

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)  
May 2023

This paper is also taken for the relevant examination for the  
Associateship of the Royal College of Science

**Multivariate Analysis**

Date: 5 May 2023

Time: 10:00 – 11:30 (BST)

Time Allowed: 1.5 hrs

**This paper has 3 Questions.**

**Please Answer All Questions in 1 Answer Booklet**

Candidates should start their answers to each question on a new sheet of paper.

Supplementary books may only be used after the relevant main book(s) are full.

Any required additional material(s) will be provided.

Allow margins for marking.

Credit will be given for all questions attempted.

**DO NOT OPEN THIS PAPER UNTIL THE INVIGILATOR TELLS YOU TO**

**The open-book material allowed during the examinations consists of any  
material provided by the lecturers and annotated by the students, i.e.  
annotated lecture notes, annotated slides, and annotated problem class sheets.  
Books and electronic devices are not allowed.**

In this exam you may state without any proof any results you take from the lecture notes.

1. (a) Consider the random vector  $\mathbf{X} \equiv (X_1, X_2)^T$  where  $\mathbf{X}$  is  $N_m(\boldsymbol{\mu}_X, \Sigma_X)$  with  $\boldsymbol{\mu}_X = (2, -2)^T$  and  $\Sigma_X = \text{Cov}(\mathbf{X}) = \begin{pmatrix} 4 & -1 \\ -1 & 2 \end{pmatrix}$ . Let

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = \begin{pmatrix} X_1 + 1 \\ X_2 - 1 \\ X_1 + X_2 \end{pmatrix}.$$

- (i) Derive the mean and covariance matrix of  $\mathbf{Y}$ .
- (ii) What distribution does  $\mathbf{Y}$  follow? Justify your answer.
- (iii) What is the conditional distribution of  $(Y_1, Y_2)^T | Y_3 = 0$ ?
- (iv) Sketch two separate plots showing how  $\mathbf{X} \equiv (X_1, X_2)^T$  and  $(Y_1, Y_2)^T | Y_3 = 0$  will be distributed (one plot for each). You can show contours of equal density if applicable, and draw lines to show how the principal axes will be aligned. You are not required to find the precise angles and ratios of the axes orientation and lengths (or any interception points), but please identify the distributional mean in each plot and comment on the fundamental difference between the two plots.

- (b) Consider the following  $3 \times 3$  matrix:

$$A = \begin{pmatrix} a & 0 & 0 \\ 0 & b & d \\ 0 & d & c \end{pmatrix}$$

where  $\{a, b, c, d\}$  are real numbers. For what range of values of  $\{a, b, c, d\}$  is the matrix  $A$  a valid covariance matrix? Justify your answer.

- (c) Consider the following  $3 \times 3$  matrix:

$$B = \begin{pmatrix} 1 & a & a \\ a & 1 & a \\ a & a & 1 \end{pmatrix}$$

where  $a$  is a real number. For what range of values of  $a$  is the matrix  $B$  a valid covariance matrix? Justify your answer. Comment on whether your answer would change if  $B$  were just a  $2 \times 2$  matrix given by  $B = \begin{pmatrix} 1 & a \\ a & 1 \end{pmatrix}$ .

[Total 25 marks]

2. (a) Suppose  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N$  are independent  $N_m(\boldsymbol{\mu}, \Sigma)$  random vectors where  $\Sigma$  is positive definite and  $N > m$ . Ignoring the constant of proportionality, the likelihood function can be written as

$$L(\boldsymbol{\mu}, \Sigma) = (\det(\Sigma))^{-N/2} \exp\left(-\frac{1}{2} \Sigma^{-1} A\right) \exp\left(-\frac{1}{2} N(\bar{\mathbf{X}} - \boldsymbol{\mu})^T \Sigma^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu})\right)$$

where  $\bar{\mathbf{X}} = \frac{1}{N} \sum_{i=1}^N \mathbf{X}_i$ ,  $A = \sum_{i=1}^N (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T$ , and  $\text{etr}(A)$  denotes  $\exp(\text{tr}(A))$ .

- (i) Write down the maximum likelihood estimates of  $\boldsymbol{\mu}$  and  $\Sigma$  in terms of  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N$ , and write down the distributions of these estimates in terms of  $\boldsymbol{\mu}$ ,  $\Sigma$  and  $N$  only. What is the bias of the maximum likelihood estimate of  $\Sigma$ ?

- (ii) Show

$$\sup_{\boldsymbol{\mu} \in \mathbb{R}^m, \Sigma > 0} L(\boldsymbol{\mu}, \Sigma) = N^{mN/2} \exp(-mN/2) (\det(A))^{-N/2}.$$

- (iii) Consider partitioning the covariance matrix as  $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$ , where  $\Sigma_{11}$  is a  $k \times k$  matrix and  $\Sigma_{22}$  is a  $(m - k) \times (m - k)$  matrix,  $1 \leq k < m$ , and testing the null hypothesis

$$H : \Sigma = \begin{pmatrix} \lambda I_k & 0 \\ 0 & \Sigma_{22} \end{pmatrix}, \quad \lambda > 0 \text{ and } \Sigma_{22} > 0 \text{ unspecified}$$

against the alternative hypothesis  $K$  that says  $H$  is not true. The likelihood ratio statistic  $\Lambda$  for the null hypothesis  $H$  is defined as

$$\Lambda = \frac{\sup_{\boldsymbol{\mu} \in \mathbb{R}^m, \lambda > 0, \Sigma_{22} > 0} L\left(\boldsymbol{\mu}, \begin{pmatrix} \lambda I_k & 0 \\ 0 & \Sigma_{22} \end{pmatrix}\right)}{\sup_{\boldsymbol{\mu} \in \mathbb{R}^m, \Sigma > 0} L(\boldsymbol{\mu}, \Sigma)}.$$

Show

$$\Lambda^{2/N} = \frac{\det(A)}{(\frac{1}{k} \text{tr}(A_{11}))^k (\det(A_{22}))},$$

where  $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$ . [Hint:  $\sup_{\lambda > 0} \lambda^{-q} \exp(-\frac{a}{2\lambda}) = (\frac{a}{2q})^{-q} \exp(-q)$  for  $q, a > 0$ .]

- (b) The characteristic function  $\gamma : \mathbb{R}^{m \times m} \rightarrow \mathbb{C}$  of a  $m \times m$  symmetric random matrix  $A$  is defined as  $\gamma_A(\Theta) = E\{\exp(i \text{tr}(A^T \Theta))\}$  with argument  $\Theta$  itself symmetric. If  $A \sim W_m(n, \Sigma)$  then the characteristic function of  $A$  is  $\gamma_A(\Theta) = \det(I_m - 2i\Theta\Sigma)^{-n/2}$ .

- (i) Using the characteristic function prove that if  $A \sim W_m(n, \Sigma)$  and  $M$  is a  $q \times m$  matrix of rank  $q$ , then

$$MAM^T \sim W_q(n, M\Sigma M^T).$$

[Hint: Use the fact that for  $B$  ( $m \times n$ ),  $C$  ( $n \times m$ ) then  $\det(I_m + BC) = \det(I_n + CB)$ .]

- (ii) Making the partition

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \quad A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

with  $A_{11}$  and  $\Sigma_{11}$  both of size  $k \times k$  ( $k < m$ ), show  $A_{22} \sim W_{m-k}(n, \Sigma_{22})$ .

[Total 25 marks]

3. (a) Let  $\mathbf{X} = (X_1, X_2)^T$  be a random 2-dimensional vector with mean  $\boldsymbol{\mu} = (4, 2)^T$  and covariance

$$\Sigma = \begin{pmatrix} 5 & \sqrt{3} \\ \sqrt{3} & 3 \end{pmatrix}.$$

If principal component analysis is performed, what percentage of the total variance ( $\text{Var}(X_1) + \text{Var}(X_2)$ ) is captured by the first and second principal components respectively? What are the respective angles of the projections of the first and second principal components with respect to the  $x$ -axis when  $\mathbf{X}$  is plotted on the  $\{x, y\}$  plane?

(b) Let  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N$  be independent and identically distributed as  $N_m(\boldsymbol{\mu}, \Sigma)$ , with  $\Sigma$  positive definite. Let  $\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_N$  be the corresponding principal components. Find the distribution of  $\sum_{i=1}^N \mathbf{W}_i \mathbf{W}_i^T$ .

(c) Let  $\mathbf{X} = (X_1, \dots, X_m)^T$  be an  $m$ -dimensional random vector with mean  $\mathbf{0}$  and covariance matrix  $\Sigma = \{\sigma_{ij}\}$ , and let  $W_j$  be the  $j$ th principal component of  $\mathbf{X}$  with variance  $\lambda_j$  corresponding to the  $j$ th largest eigenvalue. Show the (ordinary) correlation  $\rho$  between  $X_i$  and  $W_j$ , is given by

$$\rho = \gamma_{ij} \sqrt{\frac{\lambda_j}{\sigma_{ii}}},$$

where  $\gamma_{ij}$  is the  $i$ th element of the corresponding eigenvector  $\boldsymbol{\gamma}_j$ .

[Hint: Set  $X_i = \mathbf{e}_i^T \mathbf{X}$ , where  $\mathbf{e}_i \in \mathbb{R}^m$  is the  $i$ th standard basis vector.]

(d) Consider a population comprised of two classes  $C_1$  and  $C_2$ , with a proportion  $\pi_1$  in  $C_1$  and  $\pi_2 = 1 - \pi_1$  in  $C_2$ . Objects from class  $C_i$  are multivariate normal with distribution  $N_m(\boldsymbol{\mu}_i, \Sigma_i)$ , and corresponding density function  $f_i(\mathbf{x})$ ,  $i = 1, 2$ . The region in which we assign new objects to  $C_1$  is defined by

$$\log f_1(\mathbf{x}) - \log f_2(\mathbf{x}) > \log \frac{\pi_2}{\pi_1}$$

and otherwise we assign to  $C_2$ .

- (i) In the case where  $\Sigma_1 = \Sigma_2 = \Sigma$  show the decision boundary is of the form  $\boldsymbol{\alpha}^T \mathbf{x} = c$ , giving expressions for  $\boldsymbol{\alpha} \in \mathbb{R}^m$  and  $c \in \mathbb{R}$  in terms of  $\Sigma, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \pi_1$  and  $\pi_2$ .
- (ii) Let  $\pi_1 = 3/4, \pi_2 = 1/4, \Sigma = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}, \boldsymbol{\mu}_1 = (4, 0)^T$  and  $\boldsymbol{\mu}_2 = (0, 4)^T$ . Find  $\boldsymbol{\alpha}$  and  $c$  and then on a single set of axes sketch contours of equal density for the distribution of classes  $C_1$  and  $C_2$  respectively, as well as the decision boundary. How would the decision boundary change if instead we had  $\pi_1 = 1/4, \pi_2 = 3/4$ ?
- (iii) Explain in a maximum of 2 sentences a method you could use if  $\Sigma_1 \neq \Sigma_2$  and what shape the decision boundary would take.

[Total 25 marks]



Module: MATH70092  
Setter: Sykulski  
Checker: Ernst  
Editor: Varty  
External: Woods  
Date: May 15, 2023

MSc EXAMINATIONS (STATISTICS)  
May 2023

MATH70092 Multivariate Analysis  
Time: 1 hour and 30 minutes

Setter's signature

.....

Checker's signature

.....

Editor's signature

.....

In this exam you may state without any proof any results you take from the lecture notes.

1. (a) Consider the random vector  $\mathbf{X} \equiv (X_1, X_2)^T$  where  $\mathbf{X}$  is  $N_m(\boldsymbol{\mu}_X, \Sigma_X)$  with  $\boldsymbol{\mu}_X = (2, -2)^T$  and  $\Sigma_X = \text{Cov}(\mathbf{X}) = \begin{pmatrix} 4 & -1 \\ -1 & 2 \end{pmatrix}$ . Let

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = \begin{pmatrix} X_1 + 1 \\ X_2 - 1 \\ X_1 + X_2 \end{pmatrix}.$$

- (i) Derive the mean and covariance matrix of  $\mathbf{Y}$ .
- (ii) What distribution does  $\mathbf{Y}$  follow? Justify your answer.
- (iii) What is the conditional distribution of  $(Y_1, Y_2)^T | Y_3 = 0$ ?
- (iv) Sketch two separate plots showing how  $\mathbf{X} \equiv (X_1, X_2)^T$  and  $(Y_1, Y_2)^T | Y_3 = 0$  will be distributed (one plot for each). You can show contours of equal density if applicable, and draw lines to show how the principal axes will be aligned. You are not required to find the precise angles and ratios of the axes orientation and lengths (or any interception points), but please identify the distributional mean in each plot and comment on the fundamental difference between the two plots.

**ANSWER: (Seen Similar)** (i) We can write  $\mathbf{Y} = B\mathbf{X} + \mathbf{b}$  where

$$B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}$$

and  $\mathbf{b} = (1, -1, 0)^T$ . So from p19 of notes we have  $\boldsymbol{\mu}_Y = B\boldsymbol{\mu}_X + \mathbf{b} = (3, -3, 0)^T$  and

$$\Sigma_Y = \text{Cov}(\mathbf{Y}) = B\text{Cov}(\mathbf{X})B^T = \begin{pmatrix} 4 & -1 & 3 \\ -1 & 2 & 1 \\ 3 & 1 & 4 \end{pmatrix}$$

(ii)  $\mathbf{Y}$  follows a multivariate normal specified by  $\boldsymbol{\mu}_Y$  and  $\Sigma_Y$  as any linear transformation of a multivariate normal distribution is itself multivariate normal (Theorem 4.4 of notes).

(iii) Partition  $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)^T$  and  $\Sigma_Y$  as

$$\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

[This question continues on the next page ...]

such that from part (i) we have  $\mu_1 = (3, -3)^T$ ,  $\mu_2 = 0$ ,  $\Sigma_{12} = (3, 1)^T$ ,  $\Sigma_{2,2} = 4$  and

$$\Sigma_{11} = \begin{pmatrix} 4 & -1 \\ -1 & 2 \end{pmatrix}$$

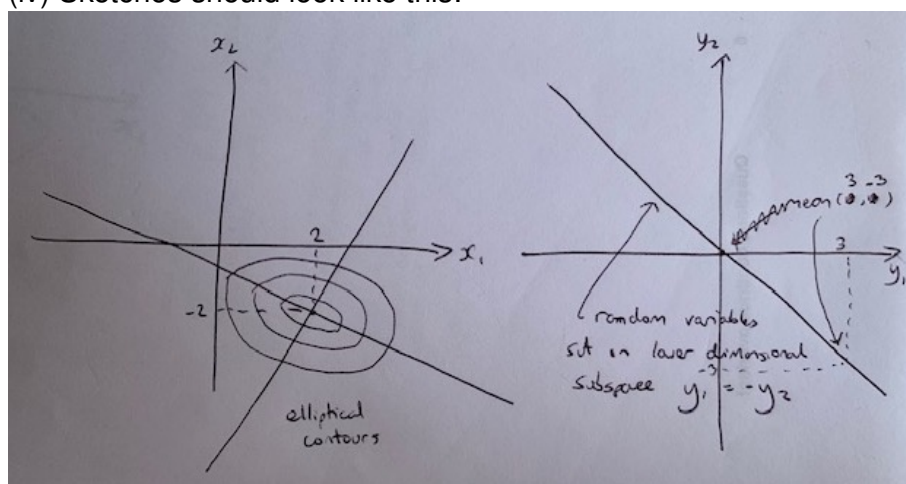
Let  $\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ , then the conditional distribution of  $(Y_1, Y_2)^T$  given  $Y_3$  is (p32 of notes, Theorem 4.8)

$$N_2(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(Y_3 - \mu_2), \Sigma_{11.2}),$$

where  $Y_3 = 0$  and  $\Sigma_{11.2} = \begin{pmatrix} 4 & -1 \\ -1 & 2 \end{pmatrix} - (3, 1)^T \frac{1}{4} (3, 1) = \begin{pmatrix} 4 & -1 \\ -1 & 2 \end{pmatrix} - \frac{1}{4} \begin{pmatrix} 9 & 3 \\ 3 & 1 \end{pmatrix} = \frac{7}{4} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$

Therefore the conditional distribution of  $(Y_1, Y_2)^T | Y_3 = 0$  is  $N\left(\begin{pmatrix} 3 \\ -3 \end{pmatrix}, \frac{7}{4} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\right).$

(iv) Sketches should look like this:



The sketch for  $(X_1, X_2)^T$  should clearly show elliptical contours that stretch more in the  $X_1$  direction (as this has bigger variance), the mean must show as  $(2, -2)^T$ . The sketch for  $(Y_1, Y_2)^T | Y_3 = 0$  should show the data sitting on a lower dimensional subspace (the line  $Y_1 = Y_2$ ) with the mean going through  $(3, -3)^T$ . The fundamental difference is that in the right plot the observations will sit on a line, whereas in the left plot the observations will be elliptically distributed in  $\mathbb{R}^2$ . [13 marks]

(b) Consider the following  $3 \times 3$  matrix:

$$A = \begin{pmatrix} a & 0 & 0 \\ 0 & b & d \\ 0 & d & c \end{pmatrix}$$

where  $\{a, b, c, d\}$  are real numbers. For what range of values of  $\{a, b, c, d\}$  is the matrix  $A$  a valid covariance matrix? Justify your answer.

[This question continues on the next page ...]

**ANSWER: (Seen Similar)** Require  $\{a, b, c\} \geq 0$  as variances must be non-negative. For  $d$ , need to show  $A$  is non-negative definite: i.e. for all  $(x_1, x_2, x_3) \in \mathbb{R}^3$  we require

$$\begin{aligned} (x_1, x_2, x_3)A(x_1, x_2, x_3)^T &\geq 0 \\ \Rightarrow ax_1^2 + bx_2^2 + cx_3^2 + 2dx_2x_3 &\geq 0 \\ \Rightarrow ax_1^2 + \left(x_2 + \frac{b}{d}x_3\right)^2 + \left(\frac{c}{b} - \frac{d^2}{b^2}\right)x_3^2 &\geq 0 \quad (b > 0) \end{aligned}$$

which is true only when  $\frac{c}{b} - \frac{d^2}{b^2} \geq 0$ , i.e.  $d^2 \leq bc$ . When  $b = 0$ , clearly we require  $d = 0$  for non-negative definiteness. Therefore final answer is  $\{a, b, c\} \geq 0$  and  $d^2 \leq bc$ . Also accept justification by showing eigenvalues are non-negative (via Lemma 2.6) for this same range of values, or by similar valid reasoning. [5 marks]

(c) Consider the following  $3 \times 3$  matrix:

$$B = \begin{pmatrix} 1 & a & a \\ a & 1 & a \\ a & a & 1 \end{pmatrix}$$

where  $a$  is a real number. For what range of values of  $a$  is the matrix  $B$  a valid covariance matrix? Justify your answer. Comment on whether your answer would change if  $B$  were just a  $2 \times 2$  matrix given by  $B = \begin{pmatrix} 1 & a \\ a & 1 \end{pmatrix}$ .

**ANSWER: (Unseen)** Need to show  $B$  is non-negative definite. This is best shown by finding range of values of  $a$  for which eigenvalues are non-negative. Eigenvalues must satisfy:

$$(1 - \lambda)^3 - 3a^2(1 - \lambda) + 2a^3 = 0.$$

Replacing  $x = 1 - \lambda$  this becomes:

$$x^3 - 3a^2x + 2a^3 = 0,$$

and we can see by inspection that  $x = a$  and  $x = -2a$  are solutions and the polynomial therefore factorises to:

$$(x - a)^2(x + 2a) = 0.$$

Therefore the eigenvalues are  $1 - a$  (twice) and  $1 + 2a$ . As all eigenvalues must be non-negative, the valid range of  $a$  values is therefore  $-0.5 \leq a \leq 1$ . If  $B$  were just a  $2 \times 2$  matrix then from part (a) we can see that  $-1 \leq a \leq 1$ , and therefore the answer does change. [7 marks]

[Total 25 marks]

2. (a) Suppose  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N$  are independent  $N_m(\boldsymbol{\mu}, \Sigma)$  random vectors where  $\Sigma$  is positive definite and  $N > m$ . Ignoring the constant of proportionality, the likelihood function can be written as

$$L(\boldsymbol{\mu}, \Sigma) = (\det(\Sigma))^{-N/2} \exp\left(-\frac{1}{2} \text{tr}(\Sigma^{-1} A)\right) \exp\left(-\frac{1}{2} N(\bar{\mathbf{X}} - \boldsymbol{\mu})^T \Sigma^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu})\right)$$

where  $\bar{\mathbf{X}} = \frac{1}{N} \sum_{i=1}^N \mathbf{X}_i$ ,  $A = \sum_{i=1}^N (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T$ , and  $\text{etr}(A)$  denotes  $\exp(\text{tr}(A))$ .

- (i) Write down the maximum likelihood estimates of  $\boldsymbol{\mu}$  and  $\Sigma$  in terms of  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N$ , and write down the distributions of these estimates in terms of  $\boldsymbol{\mu}$ ,  $\Sigma$  and  $N$  only. What is the bias of the maximum likelihood estimate of  $\Sigma$ ?
- (ii) Show

$$\sup_{\boldsymbol{\mu} \in \mathbb{R}^m, \Sigma > 0} L(\boldsymbol{\mu}, \Sigma) = N^{mN/2} \exp(-mN/2) (\det(A))^{-N/2}.$$

- (iii) Consider partitioning the covariance matrix as  $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$ , where  $\Sigma_{11}$  is a  $k \times k$  matrix and  $\Sigma_{22}$  is a  $(m-k) \times (m-k)$  matrix,  $1 \leq k < m$ , and testing the null hypothesis

$$H : \Sigma = \begin{pmatrix} \lambda I_k & 0 \\ 0 & \Sigma_{22} \end{pmatrix}, \quad \lambda > 0 \text{ and } \Sigma_{22} > 0 \text{ unspecified}$$

against the alternative hypothesis  $K$  that says  $H$  is not true. The likelihood ratio statistic  $\Lambda$  for the null hypothesis  $H$  is defined as

$$\Lambda = \frac{\sup_{\boldsymbol{\mu} \in \mathbb{R}^m, \lambda > 0, \Sigma_{22} > 0} L\left(\boldsymbol{\mu}, \begin{pmatrix} \lambda I_k & 0 \\ 0 & \Sigma_{22} \end{pmatrix}\right)}{\sup_{\boldsymbol{\mu} \in \mathbb{R}^m, \Sigma > 0} L(\boldsymbol{\mu}, \Sigma)}.$$

Show

$$\Lambda^{2/N} = \frac{\det(A)}{(\frac{1}{k} \text{tr}(A_{11}))^k (\det(A_{22}))},$$

where  $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$ . [Hint:  $\sup_{\lambda > 0} \lambda^{-q} \exp(-\frac{a}{2\lambda}) = (\frac{a}{2q})^{-q} \exp(-q)$  for  $q, a > 0$ .]

**ANSWER: (Seen or seen similar)** (i) The MLE of the mean is

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{X}} = \frac{1}{N} \sum_{i=1}^N \mathbf{X}_i \sim N_m\left(\boldsymbol{\mu}, \frac{1}{N} \Sigma\right)$$

The MLE of the variance is

$$\hat{\Sigma} = \frac{1}{N} A = \frac{1}{N} \sum_{i=1}^N (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T \sim W_m\left(N-1, \frac{1}{N} \Sigma\right)$$

[This question continues on the next page ...]

Which can both be written down by combining Theorems 5.4 and 5.6 in the notes (or otherwise). The bias of  $\hat{\Sigma}$  is therefore  $\Sigma - E(\hat{\Sigma}) = \Sigma - \frac{N-1}{N}\Sigma = \frac{1}{N}\Sigma$ .

(ii)  $\bar{\mathbf{X}}$  and  $\frac{1}{N}A$  are the ML estimators therefore

$$\begin{aligned}\sup_{\mu \in \mathbb{R}^m, \Sigma > 0} L(\mu, \Sigma) &= L\left(\bar{\mathbf{X}}, \frac{A}{N}\right) \\ &= (\det(N^{-1}A))^{-N/2} \text{etr}\left(-\frac{N}{2}A^{-1}A\right) \\ &= (N^{-m} \det(A))^{-N/2} \text{etr}\left(-\frac{N}{2}I\right) \\ &= N^{mN/2} \exp(-mN/2) (\det(A))^{-N/2}.\end{aligned}$$

(iii) The denominator is

$$\sup_{\mu \in \mathbb{R}^m, \Sigma > 0} L(\mu, \Sigma) = N^{mN/2} \exp(-mN/2) (\det(A))^{-N/2}$$

from (ii). The numerator is

$$\begin{aligned}\sup_{\mu \in \mathbb{R}^m, \lambda > 0, \Sigma_{22} > 0} L\left(\mu, \begin{pmatrix} \lambda I_k & 0 \\ 0 & \Sigma_{22} \end{pmatrix}\right) &= \sup_{\lambda > 0, \Sigma_{22} > 0} \left(\det \begin{pmatrix} \lambda I_k & 0 \\ 0 & \Sigma_{22} \end{pmatrix}\right)^{-N/2} \text{etr}\left(-\frac{1}{2} \begin{pmatrix} \lambda I_k & 0 \\ 0 & \Sigma_{22} \end{pmatrix}^{-1} A\right) \\ &= \sup_{\lambda > 0, \Sigma_{22} > 0} (\det(\lambda I_k))^{-N/2} (\det(\Sigma_{22}))^{-N/2} \text{etr}\left(-\frac{1}{2\lambda} A_{11}\right) \text{etr}\left(-\frac{1}{2} \Sigma_{22}^{-1} A_{22}\right) \\ &= \sup_{\lambda > 0, \Sigma_{22} > 0} \lambda^{-kN/2} \text{etr}\left(-\frac{1}{2\lambda} A_{11}\right) (\det(\Sigma_{22}))^{-N/2} \text{etr}\left(-\frac{1}{2} \Sigma_{22}^{-1} A_{22}\right) \\ &= \left(\frac{1}{kN} \text{tr} A_{11}\right)^{-kN/2} e^{-kN/2} N^{(m-k)N/2} e^{-(m-k)N/2} (\det(A_{22}))^{-N/2}\end{aligned}$$

from part (ii) and hint. Therefore

$$\Lambda = \left(\frac{1}{k} \text{tr} A_{11}\right)^{-kN/2} (\det(A_{22}))^{-N/2} (\det(A))^{N/2}$$

and

$$\Lambda^{2/N} = \frac{\det(A)}{\left(\frac{1}{k} \text{tr} A_{11}\right)^k \det(A_{22})}.$$

[18 marks]

(b) The characteristic function  $\gamma : \mathbb{R}^{m \times m} \rightarrow \mathbb{C}$  of a  $m \times m$  symmetric random matrix  $A$  is defined as  $\gamma_A(\Theta) = E\{\exp(\text{itr}(A^T \Theta))\}$  with argument  $\Theta$  itself symmetric. If  $A \sim W_m(n, \Sigma)$  then the characteristic function of  $A$  is  $\gamma_A(\Theta) = \det(I_m - 2i\Theta\Sigma)^{-n/2}$ .

(i) Using the characteristic function prove that if  $A \sim W_m(n, \Sigma)$  and  $M$  is a  $q \times m$  matrix of rank  $q$ , then

$$MAM^T \sim W_q(n, M\Sigma M^T).$$

[Hint: Use the fact that for  $B$  ( $m \times n$ ),  $C$  ( $n \times m$ ) then  $\det(I_m + BC) = \det(I_n + CB)$ .]

[This question continues on the next page ...]

(ii) Making the partition

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \quad A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

with  $A_{11}$  and  $\Sigma_{11}$  both of size  $k \times k$  ( $k < m$ ), show  $A_{22} \sim W_{m-k}(n, \Sigma_{22})$ .

**ANSWER: (Seen or seen similar)** (i) From the definition, the characteristic function of  $MAM^T$  is

$$\begin{aligned} E\{\exp(\text{itr}(MAM^T\Theta))\} &= E\{\exp(\text{itr}(AM^T\Theta M))\} \\ &= \gamma_A(M^T\Theta M) \\ &= \det(I_m - 2iM^T\Theta M\Sigma)^{-n/2} \\ &= \det(I_q - 2i\Theta M\Sigma M^T)^{-n/2} \end{aligned}$$

and hence  $MAM^T \sim W_q(n, M\Sigma M^T)$ .

(ii) Let  $M = [0; I_{m-k}]$  where 0 represents the  $(m-k) \times k$  zero matrix. The result follows from (i) with  $q = m - k$ . [7 marks]

[Total 25 marks]

3. (a) Let  $\mathbf{X} = (X_1, X_2)^T$  be a random 2-dimensional vector with mean  $\boldsymbol{\mu} = (4, 2)^T$  and covariance

$$\Sigma = \begin{pmatrix} 5 & \sqrt{3} \\ \sqrt{3} & 3 \end{pmatrix}.$$

If principal component analysis is performed, what percentage of the total variance ( $\text{Var}(X_1) + \text{Var}(X_2)$ ) is captured by the first and second principal components respectively? What are the respective angles of the projections of the first and second principal components with respect to the  $x$ -axis when  $\mathbf{X}$  is plotted on the  $\{x, y\}$  plane?

**ANSWER: (Unseen)** The eigenvalues are found by solving  $(5 - \lambda)(3 - \lambda) - 3 = 0$  which yields  $(\lambda - 6)(\lambda - 2) = 0$  which gives  $\lambda = 6$  or  $2$ . The largest eigenvalue of  $6$  corresponds to the first principal component, and this therefore captures  $6/8 = 75\%$  of the total variance. It immediately follows that the second principal component captures the remaining  $25\%$  of the variance, as total variance is preserved by principal component analysis.

The angle of the projection of the first principal component is found by first finding the eigenvector  $(v_1, v_2)^T$  corresponding to  $\lambda = 6$  which corresponds to  $-v_1 + \sqrt{3}v_2 = 0$  which after normalising gives the eigenvector  $(\sqrt{3}/2, 1/2)^T$ . The angle to the  $x$ -axis is therefore  $\cos^{-1}(\sqrt{3}/2) = \pi/6$ . It immediately follows that the projection of the second principal component has angle of  $\pi/6 + \pi/2 = 2\pi/3$  due to the orthogonality of the two components. [6 marks]

- (b) Let  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N$  be independent and identically distributed as  $N_m(\boldsymbol{\mu}, \Sigma)$ , with  $\Sigma$  positive definite. Let  $\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_N$  be the corresponding principal components. Find the distribution of  $\sum_{i=1}^N \mathbf{W}_i \mathbf{W}_i^T$ .

**ANSWER: (Unseen)**  $\mathbf{W}_i = \Gamma^T(\mathbf{X}_i - \boldsymbol{\mu})$ , therefore because  $\text{Cov}(\mathbf{W}_i) = \Gamma^T \Sigma \Gamma = D$ ,  $\mathbf{W}_i \sim N_m(\mathbf{0}, D)$  where  $D$  is a diagonal matrix of descending eigenvalues. It follows that  $\sum_{i=1}^N \mathbf{W}_i \mathbf{W}_i^T \sim W_m(N, D)$ . [3 marks]

- (c) Let  $\mathbf{X} = (X_1, \dots, X_m)^T$  be an  $m$ -dimensional random vector with mean  $\mathbf{0}$  and covariance matrix  $\Sigma = \{\sigma_{ij}\}$ , and let  $W_j$  be the  $j$ th principal component of  $\mathbf{X}$  with variance  $\lambda_j$  corresponding to the  $j$ th largest eigenvalue. Show the (ordinary) correlation  $\rho$  between  $X_i$  and  $W_j$ , is given by

$$\rho = \gamma_{ij} \sqrt{\frac{\lambda_j}{\sigma_{ii}}},$$

where  $\gamma_{ij}$  is the  $i$ th element of the corresponding eigenvector  $\boldsymbol{\gamma}_j$ .

[Hint: Set  $X_i = \mathbf{e}_i^T \mathbf{X}$ , where  $\mathbf{e}_i \in \mathbb{R}^m$  is the  $i$ th standard basis vector.]

[This question continues on the next page ...]

**ANSWER: (Unseen)**

$$\begin{aligned}
\text{Cov}(X_i, W_j) &= \text{Cov}(\mathbf{e}_i^T \mathbf{X}, \boldsymbol{\gamma}_j^T \mathbf{X}) \\
&= E\{(\mathbf{e}_i^T \mathbf{X})(\boldsymbol{\gamma}_j^T \mathbf{X})\} \quad (\text{as } E(\mathbf{X}) = \mathbf{0}) \\
&= E\{\mathbf{e}_i^T \mathbf{X} \mathbf{X}^T \boldsymbol{\gamma}_j\} \\
&= \mathbf{e}_i^T \boldsymbol{\Sigma} \boldsymbol{\gamma}_j \\
&= \lambda_j \mathbf{e}_i^T \boldsymbol{\gamma}_j \\
&= \lambda_j \gamma_{ij}.
\end{aligned}$$

The correlation follows as  $\rho = \frac{\lambda_j \gamma_{ij}}{\sqrt{\sigma_{ii} \lambda_j}} = \gamma_{ij} \sqrt{\frac{\lambda_j}{\sigma_{ii}}}$ . [4 marks]

(d) Consider a population comprised of two classes  $C_1$  and  $C_2$ , with a proportion  $\pi_1$  in  $C_1$  and  $\pi_2 = 1 - \pi_1$  in  $C_2$ . Objects from class  $C_i$  are multivariate normal with distribution  $N_m(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ , and corresponding density function  $f_i(\mathbf{x})$ ,  $i = 1, 2$ . The region in which we assign new objects to  $C_1$  is defined by

$$\log f_1(\mathbf{x}) - \log f_2(\mathbf{x}) > \log \frac{\pi_2}{\pi_1}$$

and otherwise we assign to  $C_2$ .

- (i) In the case where  $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$  show the decision boundary is of the form  $\boldsymbol{\alpha}^T \mathbf{x} = c$ , giving expressions for  $\boldsymbol{\alpha} \in \mathbb{R}^m$  and  $c \in \mathbb{R}$  in terms of  $\boldsymbol{\Sigma}, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \pi_1$  and  $\pi_2$ .
- (ii) Let  $\pi_1 = 3/4, \pi_2 = 1/4, \boldsymbol{\Sigma} = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}$ ,  $\boldsymbol{\mu}_1 = (4, 0)^T$  and  $\boldsymbol{\mu}_2 = (0, 4)^T$ . Find  $\boldsymbol{\alpha}$  and  $c$  and then on a single set of axes sketch contours of equal density for the distribution of classes  $C_1$  and  $C_2$  respectively, as well as the decision boundary. How would the decision boundary change if instead we had  $\pi_1 = 1/4, \pi_2 = 3/4$ ?
- (iii) Explain in a maximum of 2 sentences a method you could use if  $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$  and what shape the decision boundary would take.

**ANSWER: (Seen or seen similar)** (i) The boundary is defined by

$$\begin{aligned}
\log f_1(\mathbf{x}) - \log f_2(\mathbf{x}) &= -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_1) + \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_2) \\
&= (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1} \mathbf{x} - \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2).
\end{aligned}$$

Let  $\boldsymbol{\alpha} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ , then the decision boundary is defined by

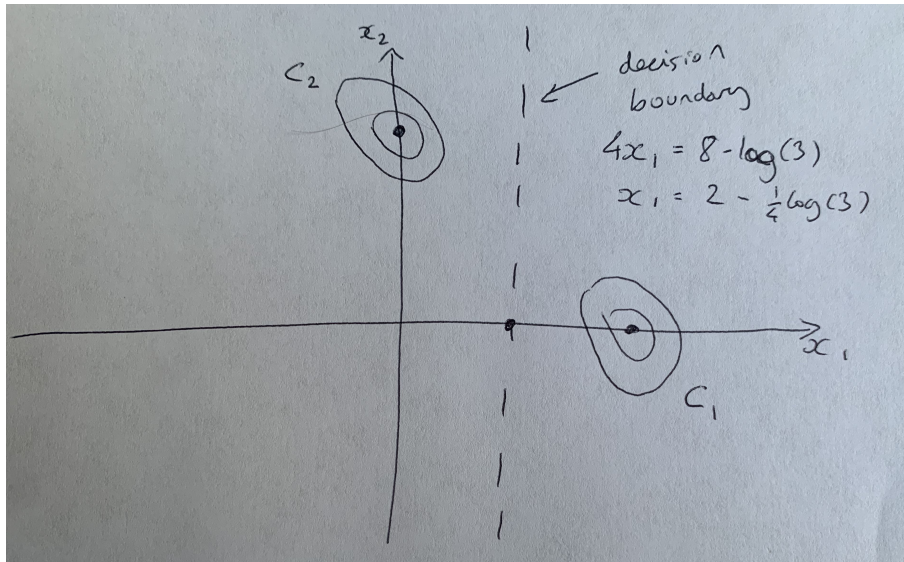
$$\begin{aligned}
\boldsymbol{\alpha}^T \left( \mathbf{x} - \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) \right) &= \log \frac{\pi_2}{\pi_1} \\
\boldsymbol{\alpha}^T \mathbf{x} &= c
\end{aligned}$$

[This question continues on the next page ...]

where  $c = \log \frac{\pi_2}{\pi_1} + \frac{1}{2} \alpha^T (\mu_1 + \mu_2)$ .

(ii)  $\alpha = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} (4, -4)^T = (4, 0)^T$  and  $c = \log(1/3) + 0.5 * (4, 0)(4, -4)^T = 8 - \log(3)$ .

Plot should therefore look like:



If  $\pi_1 = 1/4, \pi_2 = 3/4$  then the decision boundary shifts to the right and is  $x_1 = 2 + \log(3)/4$ .

(iii) Quadratic discriminant analysis could be used and the decision boundary would then become quadratic. [12 marks]

[Total 25 marks]