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. Introduction & Motivation

- Expectation-Maximization (EM) is one a method to obtain maximum likelihood estimates in incomplete data problems
- Problem: Provides no asymptotic variance-covariance matrix for the parameters
- However, this matrix is required for inference
- SEM algorithm provides an asymptotic variance-covariance matrix for EM estimates (Meng & Rubin, 1991)
- Disclaimer: SEM did not converge in our study.
- Possible reasons: Numerical instability, coding errors
- To produce similar (but adequate) results, we additionally implemented bootstrapping (see 5. Results)

2. EM algorithm

- In latent variable problems, (e.g. incomplete data problems), maximum likelihood estimation is analytically not possible
- EM algorithm convergences to a possibly 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

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

M-step:

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

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

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 $i = 1, \dots, d$

FOR $j = 1, \dots, d$

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

Calculate element in rate-of-convergence matrix DM

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

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

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

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 are allowed to convergence with different number of steps
  - \+ Parallel computing possible

4. Simulation study

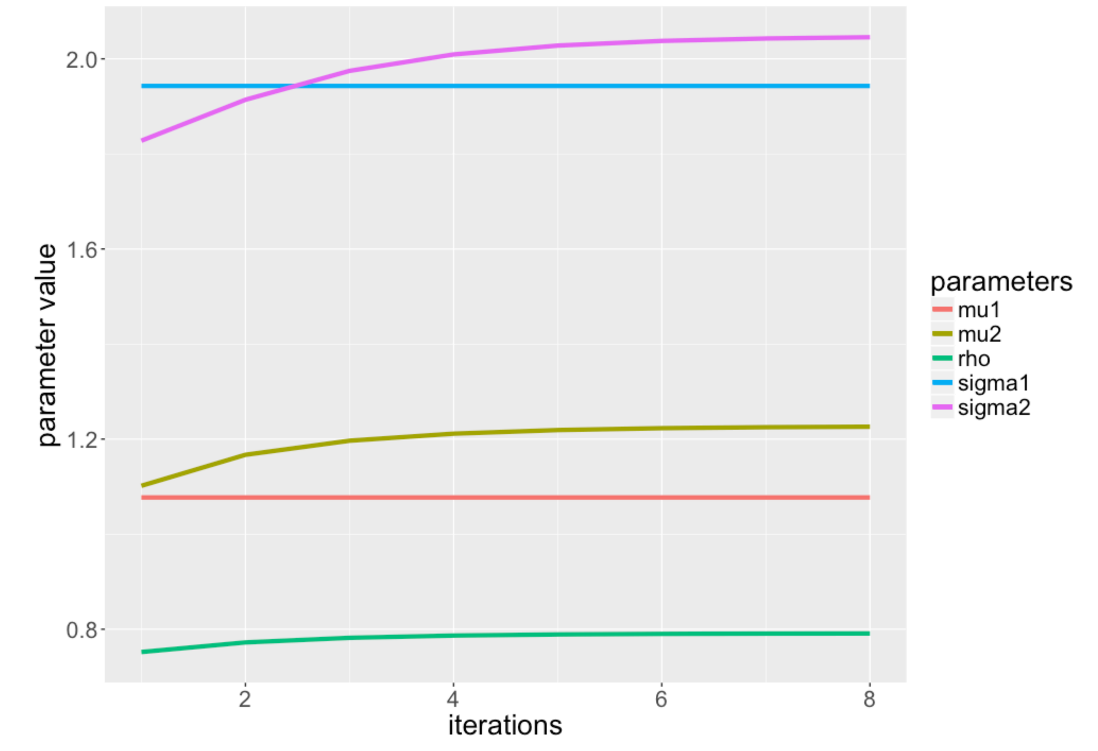
- Data simulation by random sampling from bivariate normal distribution:

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 1 \\ 1.5 \end{pmatrix}, \begin{pmatrix} 2 & 1.8 \\ 1.8 & 2.8 \end{pmatrix}\right)$$

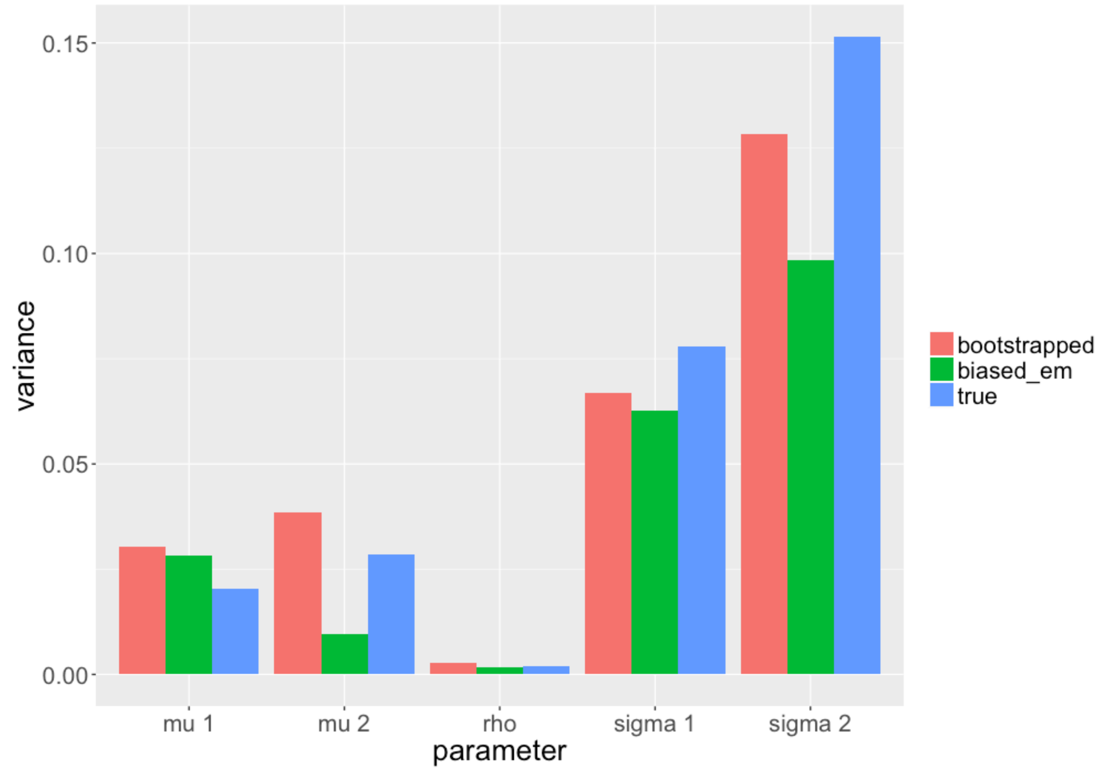
- MNAR pattern of missing data in Y_2, fraction $\tau = 0.3$
- Parameter vector $\theta = (\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, \rho)$
- True variance-covariance matrix (consistent MC estimation) is compared with biased EM matrix and bootstrapping matrix

5. Results

- Convergence of EM parameters



- Comparison of parameter variances by calculation procedure



6. Discussion

- 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. Conclusion

- In theory: SEM as an easy and convenient method to obtain variance-covariance estimates for ML parameters in missing data problems
- In practice: Implementing a stable solution may be difficult, depending on the model
- 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.