Missing values and imputation

data matrix $\underset{n\times p}{D}$

rows: $T_1, T_2, \dots, T_n$ i.i.d. $\sim \mathcal{N}_p(\mu, \Sigma)$

$\theta = (\mu, \Sigma)$

for every i: $T_i = (\underset{\substack{\text{observed}\\ d_1\times 1}}{Y_i}, \underset{\substack{\text{missing}\\ d_2\times 1}}{Z_i})$ $\quad d_1 + d_2 = p$

decompose for every i:

$$\mu_i = (\mu_{obs,i}, \mu_{mis,i})$$

$$\Sigma_i = \begin{pmatrix} \underset{d_1\times d_1}{\Sigma_{obs,i}} & \underset{d_1\times d_2}{\Sigma_{cross,i}} \\ \underset{d_2\times d_1}{\Sigma^T_{cross,i}} & \underset{d_2\times d_2}{\Sigma_{mis,i}} \end{pmatrix}$$

– log-likelihood:

$$-\sum_{i=1}^{n} \log \underbrace{f_\theta(Y_i)}_{\substack{\text{complicated expression}\\ \text{depending on missingness pattern of ith row of } D}}$$

complete likelihood is explicite

$\longrightarrow$ EM-algorithm is "suitable"

<u>sketch of EM-algorithm:</u>

1) start with $\hat{\mu}^{(0)}$, $\hat{\Sigma}^{(0)}$

2) E-step:

$$\mathbb{E}[Z_i \mid Y_i, \hat{\mu}^{(m)}, \hat{\Sigma}^{(m)}]$$

$$\underset{\text{Gaussian assumption}}{=} \hat{\mu}^{(m)}_{mis,i} + \left(\hat{\Sigma}^{(m)}_{cross,i}\right)^T \left(\hat{\Sigma}^{(m)}_{obs,i}\right)^{-1} \left(Y_i - \hat{\mu}^{(m)}_{obs,i}\right)$$

$$=: \hat{Z}_i$$

3) M-step: $\hat{\mu}^{(m+1)} = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n T_i \mid \underline{Y}, \hat{\theta}^{(m)}\right]$

$$= \left(\frac{1}{n}\sum_{i=1}^n Y_i, \frac{1}{n}\sum_{i=1}^n \hat{Z}_i\right)$$

(up to re-ordering)

$$\hat{\Sigma}^{(m+1)} = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n T_i T_i^T - \hat{\mu}^{(m+1)}\left(\hat{\mu}^{(m+1)}\right)^T \mid \underline{Y}, \hat{\theta}^{(m)}\right]$$

$\rightsquigarrow$ have to compute

$$\mathbb{E}[Y_i Y_i^T \mid \underline{Y}, \hat{\theta}^{(m)}] = Y_i Y_i^T$$

$$\mathbb{E}[Y_i Z_i^T \mid \underline{Y}, \hat{\theta}^{(m)}] = Y_i \hat{Z}_i^T$$

$$\mathbb{E}[Z_i Z_i^T \mid \underline{Y}, \hat{\theta}^{(m)}] = \hat{C}_i + \hat{Z}_i \hat{Z}_i^T$$

$$\hat{C}_i = \hat{\Sigma}^{(m)}_{mis,i} - \left(\hat{\Sigma}^{(m)}_{cross,i}\right)^T \left(\hat{\Sigma}^{(m)}_{obs,i}\right)^{-1} \hat{\Sigma}^{(m)}_{cross,i}$$

R-package: mvnmle