5th exercises for DAIM'2014

Ex. 1
Draw a sample of 1000 observations from three-dimensional multinormal distribution. Vector of expected values is $\boldsymbol{\mu} = (1, 2, 3)$ and covariance matrix is

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & 0.5 & 1.25 \\ 0.5 & 2 & 1.75 \\ 1.25 & 1.75 & 3 \end{bmatrix}$$

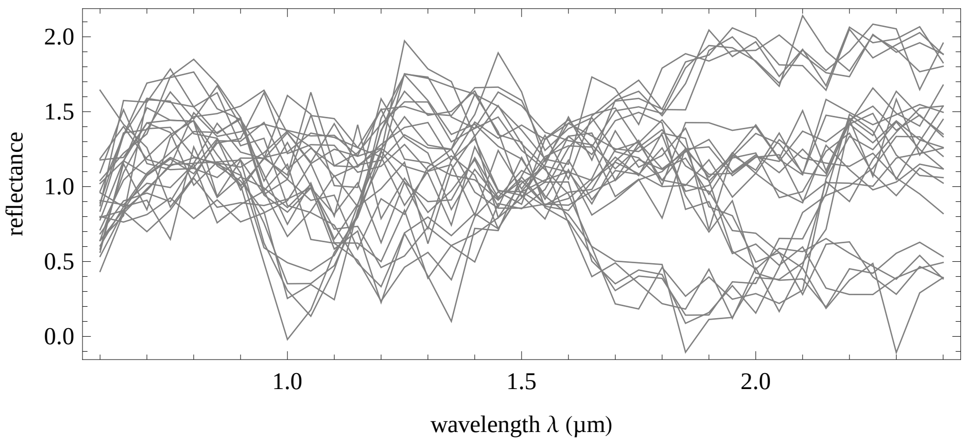
Use Eq. (6.7). When done, do scatterplots of Y_1 against Y_2, Y_1 against Y_3, and Y_2 against Y_3.

Ex. 2
Load three datafiles `MN-data-i.dat` from the course webpage. Each file has 500 observations from two-dimensional multinormal distribution. The covariance matrix is

$$\mathbf{\Sigma} = \begin{bmatrix} 10 & -4 \\ -4 & 2 \end{bmatrix}$$

for all the sets. Plot the data and compute mean vectors for the three sets. Compute so-called distance matrix between the means using both Euclidean distances and Mahalanobis distances. Distance matrix $\mathbf{D}$ is such that $[\mathbf{D}]_{ij}$ gives the distance between elements i and j.

Ex. 3
Datafile `spectra.dat` has 26 rows — first row has the wavelengths from 0.6 µm to 2.4 µm, and the rest are reflectancies of 25 "asteroids" at those wavelengths (plotted below).



Use principle component analysis to find underlying five groups of asteroids based on their spectra. The first three principle components should be enough to distinct the groups. Plot the observations in the new PCA variable space, e.g. PCA-1 against PCA-2 and PCA-1 against PCA-3.

If you succeed, you should notice that the first 5 rows of data belong to the first group, the second 5 rows to the second group etc. The spectra of those groups correspond to common minerals in asteroids — nickel-iron (obs. 1–5), olivine (obs. 6–10), orthopyroxene (obs. 11–15), plagioclase feldspar (obs. 16–20) and spinel-bearing Allende inclusion (obs. 21–25) (from Binzel et al. (eds.) Asteroids II).