MAS465 Multivariate Data Analysis
MAS6011(1) Dependent Data

1. Some background mathematics

School of Mathematics and Statistics

Dr Frazer Jarvis (J12)
a.f.jarvis@sheffield.ac.uk

Plan

- Some background mathematics. . .

Some background mathematics. . .

The lecture notes contain (in Chapter 1) background material on:

- Basic properties of eigenvectors and eigenvalues (§1.3)
- Differentiating with respect to vectors (§1.4)
- Constrained optimisation (§1.5)
- The multivariate normal distribution (§1.6)

All these are, to some extent, part of the Sheffield undergraduate degree, and I don't want to spend too long on these topics. The notes are provided for reference, so you should read them now if you aren't confident about the ideas.

Some background mathematics. . .

However, I will try to remind you about them when we get to the relevant parts, and to introduce these ideas reasonably slowly!

We will, however, now look at §1.1 and §1.2, which largely concerns notation.

You don't need to know any of the details/proofs for the exam, but the ideas do underpin all the rest of the course!

Some background mathematics. . .

Basic Notation

- Observations are column vectors $x = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}$, so each observation consists of p numbers.
    - We say that the data is p-dimensional
- We denote the operation of transposing a matrix with a dash $'$, so $x' = (x_1 \; \dots \; x_p)$ is a row vector.
- If you aren't confident about using matrices, you will need to get used to them – they occur a lot in the module! There are some background notes available (on MOLE).

Some background mathematics...

- Usually, we regard the number of observations as n, and each observation records p values.
- The $n \times p$ data matrix X' is

$$X' = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} = \begin{pmatrix} x_1' \\ x_2' \\ \vdots \\ x_n' \end{pmatrix}.$$

  - So *rows* of X' are the observations, while the *columns* correspond to individual variables.
  - Some authors call the $p \times n$ matrix X the data matrix; some authors use X where we use X'; the choice above corresponds to the notation in R (so be aware when you read books/articles that conventions differ!).

Some background mathematics. . .

- The *sample mean* vector $\overline{x}$ is defined by

$$\overline{x}' = (\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_p) = \frac{1}{n}\mathbf{1}'X',$$

  where $\mathbf{1}$ is the $n \times 1$-column vector with n 1s.
  - Notice that the sample mean vector is a column p-vector, i.e., a $p \times 1$-matrix.
  - Using properties of matrix transposition ($(AB)' = B'A'$), we have

$$\overline{x} = \frac{1}{n}X\mathbf{1}.$$

    (But we will tend to think of the sample mean vector as a row vector, $\overline{x}'$.)

Some background mathematics...

- The *sample variance* $S = \mathrm{var}(X')$ is given by

$$S = \frac{1}{n-1}(X - \overline{X})(X - \overline{X})',$$

  where $\overline{X} = (\overline{x}, \overline{x}, \ldots, \overline{x})$ is the $p \times n$ matrix with all columns equal to $\overline{x}$.
    - Notice that the sample variance matrix is a $p \times p$ matrix.

Some background mathematics. . .

Let's compute S_{ij}:

$$\begin{aligned} S_{ij} &= \frac{1}{n-1}\sum_{k=1}^{n}(X-\overline{X})_{ik}(X-\overline{X})'_{kj} \\ &= \frac{1}{n-1}\sum_{k=1}^{n}(X-\overline{X})_{ik}(X-\overline{X})_{jk} \\ &= \frac{1}{n-1}\sum_{k=1}^{n}(x_{ik}-\overline{x}_i)(x_{jk}-\overline{x}_j) \end{aligned}$$

If $i = j$, this is just

$$\mathrm{var}(x_i) = \frac{1}{n-1}\sum_{k=1}^{n}(x_{ik}-\overline{x}_i)^2,$$

the *variance* of the ith variable.

Similarly, if $i \neq j$, we get the *covariance* of the ith and jth variables.

Some background mathematics. . .

So the variance matrix encodes all the variances and covariances of the individual variables:

- to read off the variance of variable i, look at the ith diagonal entry s_{ii}
- to read off the covariance of variables i and j, look at the ijth (or jith) entry s_{ij}

Some background mathematics. . .

Note that if w is any p-vector, then

$$\mathrm{var}(X'w) = w'\mathrm{var}(X)w = w'Sw$$

gives the variance of a *linear combination* of the variables.

(Note that it is a $(1 \times p) \times (p \times p) \times (p \times 1) = 1 \times 1$ matrix, so a scalar.)

Similarly, if A is any $p \times q$ matrix, then

$$\mathrm{var}(X'A) = A'\mathrm{var}(X')A = A'SA,$$

a $q \times q$ matrix.

Some background mathematics...

In R, type

- `var(.)` to obtain the variance matrix
- `apply(.,2,mean)` to get the means of the individual variables (not `mean(.)`)

Some background mathematics. . .

The variance matrix is:

- symmetric
- real
- positive definite (as long as none of the variables is a linear combination of any of the others)
  - The variables may be linearly related, e.g.:
    - one column could be summarising others (as an average, perhaps)
    - one column could be the difference of others
    - we might have $n < p$
- an unbiased estimator of the population variance (and $\overline{x}$ is an unbiased estimator of the population mean)

Some background mathematics. . .

Just as in the univariate case, one has a notion of correlation.

The *sample correlation matrix* is the matrix R whose entries are normalised versions of the variance matrix:

$$r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}s_{jj}}}.$$

Then $r_{ii} = 1$ for all i, and $-1 \leq r_{ij} \leq 1$ for all $i \neq j$.

This coefficient is the usual correlation between the variables i and j.

If L denotes the diagonal matrix whose entries are $s_{11}, \ldots, s_{pp}$, and $L^{1/2}$ is the diagonal matrix with entries $\sqrt{s_{11}}, \ldots, \sqrt{s_{pp}}$, then

$$S = L^{1/2} R L^{1/2}.$$

R is also real and symmetric, and is positive definite when S is.

Some background mathematics. . .

Recall the following facts about real symmetric $p \times p$ matrices A (see §1.3):

- They have p linearly independent eigenvectors
- The eigenvalues are all real
- If λ and μ are two distinct eigenvalues, then the corresponding eigenvectors are orthogonal
- If λ is an eigenvalue of A with multiplicity k (i.e., if $(t - \lambda)^k$ divides the characteristic equation), then we can pick k orthogonal eigenvectors of A with eigenvalue λ.

If also the matrix is positive definite, then

- Every eigenvalue is positive

Some background mathematics. . .

As a corollary:

- If A is a real symmetric matrix, then there is a $p \times p$ matrix X whose columns are normalised (i.e., scaled to have length 1) eigenvectors of A, which is an orthogonal matrix (i.e., $XX' = X'X = I_p$)

Some background mathematics...

Suppose we have two multivariate observations, x and y, whose entries are numeric.

We can measure how "far apart" they are in various ways (e.g., Euclidean distance). However:

- If one variable has a large variance, it shouldn't be so surprising that x and y should differ a lot in that variable
- If two variables are strongly correlated, and x and y differ a lot in one of these variables, it shouldn't be surprising that they differ a lot in the other variable too.

A much better statistical measure of how far x and y are apart ought to take account of this; such a measure is given by the *Mahalanobis distance*, given by

$$\sqrt{(x-y)'S^{-1}(x-y)}.$$