Computational Methods (16:540:540)
Homework 3: Solution

1. Generate 2000 Multivariate Normal(MVN) data, e.g. MVN with $p = 3$, thus,

$$\mathbf{X} = (X_1, X_2, X_3)^T \sim \mathbf{N}_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

where

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \\ 2 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} 20 & 2 & -2 \\ 2 & 2 & 3 \\ -2 & 3 & 7 \end{bmatrix}$$

(a) Draw the histogram for X_1, X_2 and X_3.

(b) Find the sample mean of X_1, X_2 and X_3.

(c) Find the sample covariance matrix $\mathbf{S}$.

(d) Calculate the correlation ρ_{12} between X_1 and X_2.

(e) Find the conditional distribution of X_1 given $X_2 = x_2$

(f) Find $E[X_1|X_2 = 6]$ and $Var[X_1|X_2 = 6]$, $E[X_3|X_1 = 10]$ and $Var[X_3|X_1 = 10]$.

(g) Derive the joint p.d.f. of X_1, X_2 and X_3, and evaluate $f(0, 0, 1.2), f(0, 0, 0)$

Note : Try to manually calculate all those sample statistics rather than use functions of Matlab or Python.

Solution:
Use $x = mvnrnd(mu, sigma, n)$; to generate samples in Matlab. The data is shown as below.

Observation	X_1	X_2	X_3
1	7.405	1.638	3.842
2	13.201	3.431	5.081
3	-5.102	1.452	4.291
$\vdots$	$\vdots$	$\vdots$	$\vdots$
2000	8.856	0.904	0.569
mean	5.010	1.045	2.062
variance	19.956	2.023	7.025

(a) Figure 1 shows the histogram for X_1, X_2 and X_3. The peaks of the variables X_1, X_2 and X_3 are around their mean, and distributed as normal distribution.

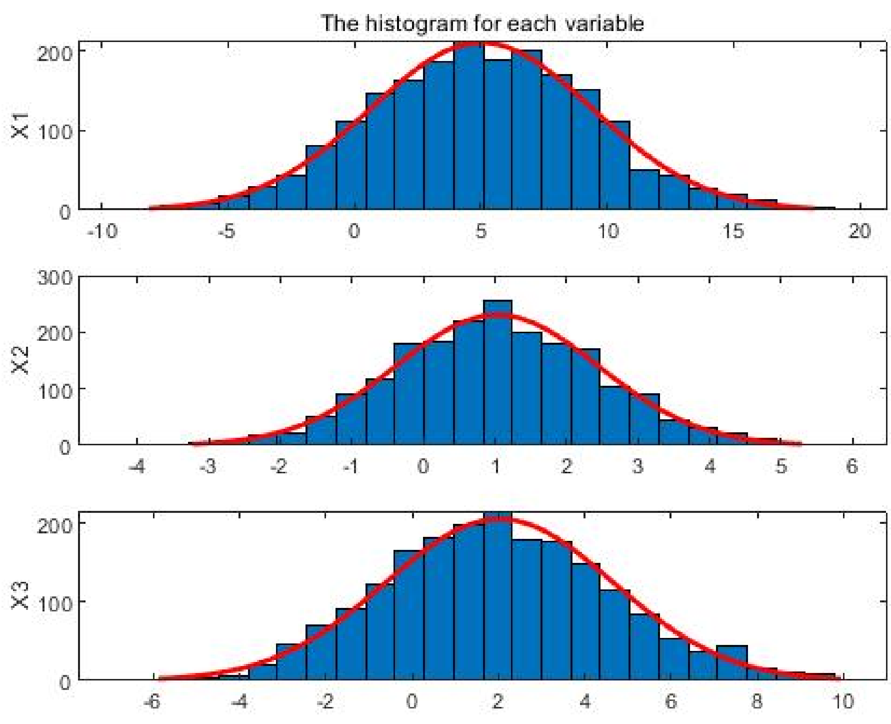

Figure 1: The histogram for X_1, X_2 and X_3.

(b) The sample mean is

$$\bar{\mathbf{X}} = \frac{1}{2000}\sum_{i=1}^{2000}\mathbf{X}_i = \begin{bmatrix}\bar{X}_1\\ \bar{X}_2\\ \bar{X}_3\end{bmatrix} = \begin{bmatrix}\frac{x_{1,1}+x_{2,1}+...+x_{2000,1}}{2000}\\ \frac{x_{1,2}+x_{2,2}+...+x_{2000,2}}{2000}\\ \frac{x_{1,3}+x_{2,3}+...+x_{2000,3}}{2000}\end{bmatrix} = \begin{bmatrix}5.010\\ 1.045\\ 2.062\end{bmatrix}$$

(c) The sample variance is

$$s_j^2 = s_{jj} = \frac{1}{n-1}\sum_{i=1}^{5}(x_{ij} - \bar{\mathbf{x}}_j)^2, j = 1, 2, 3$$

$$s_{jk} = \frac{1}{n-1}\sum_{i=1}^{5}(x_{ij} - \bar{\mathbf{x}}_j)(x_{ik} - \bar{\mathbf{x}}_k)$$

$$\mathbf{S_n} = \begin{bmatrix} s_1^2 & s_{12} & s_{13}\\ s_{21} & s_2^2 & s_{23}\\ s_{31} & s_{32} & s_3^2\end{bmatrix} = \begin{bmatrix}19.955 & 2.113 & -1.825\\ 2.113 & 2.023 & 3.026\\ -1.825 & 3.026 & 7.025\end{bmatrix}$$

(d) The sample correlation ρ_{ij} between X_i and X_j is

$$\rho_{11} = \rho_{22} = \rho_{33} = 1$$

$$\rho_{12} = \rho_{21} = \frac{s_{12}}{\sqrt{s_{11}}\sqrt{s_{22}}} \approx 0.333$$

$$\rho_{13} = \rho_{31} = \frac{s_{13}}{\sqrt{s_{11}}\sqrt{s_{33}}} \approx -0.154$$

$$\rho_{23} = \rho_{32} = \frac{s_{23}}{\sqrt{s_{22}}\sqrt{s_{33}}} \approx 0.803$$

$$\mathbf{R} = \begin{bmatrix}1 & \rho_{12} & \rho_{13}\\ \rho_{21} & 1 & \rho_{23}\\ rho_{31} & \rho_{32} & 1\end{bmatrix} = \begin{bmatrix}1 & 0.333 & -0.154\\ 0.333 & 1 & 0.803\\ -0.154 & 0.803 & 1\end{bmatrix}$$

The correlation ρ_{12} between X_1 and X_2 is 0.333, which mean they are positively correlated.

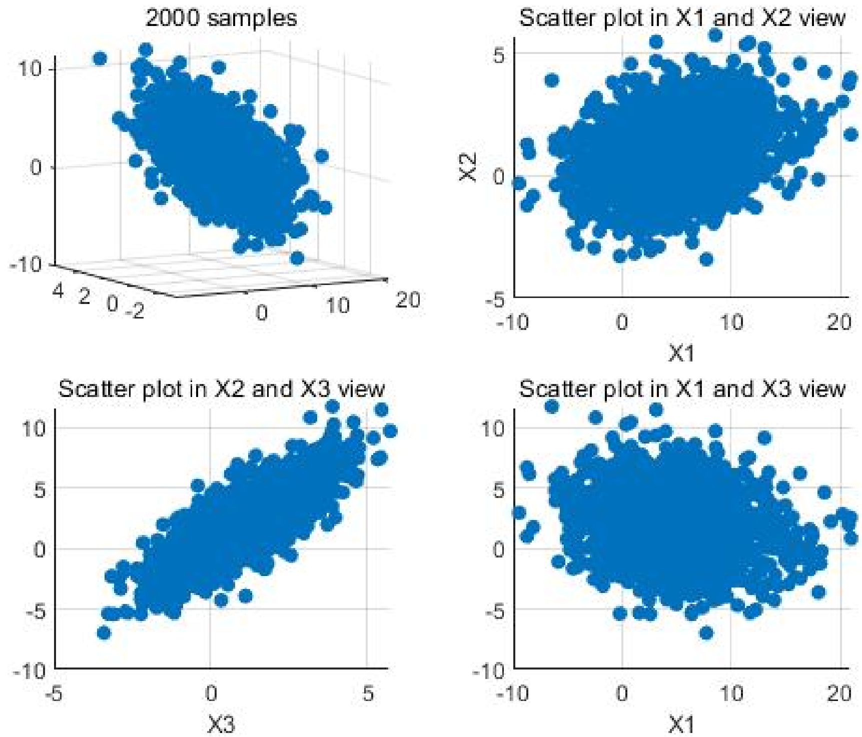

Figure 2: Scatter plot the 2000 samples.

(e) Since $\mathbf{X} \sim \mathbf{N}_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}' = \begin{bmatrix} 5 & 1 & 2 \end{bmatrix}$ and

$$\boldsymbol{\Sigma} = \begin{bmatrix} 20 & 2 & -2 \\ 2 & 2 & 3 \\ -2 & 3 & 7 \end{bmatrix}$$

Let $\mathbf{Y_1} = (X_1, X_3), \mathbf{Y_2} = X_2$, we have $\mathbf{Y} = (\mathbf{Y_1}, \mathbf{Y_2}) \sim N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu'} = [5, 2, 1]$ and

$$\boldsymbol{\Sigma} = \begin{bmatrix} 20 & -2 & 2 \\ -2 & 7 & 3 \\ 2 & 3 & 2 \end{bmatrix}$$

The conditional distribution of (X_1, X_3), given that $X_2 = x_2$ for the joint distribution follows,

$$N(\boldsymbol{\mu_1} + \Sigma_{12}\Sigma_{22}^{-1}(y_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$$

$$\Rightarrow N(\begin{pmatrix} 5 \\ 1 \end{pmatrix} + \begin{pmatrix} 2 \\ 3 \end{pmatrix} \cdot \frac{1}{2} \cdot (x_2 - 1), \begin{pmatrix} 20 & -2 \\ -2 & 4 \end{pmatrix} - \begin{pmatrix} 2 \\ 3 \end{pmatrix} \cdot \frac{1}{2} \cdot \begin{pmatrix} 2 & 3 \end{pmatrix})$$

$$\Rightarrow N(\begin{pmatrix} x_2+4 \\ \frac{3x_2-1}{2} \end{pmatrix}, \begin{pmatrix} 18 & -5 \\ -5 & -\frac{1}{2} \end{pmatrix}).$$

So, the conditional distribution of (X_1, X_3), given that $X_2 = x_2$ is,

$$\phi(\mathbf{x}) = f((X_1, X_3)|(X_2 = x_2)) = \frac{1}{(2\pi)^{2/2}|\mathbf{\Sigma}|^{1/2}} exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

where

$$\boldsymbol{\mu} = \begin{pmatrix} x_2+4 \\ 1.5x_2 - 0.5 \end{pmatrix}, \quad \text{and} \quad \mathbf{\Sigma} = \begin{pmatrix} 18 & -5 \\ -5 & -\frac{1}{2} \end{pmatrix}).$$

The conditional distribution of X_1, given that $X_2 = x_2$ for the joint distribution follows,

$$f(X_1|X_2 = x_2) = \int_{-\infty}^{\infty} \phi(\mathbf{x}) dx_3$$

(f) Based on conditional distribution above, we have

$$E[X_1|X_2 = 6] = x_2 + 4 = 10, \text{and } Var[X_1|X_2 = 6] = 18$$

The conditional distribution of X_3, given that $X_1 = x_1$ for the joint distribution follows,

$$N(\mu_3 + \frac{\sigma_{31}}{\sigma_{11}}(x_1 - \mu_1), \sigma_{33} - \frac{\sigma_{31}^2}{\sigma_{11}}) = N(2 + \frac{-2}{20}(x_1 - 5), 7 - \frac{4}{20}) = N(\frac{-x_1+25}{10}, 6.8).$$

So, $E[X_3|X_1 = 10] = 1.5$ and $Var[X_3|X_1 = 6.8]$.

(g) The joint p.d.f. of X_1, X_2 and X_3 is,

$$\phi(\mathbf{x}) = f(X_1, X_2, X_3) = \frac{1}{(2\pi)^{3/2}|\mathbf{\Sigma}|^{1/2}} exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

where

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \\ 2 \end{bmatrix}, \quad \text{and} \quad \mathbf{\Sigma} = \begin{bmatrix} 20 & 2 & -2 \\ 2 & 2 & 3 \\ -2 & 3 & 7 \end{bmatrix}.$$

$$|\mathbf{\Sigma}|^{1/2} = \sqrt{40}, \quad (\begin{bmatrix} 5 & 1 & 2 \end{bmatrix} - \begin{bmatrix} 5 & 1 & 2 \end{bmatrix})\mathbf{\Sigma}^{-1}(\begin{bmatrix} 5 \\ 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 5 \\ 1 \\ 2 \end{bmatrix}) = 0, (\begin{bmatrix} 0 & 0 & 0 \end{bmatrix} - \begin{bmatrix} 5 & 1 & 2 \end{bmatrix})\mathbf{\Sigma}^{-1}(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 5 \\ 1 \\ 2 \end{bmatrix}) = \frac{149}{40}$$

So,

$$f(5,1,2) = \frac{1}{(2\pi)^{3/2}\sqrt{40}} exp(0) \approx 0.01003, \text{and } f(0,0,0) = \frac{1}{(2\pi)^{3/2}\sqrt{40}} exp(\frac{149}{40}) \approx 0.00155$$