Part A

Matrix Algebra Refresher

1 Addition

If two matrices have the same dimension (both are r × c), then matrix addition and subtraction simply follows by adding (or subtracting) on an element by element basis.
Matrix addition: $(A+B)_{ij} = A_{ij} + B_{ij}$
Matrix subtraction: $(A-B)_{ij} = A_{ij} - B_{ij}$

Examples:
$\boldsymbol{A} = \begin{pmatrix} 3 & 1 \\ 2 & 5 \end{pmatrix}$ and $\boldsymbol{B} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$, then $\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{B} = \begin{pmatrix} 4 & 3 \\ 2 & 6 \end{pmatrix}$ and $\boldsymbol{D} = \boldsymbol{A} - \boldsymbol{B} = \begin{pmatrix} 2 & -1 \\ 2 & 4 \end{pmatrix}$

```
> A = matrix(c(3, 1, 2, 5), nrow = 2, ncol = 2, byrow = TRUE)
> B = matrix(c(1, 2, 0, 1), nrow = 2, ncol = 2, byrow = TRUE)
> C = A + B
> C
```

```
     [,1] [,2]
[1,]    4    3
[2,]    2    6
```

```
> D = A - B
> D
```

```
     [,1] [,2]
[1,]    2   -1
[2,]    2    4
```

1.1 Partitioned Matrices

It will often prove useful to divide (or partition) the elements of a matrix into a matrix whose elements are itself matrices.

$$\boldsymbol{C} = \begin{pmatrix} 3 & 1 & 2 \\ 2 & 5 & 4 \\ 1 & 1 & 2 \end{pmatrix} = \begin{pmatrix} 3 & \vdots & 1 & 2 \\ \cdots & \cdots & \cdots & \cdots \\ 2 & \vdots & 5 & 4 \\ 1 & \vdots & 1 & 2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{a} & \boldsymbol{b} \\ \boldsymbol{d} & \boldsymbol{B} \end{pmatrix}$$

$$\boldsymbol{a} = \begin{pmatrix} 3 \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} 1 & 2 \end{pmatrix}, \boldsymbol{d} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \boldsymbol{B} = \begin{pmatrix} 5 & 4 \\ 1 & 2 \end{pmatrix}$$

One useful partition is to write the matrix as either (1) a column vector of row vectors, or (2) a row vector of column vectors:

$$(1)\ \boldsymbol{C} = \begin{pmatrix} 3 & 1 & 2 \\ 2 & 5 & 4 \\ 1 & 1 & 2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{r}_1 \\ \boldsymbol{r}_2 \\ \boldsymbol{r}_3 \end{pmatrix}, \text{ where } \boldsymbol{r}_1 = \begin{pmatrix} 3 & 1 & 2 \end{pmatrix}, \boldsymbol{r}_2 = \begin{pmatrix} 2 & 5 & 4 \end{pmatrix}, \boldsymbol{r}_3 = \begin{pmatrix} 1 & 1 & 2 \end{pmatrix}$$

(2) $\boldsymbol{C} = \begin{pmatrix} 3 & 1 & 2 \\ 2 & 5 & 4 \\ 1 & 1 & 2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{c}_1 & \boldsymbol{c}_2 & \boldsymbol{c}_3 \end{pmatrix}$, where $\boldsymbol{c}_1 = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}, \boldsymbol{c}_2 = \begin{pmatrix} 1 \\ 5 \\ 1 \end{pmatrix}, \boldsymbol{c}_3 = \begin{pmatrix} 2 \\ 4 \\ 2 \end{pmatrix}$

```
> C = matrix(c(3, 1, 2, 2, 5, 4, 1, 1, 2), nrow = 3, ncol = 3, byrow = TRUE)
> a = C[1, 1]
> b = C[1, 2:3]
> d = C[2:3, 1] # not as expected
> d = C[2:3, 1, drop = FALSE] # to keep the same "object/structure" as C
> B = C[2:3, 2:3]
```

2 Multiplication

2.1 Dot product

The dot (or inner) product of two vectors (both of length n) is defined as follows:

$$\boldsymbol{a} \cdot \boldsymbol{b} = \sum_{i=1}^{n} a_i b_i$$

Example:

$\boldsymbol{a} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$ and $\boldsymbol{b} = \begin{pmatrix} 4 & 5 & 7 & 9 \end{pmatrix}$, then $\boldsymbol{a} \cdot \boldsymbol{b} = 1 \times 4 + 2 \times 5 + 3 \times 7 + 4 \times 9 = 71$

```
> a = matrix(c(1, 2, 3, 4), ncol = 1)
> b = c(4, 5, 7, 9)
> a*b
```

```
     [,1]
[1,]    4
[2,]   10
[3,]   21
[4,]   36
```

```
> sum(a*b)
```

```
[1] 71
```

2.2 Multiplication of matrices

The order in which matrices are multiplied affects the matrix products, e.g. $\boldsymbol{AB} \neq \boldsymbol{BA}$. For the product of two matrices to exist, the matrices must conform. For $\boldsymbol{AB}$, the number of columns of $\boldsymbol{A}$ must equal the number of rows of $\boldsymbol{B}$. The matrix $\boldsymbol{C}$ has the same number of rows as $\boldsymbol{A}$ and the same number of columns as $\boldsymbol{B}$.

$$\boldsymbol{C}_{(r\times c)} = \boldsymbol{A}_{(r\times k)}\boldsymbol{B}_{(k\times c)}$$

The ij-th element of $\boldsymbol{C}$ is given by

$$C_{ij} = \sum_{l=1}^{k} A_{il}B_{lj}$$

Outer indices give the dimension of the resulting matrix, with r rows and c columns; inner indices must match (columns of $\boldsymbol{A}$ = rows $\boldsymbol{B}$).

Example: Is the product $\boldsymbol{ABCD}$ defined? If so, what is its dimensionality? Suppose $\boldsymbol{A}_{3\times 5}\boldsymbol{B}_{5\times 9}\boldsymbol{C}_{9\times 6}\boldsymbol{D}_{6\times 23}$ Yes, it is defined, as inner indices match. Result is a 3×23 matrix (3 rows, 23 columns).

More formally, consider the product $\boldsymbol{L} = \boldsymbol{MN}$. Express $\boldsymbol{M}$ as a column vector of row vectors and $\boldsymbol{N}$ as a row vector of columns:

$$\boldsymbol{M} = \begin{pmatrix} m_1 \\ m_2 \\ \vdots \\ m_r \end{pmatrix}, \text{ where } m_i = \begin{pmatrix} M_{i1} & M_{i2} & \cdots & M_{ic} \end{pmatrix} \; \boldsymbol{N} = \begin{pmatrix} n_1 & n_2 & \vdots n_b \end{pmatrix}, \text{ where } n_j = \begin{pmatrix} N_{1j} \\ N_{2j} \\ \cdots \\ N_{cj} \end{pmatrix}$$

The ij-th element of $\boldsymbol{L}$ is the inner product of $\boldsymbol{M}$'s row i with $\boldsymbol{N}$'s column j: $\boldsymbol{L} = \begin{pmatrix} m_1 \cdot n_1 & m_1 \cdot n_2 & \cdots & m_1 \cdot n_b \\ m_2 \cdot n_1 & m_2 \cdot n_2 & \cdots & m_2 \cdot n_b \\ \vdots & \vdots & \ddots & \vdots \\ m_r \cdot n_1 & m_r \cdot n_2 & \cdots & m_r \cdot n_b \end{pmatrix}$

Example:

$$\boldsymbol{AB} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} ae+bg & af+bh \\ ce+dg & cf+dh \end{pmatrix}$$

Likewise

$$\boldsymbol{BA} = \begin{pmatrix} ae+cf & eb+df \\ ga+ch & gd+dh \end{pmatrix}$$

Order of multiplication matters! Indeed, consider $\boldsymbol{C}_{3\times 5}\boldsymbol{D}_{5\times 5}$ which gives a 3×5 matrix, versus $\boldsymbol{D}_{5\times 5}\boldsymbol{C}_{3\times 5}$ which is not defined.

```
> A = matrix(c(3, 1, 2, 5), nrow = 2, ncol = 2, byrow = TRUE)
> B = matrix(c(1, 2, 0, 1), nrow = 2, ncol = 2, byrow = TRUE)
> A %*% B
```

```
     [,1] [,2]
[1,]    3    7
[2,]    2    9
```

```
> B %*% A
```

```
     [,1] [,2]
[1,]    7   11
[2,]    2    5
```

```
> C = matrix(c(3, 1, 2, 2, 5, 4, 1, 1, 2), nrow = 3, ncol = 3, byrow = TRUE)
> D = matrix(c(2, 0, 5, 0, 4, 21), nrow = 3, ncol = 2)
> # C
> # D
> C %*% D
```

```
     [,1] [,2]
[1,]   16   46
[2,]   24  104
[3,]   12   46
```

```
> # D %*% C # non-conformable arguments
```

3 Transpose of a matrix

The transpose of a matrix exchange the rows and columns, $\boldsymbol{A}_{ij}^{\top}$. You can also encounter $\boldsymbol{A}_{ij}'$. Useful identities:

$$(\boldsymbol{AB})^{\top} = \boldsymbol{B}^{\top}\boldsymbol{A}^{\top}$$

$$(\boldsymbol{ABC})^{\top} = \boldsymbol{C}^{\top}\boldsymbol{B}^{\top}\boldsymbol{A}^{\top}$$

Examples:

$$\boldsymbol{C} = \begin{pmatrix} 3 & 1 & 2 \\ 2 & 5 & 4 \\ 1 & 1 & 2 \end{pmatrix}, \boldsymbol{C}^{\top} = \begin{pmatrix} 3 & 2 & 1 \\ 1 & 5 & 1 \\ 2 & 4 & 2 \end{pmatrix}$$

$$\boldsymbol{D} = \begin{pmatrix} 2 & 0 \\ 5 & 0 \\ 4 & 21 \end{pmatrix}, \boldsymbol{D}^{\top} = \begin{pmatrix} 2 & 0 & 5 \\ 0 & 4 & 21 \end{pmatrix}$$

```
> C = matrix(c(3, 1, 2, 2, 5, 4, 1, 1, 2), nrow = 3, ncol = 3, byrow = TRUE)
> D = matrix(c(2, 0, 5, 0, 4, 21), nrow = 3, ncol = 2)
> C
> t(C)
> D
> t(D)
```

4 The identity matrix

The identity matrix serves the role of the number 1 in matrix multiplication: $\boldsymbol{AI} = \boldsymbol{IA} = \boldsymbol{A}$. $\boldsymbol{I}$ is a square diagonal matrix, with all diagonal elements being one, all off-diagonal elements being zero.

$$\boldsymbol{I}_{3\times 3} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

```
> I = diag(3)
> I
```

5 The inverse matrix

For a square matrix $\boldsymbol{A}$, define its inverse $\boldsymbol{A}^{-1}$, as the matrix satisfying

$$\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}$$

For $\boldsymbol{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, $\boldsymbol{A}^{-1} = \frac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$.

If $ad - bc$ (the determinant of $\boldsymbol{A}$, $\det(\boldsymbol{A})$) is zero, the inverse does not exist. If $\det(\boldsymbol{A})$ is not zero, $\boldsymbol{A}^{-1}$ exists and A is said to be non-singular. If $\det(\boldsymbol{A})$=0, $\boldsymbol{A}$ is singular, and no *unique* inverse exists (generalised inverses do).

Generalised inverses, and their uses in solving systems of equations, are discussed in Appendix 3 of Lunch & Walsh. $\boldsymbol{A}^{-1}$ is the typical notation to denote the G-inverse of a matrix. When a G-inverse is used, provided the system is consistent, then some of the variables have a family of solutions (e.g. $x_1 = 2$, but $x_2 + x_3 = 6$).

For a diagonal matrix $\boldsymbol{D}$, $\det(\boldsymbol{D})$, which is also denoted $|\boldsymbol{D}|$, is the product of the diagonal elements. For any square matrix $\boldsymbol{A}$, $\det(\boldsymbol{A})$ is the product of the eigenvalues λ of $\boldsymbol{A}$, which satisfy

$$\boldsymbol{A}\boldsymbol{e} = \lambda\boldsymbol{e}.$$

If $\boldsymbol{A}$ is a n $\times$ n matrix, solutions to λ are a n-degree polynomial, and $\boldsymbol{e}$ is the eigenvector associated with λ. If any of the roots to the equation are zero, $\boldsymbol{A}^{-1}$ is not defined. In this case, for some linear combination $\boldsymbol{b}$, we have $\boldsymbol{A}\boldsymbol{b} = 0$.

Useful identities:

$$(\boldsymbol{A}^\top)^{-1} = (\boldsymbol{A}^{-1})^\top$$

$$(\boldsymbol{A}\boldsymbol{B})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$$

```
> B = matrix(c(1, 2, 0, 1), nrow = 2, ncol = 2, byrow = TRUE)
> solve(B)
```

```
     [,1] [,2]
[1,]    1   -2
[2,]    0    1
```

```
> B %*% solve(B)
```

```
     [,1] [,2]
[1,]    1    0
[2,]    0    1
```

```
> C = matrix(c(3, 1, 2, 2, 5, 4, 1, 1, 2), nrow = 3, ncol = 3, byrow = TRUE)
> solve(C)
```

```
      [,1]       [,2]       [,3]
[1,]  0.50  0.0000000 -0.5000000
[2,]  0.00  0.3333333 -0.6666667
[3,] -0.25 -0.1666667  1.0833333
```

```
> C %*% solve(C) # comment
```

```
             [,1] [,2]          [,3]
[1,] 1.000000e+00    0  4.440892e-16
[2,] 2.220446e-16    1 -8.881784e-16
[3,] 1.110223e-16    0  1.000000e+00
```

6 Solving equations

Matrices are compact ways to write systems of equation

$$\begin{aligned} 5x_1 + 6x_2 + 4x_3 &= 6 \\ 7x_1 - 3x_2 + 5x_3 &= -9 \\ -x_1 - x_2 + 6x_3 &= 12 \end{aligned}$$

$$\begin{pmatrix} 5 & 6 & 4 \\ 7 & -3 & 5 \\ -1 & -1 & 6 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 6 \\ -9 \\ 12 \end{pmatrix}$$

To solve $\boldsymbol{A}x = \boldsymbol{c}$, similar to scalar algebra:

- multiply both side by $\boldsymbol{A}^{-1}$ (on the left) $\boldsymbol{A}^{-1}\boldsymbol{A}x = \boldsymbol{A}^{-1}\boldsymbol{c}$
- by definition of the inverse, $\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$, thus $\boldsymbol{I}x = x = \boldsymbol{A}^{-1}c$

Exercise 1. Put the following system of equations in matrix form, and solve using R.

$$\begin{cases} 3x_1 + 4x_2 + 4x_3 + 6x_4 = -10 \\ 9x_1 + 2x_2 - x_3 - 6x_4 = 20 \\ x_1 + x_2 + x_3 - 10x_4 = 0 \\ 2x_1 + 9x_2 + 2x_3 - x_4 = -10 \end{cases}$$

7 The trace

The trace, $\text{tr}(\boldsymbol{A})$ or $\text{trace}(\boldsymbol{A})$, of a square matrix $\boldsymbol{A}$ is simply the sum of its diagonal elements. The importance of the trace is that it equals the sum of the eigenvalues of $\boldsymbol{A}$, $\text{tr}(\boldsymbol{A}) = \sum \lambda$. For a covariance matrix $\boldsymbol{V}$, $\text{tr}(\boldsymbol{V})$ measures the total amount of variation in the variables.

```
> B = matrix(c(1, 2, 0, 1), nrow = 2, ncol = 2, byrow = TRUE)
> diag(B)
```

```
[1] 1 1
```

```
> sum(diag(B))
```

```
[1] 2
```

```
> C = matrix(c(3, 1, 2, 2, 5, 4, 1, 1, 2), nrow = 3, ncol = 3, byrow = TRUE)
> diag(C)
```

```
[1] 3 5 2
```

```
> sum(diag(C))
```

```
[1] 10
```

8 Quadratic and bilinear forms

Quadratic product:
for $\boldsymbol{A}_{n\times n}$ and $\boldsymbol{x}_{n\times 1}$

$$\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} \boldsymbol{a}_{ij}\boldsymbol{x}_i\boldsymbol{x}_j$$

Bilinear form (generalisation of quadratic product):
for $\boldsymbol{A}_{m\times n}$, $\boldsymbol{a}_{n\times 1}$ and $\boldsymbol{b}_{m\times 1}$, their bilinear is $\boldsymbol{b}_{1\times m}^\top \boldsymbol{A}_{m\times n}\boldsymbol{a}_{n\times 1}$

$$\boldsymbol{b}^\top \boldsymbol{A}\boldsymbol{a} = \sum_{i=1}^{m}\sum_{j=1}^{n} \boldsymbol{A}_{ij}\boldsymbol{b}_i\boldsymbol{a}_j$$

Note that $\boldsymbol{b}^\top \boldsymbol{A}\boldsymbol{a} = \boldsymbol{a}^\top \boldsymbol{A}^\top \boldsymbol{b}$

9 Variance-Covariance matrix

- A very important square matrix is the variance-covariance matrix $\boldsymbol{V}$ associated with a p-dimensional random vector X.
- $V_{ij} = \text{Cov}(X_i, X_j)$ so that the i-th diagonal element of $\boldsymbol{V}$ is the variance of X_i, and off-diagonal element are covariances
- $\boldsymbol{V}$ is a symmetric, square matrix. The number of unique elements of $\boldsymbol{V}$ of size $p \times p$ matrix is $(p^2 - p)/2 = p(p-1)/2$.

A few properties.

If a, b, c, d are constant (non-random) and $\boldsymbol{A}$ is a $r \times p$ matrix and $\boldsymbol{B}$ a $s \times p$ matrix (non-random), then

- $\text{Cov}(X_i, a) = 0$
- $\text{Cov}(X_i, X_i) = \text{Var}(X_i)$
- $\text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i)$
- $\text{Cov}(aX_i, bX_j) = ab\text{Cov}(X_i, X_j)$
- $\text{Cov}(X_i + a, X_j + b) = \text{Cov}(X_i, X_j)$
- $\text{Cov}(aX_i + bX_j, cX_k + dX_l) = ac\text{Cov}(X_i, X_k) + ad\text{Cov}(X_i, X_l) + bc\text{Cov}(X_j, X_k) + bd\text{Cov}(X_j, X_l)$
- $\text{Var}(aX_i + bX_j) = a^2\text{Var}(X_i) + b^2\text{Var}(X_j) + 2ab\text{Cov}(X_i, X_j)$
- $\text{Var}(X) = \text{Cov}(X, X^\top)$
- $\text{Var}(\boldsymbol{A}X) = \text{Cov}(\boldsymbol{A}X, (\boldsymbol{A}X)^\top) = \boldsymbol{AVA}^\top$
- $\text{Cov}(\boldsymbol{A}X, (\boldsymbol{B}X)^\top) = \boldsymbol{AVB}^\top$

Example: Suppose X is a 3-dimensional random vector and that the variances of X_1, X_2 and X_3 are $10, 20$ and 30. X_1 and X_2 have a covariance of -5, X_1 and X_3 of 10, while X_2 and X_3 are uncorrelated.
What is the variance of the random variables $Y_1 = X_1 - 2X_2 + 5X_3$ and $Y_2 = 6X_2 - 4X_3$, where

$$\boldsymbol{V} = \begin{pmatrix} 10 & -5 & 10 \\ -5 & 20 & 0 \\ 10 & 0 & 30 \end{pmatrix}, \boldsymbol{c}_1 = \begin{pmatrix} 1 \\ -2 \\ 5 \end{pmatrix}, \boldsymbol{c}_2 = \begin{pmatrix} 0 \\ 6 \\ -4 \end{pmatrix}$$

$\text{Var}(Y_1) = \text{Var}(\boldsymbol{c}_1^\top X) = \boldsymbol{c}_1^\top \text{Var}(X)\boldsymbol{c}_1 = 960$
$\text{Var}(Y_2) = \text{Var}(\boldsymbol{c}_2^\top X) = \boldsymbol{c}_2^\top \text{Var}(X)\boldsymbol{c}_2 = 1200$
$\text{Cov}(Y_1, Y_2) = \text{Cov}(\boldsymbol{c}_1^\top X, \boldsymbol{c}_2^\top X) = \boldsymbol{c}_1^\top \text{Var}(X)\boldsymbol{c}_2 = -910$

```
> V = matrix(c(10, -5, 10, -5, 20, 0, 10, 0, 30), nrow = 3, ncol = 3, byrow = TRUE)
> c1 = matrix(c(1, -2, 5), ncol = 1)
> c2 = matrix(c(0, 6, -4), ncol = 1)
> t(c1) %*% V %*% c1
```

```
     [,1]
[1,]  960
```

```
> t(c2) %*% V %*% c2
```

```
     [,1]
[1,] 1200
```

```
> t(c1) %*% V %*% c2
```

```
     [,1]
[1,] -910
```

10 Normal Distribution

The continuous random variable X follows a normal distribution if its probability density function is defined as:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$$

for $\mu \in \mathbb{R}$, and $0 < \sigma$. The mean of X is μ and the variance of X is σ^2. We say $X \sim \mathcal{N}(\mu, \sigma^2)$.

We next simulate 1000 observations of a random variable X of mean 10 and variance 2. 10 is the true (population) mean and 2 is the true (population) variance. From the 1000 observations of X we can calculate the sample mean and the sample variance, which are estimates of the true mean and variance.

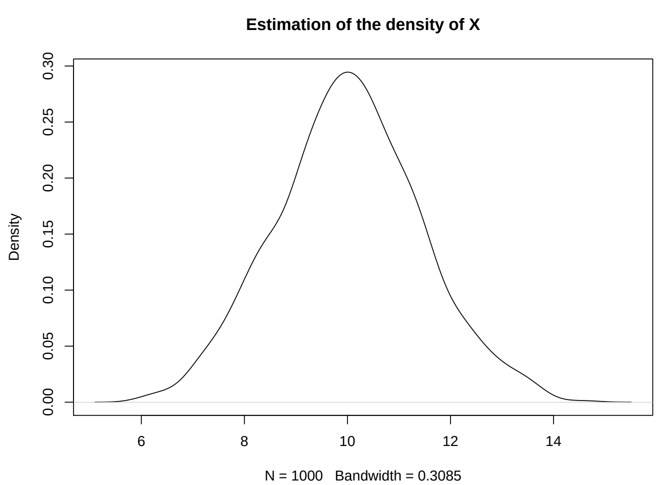
```
> # simulate 1000 random numbers from a normal distribution with mean 10 and variance 2
> set.seed(123) # for reproducibility
> num = rnorm(1000, mean = 10, sd = sqrt(2))
> mean(num) # sample mean
```

```
[1] 10.02281
```

```
> var(num) # sample variance
```

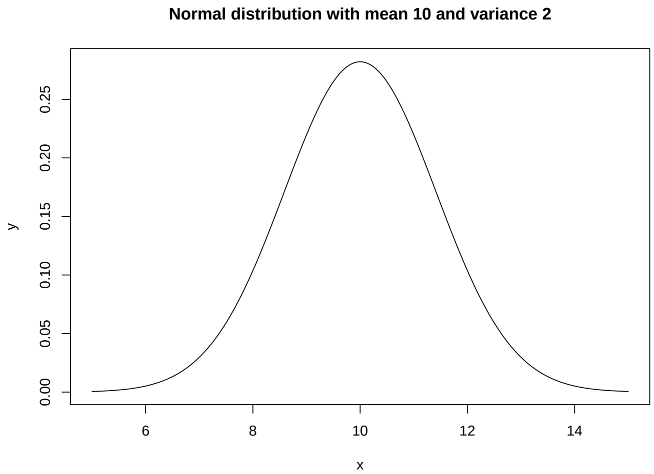
```
[1] 1.966918
```

```
> # plot the density of the normal distribution
> plot(density(num), main = "Estimation of the density of X")
```



We can also display the true density of X.

```
> #### OR
> x   <- seq(5,15,length=1000)
> y   <- dnorm(x,mean=10, sd=sqrt(2))
> plot(x,y, type="l", lwd=1, main ="Normal distribution with mean 10 and variance 2")
```



11 Multivariate Normal Distribution (MVN)

11.1 Definition

Consider p independent normal random variables $\{X_1, \ldots, X_p\}$, the i-th of which has mean μ_i and variance σ_i^2, then X has mean $\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix}$ and its covariance matrix is

$$\boldsymbol{V} = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & \ldots & \sigma_n^2 \end{pmatrix}, \quad \text{with} |\boldsymbol{V}| = \prod_{i=1}^{p} \sigma_i^2$$

The joint density of X is then,

$$\begin{aligned} p(\boldsymbol{x}) &= \prod_{i=1}^{p} (2\pi)^{-1/2} \sigma_i^{-1} \exp\left(-\frac{(x_i - \mu_i)^2}{2\sigma_i^2} \right) \\ &= (2\pi)^{-p/2} \left(\prod_{i=1}^{p} \sigma_i \right)^{-1} \exp\left(-\sum_{i=1}^{p} \frac{(x_i - \mu_i)^2}{2\sigma_i^2} \right) \end{aligned}$$

Using $\sum_{i=1}^{p} \dfrac{(x_i - \mu_i)^2}{\sigma_i^2} = (\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{V}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})$, this can be expressed more compactly in matrix form

$$p(\boldsymbol{x}) = (2\pi)^{-n/2} |\boldsymbol{V}|^{-1/2} \exp\left[-\frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{V}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \right],$$

which is also the matrix form in the case of non-independent random variables (off-diagonal of $\boldsymbol{V}$ is non zero), i.e. this holds for any vector $\boldsymbol{\mu}$ and symmetric positive-definite matrix $\boldsymbol{V}$, as $|\boldsymbol{V}| > 0$.

Just as a univariate normal is defined by its mean and variance, a multivariate normal is defined by its mean vector $\boldsymbol{\mu}$ (also called the centroid) and variance-covariance matrix $\boldsymbol{V}$.

11.2 Properties of the MVN

- if X is MVN, any subset of X is also MVN
- if X is MVN, any linear combination of the elements of X is also MVN. If $X \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{V})$:
  $Y = X + \boldsymbol{a} \sim \mathcal{N}(\boldsymbol{\mu} + \boldsymbol{a}, \boldsymbol{V})$
  $Y = \boldsymbol{a}^\top X = \sum_{k=1}^{n} a_i X_i \sim \mathcal{N}(\boldsymbol{a}^\top \boldsymbol{\mu}, \boldsymbol{a}^\top \boldsymbol{V}\boldsymbol{a})$
  $Y = \boldsymbol{A}X \sim \mathcal{N}(\boldsymbol{A}\boldsymbol{\mu}, \boldsymbol{A}^\top \boldsymbol{V}\boldsymbol{A})$
- Conditional distributions are also MVN.
  Partition X into two components, X_A (m dimensional column vector) and X_B (n-m dimensional column vector): $X = \begin{pmatrix} X_A \\ X_B \end{pmatrix}, \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_A \\ \boldsymbol{\mu}_A \end{pmatrix}$ and $\boldsymbol{V} = \begin{pmatrix} \boldsymbol{V}_{AA} & \boldsymbol{V}_{AB} \\ \boldsymbol{V}_{AB}^\top & \boldsymbol{V}_{BB} \end{pmatrix}$.
  Note that the marginal distributions are $X_A \sim \mathcal{N}(\boldsymbol{\mu}_A, \boldsymbol{V}_{AA})$ and $X_B \sim \mathcal{N}(\boldsymbol{\mu}_B, \boldsymbol{V}_{BB})$.
  $X_A|X_B$ is MVN with m-dimensional mean vector

  $$\boldsymbol{\mu}_{A|B}(\boldsymbol{x}_B) = \boldsymbol{\mu}_A + \boldsymbol{V}_{AB} V_{BB}^{-1}(\boldsymbol{x}_B - \boldsymbol{\mu}_B)$$

  and $m \times m$ covariance matrix

  $$\boldsymbol{V}_{A|B} = \boldsymbol{V}_{AA} - \boldsymbol{V}_{AB} V_{BB}^{-1} V_{AB}^\top$$

- if X is MVN, the regression of any subset of X on another subset is linear and homoscedastic

  $$\begin{aligned} \boldsymbol{x}_A &= \boldsymbol{\mu}_{A|B}(\boldsymbol{x}_B) + \boldsymbol{e} \\ &= \boldsymbol{\mu}_A + \boldsymbol{V}_{AB}\boldsymbol{V}_{BB}^{-1}(\boldsymbol{x}_B - \boldsymbol{\mu}_B) + \boldsymbol{e} \end{aligned}$$

  where $\boldsymbol{e}$ is MVN with mean vector $\boldsymbol{0}$ and variance-covariance matrix $\boldsymbol{V}_{A|B}$.
  The regression is linear because it is a linear function of $\boldsymbol{x}_B$ ($\boldsymbol{x}_B - \boldsymbol{\mu}_B$)
  The regression is homoscedastic because the variance-covariance matrix for $\boldsymbol{e}$ does not depend on the value of the $\boldsymbol{x}$'s ($\boldsymbol{V}_{A|B}$).

Example 1. *Example: Regression of Offspring value on Parental values.*

Assume the vector of offspring value and the values of both its parents is MVN. Then from the correlations among (outbred) relatives,

$$\begin{pmatrix} z_0 \\ z_s \\ z_d \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \mu_0 \\ \mu_s \\ \mu_d \end{pmatrix}, \sigma_z^2 \begin{pmatrix} 1 & h^2/2 & h^2/2 \\ h^2/2 & 1 & 0 \\ h^2/2 & 0 & 1 \end{pmatrix}\right]$$

Let $X_A = (z_0)$, $X_B = \begin{pmatrix} z_s \\ z_d \end{pmatrix}$, $\boldsymbol{V}_{AA} = \sigma_z^2$, $\boldsymbol{V}_{AB} = \frac{h^2\sigma_z^2}{2}\begin{pmatrix} 1 & 1 \end{pmatrix}$, $\boldsymbol{V}_{BB} = \sigma_z^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

$\boldsymbol{\mu}_{A|B}(\boldsymbol{x}_B) = \boldsymbol{\mu}_A + \boldsymbol{V}_{AB}V_{BB}^{-1}(\boldsymbol{x}_B - \boldsymbol{\mu}_B)$

Hence,

$$\begin{aligned} z_0 &= \mu_0 + \frac{h^2\sigma_z^2}{2}\begin{pmatrix} 1 & 1 \end{pmatrix}\sigma_z^{-2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} z_s - \mu_s \\ z_d - \mu_d \end{pmatrix} + e \\ &= \mu_0 + \frac{h^2}{2}(z_s - \mu_s) + \frac{h^2}{2}(z_d - \mu_d) + e \end{aligned}$$

where e is normal with mean zero and variance

$$\boldsymbol{V}_{A|B} = \boldsymbol{V}_{AA} - \boldsymbol{V}_{AB}V_{BB}^{-1}V_{AB}^{\top}$$

$$\begin{aligned} \sigma_e^2 &= \sigma_z^2 - \frac{h^2\sigma_z^2}{2}\begin{pmatrix} 1 & 1 \end{pmatrix}\sigma_z^{-2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\frac{h^2\sigma_z^2}{2}\begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= \sigma_z^2\left(1 - \frac{h^4}{2}\right) \end{aligned}$$

Hence, the regression of offspring trait value given the trait values of its parents is

$$z_0 = \mu_0 + h^2/2(z_s - \mu_s) + h^2/2(z_d - \mu_d) + e$$

where the residual e is normal with mean zero and variance $\mathrm{Var}(e) = \sigma_z^2(1 - h^4/2)$

Similar logic gives the regression of offspring breeding value on parental breeding value as

$$\begin{aligned} A_0 &= \mu_0 + (A_s - \mu_s)/2 + (A_d - \mu_d)/2 + e \\ &= A_s/2 + A_d/2 + e \end{aligned}$$

where the residual e is normal with mean zero and variance $\mathrm{Var}(e) = \sigma_A^2/2$.

Writing to file Additional Files For Students/Rcode/PartA - Matrix Reminder.R