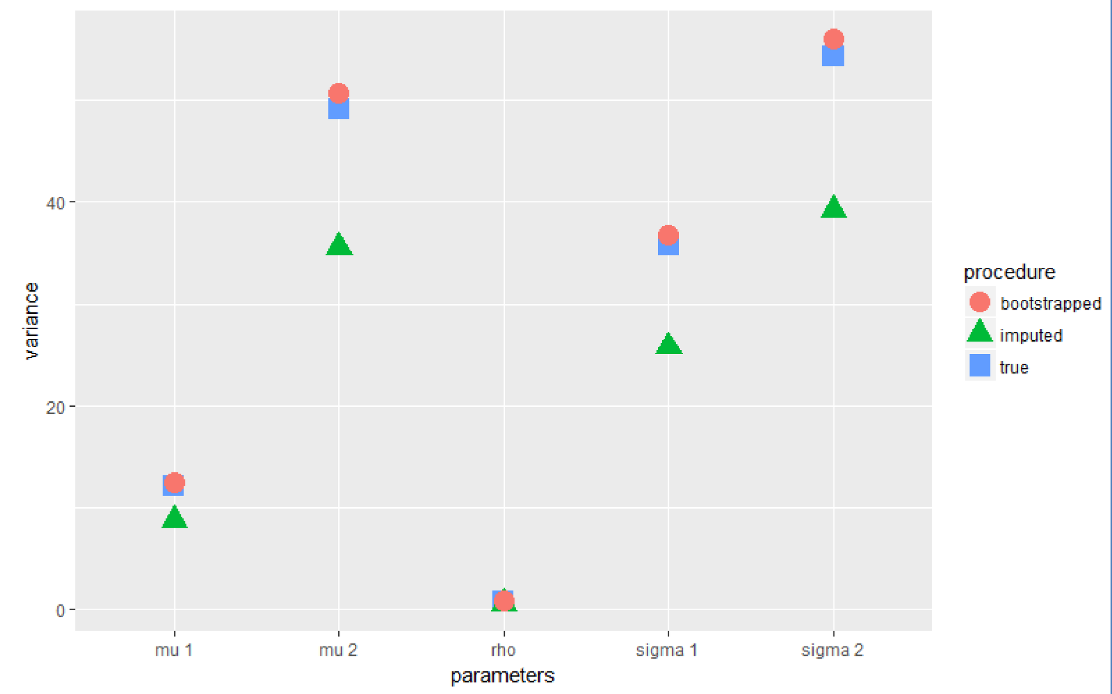
- MNAR pattern of missing data in Y_2, fraction $\tau = 0.4$
- Parameter vector $\theta = (\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, \rho)$
- True variance-covariance matrix is compared with EM imputed matrix an bootstrapped matrix

5. Results

- Convergence of EM parameters



- Comparison of parameter variances by calculation procedure



6. Results

- EM works well in our multi-parameter problem and algorithm converges fast
- Complexity of SEM implementation depends on type of model
- Numerical stability could not be achieved in our case
- In contrast, an asymptotic variance-covariance matrix can relatively easily be obtained from bootstrapping
- Bootstrapping delivers accurate results in our simulation study
- Bootstrapping computationally by far more costly than SEM, however the performance increase of modern hardware marginalizes this problem

7. Discussion

- In theory: SEM as an easy and convenient tool to obtain variance-covariance estimates for ML parameters in missing data problems
- In practice: Difficulties implementing a numerically stable solution
- Few implementations of SEM, like for item response problems (Pritikin, 2016)
- Comparison of SEM and Bootstrapping in small samples beyond the scope of this study

References

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Series B (methodological)*, 1-38.

McLachlan, G., & Krishnan, T. (2007). *The EM algorithm and extensions* (Vol. 382). John Wiley & Sons.

Meng, X. L., & Rubin, D. B. (1991). Using EM to obtain asymptotic variance-covariance matrices: The SEM algorithm. *Journal of the American Statistical Association*, *86*(416), 899-909.

Little, R. J., & Rubin, D. B. (2014). *Statistical analysis with missing data* (Vol. 333). John Wiley & Sons.

Pritikin, J. N. (2016). A computational note on the application of the Supplemented EM algorithm to item response models. arXiv preprint arXiv:1605.00860.